

ECONOMIC GROWTH MODELS: A PRIMER / STUDENT'S GUIDE

Miguel Lebre de Freitas (afreitas@novasbe.pt)

Nova School of Business and Economics

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Contents

Introduction: The growth question

Part I – Factor accumulation

1. Population and Economic Development
2. The basic Solow model
3. Exogenous Growth
4. The Neoclassical model with Human Capital
5. The AK model

Part II – Technology and its diffusion

6. External economies and Learning by Doing
7. Research and development
8. The extent of the market
9. Technology adoption

Part III – Fundamentals

10. Government inputs
11. Distortions
12. Inequality
13. Corruption
14. Geography or Institutions?

Epilogue: what have we learned?

Detailed Contents

ECONOMIC GROWTH MODELS: A PRIMER / STUDENT'S GUIDE.....	1
Contents	2
Detailed Contents.....	3
0 Introduction	15
0.1 Outline of the book	15
0.1.1 General organization.....	15
Part I – Basic models	16
Part II – Technology and its diffusion	17
Part III – Getting the prices right	18
0.2 How to use this book?.....	20
0.2.1 One semester course	20
0.2.2 Half term course in economic growth.....	21
0.2.3 Other courses	21
0.3 Symbols and notation.....	22
Symbols used	22
0.4 Mathematical notation	23
Part I – Factor accumulation	25
1 Population and stages of development	26
1.1 Introduction.....	26
1.2 The basic Malthusian model	28
1.2.1 Diminishing Returns	29
1.2.2 The Malthus theory of population.....	30
1.2.3 Dynamics and equilibrium.....	31
1.2.4 What happens when technology improves?.....	33
1.2.5 Smith' mark of prosperity	35
1.3 Population and technology.....	36
1.3.1 Population and technological change.....	36
1.3.2 Path dependence and initial conditions.....	37
Box 1.4. Technology and population density: an historical experiment.....	38

1.3.3	Race between technology and diminishing returns	39
1.3.4	Escaping the trap.....	41
1.3.5	Modelling the relationship between technology and population	42
1.4	The demographic transition	46
1.4.1	The three phases of economic development	46
1.4.2	Birth rates and death rates.....	50
1.4.3	Why birth rates decline with economic development?	51
1.4.4	Missing institutions and risk premium.....	52
1.4.5	Income and substituting effects	53
1.4.6	The stubbornness of birth rates	55
1.4.7	Why are birth rates so persistent?	56
1.5	Globalization, fertility decisions and the Great Divergence.....	59
	Box 1.7. The Great Divergence	60
1.5.1	Industrialization and the demographic transition.....	62
1.5.2	From the Agriculture Revolution to the Industrial Revolution.....	63
1.5.3	International trade and industrialization	66
1.5.4	Trading population for productivity.....	68
1.5.5	Static and dynamic gains from international trade	69
1.6	Key ideas of Chapter 1.....	70
1.7	Review questions and exercises.....	71
2	The basic Solow model.....	75
2.1	Introduction.....	75
2.2	The Solow model	76
2.2.1	The neo-classical production function.....	76
2.2.2	Main assumptions of the Solow model.....	79
2.2.3	Factor prices and factor income shares.....	81
2.2.4	Factor income shares.....	82
2.2.5	The flow income chart	82
2.2.6	The Fundamental Dynamic Equation	83
2.2.7	Graphical illustration	84
2.2.8	The steady state.....	85
2.3	The Solow model and the facts of economic growth.....	86
2.3.1	The Solow model and the Kaldor acts	86
2.3.2	Savings, population and per capita income in the real world	88
2.4	Transitional Dynamics	92

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

2.4.1	What happens if the savings rate increases?	92
2.4.2	What happens if population growth or the depreciation rate decline?.....	93
2.4.3	What happens if the level of technology improves?	94
2.5	The Golden Rule	95
2.5.1	The Golden Rule of capital accumulation	95
2.5.2	Dynamic efficiency	98
	Box 2.3. Using taxes and subsidies to meet the golden rule.....	100
2.6	The case with endogenous savings	101
2.6.1	An optimal consumption rule.....	101
2.6.2	What happens when the rate of time preference decreases?	101
2.6.3	The modified golden rule.....	102
2.6.4	Exogenous or endogenous savings?.....	103
2.7	The Solow Residual	104
2.8	Key ideas in Chapter 2	106
	Appendix 2.1: The optimal consumption path in a simple 2-period model.....	107
	Problems and Exercises	108
3	Exogenous Growth	111
3.1	Introduction.....	111
3.2	Perfect technological diffusion	112
3.3	The extended Solow model.....	113
3.3.1	Labour augmenting technological progress	113
3.3.2	The two components of “technology”	113
3.3.3	Other assumptions.....	114
3.3.4	Labour in efficiency units	114
3.3.5	The Fundamental Dynamic Equation revisited.....	115
3.3.6	The steady state.....	116
3.4	Transitional Dynamics	117
3.4.1	What happens if the Savings Rate increases?	117
3.5	The extended Solow model meeting the real-world facts.....	120
3.5.1	Revisiting the Kaldor’ facts	120
3.5.2	Absolute convergence.....	121
3.5.3	Interpreting growth patterns using the Solow model.....	126
3.6	Growth accounting revisited	129
3.6.1	Autonomous versus induced contributions.....	129
3.6.2	An alternative approach	130
3.7	Discussion: What we have achieved?	134

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

3.8	Key ideas of Chapter 3	135
Appendix 3.1	Transition dynamics in the Solow model.....	136
	Problems and Exercises	137
4	The Neoclassical model with Human Capital	140
4.1	Introduction.....	140
4.2	The Lucas' 1990 Paradox	141
4.2.1	Explaining cross-country income differences using the Solow model..	141
4.2.2	Using the production function	143
4.2.3	Why doesn't capital flow from rich to poor counties?	145
4.2.4	So, how can we explain these inconsistencies?	150
4.3	Human capital	151
4.3.1	What is human capital?	151
4.3.2	Human capital as an input to production	152
4.3.3	Human capital and the productivity of physical capital.....	154
4.4	The model of Mankiw, Romer and Weil (MRW).....	155
4.4.1	Main assumptions	155
4.4.2	The steady state in the MRW model.....	156
4.4.3	Factor income shares in the MRW model.....	158
4.4.4	Cross-country income differences revisited.....	159
4.4.5	Conditional convergence again.....	162
4.5	Productivity matters!.....	166
4.6	Discussion: where are we standing.....	170
4.7	Key ideas of Chapter 4.....	170
	Problems and Exercises	171
5	The AK model	174
5.1	Introduction.....	174
5.2	The basic AK model	175
5.2.1	Getting rid of diminishing returns.....	175
5.2.2	A Graphical Illustration	176
5.2.3	What happens when the saving rate increase?	177
5.2.4	The Harrod-Domar equation.....	178
5.2.5	No convergence	179
5.3	Incarnations of the AK model.....	180
5.3.1	The Harrod-Domar model.....	180
5.3.2	A one sector model with Physical and Human Capital.....	185
5.3.3	A two-sector model of endogenous growth.....	187

5.4	The AK model with endogenous savings	191
5.4.1	Adding the optimal consumption rule to the AK model.....	191
5.4.2	Transpiration responds to inspiration!	192
5.4.3	Proximate versus fundamental causes of economic growth	194
5.5	Empirical controversies	198
5.5.1	Level effects or growth effects?.....	198
5.5.2	Tests on conditional convergence.....	200
5.6	Discussion	207
5.7	Key ideas of Chapter 5.....	209
	Problems and Exercises	209
	Part II – Technology and its diffusion	213
6	External economies and learning by doing.....	214
	Learning Goals:.....	214
6.1	Introduction.....	214
6.2	Externalities on capital accumulation	216
6.2.1	A “development modifier”.....	216
6.2.2	Factor prices in the competitive equilibrium	217
6.2.3	The aggregate production function	218
6.2.4	The AK model again.....	219
6.3	The market failure and optimal intervention	220
6.3.1	Growth accounting revisited	220
6.3.2	The social return of capital	221
6.3.3	Optimal intervention	221
6.3.4	Growth effects.....	222
6.4	The case with increasing returns.....	223
6.4.1	External economies of scale.....	223
6.4.2	Alfred Marshall and the theory of economic geography	225
6.4.3	The Arrow model.....	229
6.4.4	Lucas: externalities on Human Capital	234
6.4.5	Localized versus global technological spillovers	236
6.5	Learning by doing and international trade	237
6.5.1	Comparative advantages locked in	237
6.5.2	Static versus dynamic benefits of trade	239
6.5.3	Terms of trade effects	241
6.6	Key ideas of Chapter 6.....	243

6.7	Problems and Exercises	243
7	Research and development	246
	Learning Goals:.....	246
7.1	Introduction.....	246
7.2	R&D Taxonomy.....	247
7.2.1	Basic research versus R&D	247
7.2.2	Horizontal and vertical innovations	248
7.2.3	Process innovations and product innovations	249
7.3	The basic R&D model	249
7.3.1	A production function for final goods	249
7.3.2	Production function for intermediate goods	250
7.3.3	Two sources of technological change	251
	Box 7.1. The division of labour effect	251
7.3.4	The trade-off between production and R&D	252
7.4	Market structure and operating profits.....	253
7.4.1	Demand for intermediate inputs.....	253
7.4.2	The case with an horizontal innovation	253
7.4.3	Static efficiency versus dynamic efficiency	256
7.4.4	The equilibrium wage rate and operating profits	257
7.5	Competition through innovation	258
7.5.1	Limit pricing	258
7.5.2	The case with a more efficient vertical innovation	259
7.5.3	Drastic versus non-drastic vertical innovations	260
7.5.4	Creative destruction	262
7.5.5	Neck-and-neck competition	265
7.6	R&D and an investment decision	269
7.6.1	The market value of an innovation	269
7.6.2	The net present value of investment in R&D.....	271
7.6.3	The break-even value of the innovation.....	272
7.6.4	The equilibrium level of R&D (partial equilibrium)	272
7.7	Making knowledge excludable	276
7.7.1	Excludability sources	276
7.7.2	Patents, copyrights and trade marks.....	278
7.7.3	The economics of patents.....	279
7.7.4	The case against patents.....	281

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

7.8	Financing research and development.....	285
7.8.1	Financing constraints	285
7.8.2	Subsidies to private R&D	288
7.8.3	Government funded R&D.....	289
7.9	Key ideas of chapter 7.....	290
	Appendix 7.1. The crowding out effect	291
	Appendix 7.2. The equilibrium level of R&D	293
	Problems and Exercises	295
8	The extent of the market.....	301
	Learning Goals:.....	301
8.1	Introduction.....	301
8.2	The Big Push model.....	302
8.2.1	The demand side	302
8.2.2	Traditional production	303
8.2.3	Equilibrium under cottage production	303
8.2.4	Modern production.....	304
8.2.5	Industrialization and the extent of the market.....	305
8.2.6	Why the decentralized economy may fail.....	306
8.2.7	The profit function again	307
8.2.8	What do we mean by equilibrium?	308
8.2.9	Cottage for sure.....	308
8.2.10	Industrialization for sure.....	309
8.2.11	Multiple equilibrium	310
8.2.12	Graphical illustration	310
8.2.13	The coordination failure.....	311
8.2.14	Expectations.....	312
8.2.15	Infrastructures	312
8.3	The extent of the market and the division of labour	315
8.3.1	Expanding varieties.....	315
8.3.2	The new trade theory	320
8.4	The division of labour and the extent of the market	322
8.4.1	Vertical complementarities and circular causation	322
8.4.2	Traps and trade.....	326
8.4.3	Multinationals and linkages	328
8.5	Centripetal and centrifugal forces.....	331
8.5.1	International factor mobility	331

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

8.5.2	The core-periphery model.....	333
8.5.3	The Krugman-Venables theory.....	335
8.6	Discussion.....	337
8.7	Key ideas of chapter 8.....	338
	Problems and Exercises	339
9	Technology adoption	342
	Learning Goals:.....	342
9.1	Introduction.....	342
9.2	The key role of openness	345
9.2.1	A necessary condition	345
9.2.2	The critical role of international trade	346
9.2.3	Foreign Direct Investment	347
9.3	Permeability to technological diffusion.....	350
9.3.1	Complementarities	351
9.3.2	The vintage capital theory.....	352
9.3.3	Switching costs	353
9.3.4	Leapfrogging.....	355
9.3.5	Barriers to technological adoption.....	357
9.4	Matching specific needs.....	359
9.4.1	Self-discovery	359
9.4.2	Mastering foreign technologies.....	360
9.4.3	Institutions do not travel well	361
9.5	A simple model of technology adoption.....	364
9.5.1	Modelling technology adoption	364
9.5.2	The Steady state	367
9.5.3	Transition dynamics.....	369
9.5.4	What happens if the technological adoption effort increases?.....	370
9.5.5	What happens when barriers to technology adoption decline?.....	371
9.5.6	What happens when the world rate of technological progress increases? 372	
9.5.7	The great divergence revisited.....	374
9.5.8	Proximate causes versus ultimate causes once again.....	374
9.6	Key ideas of chapter 9.....	377
	Problems and Exercises	378
	Part III – Government, policies and institutions	380
10	Government services.....	381

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Learning Goals:.....	381
10.1 Introduction.....	381
10.2 The role of government.....	382
10.2.1 Why do we need a government?	382
10.2.2 Public goods and publicly provided goods	384
10.2.3 Intervention options	387
Box 10.1. Normative versus positive economics	388
10.3 A simple growth model with government spending	388
10.3.1 Private productivity and public inputs	388
10.3.2 Congestion versus non-congestion	389
10.3.3 The aggregate production function	390
10.3.4 Factor income shares.....	391
10.3.5 The market failure.....	391
10.3.6 Getting the prices right.....	391
10.4 Intervention trade-offs	392
10.4.1 Unproductive public expenditures	392
10.4.2 Income and expenditure	396
10.4.3 The flow income chart	397
10.4.4 The steady state.....	398
10.4.5 The golden rule	398
10.4.6 Efficient bureaucracy	399
10.4.7 A Graphical illustration.....	400
Box 10.4. The case with an AK production function	401
10.5 Key ideas of chapter 10.....	403
Problems and Exercises	403
11 Distortions.....	406
Learning Goals:.....	406
11.1 Introduction.....	406
11.2 Distortion in the consumption-investment decisions	407
11.2.1 The case with endogenous savings	409
11.2.2 Example: Financial intermediation and economic growth	410
11.3 Discriminatory taxation	413
11.3.1 The efficient allocation	413
11.3.2 Distortionary taxes	415
11.3.3 Self-financed subsidies	417
11.3.4 Tax evasion	420

11.3.5 Externalities	425
11.4 Policies and growth.....	426
Key ideas of chapter 11	431
Problems and Exercises	433
12 Inequality	436
12.1 Introduction.....	436
12.2 Facts on inequality and growth.....	437
12.2.1 Inequality in the personal distribution of income	437
12.2.2 Measurement problems.....	438
12.2.3 Inequality and economic performance in the real world	442
12.3 Inequality, efficiency and redistribution	447
12.3.1 Fair inequality	448
12.3.2 Unfair redistribution.....	450
12.3.3 Unfair inequality	451
12.3.4 The trade-off between efficiency and equity	452
12.3.5 Redistribution and aggregate savings	453
12.4 Financial market imperfections.....	455
12.4.1 A inter-generational model with bequests	456
12.4.2 The case with perfect financial markets.....	457
12.4.3 The case with borrowing constraints	458
12.4.4 Redistribution and efficiency revisited	461
12.4.5 Human versus physical capital.....	462
12.5 Inequality and political institutions.....	465
12.6 Key ideas of Chapter 12.....	467
12.7 Problems and Exercises	468
13 Corruption	470
Learning Goals:.....	470
13.1 Introduction.....	470
13.2 Corruption	471
13.2.1 What is corruption?.....	471
13.2.2 What is rent seeking?	472
13.2.3 What bribes are for?.....	473
13.2.4 The grease in the wheel's argument.....	474
13.3 A model of centralized corruption	478
13.3.1 Main assumptions	478
13.3.2 The clever Kleptocrat.....	479

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

13.3.3	Dynamic considerations.....	480
13.3.4	Shaping the incentives of the leader	482
13.4	The model with decentralized corruption	483
13.4.1	Decentralized corruption and the “tragedy of the commons”	483
13.4.2	Rent seeking as a diversion activity.....	484
13.4.3	The rent-seekers’ production function.....	485
13.4.4	Optimal rent seeking at the individual level	485
13.4.5	The equilibrium level of rent seeking	486
13.4.6	What happens if the effectiveness of rent seeking time declines?.....	487
13.4.7	What happens if the tax rate increases?	488
13.4.8	Other costs of corruption	491
13.5	Coping with decentralized corruption.....	492
13.5.1	Rules versus discretion	492
13.5.2	Sticks and carrots	492
13.5.3	Social norms.....	494
13.5.4	Optimal corruption.....	496
13.6	When institutions become dysfunctional	496
13.6.1	Why corruption brings more corruption?	497
13.6.2	Multiple equilibria	498
13.6.3	Escaping the trap.....	498
13.7	Discussion	500
13.8	Key ideas of Chapter 13.....	501
13.9	Problems and Exercises	502
14	Institutions and Geography	506
	Learning Goals:.....	506
14.1	Introduction.....	506
14.2	The Institutions hypothesis	507
14.2.1	What are institutions?	507
14.2.2	How have institutions evolved over time?.....	509
14.2.3	Economic institutions.....	510
14.2.4	Political institutions	512
14.2.5	No natural gravitation	514
14.3	The Geography hypothesis	516
14.3.1	Cumulative causation.....	516
14.3.2	Latitude and per capita income	518

14.3.1 The mark of geography	519
14.3.2 Long lasting effects.....	522
14.4 The geography vs institutions debate.....	523
14.4.1 Natural experiments	523
14.4.2 Reversals of fortune	525
14.4.3 The indirect influence of geography	527
14.5 Key ideas of Chapter 14.....	529
Problems and Exercises	530

0 Introduction

0.1 Outline of the book

0.1.1 General organization

The book comprises three parts.

Part I introduces the basic models of economic growth, namely the Malthusian model, the Solow model and the AK model, as well as some of their variants. These models focus mainly on the contribution of inputs to production, while technology is in general assumed exogenous.

Part II is dedicated to productivity in the “engineering sense”. The main question is how the technological level is determined and why it differs across the space. The theories addressed include those that see technological changes as a natural outcome of productive experience (learning by doing), those that model R&D and a rent seeking activity (the Schumpeterian view), theories that stress the role of strategic complementarities (big-push), and a model of technological catch up.

Part III is devoted to productivity in the sense of resource allocation. The main concern is the effectiveness with which factors of production and a given technology are combined to produce valuable output. The chapters discuss the role of the government in providing public goods and infrastructure, as well as the pervasive role of market failures or government failures in distorting the allocation of resources. This part of the book also addresses the key role of institutions in shaping the incentives of policymakers, and it ends up with the debate on geography vs institutions as the primary determinant of economic growth.

The proposed structure is an attempt to organize the narrative along a sequence of models or versions of a baseline model. To take full opportunity of the analytical tools and concepts as they are introduced in each chapter, we occasionally override the thematic boundaries the chapter is said to address. For instance, questions like endogenous technological change, cumulative causation and static vs. dynamic efficiency arrive as early as in Chapter 1. In Chapter 6, where the main topics are knowledge spillovers and learning by doing, we take opportunity to discuss other externalities and the implications for international <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

trade. The implication of pursuing essentially a model-based narrative is that some important topics are scattered along the book, instead of systematically addressed in purposeful chapters. This includes, for instance, the relationship between international trade and growth.

Part I – Basic models

Chapter 1 focuses on the relationship between population and economic development. This chapter offers a first approach to the problem of diminishing returns and explains the importance of technological progress in overcoming it. The chapter introduces briefly the mutual causality between the size of the population and technological change. Finally, the chapter reviews the theory of demographic transition and presents a theory to explain why some countries entered in modern growth later than others.

Chapter 2 is devoted to the basic Solow model. It is shown that introducing physical capital in the production function, wages no longer need to converge to a low-income equilibrium trap. Still, because physical capital is itself subject to diminishing returns, this model is not capable of generating sustained economic growth. The chapter includes a discussion on the optimality of the saving rate and shows that making the saving rate endogenous does not alter qualitatively the conclusions. Finally, with the help of a simple growth accounting exercise, it is argued that assuming an invariant technology is at odds with real world facts.

Chapter 3 extends the basic Solow model to the case where technology expands over time. It is shown that this modification rescues the model from its main limitation and makes it capable of describing the main stylized facts of economic growth. It is explained why exogenous technological progress is a natural implication of the other assumptions of the model.

Chapter 4 explains the failure of the Solow model in accounting for cross-country differences in per capita income in terms of magnitudes, in light of the Lucas (1990) Paradox. The key role of country-specific factors affecting productivity is unrevealed. The chapter then extends the model so as to account for human capital accumulation, and shows that this extension improves the ability of the model to account for cross-country differences in per capita income. Even though, it is argued, accounting for human capital is not enough. Thus, one needs to account for the possibility of productivity to differ across countries. The chapter

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

ends up with a discussion on the need to better understand what drives total factor productivity, distinguishing the two main components: “efficiency in resource allocation”, which we relate to the level of productivity (to be addressed Part III); and “technology”, which is assumed to drive the growth rate of technology (to be addressed in Part II).

Chapter 5 reviews different incarnations of the AK model to show that, getting rid of diminishing returns, factor accumulation could lead to unceasing growth, without the need to postulate exogenous technological progress. It is argued that the empirical evidence has not been too favourable to the view that factor accumulation alone is enough to induce sustained growth. It is pointed out, however, that the linear structure of the AK model underlies the models of endogenous technological change that we address later in the book.

Part II – Technology and its diffusion

Chapter 6 motivates the AK model with the argument that externalities related to capital accumulation can be strong enough to overcome diminishing returns. The chapter reviews the theories of Marshallian externalities and learning by doing. The chapter introduces the concept of cumulative causation and addresses briefly the localized-versus-global knowledge spillovers controversy. The implications of learning by doing for international trade policy are also discussed.

Chapter 7 addresses the question of why some people devote time and effort to develop new products and production processes. The chapter explains the role of knowledge excludability in providing economic agents with market incentives to invent new technology and explains the Schumpeterian trade-off between static efficiency and dynamic efficiency. The role of patents and other mechanisms of knowledge exclusion in securing the benefits of research is addressed. It is argued that, even with patent coverage, inventors do not in general fully appropriate the benefits of their inventions, so a case may exist for government intervention. The chapter includes a discussion on the relationship between product market competition and innovation. A simplified Schumpeterian model is introduced, to illustrate the optimal allocation of time between working time and time devoted to research.

Chapter 8 returns to the case with imperfect competition, to explore the argument that the dimension of the market may influence the pace of technological adoption. Drawing on theories of big push, it is shown that the extent of the market and modernization can be

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

reciprocally determined, giving rise to cumulative causation and cycles. The question of international trade and comparative advantages is re-examined, now in terms of a model with pecuniary externalities. The chapter ends up with a discussion of the role of geography as a fundamental determinant of economic performance.

Chapter 9 deals with the problem of a developing country which main challenge is to adopt technologies developed abroad. The chapter starts with a discussion on the factors that delay the pace of cross-country technological diffusion. Then a simple model of technological catch up is presented. In this model, there is a World technological frontier and the country's policies determine how close the country gets to that frontier. In this model, backwardness provides the poor country with the potential to catch up, but whether this advantage materializes or not, it depends on the country economic and political circumstances.

Part III – Getting the prices right

Chapter 10 focuses on the role of government in providing essential inputs to production, such as the rule of law and public infrastructure. A simple model with a non-excludable good is used to explain the trade-off between the benefits of public provision and the costs of taxation. The model is also used to illustrate the pervasive role of government failures.

Chapter 11 addresses the effect of different types of distortions, in a context of constant returns to scale. The chapter starts with distortions affecting the consumption savings decision and exemplifies with transport costs and financial market imperfections. Then, it addresses the case of misallocation in factor markets caused by distortionary taxation, tax evasion, monopoly, externalities, high inflation, and human rural migration. Intervention trade-offs are again discussed, now adding the idea of second-best decision-making. The chapter ends up with a brief introduction to the debate on economic reform in emerging economies, and the controversy surrounding the Washington Consensus.

Chapter 12 turns to the relationship between inequality and growth. After presenting some empirical evidence, it is argued that inequality is inherent to the market mechanism of producing incentives. To the extent that inequality arises in result of different opportunities, societies may abdicate some efficiency, to achieve more equity. Finally, it is shown that there

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

might be no trade-off between efficiency and equity in case information is asymmetric and agents face borrowing constraints. The chapter also discusses the indirect effect of inequality on growth, through pressures for redistribution and conflict.

Chapter 13 addresses the implications of corruption, using an extended Solow model with public inputs. Three cases of corruption are distinguished: centralized corruption, decentralized corruption and the case of generalized corruption. In the later case, strategic complementarities in corruption give rise to virtuous and vicious cycles. The chapter describes the importance of institutions in getting the incentives right.

0.2 How to use this book?

The book allows some flexibility in terms of reading. However, some chapters that depend on preceding chapters should not be read in isolation. The core chapters are the Solow model and the AK model (chapters 2, 3 and 5). In the table that follows, next to each chapter is indicated the chapter that should be read before:

The book is designed to cover a one-semester course fully devoted to economic growth. In principle, the lecturer should be able to cover all chapters, skipping non-essential sections in each chapter at his own discretion.

For shorter courses, possible options are the following:

0.2.1 One semester course

The book is basically designed for a course on economic growth with development concerns. For a semester course on Economic Growth, the following sequence is appropriated:

2. The Basic Solow model
3. Exogenous Growth (2)
4. The Neoclassical model with Human Capital (3)
5. Endogenous Growth (2)
6. Learning by doing (5)
7. Research and development
9. Technology adoption (3)
10. Government Services (2)
13. Corruption and rent seeking (10)

A course more concerned with development issues and the role of government could try the following:

1. The Malthusian model
2. The Basic Solow model
3. Exogenous Growth (2)

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

- 5. Endogenous Growth (2)
- 8. The extent of the market
- 9. Technology adoption (3)
- 10. Government Services (2)
- 11. Distortions (10)
- 12. Inequality
- 13. Corruption and rent seeking (10)

0.2.2 Half term course in economic growth

The book can also be used in half term courses, skipping non-core sections.

A possible selection is the following:

- 2. The Basic Solow model
- 3. Exogenous Growth
- 4.2. The Lucas Paradox
- 5.2. The simple AK model
- 5.3. The AK model with endogenous savings
- 5.4.4. A two-sector model of endogenous growth
- 6.2. Externalities on capital accumulation
- 7. Research and development
- 9. Technology adoption

0.2.3 Other courses

Specific chapters can also be adopted in other courses. Chapters 2, 3, and parts of 5 and 6 can be adopted in intermediate courses in macroeconomics. Policy oriented courses on macro or on public economics can find useful the chapters 10, 11 and 13. Chapters 7 and 8 may be of interest for undergraduate courses on Industrial Economics. Chapter 8 may be useful for a course on international trade. Chapter 1 alone may be of interest for a course on Economic History. The book can also provide a brief introduction to growth models, for students engaged in more advanced studies.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

0.3 Symbols and notation

Symbols used

Y = Output

Y_d = Households' disposable income

T = Land

N = Labour, population

y = Per-capita output

C = Private consumption

c = Private consumption per capita

β = Physical capital-output elasticity

α = Human capital-output elasticity

K = Physical Capital

I = Gross investment in Physical Capital

k = Capital per worker

π = Profits

L = Labour measured in efficiency units

λ = Effective labour input per worker.

$\tilde{k}$ = Capital per unit of efficiency labour

$\tilde{y}$ = Output per unit of efficiency labour

H = Human Capital

I^H = Gross investment in Human Capital

h = Human Capital per worker

$\tilde{h}$ = Human capital per unit of efficiency labour

s = Fraction of disposable income devoted to physical capital accumulation

s_H = Fraction of disposable income devoted to human capital accumulation

$\tilde{k}$ = Physical capital per unit of Human Capital

$\tilde{y}$ = Output per unit of Human Capital

s_R = Fraction of disposable income devoted to Research and Development

v = Speed of adjustment to the steady state in the neoclassical growth model

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

δ = Depreciation rate

γ = Growth rate of per capita income/Growth rate of Harrod Neutral TFP

g = Hick Neutral rate of technological progress

ε = Externality

η = External effect of public inputs

ρ = Subjective discount rate

ψ = Fraction of working time devoted to rent-seeking

b = Effectiveness of the rent seeking effort

ϕ = Fraction of public expenditures which are unproductive

$1-u$ = Fraction of human capital devoted to human capital accumulation

μ = Fraction of the labour force devoted to R&D

r = Real Interest rate

w = Real wage-rate

G = Productive government expenditures

τ = Production tax / income tax

τ_H = Tax on human capital income

τ_K = Tax on physical capital income

x_j = Production of intermediate input j

X = Composite measure of intermediate inputs

N_j = Raw labour used in production of intermediate input j

N_Y = Labour used in the production of Y

F = Fixed cost

t = Time index

0.4 Mathematical notation

A dot over a variable denotes time variation:

$$\dot{X} = \partial X / \partial t .$$

The time variation divided by the level is the growth rate:

$$\hat{X} = \dot{X} / X$$

When a variable grows at a constant rate – say g – over time, the relationship between the value of X at time zero and at time t , is:

$$X_t = X_0 e^{gt}.$$

In logs, a linear equation arises:

$$\ln X_t = \ln X_0 + gt$$

In many figures, economic variables are represented in logs, so that we can read the growth rate in the slope of a linear regression.

Part I – Factor accumulation

1 Population and stages of development

“The most decisive mark of the prosperity of any country is the increase in the number of its inhabitants”. [Adam Smith].

Learning Goals:

- Acknowledge the importance of the Malthus theory of population to explain historical facts in the ultra-long run.
- Understand the challenges posed by the Law of Diminishing Returns
- Understand the critical role of technology in overcoming diminishing returns to labour.
- Acknowledge that a larger population may also bring benefits, through faster technological change.
- Understand the factors that drive the changing attitude towards fertility along the process of demographic transition.
- Use the Malthusian theory and the theory of demographic transition to interpret the rising cross-country economic disparities that followed the industrial revolution.

1.1 Introduction

The world population has been expanding at impressive rates. Along the last two centuries, the World population increased from 1 billion to more than 7 billion. Although population growth is decelerating, population is still increasing and is expected to reach 9 billion in 2050.

A question that arises is whether the continuing population expansion will challenge our living standards, by overwhelming the existing resources. Such question was first formulated by the British philosopher Reverend Thomas Malthus, in its famous book “An essay on the Principle of Population”, published in 1798. Malthus contended that a fixed amount of natural resources could not feed a constantly increasing population. Thus, the population explosion that was already becoming evident in the 18th century could not be sustained forever. Malthus observed that societies throughout history had experienced at one time or another different types of “checks” that prevented population to overstretch their

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

resource limitations. This includes epidemics, famines, and wars, but also abstinence and birth control.

This chapter addresses the Malthus theory of population, as well as its limitations. As we will see, the Malthus ideas provide a useful tool to interpret the almost constant living standards that characterized the pre-industrial era. The Malthusian model might also be thought to provide a reasonable narrative for some of the world's today poorest countries and regions. But fortunately, the Malthus pessimism regarding the future of the human kind did not materialize: somehow ironically, at the time Malthus was writing his book, a set of countries in West Europe entered in a new phase of economic development, in which population and living standards were expanding together. The failure of the Malthus theory to describe modern growth is explained by three main limitations in the basic model. First, Malthus overlooked the potential of technological progress in mitigating the challenges posed by resource constraints. Second, he ignored the potential benefits of a larger population in speeding up the rate of technological progress. Third, he ignored the possibility of technological change impacting on the human attitude towards fertility.

This chapter reviews these questions, focusing on the relationship between the size of population, per capita income, and technology. The chapter is organized as follows. Section 1.2 introduces the Malthus theory in its basic formulation, where the pervasive role of diminishing returns is stressed. In Section 1.3, we discuss the potential role of technology in overcoming the limitations imposed by diminishing returns. This chapter also introduces the possibility of technological change being a positive function of the size of population. Section 1.4 addresses the question of why the Malthus theory of population no longer applies in our days. The section reviews some theories that explain why the attitude towards fertility has changed with economic development. In Section 1.5, we put the all pieces together to offer an interpretation of why cross country income disparities increased significantly in the two hundred years that followed the Industrial Revolution.

Starting a course on economic growth with a chapter focusing on population is most useful. First, because it illustrates the problem of diminishing returns in its simpler form, and also the key role of technological progress to explain long term economic growth. Second, because the model provides a credible narrative for the history of the human kind along the long period that preceded the industrial revolution. Understanding the forces that triggered the changes in the demographic behaviour in the now developed countries may shed some

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

light on the challenges posed to the poorest countries that are still to make such change. Third, because it brings to the analysis the positive relation between the size of population and technological, that underlies many theories of technology and growth, in its simplest possible manner. Finally, the chapter provides a good opportunity to introduce key concepts and ideas that will be further explored along the book. This includes the notion of per capita income, equilibrium, stability, transitional dynamics, static and dynamic efficiency, and knowledge diffusion, among others.

1.2 The basic Malthusian model

The original theory of Thomas Malthus was essentially descriptive, and of course much richer than the simplification adopted here. We turn, however, to a very simple model to sketch out the basic mechanics underlying his central argument.

Consider a closed economy (i.e. one with no international trade), with no government, and basically devoted to agriculture. In this economy, output takes the form of a single homogeneous good (Z), which is produced using labour (N) and land (T). In this simple model, population and the labour force are the same. Since we are concerned with long term growth, we assume that prices are flexible, so that full employment holds each moment in time.

The relationship between inputs and output is described by an aggregate production function of the form¹:

$$Z_t = BT_t^\beta N_t^{1-\beta}. \quad (1.1)$$

In equation (1.1) the subscript t is a time index. The term B is called Total Factor Productivity (TFP). This term measures the state of technology, or the “efficiency” in which

¹ Specification (1.1) corresponds to a well-known class of production functions, named Cobb-Douglas. The main properties of the Cobb-Douglas production function are diminishing returns and a unit elasticity of substitution between inputs. The Malthus theory is consistent with more general production functions, but we stick with this particular formulation, for simplicity.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

the resources are being combined for production. This parameter is assumed exogenous and equally available to all agents in the economy (that is, technology is assumed to spill-over across individuals instantaneously and at no cost).

To capture the existence of physical limits to land expansion, we assume that the amount of land available to agriculture is fixed ($T = \bar{T}$). The labour force, on the contrary, is endogenously determined, as explained below.

1.2.1 Diminishing Returns

The Law of Diminishing Returns (LDR) is one of the oldest and more important postulates in Economics. In short, it states that, increasing one ingredient of production keeping all other ingredients constant has a decreasing marginal impact on output.

In the case at hand, if the availability of land is fixed and technology is held constant, the only way of expanding production is by increasing the use of labour. However, increasing the use of labour will come along with a lower level of output per worker. To see this, let's divide both terms in (1.1) by the number of workers, N , obtaining:

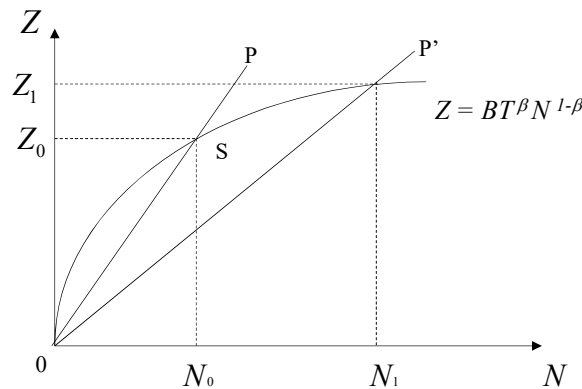
$$y = B \left(\frac{\bar{T}}{N} \right)^\beta, \quad (1.2)$$

where $y = Z/N$ denotes for output per worker, or per capita output (remember that in this model the population and the work force are the same).

The LDR is illustrated in Figure 1.1. In the figure, the vertical axis measures the level of output (Z) and the horizontal axis measures the size of the labour force, N . The average product of labour is measured by the slope of the ray that departs from the origin and crosses the production function in each point. For instance, when the workforce is equal to N_0 , output will be Z_0 . The corresponding average product of labour is Z_0/N_0 (i.e, the slope of OP). Given the shape of the production function, when the number of workers rises to N_1 , output expands less than proportionally, so that the average product of labour declines to OP' .

Figure 1.1: Output, population, and productivity

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>



The figure illustrates the Law of Diminishing Returns, referring to the production function (1.1). The slopes of the rays OP and OP' measure output per worker, as given by equation (1.2). All else equal, when the number of workers increases from N_0 to N_1 , output per worker declines (the slope of OP' is lower than that of OP). Note that marginal products are measured by the slopes of the production function in each point, and are also decreasing.

1.2.2 The Malthus theory of population

Malthus formulated his theory of population observing first, biological facts in the wild life. He noted that, in nature, animals and plants are “impelled by a powerful instinct to the increase of their species”. He also pointed out that “superabundant effects” are repressed by “want of room and nourishment”².

Malthus then argued that a similar mechanism holds for human beings: the “passion between sexes” compels humans to proliferate. However, resources are limited. Hence, a point will come when a human population expanding without control reaches the limit up to which food sources can support it. Beyond this point, any further increase in population would result in food shortage and henceforth in death caused by starvation, diseases, and violence. To these *natural* barriers, Malthus added more *conscientious* prevention mechanisms, which operate through birth control: the Human kind differs from other specimens in that it may deliberately reduce fertility to avoid resource shortages. In his terminology, Malthus distinguished two types of checks that prevent population to expand

² Malthus, T., 1798. An Essay on the Principle of Population as it Affects the Future Improvement of the Society. London: J. Johnson.

beyond the land carrying capacity: “*positive checks*”, like epidemics, famines and wars, which cause death; and “*preventive checks*”, like abortion, birth control and postponement of marriage, which refrain birth.

In terms of our simple model, the Malthus theory of population is accounted for introducing a “subsistence level of per capita income” ($\bar{y}$) defined as the minimum income necessary to sustain the life of a human being. Thus, whenever per capita income rises above that subsistence level, population expands; when per capita income fall below the subsistence level, population declines³.

1.2.3 Dynamics and equilibrium

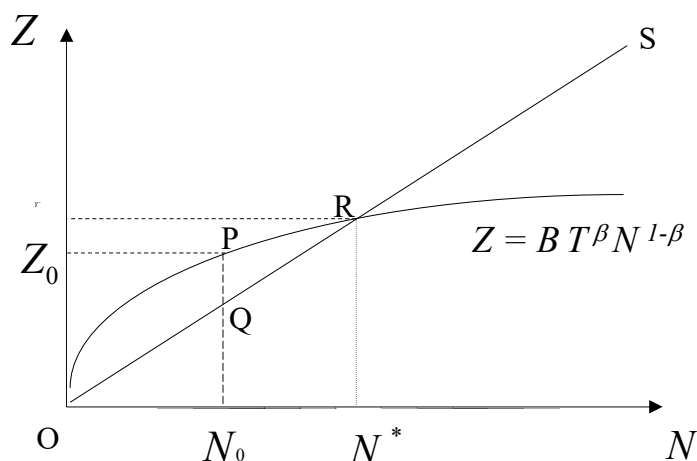
Summing up, the model has two basic ingredients: the Law of Diminishing Returns (LDR) and the Malthusian theory of population. With these two ingredients, the model solves in an intuitive manner: in the absence of technological progress, the LDR implies that a growing labour force will lead to a more intensive use of land and thereby to a decline in output per worker. As output per worker declines, population expansion decelerates. At the time output per worker equals the “subsistence” level, both population and output stop expanding. This is the equilibrium of the model.

Thus, given the land availability and the level of technology, the size of the population in this model is self-equilibrating. Furthermore, any technological improvement will be offset, in the long run, by an increase in the size of the population, without delivering any positive impact on living standards.

Figure 1.2: Dynamics and equilibrium in the Malthus Model

³ Formally, a possible way of describing the population dynamics will be setting $\dot{n}_t = \dot{N}/N = \nu[y_t - \bar{y}]$, where ν is some positive parameter measuring the speed at which the size of population responds to deviations of per capita income relative to the subsistence level.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>



The exogenous subsistence level of per capita income ($\bar{y}$) is represented by the slope of the line OS. Since at point P per capita income Z_0/N_0 is higher than the subsistence level, there will be a tendency for population to expand (move to the right). This process ends up at point R, where output per worker is equal to the subsistence level. This is the steady state of the model.

Figure 1.2 illustrates the dynamics of the model. In the figure, the exogenous level of subsistence income ($\bar{y}$) is represented by the slope of the line OS. Suppose that initially there are N_0 workers producing Z_0 . Since in this point per capita income Z_0/N_0 is higher than the subsistence level, there will be a tendency for population to expand. As population expands, land will be used more intensively implying a decline in output per worker. This process ends up at point R, where output per worker is again equal to the subsistence level. At this point, there is no tendency for population to expand: any further increase in population would result in famine, disease, and birth control. This is the equilibrium of the model.

Box 1.1. Key concept: Stable Steady State

Technically, point R in Figure 1.2 is an equilibrium, or *steady state*, because once it is reached, there will be no tendency for the economy to move away from it.

An important property of this equilibrium is that the dynamic forces of the model are such that this equilibrium will be met, irrespectively of the departing point: for instance, if population is initially smaller than N^* , per capita income will be above the subsistence level and population will expand. Conversely, if population is initially larger than N^* , then the level of per capita income will fall below the subsistence level and population will decline. Because the economy will approach the equilibrium irrespectively of the departing point, this equilibrium is said to be *stable*.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Box 1.2. Historical episode: The Black-Death

The Black Death was one of the most devastating pandemics in human history. It peaked in Europe between 1346 and 1353, killing almost one half of the population at that time. In the case of England, for instance, the Black Death led to a decline in population from about 6 million to 3.5 million people in two years only (1348-1349).

Because population naturally takes time to expand again, there was a long period after the Black Death that European populations remained below trend. If the theory of diminishing returns was right, the higher availability of land per worker after the Black Death should have resulted in higher per capita incomes (such as from R to P in Figure 1.2). The historical data actually confirms this prediction: Along the 150 years that followed the black death, GDP per capita jumped ahead of its previous trend, reflecting a lower population and diminishing returns, in a context of sluggish technological change⁴.

1.2.4 What happens when technology improves?

We now turn to the question of the equilibrium of the model changes when the level of technology improves. Technological progress means that inputs become more productive. As an example, suppose that people in this economy discovered the plough and started using water power. In terms of our model, these innovations are captured by an increase in the productivity parameter B .

Figure 1.3 describes the impact of a technological change in this economy. The figure displays the production function in the *intensive form* (equation 1.2) rather than in the *extensive form* (equation 1.1). Accordingly, the vertical axis measures per capita income, instead of total production. All the rest is the same.

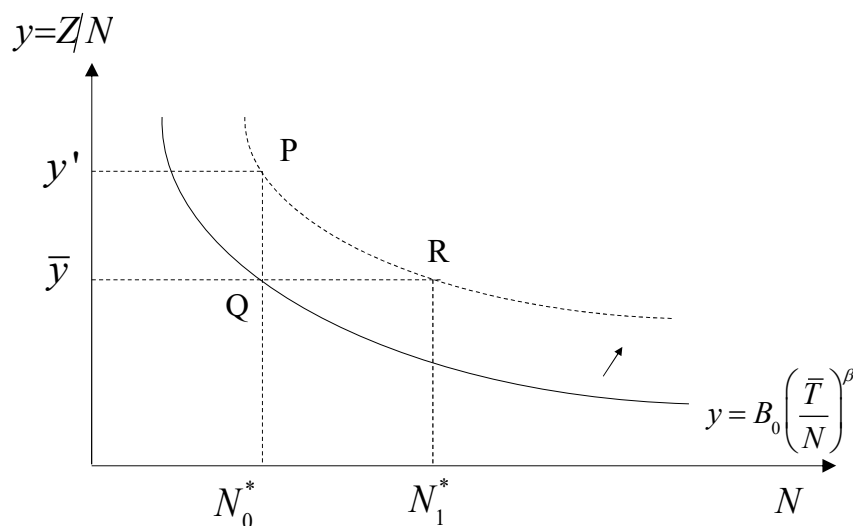
⁴ Clark, G., 2001, "The Secret History of the Industrial Revolution". UC Davis.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Suppose that the economy starts out in Q, with per capita income equal to the subsistence level, $\bar{y}$. Then, a technological improvement from B_0 to $B_1 > B_0$ causes the schedule describing output per capita to shift upwards. At the impact (with the size of population unchanged), per capita output increase to y' . Since at point P per capita income is higher than the subsistence level, population in this economy starts expanding (this is the Malthus theory of population). Then, as population slowly expands, diminishing returns show up, driving per capita income back to the subsistence level. At the time the new equilibrium, R, is reached, all gains from technological progress had been channelled to the increase in the size of population, to N_1^* . In the long run, the initial improvement in living standards has been completely offset by population expansion.

Similar results hold when the availability of land increases. If, for instance, a swampy stretch of land was drained, in terms of equation (1.1) that would be accounted for a rise in parameter T . In terms of Figure 1.3, the change is equivalent to when B increase: after all, what matters is that the “carrying capacity” of the economy (in terms of number of humans) has increased.

Figure 1.3: Productivity shift in the Malthus Model



Departing from Q, suppose that a technological improvement (or an increase in the availability of land) moves the per capita output schedule to the right (dashed curve). With a constant population, production jumps to point P. Since in P per capita income is above the subsistence level, population in this economy starts expanding until the new equilibrium, R, is reached. In the long run, the gains from technological progress are totally channelled to the increase in the size of population, to N_1^* .

Box 1.3. Key concept: Transition dynamics vs. change in the steady state

Both Figure 1.2 and Figure 1.3 describe how the Malthusian economy evolves along time until reaching the steady state. This adjustment process is called *transition dynamics*. There is however a critical distinction between the case in Figure 1.2 and that of Figure 1.3: in Figure 1.2, the economy is not initially in the steady state (point P) and approaches the steady state (point R). In Figure 1.3, the steady state of the model changes: because an exogenous parameter of the model changed, the initial equilibrium (Q) no longer holds. Hence, the economy engages in a transition dynamics until the new steady state (R) is reached.

1.2.5 Smith' mark of prosperity

The model just described reveals, in a simple manner, the dramatic implications of the LDR: for any given state of technology, the growth of the economy settles at a point where income per person is constant at the very low subsistence level.

Formally, the equilibrium level of population is found by setting the level of per capita output (1.2) equal to the subsistence income, $\bar{y}$. Solving for N , one obtains:

$$N^* = \left(\frac{B}{\bar{y}} \right)^{\frac{1}{\beta}} \bar{T} \quad (1.3)$$

This equation states that the equilibrium level of population increases with the availability of land and with the level of the technology. For instance, countries with superior technology, B , should exhibit higher *population densities*, defined as the number of inhabitants per unit of available land (N/T) - note how well this prediction of the model fits with the Adam Smith claim quoted at the beginning of this chapter!

The prediction that cross-country differences in technology should give rise to differences in population density rather than to differences in living standards was investigated empirically by some authors. Among these, Robert Lucas Jr., from the University of Chicago, found that, prior to the Industrial Revolution, differences in living standards across regions in the world were indeed much smaller than differences in technology⁵.

1.3 Population and technology

1.3.1 Population and technological change

The Malthus theory implies that population expansion exerts a *negative* influence on living standards. This is because of diminishing returns: since the amount of arable land is fixed, for any given level of technology, a larger population implies that less output is available per person. One may argue, however, that the size of population also exerts a *positive* effect on living standards, through its influence on technology⁶.

In contrast to most other goods, technology is non-rival: that is, an invention can be shared by many people without losing its effectiveness. Hence, as long as technology is free to spill over across the society, each single agent will benefit from everybody else discoveries. Thus, a society with a larger population should enjoy a faster technological change, just because a larger population has the potential to contain a larger number of potential inventors.

⁵ Lucas, R., 2004, The industrial revolution: past and future, 2003 Annual Report Essay, Federal Reserve Bank of Minneapolis.

⁶ Pioneers on this idea include Kuznets, S., 1960, Population change and aggregate output, in A.J. Coale (ed.), *Demographic and economic change in developed countries*, Princeton University Press. Simon, J., 1977, *The Economics of Population Growth*, Princeton University Press. Simon, J., 1981, *The Ultimate Resource*, Princeton University Press.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

As once William Petty, a 17th century economist and philosopher, said: “It is more likely that one ingenious curious man may rather be found among 4 million than 400 persons”. If each person has a given probability of inventing something, all else equal, a larger and more *diverse* population should, in principle, be capable of generating more inventions per unit of time.

The implication of allowing the level of technology to depend on the size of population in the Malthus model is straightforward: *population and technological progress will reinforce each other*. A larger population will bring faster technological progress, and faster technological progress will cause population to expand faster.

1.3.2 Path dependence and initial conditions

The possibility of population and technology reinforcing each other opens an avenue for *path dependence*: the technological level at a given moment in time in a given territory will be determined by the size of population in the period before, which in turn would have been determined by past technology, and so on, until the beginning of times. That being so, *initial conditions* should have played a key role in shaping cross-region technological differences along time.

Referring to figure 1.3, consider two different regions, Q and R, differing only by the quality of national resources at the very beginning. That is, technology was the same (B_0), but parameter T was larger region R, implying that the productivity curve of region R (dashed) was above that of region Q (solid). Further assume that these two regions were completely isolated from each other, so that any technological improvements in one region (change in B) could only spill over inside that region's borders.

Once the geographical conditions determined that region R could feed a larger population than region Q, the Malthusian mechanism of population dynamics would imply a larger population in region R (N_1^*) than in region Q (N_0^*). Then, because region R had a larger population, an asymmetric path of technological change would have been triggered: region R should have a higher chance of achieving technological progress just because the pool of human beings was larger, which then would bring more population, and by then faster technological change, in a virtuous cycle. Box 1.4 shows how well this prediction of the model fits the real world facts before the 15th century.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Box 1.4. Technology and population density: an historical experiment

The Malthusian prediction that the level of technology impacts positively on the size of its population was investigated by a professor from Harvard, Michael Kremer⁷. Kremer added to the simpler Malthusian formulation a mechanism of reverse causality from population to technology: he argued that regions with larger populations should exhibit faster technological progress than regions with smaller populations. The reason is intuitive: if the probability of inventing something is the same for any single person, then a region with more inhabitants should, in principle, be better endowed to generate ideas and enjoy fast technological progress than a region with less inhabitants.

A problem in this reasoning arises in that knowledge does not recognize borders: arguably, technological diffusion across regions should help mitigate technological differences, blessing the laggard regions with the opportunity to free ride on the other regions' discoveries. To abstract from this possibility, Kremer examined a particular period of the human history, where populations in different areas were effectively isolated from each other. This period provides a "natural experiment". Natural experiments are studies where the experimenter cannot manipulate the independent variable to access its effect on the dependent variable, but it can observe the effects of exogenous changes in the independent variable.

The author first observed that, before the end of the last ice age (about 10.000 B.C.) ocean levels were so low that humans could easily migrate across continents, including through the Bering Strait, which connects Asia to the Americas. Hence, at that time, technology had the potential to diffuse across regions. It is thus plausible to assume that – say - by 12.000 B.C., the known technologies were pretty similar across the all humanity. Note that in these times human were basically hunter-gatherers.

With the melting of the polar ice caps, around 10,000 B.C, land bridges were flooded. In consequence, the Old World (Europe, Asia, Africa), the Americas, Australia, Tasmania

⁷ Kremer, M. 1993. "Population growth and technological change: one million B.C. to 1990". Quarterly Journal of Economics, 108, 681-716.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

and the Flinders Island became isolated from each other, for an exogenous reason. If the theoretical model was right, one would expect that, at the time connections were re-established - with the European explorations of the 15th century - technological levels, population densities and land sizes were all positively correlated. Why? Because larger regions would have built bigger populations and therefore technology would have developed and diffuse quicker, causing in turn faster population expansions.

Kremer showed that the historical facts confirm these conjectures. By the year 1500, population densities were much higher in Eurasia-Africa (4.85/km²) - the region with larger area - than in the Americas (0.36/Km²), Australia (0.026/Km²), Tasmania (0.018/Km² to 0.074/Km²) and the Flinders Island (0.0/Km²). Accordingly, the Old World had the highest level of technological sophistication, followed by the Americas (the Aztec and the Mayan had already discovered agriculture). Australia was in an intermediate stage, having developed some artefacts like the boomerang, but with a population that was still of hunters and gatherers. Tasmania registered technological regression: its inhabitants lacked basic tools such as fire-making and lost the ability to make bone tools. Finally, the Flinders Island, with 680 square kilometres of land and only 500 inhabitants initially, lost all its inhabitants by around 4,700 BC.

1.3.3 Race between technology and diminishing returns

The idea that technological improvements impact only on the size of population without improving living standards does not square well with the historical evidence at all times. At some stage in human history, per capita income and population started growing together, implying a departure from the low level subsistence trap. To account for this possibility, let's return to our basic model.

In Figure 1.3, we examined the implications of a “once and for all” improvement in technology. In that case, the living standards increase only temporarily: as population expands to match the increased “carrying capacity” of the economy, per capita income falls back to the subsistence level.

Now, let's consider a sequence of technological improvements. In the real world, economies are exposed to continuous technological progress. So the question arises as to whether a fast enough rate of technological progress would be capable of driving per capita

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

income away from the subsistence level on a permanent basis. The answer to this question depends: (a) on the pace of technological progress; and (b) on how fast the population will respond to changes in living standards.

In terms of Figure 1.3, assume that a second invention tilted the production function while population was still on its way from P to R. And again, before the new long run equilibrium was reached, a third technological change took place, and so on. Clearly, if the economy was continuously hit by technological improvements and population expansion was slow enough, then per capita income would never fall back to the subsistence level, even if population was expanding according to the Malthusian rule. The economy would be permanently engaged in transition dynamics, with per capita income increasing over time.

An economy in such a stage is said to be engaged in the “Post-Malthusian Regime”: the rising per capita income implies a faster population growth (as Malthus predicted), but technological progress is so fast that living standards don’t actually fall back.

Formally, the condition for per capita income to depart from the subsistence level and expand over time is obtained log-differentiating (1.2):

$$\hat{y}_t = \hat{B}_t - \beta n_t > 0 \quad (1.4)$$

Where $\hat{y} = \dot{y}/y$, $\hat{B} = \dot{B}/B$, and $n = \dot{N}/N$ refer to the rates of change of per capita income, productivity and of population, respectively.

Equation (1.4) describes the change in per capita income as a *race between technological change and diminishing returns*. If technology expands fast enough and population never reaches the equilibrium, then per capita income will be growing over time. This equation reveals that the benign scenario with rising living standards is more likely, the less severe are diminishing returns (as captured by β).

Further extending our discussion, one may account for the possibility of the other input to production (T) eroding over time. So far, we have assumed that the availability of land is constant. Arguably, any eventual erosion on the quality of the arable land could be fixed with simple techniques, such as rotation of cultures. But if T is thought to include non-renewable natural resources (those that are depleted when used in production such as oil, natural gas, and climate), then T shall be modelled as decreasing over time. Defining $\delta = \dot{T}/T$ as the rate of erosion of the natural resources, then equation (1.4) becomes:

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

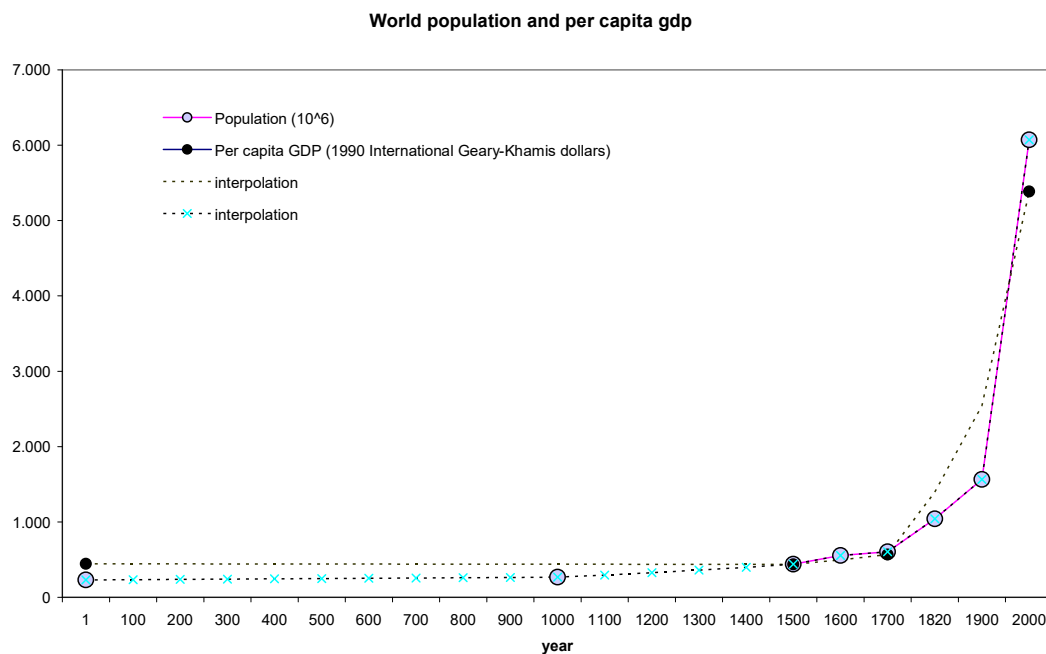
$$\hat{y}_t = \hat{B}_t - \beta(n_t + \delta_t) > 0 \quad (1.4a)$$

That is, technological progress must be fast enough to overcome both the population expansion and the erosion of natural resources along time.

1.3.4 Escaping the trap

Putting the pieces together, we now can offer a theory of why human societies managed to escape the low level subsistence trap. To motivate, Figure 1.4 plots the evolution of per capita income and of population in the world economy from year 1 to 2000. As shown in the figure, both the size of population and per capita income increased very little until the 17th century. Then, suddenly, they both accelerated sharply.

Figure 1.4 Population and per capita GDP over the last two thousand years



The figure displays the evolution of per capita income and of population along the last twenty centuries. Source: Angus Maddison, 2001. The World Economy: a Millenial Perspective. Development Centre, Paris.

This pattern has a natural interpretation in terms of the ideas sketched out above. Thousands of year ago, human societies were dispersed in small bands, with little contact with each other. At that time, the arrival of new ideas was slow. With a slow arrival of new technologies, populations had time to expand and fully match the technological change, just

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

like as described in figure 1.3. For a long period of time, as populations expanded, more and more ideas became available causing populations to expand more and more, though without departing from subsistence income.

At some stage, when populations became large enough, the arrival rate of new ideas accelerated so much that productivity growth outpaced population growth: in the figure, this is captured by the increase in per capita income after the 15th century. At that time, it looks like *population lost the race* with technological change.

1.3.5 Modelling the relationship between technology and population

So far, we have argued that a larger population should come along with faster technological progress, without modelling that relationship. In fact, modelling technological progress as a function of the size of population is not an easy task.

First, because technology is something that we don't know how to measure: shall we count ideas? Shall all ideas be counted as valuing the same? If not, how to evaluate each particular idea? Second the likelihood of an agent discovering something may depend on the existing stock of existing ideas: are previous ideas helpful for new discoveries? Or will new discoveries become more difficult as the number of ideas already discovered increases? Third, the positive relationship between the size of population and the arrival of new ideas may be mitigated by the possibility of overlapping efforts by independent inventors.

All these questions mean that the relationship between technology and the size of population is too complex to be described by a simple equation. And yet, the shape of such a "technology production function" is an essential ingredient in our model. With no surprise, the choice of an appropriate specification for the relationship between technology and the size of population became a matter of dispute in the research arena.

A specification introduced by Paul Romer (1990) is as follows⁸:

$$\dot{B} = bNB. \quad (1.5)$$

In this equation, the exogenous parameter b captures the likelihood of somebody inventing something. In light of this formulation, a larger population shall deliver a faster *rate* of technological progress, because it has the potential to contain more “ingenious curious man”.

A problem with this formulation is that it gives rise to a counterfactual “scale effect”, whereby the growth rate of per capita income becomes a positive function of the size of population. To see this, just replace (1.5) in (1.4), obtaining $\hat{y}_t = bN_t - \beta n_t$. With a growing population ($n_t > 0$), that would mean an ever-accelerating rate of economic growth⁹.

Technically, the scale effect arises because the creation of new knowledge in (1.5) is assumed proportional to the existing stock of knowledge in the economy (B). As for the rationale, you may interpret this linearity as capturing a positive externality arising from the existing stock of knowledge to new knowledge: inventors develop new ideas learning from old ideas.

As an example, consider the case of the steam engine, invented by James Watt in 1769. The idea was not inspired by watching steam rise from a teakettle’s spout, as it is commonly said. The idea was actually borne while James Watt was repairing an earlier steam engine invented 57 years before by Thomas Newcomen. The later, in turn, was an improvement of a steam engine patented in 1698 by the Englishman Thomas Savery, which followed another designed by the Frenchman Denis Papin around 1680, which in turn had

⁸ Romer, P., 1990. “Endogenous technological change”. *Journal of Political Economy* 98, s71-s102.

⁹ In alternative, if the size of population adjusted so fast that per capita income remained at the subsistence level each moment in time, then population growth in the steady state would become a positive function of the size of population, $n_t = (b/\beta)N_t$. The scale effect would just be moved to population instead as to per capita income.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

precursors in the ideas of the Dutchman Christiaan Huygens, and so on¹⁰. The cumulative nature of knowledge was coined the “standing on shoulders” effect, owing this name to a famous quotation from Isaac Newton (“If I have been able to see further, it was only because I stood on the shoulders of giants”).

Charles Jones, in a paper published in 1995, proposed an alternative specification for the production function of knowledge to get rid of the scale effect¹¹. Jones added a mechanism to counteract the linearity implied by the “standing on shoulders effect”. The main argument is that new discoveries are increasingly difficult to find. That is, as the stock of accumulated knowledge increases, researchers will find it more difficult to invent new technologies, because the easiest ideas have already been discovered. Then, as technology becomes more complex, it takes more time and effort for a researcher to learn everything it needs just to catch up with cutting edge.

As for an illustration, note that many breakthrough inventions in the 18th and 19th centuries were achieved by hobbyists or by single individuals. Thomas Edison, for instance, invented alone the light bulb, the phonograph and the motion picture. Today, advances in technology are mostly achieved by scientists engaged in research teams and focusing on very narrow problems. The assumption that new discoveries become increasingly more difficult became known as the “Fishing out effect”¹².

Formally, the “Fishing out effect” is modelled allowing past discoveries to impact on the productivity of current researchers with declining marginal returns:

$$\dot{B} = bN_t^\theta B_t^\varepsilon \quad , \text{ with } 0 < \varepsilon < 1, 0 < \theta < 1 \quad (1.6)$$

¹⁰ Diamond, J., 1998. *Guns, Germs and Steel: a short history of everybody for the last 13,000 years*. Vintage, Surrey, UK.

¹¹ Jones, C., 1995. "Time series tests of endogenous growth models", *Quarterly Journal of Economics*, 110, 495-525.

¹² The label “fishing out” arises from the classical example of the fishing pound for the Tragedy of the Commons: if the pound is stocked with a *fixed* number of fish, then it becomes increasingly difficult to catch each new fish.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

In this model, the sign and magnitude of parameter ε captures the net balance of two opposing externalities on productivity growth: the “standing on shoulders” effect (positive), whereby productivity of current research increases with the accumulated knowledge in the society and; the “fishing out effect” (negative) whereby past discoveries turn new ideas more difficult. Jones conjectured that the net effect of these two externalities leads to $0 < \varepsilon < 1$: that is, new researchers benefit from previous ideas, but there are diminishing returns to knowledge in knowledge production.

The parameter θ captures the negative externality arising from overlapping discoveries. This happens when the same piece of knowledge is invented independently by different inventors. This effect weakens the relationship between population size and technological change, and is coined “stepping on shoes”. This assumption is not, however, necessary nor sufficient to remove the scale effect¹³.

When the knowledge production function takes the form (1.6), the rate of technological progress becomes:

$$\hat{B} = \frac{\dot{B}}{B} = bN_t^\theta B_t^{\varepsilon-1}. \quad (1.7).$$

With $0 < \varepsilon < 1$, (1.7) implies a negative relationship between the growth rate of B and the level of B . This means that the model converges to a steady state, with a constant rate of technological progress, $\hat{B}$.

The steady state growth rate may be obtained by log-differentiating both sides of (1.7) and setting the result equal to zero. This gives:

$$\frac{\dot{B}}{B} = \theta n / (1 - \varepsilon) \quad . \quad (1.8)$$

¹³ Of course, in the limit case in which $\theta = 0$, all individuals would inventing exactly the same piece of knowledge each moment in time. In that case, the relationship between the size of population and the rate of technological progress would disappear, and the model become of exogenous growth. But that would be an extreme and unrealistic case.

In (1.8), the growth rate of technology is not a direct function of the population size, so the scale effect is removed. Still, a *weak scale effect* shows up: the growth rate of per capita income is a direct function of the population growth rate. This is different however from the original *strong* scale effect whereby the growth rate of technology was a function of the size of the population.

Note that in model 1.8 population growth is *necessary* to sustained technological progress: in (1.7), the assumption that new ideas become increasingly difficult to discover implies that the growth rate of B falls down to zero over time when the population is constant (in other words, once linearity is removed from the knowledge production function, a constant population will no longer be sufficient to sustain the continuing proportional increase in the stock of knowledge that is necessary to sustain long run growth). Thus, only with an ever-increasing population will be possible to maintain a constant rate of technological progress. Still, a model that displays a growth rate of per capita income as a positive function of population growth is much more realistic than a model where the growth rate of per capita income is a positive function of population size.

1.4 The demographic transition

According to the Malthus theory of population, a higher per capita income should come along with a faster population expansion. This proposition looks like squaring well with the real world facts previous to the industrial era, but it no longer applies in *modern societies*: in our days, wealthier people tend to have less children, not more.

So, a question arises as why there has been such a change in humans' attitude towards fertility. This section briefly reviews the theories of demographic transition, establishing a bridge between the Malthusian theory of population, which applies to pre-industrial societies, and theories more adequate to describe economies engaged in the Modern Growth Regime. This quest will help understand why some countries departed from the Malthusian trap sooner than others, and why cross-country income disparities today are as they are.

1.4.1 The three phases of economic development

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Looking at the relationship between population growth per capita income along the human history, Oded Galor and David Weil proposed a classification of economic development in three distinct phases (see Box 1.6 for real world data)¹⁴:

- The Malthusian Regime characterised by slow technological progress and with population responding positively to per capita income. In this regime, most of technological progress is matched by population expansion (Europe before the Industrial Revolution, Asia until the end of the nineteenth century).

- The Modern Growth regime, characterized by steady growth of per capita income and in which the relationship between income and population growth is reversed: the acceleration of per capita income translates into a slower population growth (Western Offshoots by the end of 19th century, Europe at the beginning of the 20th century, Asia after the third quarter of the twentieth century).

- The Post-Malthusian regime, an intermediate stage between the Malthusian and the Modern growth regimes. This regime shares one characteristic with each of one of the other two: the Malthusian (positive) relationship between income per capita and population still holds; but technology takes a clear lead in the race with population, so per capita income accelerates as well.

Box 1.5: Population and technological change along the last twenty centuries

Table 1.1 describes the evolution of GDP, population and per capita GDP in West Europe and in Asia along that last two thousand years. A rough estimate of the pace of *technological progress* $\dot{B}/B$ is also displayed in line (4), using equation (1.4) and postulating a value for β equal to 1/3 (details in the table).

¹⁴ Galor, O. and Weil, D., 2000. "Population, Technology and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond". The American Economic Review 90(4), 806-828.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

According to the table, in Western Europe and in Asia, GDP and population expanded very slowly from year 1 to 1.000 (lines 1 and 2), reflecting an almost stagnant technology (line 4). During this period, technological change was fully matched by population expansion, in accordance to the basic Malthus model.

In the centuries that followed, up to the Industrial Revolution (1820), technological progress was slow for modern standards. In Western Europe, TFP growth stood at rates that ranged from 0.15% to 0.29% per year (line 4). In Asia, technological progress was even slower. During this period, however, population responded less than proportionally to technological change. Although per capita GDP was improving very slowly (0.11%-0.16% in Western Europe, 0.02%-0.17% in Asia), population was already losing the race with technology.

This pattern became more apparent after 1820 in Europe: between 1820 and 1900, both the growth rates of per capita income and of population accelerated significantly. This suggests that the Malthusian mechanism whereby a higher income translates into faster population expansion was still in operation. That is, the world was still in the Post-Malthusian regime. However, the proportion of technological change that was matched with increasing population declined sharply (line 5).

Finally, the positive relationship between the standard of living and population growth vanished in the twentieth century Europe. As shown in Table 1.1, after picking up along 1820-1900, *population growth rates started declining* in Europe, even though GDP per capita was growing faster than ever. This means that the Malthusian theory of population no longer applies in this period: Europe was already engaged in *modern economic growth*.

In Asia, a similar phenomenon has occurred, though with a time lag in respect to Western Europe: the acceleration of productivity growth and of per capita income occurred in the first half of the twentieth century, only. As in Europe, this process was first accompanied by rapid population expansion, though not fast enough to prevent the acceleration of per capita income (race between technology and population). The decline in the growth rate of population in Asia only occurred in the last quarter of the twentieth century.

Table 1.1. GDP, Population and per Capita GDP, 1-2000

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

	1	1000	1500	1600	1700	1820	1900	1960	2000
29 Western European Countries									
(1) GDP									
Billions of 1990 International Geary-Khamis dollars	11	10	44	66	81	160	676	2,251	7,430
Growth Rate (% per annum)		-0.01	0.29	0.40	0.21	0.57	1.82	2.02	3.03
(2) Population									
Millions	25	25	57	74	81	133	234	326	391
Growth Rate (% per annum)		0.00	0.16	0.25	0.10	0.41	0.71	0.56	0.45
(3) Per Capita GDP									
1990 International Geary-Khamis dollars	450	400	771	890	998	1,204	2,893	6,896	19,002
Growth Rate (% per annum)		-0.01	0.13	0.14	0.11	0.16	1.10	1.46	2.57
Memo:									
(4) Total Factor Productivity (% per annum)		-0.01	0.19	0.23	0.15	0.29	1.35	1.65	2.73
(5) Population growth divided by GDP growth			0.55	0.64	0.46	0.72	0.39	0.28	0.15
Asia									
(1) GDP									
Billions of 1990 International Geary-Khamis dollars	78	82	161	217	230	413	557	1,736	13,762
Growth Rate (% per annum)		0.00	0.14	0.30	0.06	0.49	0.37	1.91	5.31
(2) Population									
Millions	174	183	284	379	402	710	873	1,687	3,605
Growth Rate (% per annum)		0.00	0.09	0.29	0.06	0.48	0.26	1.10	1.92
(3) Per Capita GDP									
1990 International Geary-Khamis dollars	449	449	568	572	571	581	638	1,029	3,817
Growth Rate (% per annum)		0.00	0.05	0.01	0.00	0.01	0.12	0.80	3.33
Memo:									
(4) Total Factor Productivity (% per annum)		0.00	0.08	0.10	0.02	0.17	0.20	1.18	4.03
(5) Population growth divided by GDP growth			0.65	0.98	1.03	0.97	0.69	0.58	0.36

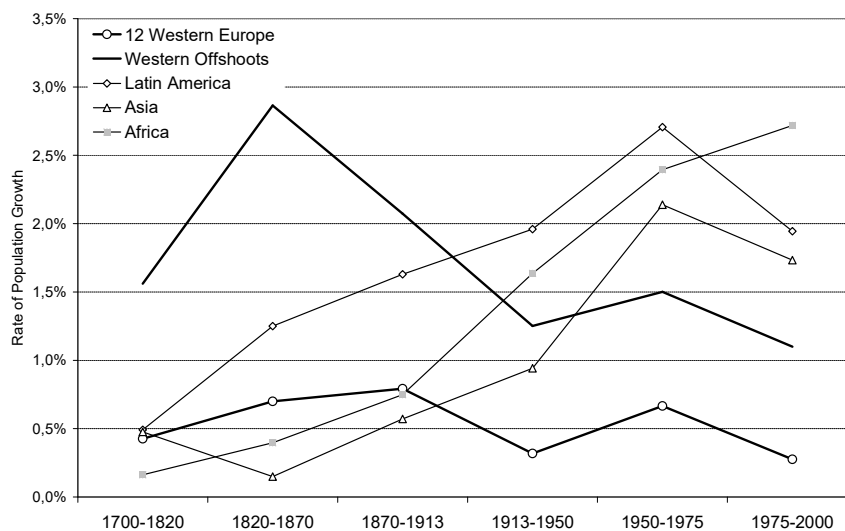
Source: (1) and (2): Maddison (2001), op. cit.. (3)=(1)/(2). Total Factor Productivity growth (4) is a measure of technological progress and is computed as a residual using $\dot{B}/B = \dot{y}/y + \beta(\dot{N}/N)$, and postulating $\beta=1/3$, that is: (4)=(3)+(1/3)*(2). Note that, since only labour is accounted for in this decomposition, the term B captures the contribution of all other eventual inputs to production.

Similar paths have been found for other regions in the world, though at distinct times. Figure 1.5 describes the evolution of population growth rates in different regions of the world along the last three centuries. As shown in the figure, population growth rates started declining by the end of the 19th century and by the beginning of the 20th century in the Western Offshoots and in Western Europe, respectively. In Asia and Latin America, the Malthusian mechanism seems to have vanished in the last quarter of the 20th century, only. In Africa, population growth rates were still increasing by the end of the 20th century.

Note that in the case of Western Europe, the fall in birth rates occurred in the 19th century, thus long before modern contraception became available. This means that, it was not availability of modern contraception but instead the willingness to have less children that marked the phenomenon of demographic transition.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Figure 1.5.- Population growth along 1700-2000



Source: Maddison (2001), op cit. Notes: The 12 Western European Countries are Austria, Belgium, Denmark, Finland, France, Germany, Italy, Netherlands, Norway, Sweden, Switzerland, United Kingdom. The Western Offshoots are Australia, New Zealand, Canada and the United States.

1.4.2 Birth rates and death rates

A question that naturally arises is what fundamental changes have occurred after the Industrial Revolution so as to reverse the relationship between per capita income and population, tilting economies towards the Modern Growth Regime. To answer this question, one shall deepen our understanding of demographic behaviour,

The process by which a country's demographic characteristics are transformed as it develops is labelled "demographic transition". To study the process of demographic transition, it is useful to introduce two demographic indicators: the birth rate and the death rate. The birth rate is defined as the number of new-borns each year per thousand of inhabitants. The death rate is defined as the number of people that die each year per thousand of inhabitants. The difference between birth rates and death rates gives precisely the growth rate of population.

In order to understand how economic development alters the relationship between living standards and population growth, one has to question how economic development impacts on birth rates and on death rates.

The impact of economic development on death rates is straightforward: death rates decline monotonically with economic development: in poor economies, death rates are high, especially among children, due to malnourishment, deficient sanitation and disease. When income increases, better nutrition, improvements in housing, public health, modern sewage, clean water, etc. cause death rates to decline (note that this is fully consistent with the Malthusian idea of “positive checks”).

Since death rates decline with economic development, death rates cannot explain why, in the Modern Growth Regime, higher per capita incomes tend to come along with lower population growth: the explanation for the demographic transition must therefore lie on *birth rates*: the decline in birth rates is the most important feature of the demographic transition.

1.4.3 Why birth rates decline with economic development?

The Malthus theory of population provides an explanation for why birth rates decline when per capita income falls below a critical level: in order to escape poverty, people refrain from giving more birth (the “preventive check”). Malthus did not account, however, the possibility of people deciding to refrain birth when per capita increases beyond a certain level.

Recent theories aiming to understand the changing attitude towards fertility have investigated the *economic incentives* underlying fertility choices¹⁵. A common aspect of these theories is that parents decide the number of offspring according to some optimal criteria. As in any economic model, that choice is formulated in terms of benefits and costs: children confer benefits to their parents, but children involve costs to parents. The nature of costs and

¹⁵ Important contributions include Becker, G., 1960. An economic analysis of fertility. In Easterlin, R. (ed), Demographic and economic change in developed countries. Universities-National Bureau Conference Series n° 11, Princeton: Princeton University Press, pp. 209-40. Becker, G., Murphy, K and Tamura, R., 1990. “Human Capital, fertility and economic growth”. Journal of Political Economy 98 (5): S12-S37. Schultz, T., 1997. The demand for children in low income countries. In Rosenzweig, M., Stark, O. (eds.) The Handbook of Population and family Economics, Amsterdam Elsevier.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

benefits varies in different models, to capture the differential role of children may have in different societies. Below, we briefly review some of these ideas.

1.4.4 Missing institutions and risk premium

A strand in the literature focuses on the *asset role* of children. The main idea is that parents may “invest” in children as a source of potential income, in particular during the old-age or in case of bad luck. Parents may decide to *invest* in the number of children, as a way of getting services. Hence, just like in our days the demand for insurance and savings increases with disposable income, one would expect the desired fertility of parents to increase in direct proportion with the respective disposable income, just like Malthus theorized.

Of course, the “investment in children” is not absent of risk. The pecuniary returns may actually prove very uncertain: a child may move to another village and decide not to look after his parents; there is uncertainty regarding the ability of each child to generate a decent income; in some societies, the earning potential depends on gender; etc. Hence, one shall expect risk averse parents, caring about the income generating potential of their total number of offspring, to decide in advance to rise more children than they would actually need, just in case.

Note that this “risk premium effect” will be more significant in a context of high mortality rates: parents decide to have more children to increase the chance of reaching a minimum number of survivals. By this mechanism, death rates have an *indirect influence* on birth rates through a risk premium effect: when mortality rates decline, birth rates will decline as well because parents will become more certain regarding the number of survivals.

In modern societies, the “asset role” of children is much less important than before. Economic development comes along with institutions that specialize in covering risks and in protection people in the old age. Workers have the opportunity to smooth consumption through the financial markets, to buy risk-protection from insurance companies and through compulsory social security systems. In such a context, children will be an expensive (dominated) form of transferring income to the future or to bad states of nature. The asset role of children became dominated and not surprisingly people responded reducing the number of offspring.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

A similar argument refers to the defence needs in small communities. Prior to the modern growth regime a small group of people could not defend a large territory from outside expropriators. Hence, there was a trade-off between diminishing returns to labour (requiring lower population) and the risks of being invaded. Social norms then emerged, creating incentives for populations to be such as to allow for the highest possible living standards, taking into account the defence needs. In modern regimes, the emergence of unified states shifted the burden of defending the territory from local communities to the central government. As the defence of the territory became less important, social norms evolved to prioritize labour productivity¹⁶.

1.4.5 Income and substituting effects

A theory of fertility alternative to the “asset view” assumes that parents derive “intrinsic pleasure” in rearing children¹⁷. In this framework, children enter in the households’ utility function as normal goods. Hence, when parents’ income increases, everything else constant, fertility will increase. On the other hand, fertility depend on the cost of rearing children: children need to be fed, clothed and schooled. When these costs increase relative to other goods, the demand for children will decline, all else constant.

In this framework, technological progress influences fertility decisions through two different channels:

- On one hand, it eases the household’s budget constraints, allowing parents to spend more resources in raising children (positive income effect). This is the pure Malthusian mechanism.
- On the other hand, it the *cost of rearing children* (negative substitution effect).

¹⁶ Parente, S. and Prescott, P. , 2005. A unified theory of the evolution of international income levels” in Aghion, P., and Durlauf, S. (eds), Handbook of Economic Growth, Volume 1, Chapter 21, pp. 1371-1416, Elsevier.

¹⁷ Becker, G., 1981. A Treatise on the Family. Harvard University Press.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

The reasons why technological progress increases the relative costs of rearing children are of two types:

First, looking after children consumes parents' time, and time involves an opportunity cost. Along the development process, the opportunity cost of time has increased. In traditional societies, where job opportunities for women were scarce and social norms dictated that women should stay at home, the opportunity cost of upbringing children was naturally low. But in modern societies, women are as engaged in the labour force as men are, at wages are higher, implying a higher opportunity cost of raising children.

Second, children must be schooled, and school requirement have increased along the development process. In the Middle Ages, people were basically devoted to agriculture, using a technology that was relatively simple and pretty stable. In such a context, children could start helping their parents at very young ages. They could acquire their skills by observing what their parents were doing. In these societies, returns to formal education were perceived to be low. In plus, in an environment plagued by high infant mortality, parents would be naturally reluctant in spending valuable resources to educate a single child: they would rather invest in *quantity*: whenever a productivity improvement eased the household budget constraints, parents would tend have more children, in a Malthusian way. In the Modern Growth Regime, in contrast, the knowledge required to operate complex machinery cannot be acquired observing what parents are doing. Technological sophistication creates a demand for technical skills, raising the returns to formal education. At the same time, child labour becomes less attractive and is socially banished. In response, parents tend to invest more in children *quality*, sending them to school. With no surprise, the departure from the Malthusian relationship between income and fertility by the turn of the 19th century in Western Europe, was accompanied by an increase in the average years of schooling.

This framework provides an interpretation for the demographic transition based on the changing balance between the income and substitution effects along time: in earlier stages of development, the first effect dominates, so population growth responded positively to the income generated by technological change, as Malthus predicted. However, the arrival of more demanding technologies gradually changed the balance between the two effects: with returns to human capital increasing, parents gradually shifted from child quantity to child quality. Then, more educated people become more likely to develop and adopt new technologies, accelerating the pace of technological progress in a virtuous cycle. At a certain

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

stage, returns to education become so high that the substitution effect became dominant in fertility decisions, explaining the entry in the Modern Growth regime¹⁸.

1.4.6 The stubbornness of birth rates

A stylized fact of demographic transitions is that birth rates tend to decline slowly, with a lag relative to death rates. Figure 1.6 illustrates this: when the level of per capita income is very low, both birth rates and mortality rates are very high, so the net population growth rate is low (bottom panel). As living standards increase, with better nutrition and health care, death rates start falling. Initially, however, the decline in death rates is not accompanied with an equally fast decline in the birth rate. Hence, in the intermediate stage, the gap between these two widens, and the growth rate of population increases. As the country gets more developed, birth rates start declining faster than death rates. This causes the growth rate of population to fall back and the economy enters in the Modern Growth regime.

Based on this stylized fact, demographers and development economists identified three stages over which the demographic transition unravels:

- 1) The first stage, pre-industrial, is characterized by high birth rates and high death rates.
- 2) The second stage is characterized by a steady decline of death rates, while birth rates remain high¹⁹. Together with 1), this implies that in this intermediate stage, population is growing faster.

¹⁸ Galor and Weil (2000), op. cit.

¹⁹ In England, for instance, the decline in mortality rate preceded that of the birth rate by 140 years. Coale, A., Treadway, R., 1986. A summary of the changing distribution of overall fertility, marital fertility, and the proportion married in the provinces of Europe. In: Coale, A., Watkins, S. (eds), *The Decline of Fertility in Europe*. Princeton University Press. Princeton.

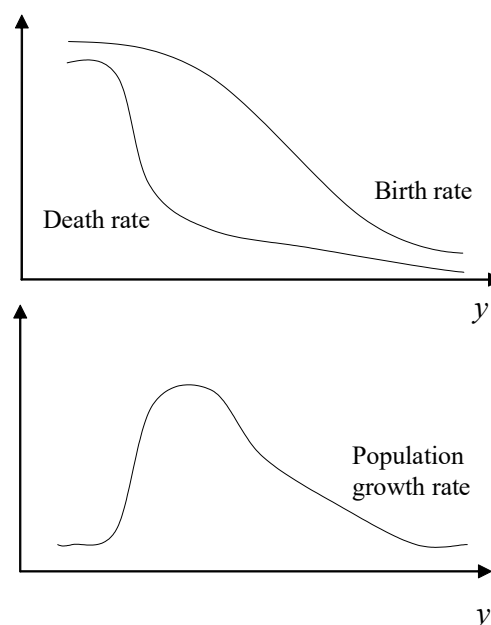
<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

- 3) In the third stage, the continuing decline in death rates is accompanied by an even faster decline in birth rates, so population expansion decelerates. In the case of Western Europe, for instance, fertility rates declined sharply by the turn of the 19th century. In Latin America and Asia this happened only in the second half of the 20th century.

Another fact of demographic transitions is that they occurred faster in the rest of the world than they had been in Europe. In the 19th century Europe, birth and death rates fell fairly gradually, accompanying the slow progresses in medicine, in sanitation, etc. Birth rates declined at a slower pace than death rates, but not so as to generate population explosions.

In more recent transition processes, in contrast, populations benefited from the fact that advances in medicine were already available from the Western World and could be copied. Thus, once the conditions in place allowed a country to take opportunity of these advances, death rates fell sharply. In result, the gap between birth rates and death rates shoot up sharply, leading to population explosions.

Figure 1.6. The stylized facts of demographic transition



The upper panel displays the stylised relationship between birth rates and death rates and per capita income. The lower panel displays the relationship between the rate of population growth (given by the difference between birth rates and death rates) and per capita income. According to this stylized view, the fact that birth rates take more time to decline than death rates imply that, in the intermediate stage, the population growth rate accelerates.

1.4.7 Why are birth rates so persistent?

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

A stylised fact of the demographic transition is that birth rates decline slowly over time, in contrast to death rates, that decline abruptly when the critical level of economic development is surpassed (figure 1.6). This phenomenon may give rise to demographic explosions. Given this, it is important to understand why birth rates decline so slowly. In fact, there are various explanations for this.

First, *economic incentives* for birth rates to decline may arrive with a lag relative to the initial improvements in living standards that cause the death rates to fall in the first place. According to this view, the arrival of a welfare state, the deepening of financial markets, the integration of women in the labour force, the raise in the value of education tend to materialize only after critical improvements in nutrition and in health care take place.

Second, family level fertility decisions are not entirely driven by private considerations: fertility decisions are also influenced by the need to conform with *social norms*: if societies demand families to give a large number of births, families desiring to conform to what is socially acceptable will refrain from reducing fertility, even if there are economic incentives to do so. Social norms evolve along time in response to economic incentives, but this process is inherently slow.

Thirds, the demographic dynamics is slow by nature, because it depends on the age structure of population, which is pre-determined each moment in time. To better understand this, note that birth rates (number of new-borns each year per thousand of inhabitants) are jointly determined by *fertility rates* (number of children per women in the reproductive age) and the structure of population (the percentage of women in the reproductive age per thousand inhabitants). Thus, even if fertility rates are already declining, the overall birth rate may remain high just because more and more young girls enter in the reproductive age. The implication for the demographic transition is straightforward: suppose that a given economy starts out with high birth rates and high death rates. Then, suddenly living standards improve causing infant mortality to decline. This means that more babies will survive childhood, causing the age structure of the population to change. With a younger age structure, the number of potential mothers in the future exceeds the existing number of mothers today. In this case, even if each new mother decides to have less offspring (that is, even if fertility rates decline responding to the low risk of child mortality), the economy' birth rate will not decline immediately because more and more women are still entering in the reproductive age. This phenomenon is known as the *population momentum*: whatever a country does, the future

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

growth rate of the population is largely determined by the existing age structure and this takes generations to change.

Box 1.6: Lant Pritchett and the Theory of States and Transitions

Along this chapter, we referred to a Malthusian Regime, a Post-Malthusian Regime, and a Modern Growth Regime. We distinguished the different attitudes towards fertility across in these two regimes and we outlined some theories attempting to explain the conditions under which a country can move from one regime to the other (Demographic Transition).

Lant Prichett (2006) offers a nice analogy for this way of structuring our thinking²⁰:

“Suppose you have a pot of water and you pick it up and turn it over. Where will the water go? The answer, that it will spill out onto the ground, is so obvious that the astute reader already realizes it is a trick question. If the water is frozen, it may stay right in the pot. If the water is vapor, then turning the pot over will trap the steam in the pot. The obvious point is that the equations of motion of water (or any other substance) depend on the state—solid, liquid, or vapor—it is in. What determines the transitions of water between states? Well, applying heat will cause water to change states, but only in a discontinuous way—water at 35° F and water at 95° F behave almost the same, while water at 32° F and at 102° F behave nothing like each other. The equations of motion of water in one state do not work at all when water is in another state, and the response of water to heat applied within a state does not work at all well when applied to transitions from one phase to another”.

Likewise – Prichett argues – “If France and Nepal can both be treated as water in a liquid state, then it is conceivable that a theory and empirics of growth that treat France and Nepal as both generic countries is adequate. I regard it as much more likely that growth dynamics are characterized as equations of motions within states and equations that

²⁰ Pritchett, L. 2006, The Quest Continues, Finance and Development, March.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

determine transitions across states” (...). “The key idea in my proposal is that economies are in different "states," and, therefore, the dynamics of output are different for economies in different states, and the dynamics of transitions between states are different from the dynamics within states”.

1.5 Globalization, fertility decisions and the Great Divergence

In the section above, we saw that under certain conditions, societies engage in a demographic transition, whereby parents’ attitudes towards fertility shift from quantity to quality. In the modern growth regime, the relationship between per capita income and the number of children is reversed, allowing living standards to improve along with technological progress. Moreover, as parents move towards child quality, the population becomes more educated and therefore more likely to experiment fast technological progress.

In the World economic history, we observe that the timing of demographic transitions differed considerably across countries and regions. At the time the Western nations was entering in Modern Growth, laggard regions were still engaged in Malthusian stages, meeting fast population expansions and sluggish per capita income: Mexico started the transition to modern economic growth during the first half of the 19th century; Japan initiated the transition in the second half of the 19th century; Brazil started in the early twentieth century, and India started its transition sometime between 1950 and 1980.²¹.

As a result of these different timings, cross-country disparities in per capita incomes increased abruptly. The fact that different countries performed the respective demographic transitions at different timings helps explain the dramatic increase in cross-country income disparities that characterized the 19th and 20th century, a phenomenon that was coined “Big divergence” (Box 1.7).

²¹ Parente and Prescott (2005), op. cit.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

A question that naturally arises is why some countries entered in the modern growth regime sooner than others. This section reviews some theories that have been proposed to address this question.

Box 1.7. The Great Divergence

Along the 19th and 20th centuries, there was a dramatic divergence in living standards across the globe. The development economist Lant Pritchett, professor at Harvard University, dubbed this period as of “Divergence, Big Time”²².

The author observed that between 1870 and 1994, one small set of countries - consisting in 12 West European countries plus 4 Western offshoots (United States, Canada, Australia and New Zealand) and Japan - managed to sustain fast economic growth, leaving the remaining regions behind.

Table 1.2: The Great Divergence

²² Pritchett, L., 1997. “Divergence, Big Time”. *Journal of Economic Perspectives* 11(3), 3-17.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

	Per Capita GDP (1990 International Geary-Khamis dollars)							Average growth rates:				
	1	1000	1500	1600	1700	1820	1960	2000	1000-1700	1700-1820	1820-1960	1960-2000
Western Europe	450	400	771	890	998	1.204	6.896	19.002	0,13	0,16	1,25	2,57
Western Offshoots	400	400	400	400	476	1.202	10.961	27.065	0,02	0,77	1,59	2,29
Latin America	400	400				692	3.133	5.838			1,08	1,57
Former USSR	400	400	499	552	610	688	3.945	4.351	0,06	0,10	1,26	0,24
7 East European Countries	400	400	496	548	606	683	3.070	5.804	0,06	0,10	1,08	1,61
Asia	449	449	568	572	571	581	1.029	3.817	0,03	0,01	0,41	3,33
16 Asian countries						581	962	3.794			0,36	3,49
26 East Asian countries						556	862	1.467			0,31	1,34
15 West Asian countries						607	2.492	5.706			1,01	2,09
Africa	430	425	414	422	421	420	1.066	1.464	0,00	0,00	0,67	0,80
World	445	436	566	595	615	667	2.777	6.012	0,05	0,07	1,02	1,95
Western Europe/ Africa	1,0	0,9	1,9	2,1	2,4	2,9	6,5	13,0				

Source: Maddison, 2001.

Notes: Western Offshoots: Australia , New Zealand, Canada, United States; 7 East European Countries: Albania, Bulgaria, Czechoslovakia, Hungary, Poland, Romania, Yugoslavia; 16 Asian countries: China, India, Indonesia, Japan, Philippines, South Korea, Thailand, Taiwan, Bangladesh, Burma, Hong Kong, Malaysia, Nepal, Pakistan, Singapore, Sri Lanka; 26 East Asian countries: Afghanistan, Cambodia, Laos, Mongolia , North Korea, Vietnam, and 20 other Small Asian Countries; 15 West Asian countries: Bahrain, Iran, Iraq , Israel , Jordan , Kuwait , Lebanon , Oman, Qatar, South Arabia , Syria , Turkey , United Arab Emirates , Yemen , Palestine and Gaza.

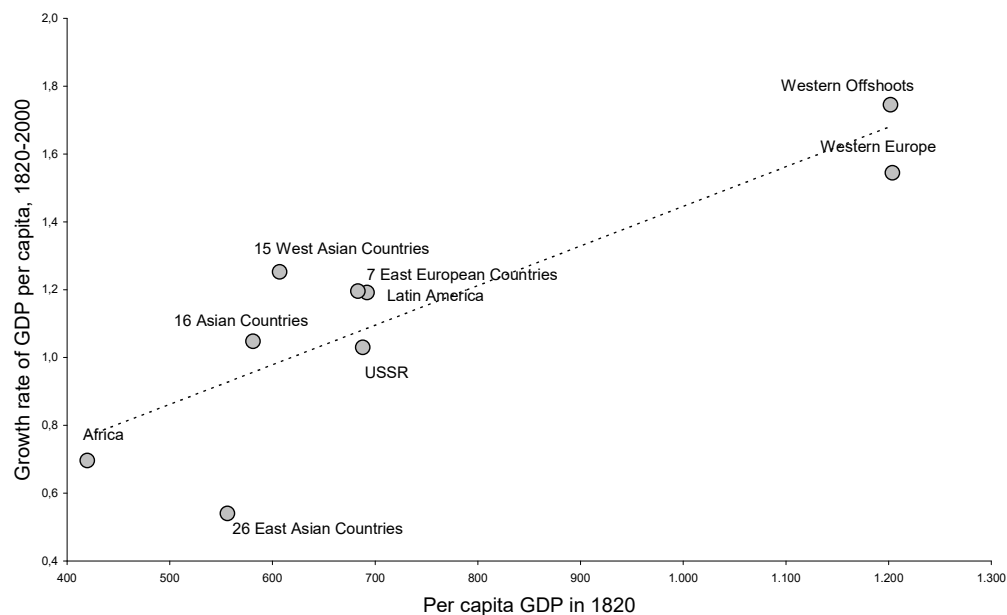
Table 1.2 illustrates the Great Divergence. The table describes the evolution of per capita incomes in some regions of the world between year 1 and year 2000. According to this data, between year 1 and year 1000, per capita income disparities remained relatively small. Starting in the 10th century, per capita income in Western Europe decoupled from those in the other regions. Still, by 1700, the ratio of per capita incomes between Western Europe and Africa was 2.4, only. Between 1820 and 2000, regional income disparities increased dramatically: per capita GDP increased 23-fold in the Western Offshoots and only 3-fold in Africa. By the year 2000, per capita income in Western Europe was 13 times higher than in Africa.

Figure 1.7 provides a graphical illustration of the Great Divergence. The figure relates the growth rates of per capita GDP to the initial levels of per capita GDP, for the period 1820-2000 (the data is the same as in Table 1.2). The positive correlation between growth and per capita incomes indicates that, along this period, initially rich countries tended to grow faster than poor countries.

The Great Divergence has vanished along the last decades. Starting in the second half of the twentieth century, a set of highly populated economies in Asia managed to accelerate their rates of economic growth, above the levels observed in the developed world. As shown in Table 1.2, between 1960 and 2000, Asia expanded much faster than Europe and the Western Offshoots. Given the population size of the converging countries in Asia, the global picture became of convergence.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Figure 1.7 The Great Divergence



Source: same as Table 1.2.

1.5.1 Industrialization and the demographic transition

The entry in Modern Growth Regime in both developed and less developed regions has been associated to a reallocation of resources from agriculture to manufactures. In the particular case of Western Europe and the Western Offshoots (Australia, Canada, New Zealand, United States), this coincided with the Industrial Revolution, which began at the end of 18th century in England and spread to other countries along the 19th century. Other countries that managed to catch up also experienced fast industrialization (for instance, Japan, South Korea, and Singapore). In general, the historical evidence points to a declining share of agricultural employment in total employment along the process of economic development²³.

²³ Clark, C., 1940. *The conditions of Economic Progress*. London: McMillan. Kuznets, S., 1966. *Modern Economic Growth*. New Haven, CT: Yale University Press. Chenery, H., Syrquin, 1975. "Patterns of Development, 1950-1970. London: Oxford University Press..

This move towards manufactures has a key role in demographic transition, because technological progress tends to be faster in manufactures than in agriculture. The acceleration of technological progress that comes along with industrialization generates an increased demand for skilled labour and theoretical knowledge, raising the returns to education and leading parents to alter their choices over their children education. In response, societies press their governments to introduce universal schooling and to ban child labour. As educational reforms succeed in inducing more children to engage in formal educational, fertility rates decline and technological change accelerates.

1.5.2 From the Agriculture Revolution to the Industrial Revolution

A natural question that arises is why some countries were able to industrialize first than others. This question is, of course, too complex to be answered with a simple theory. At this stage, however, it is useful to introduce a conventional wisdom, largely inspired in the 18th century England, according to which the modernization of agriculture played a key role in the industrialization process.

Positive links between agriculture and industry may arise at different levels: first, rising productivity in agriculture makes it possible to release workers from agriculture to industry; second, food surpluses are necessary to feed a large urban population; third, a raising income in agriculture creates a natural market for industrial products; finally, savings generated in agriculture may be used to finance investment in industry. Based on this idea, classical development theorists defended that agricultural (green) revolution is a *precondition* for industrialization²⁴. Although in our days this is no longer true, it was true for that particular period.

²⁴ Nurkse, R., 1953. "Problems of capital formation in underdeveloped countries". New York: Oxford University Press. Rostow, W., 1960. "The stages of economic growth: a non communist manifesto". Cambridge: Cambridge University Press.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

The link between technological change in agriculture and industrialization can be illustrated with the help of a simple model²⁵. Assume that there are two goods, Manufactures (Y) and Agriculture (Z) and that (homogeneous) labour is the only input to production. To simplify the algebra, it is assumed that both production functions are linear, and the total labour force in the economy is equal to 1:

$$Y = AN_Y \quad (1.9)$$

$$Z = BN_Z = B(1 - N_Y) \quad (1.10)$$

where N_Y and N_Z denote for the fractions of the labour force employed in manufactures and agriculture, respectively. Wages are assumed flexible.

The last term in equation (1.10) incorporates the resource constraint of the economy and stresses the trade-off between agricultural production and manufactures production: given the productivity parameters, A and B, the only way to expand production in agriculture is by reallocating employment from manufactures. Moving one worker from manufactures to agriculture causes the agriculture output to expand by B and the manufactures output to fall by A. Hence, the *opportunity cost* of expanding the production of Z by one unit is A/B units of Y.

Obviously, a productivity improvement in agriculture (raise in B) causes an expansion of agricultural output, Z. A difference question is whether it leads to an expansion of output in manufactures, Y. For this to happen, some workers must be reallocated from agriculture to manufactures. Such reallocation is at the centre of the argument that that modernization of agriculture is a pre-condition for industrialization.

The reallocation of labour from agriculture to manufactures is not entirely obvious: when agricultural production expands, the relative price of agricultural goods falls, so in

²⁵ The model adapts from Matsuyama, K., 1992. "Agriculture productivity, comparative advantage and economic growth". Journal of Economic Theory 58, 317-334.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

theory consumers could end up absorbing the increased production of agriculture without any extra demand for manufactures.

In the real world, however, one does not expect the demand for agriculture goods to increase that much. The reason is an indisputable statistical regularity, which tells us that, as households' income increases, the fraction of households' income spent in food tends to decline. This is the Engel's Law, owing its name to a 19th century German statistician called Ernst Engel. Such law can be incorporated in the model, postulating a utility function of the form:

$$U = \ln(Z - \bar{Z}) + \ln Y \quad , \quad (1.11)$$

where $\bar{Z}$ refers to a minimum subsistence level of agricultural consumption. When positive, this parameter implies that the income elasticity of agriculture goods is less than unitary. Substituting (1.9) and (1.10) in (1.11) and maximizing in respect to N_Y , one obtains the level of employment in manufactures that balances the demand and supply:

$$N_Y = \frac{1}{2} \left(1 - \frac{\bar{Z}}{B} \right) \quad . \quad (1.12)$$

This equation states that, when $\bar{Z} > 0$ (and only in this case), a productivity increase in agriculture (B) leads to an increase in the share of employment in manufactures. This is the result one wanted to obtain. The intuition is the following: an exogenous increase in agricultural productivity leads to an increase in per capita income; then, because of the Engel law, the relative demand for manufactures rises, implying a reallocation of labour away from agriculture towards manufactures. This captures the conventional wisdom that “modernization in agriculture is a precondition for industrialization”²⁶.

²⁶ This is a demand side argument. Hansen and Prescott (2002) propose instead a supply side theory, whereby it is the rising productivity in manufactures that makes it progressively more attractive for workers, tilting the departure of the Malthus model. [Hansen, G. and Prescott, E., 1999. “Malthus to Solow”, American Economic Review 92 (2002), 1205-17].

Note that the optimal employment shares in this economy are unaffected by changes in manufactures' productivity: an increase in A leads to an increase in Y , but the relative price of manufacture goods falls proportionally, so the demand for Z remains unchanged. Hence, technological improvements in manufactures do not cause agricultural employment to increase.

1.5.3 International trade and industrialization

The model (1.9)-(1.12) implies that an increase in agriculture productivity leads to an expansion of employment in manufactures. This is indeed what happened in Britain prior to the Industrial Revolution: the innovations introduced in agriculture caused agriculture output to expand, giving rise to positive income effects that led consumers to spend higher fractions of their income in manufactures, allowing the economy to industrialize and undergo the demographic transition.

History is however plenty of examples of countries with strong agriculture, such as Argentina, that failed to industrialize, as well as of countries with poor natural resources, such as Japan, that successfully industrialized. An explanation for this apparent paradox is that the experiences of these countries differed from that of Britain in respect to the *timing relative to globalization*²⁷.

Indeed, from the Seven Years War in 1756 until the beginning of the nineteenth century, England could be seen mainly as a closed economy. In that case, the model above applies. During the nineteenth century, however, there was a significant expansion of international trade. Even though imports were restricted, the expansion of agriculture in Argentina and the industrialization of Japan occurred at a time where globalization was already under way.

²⁷ Matsuyama (1992), op. cit.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

The key point is that, when an economy is open to international trade, the relationship between agricultural productivity and employment in manufactures is the inverse of that in a closed economy: high productivity in agriculture increases the likelihood of a country having *comparative advantages* in agriculture, in which case trade openness implies a specialization in agriculture goods and a reallocation of the labour force away from manufactures.

To see the argument formally, note that the main difference between the closed economy and the open economy concerns the determination of relative prices: in an open economy, prices are determined according to the world demands and supplies. If the domestic economy is small, prices are exogenous.

In terms of our model, let $p = P_Z/P_Y$ be the relative price of agriculture goods in terms of manufactures in the global economy. Under free trade and no frictions, the specialization pattern will be determined according to the Law of Comparative Advantages. This is equivalent to choosing the allocation of workers that maximizes the value of domestic income at world prices, given by:

$$pZ + Y = pB(1 - N_Y) + AN_Y \quad (1.13)$$

With a simple derivative with respect to N_Y you may verify that this expression is an increasing function of N_Y when $p < A/B$ and a decreasing function of N_Y when $p > A/B$. Interpreting, this means that: if the relative price of the agriculture good in the world economy is lower than the opportunity cost of producing the agriculture good in the domestic economy (that is, the domestic economy has *comparative advantages* in manufactures), then it will be optimal to expand the employment in manufactures until $N_Y = 1$; when instead the relative price of the agriculture good in the world economy is higher than the opportunity cost of producing the agriculture good domestically (the domestic economy has comparative advantages in agriculture), then it will be optimal to reduce employment in manufactures until $N_Y = 0$. Note that what matters for comparative advantages is relative productivities, A/B , not their absolute levels. That is, a country with lower productivity in manufactures than the rest of the world (low A) may still have the chance of industrializing just because it is even less productive in agriculture.

Using this reasoning, it is easy to understand why a country like Argentina failed to industrialize: in a context of trade openness, the Argentinean high productivity in agriculture induced the country to specialize in agriculture goods, retarding the industrialization process.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

In contrast, for a country like Japan, where the quality of the land is poor, it became profitable to specialize in manufactures: the low productivity in agriculture endowed the country with an abundant supply of “cheap labour” that the manufactures sector could use.

1.5.4 Trading population for productivity

Putting the pieces together, the discussion above suggests that the rapid expansion of international trade in the 19th century lead some countries to specialize in manufactures while others specialized in agriculture goods. This, in turn, may have influenced the different timing of demographic transition across countries, affecting persistently the distribution of the world population, human capital and technology. In other words, this provides an explanation for the Great Divergence.

The argument runs as follows: By the end of 19th century England and Northwest Europe became net exporters of manufacture goods and net importers of primary products, where the exports of Asia, Oceania, Latin America and Africa were overwhelming composed of primary products. In Western countries, the increasing demand for skilled labour induced by the specialization pattern caused the respective societies to press governments for educational reforms, expediting the demographic transition. With more educated populations, Western countries met faster technological progress, which further enhanced their comparative advantages in skilled intensive industries. Since these countries had escaped the Malthusian trap, the fast technological development translated into increasing living standards.

In non-industrial economies, on the contrary, international trade induced specialization in unskilled intensive goods. This generated incentives to invest in child quantity, delaying the demographic transition. In these countries, the gains from trade were channelled towards population expansions – a la Malthus –, without impacting significantly on living standards. Moreover, the growing abundance of unskilled labour reinforced the

comparative advantages in unskilled intensive products, in a vicious cycle. Galor and Mountford dubbed the emergence of North-South trade in this period as “trading population for productivity”²⁸.

The case of India provides a real life example of how specialization according to comparative advantages may have end up retarding the demographic transition. Between 1813 and 1850, India increased its trade relations with England. This opening process turned India from an exporter of manufactured goods (mainly textiles) into a supplier of primary commodities (between 1800 and 1913, industrialization in India declined by 2/3²⁹). In India, this implied a low demand for skilled workers, reducing the incentives for investment in education and delaying the demographic transition. The gains from trade were mostly channelled towards increasing population, without a significant impact on living standards. In England, on the contrary, the gains from trade were channelled towards investment in education stimulating faster technological change and faster economic growth.

1.5.5 Static and dynamic gains from international trade

The argument above suggests that the taking opportunity of comparative advantages achieving the highest possible efficiency in the short-run may not go along with welfare improving in the long run. This raises the question as to whether a poor country with initial comparative advantage in agriculture should instead impose trade restrictions on manufactures imports, rather than to engage in free trade. If that strategy helped the country to industrialize, the policy could eventually accelerate the demographic transition and tilt the economy away of the Malthusian trap. This reasoning is an incarnation of a well known

²⁸ Galor, O., Mountford, A., 2006. "Trade and the Great Divergence: The Family Connection," American Economic Review, American Economic Association, vol. 96(2), pages 299-303, May. Galor, O., Mountford, A., 2008. "Trading Population for Productivity: Theory and Evidence," Review of Economic Studies, Blackwell Publishing, vol. 75(4), 1143-1179, October.

²⁹ Bairoch, P., 1982. International Industrialization levels from 1750-1980. Journal of European Economic History 11 (2), 269-333.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

proposition in the theory of international trade and development, that “static efficiency and dynamic efficiency do not necessarily go along”.

1.6 Key ideas of Chapter 1

- The Law of Diminishing Returns (LDR) has an important role in the theory of economic growth. The Malthusian model provides a simple illustration of that role. In this model, a growing labour force leads to a more intensive use of land and thereby to a decline in per capita income. At the moment household incomes fall short a minimum subsistence level, both population and output stop growing.
- The Malthus prediction that technological improvements should translate into higher population densities without much impact on living standards provides a reasonable description of the real world facts prior to the modern era.
- The model also provides a useful tool to think about contemporaneous problems of depletion of natural resources and environmental sustainability. These problems can also be formulated in the context of a race between technology and the limitations imposed by resource constraints.
- The Malthusian model fails, however, to describe modern economic growth. On one hand, the model conflicts seriously with the stylised fact that, in modern economies, per capita incomes exhibit a tendency to growth over time, not to remain constant at the very low subsistence level. On the one hand, its predictions regarding the relationship between population growth and per capita income no longer hold in contemporaneous societies.
- The change in the human behaviour towards fertility along the process of economic development is labelled “Demographic Transition”.
- The fall in birth rates is the most important feature of demographic transition. Microeconomic theories explaining the changing attitude towards fertility view the number of offspring as being determined by individual optimization. Along time, the cost of rearing children increased relative to benefits, resulting in a reduction in the number of optimal offspring.
- Social norms and the population momentum help explain why changes in birth rates entail a lot of inertia.
- The interaction between globalization, industrialization and attitudes towards fertility may help explain the Great Divergence: countries with comparative advantages in agriculture remained basically trapped in the Malthusian regime. In countries with comparative advantages in manufactures, societies felt the pressure to switch from child quantity to child quality, investing more in education and achieving faster technological change, in a virtuous cycle.

1.7 Review questions and exercises

Key concepts

- *Positive checks vs. preventive checks. Stable vs. unstable steady state. Race between technology and diminishing returns. Stepping on shoes, standing on shoulders, fishing out effects. Post-Malthusian regime vs Modern Growth Regime. Demographic transition. Asset role of children. Population momentum. The Great Divergence. Static versus dynamic efficiency. Trading population for productivity.*

Essay questions:

- Comment: “The most decisive mark of the prosperity of any country is the increase in the number of its inhabitants”.
- Comment: “Technology and population reinforce each other”.
- By the 15th century, population density in Eurasia was much higher than in Australia. Explain.
- What drives the fall in fertility rates in the transition to the Modern Growth regime?
- Explain why birth rates exhibit a long persistence, declining much slower than death rates.
- Explain why high productivity in agriculture favoured the demographic transition in England but not in Argentina.

Exercises

- 1.1. Consider a closed economy with no government and basically devoted to agriculture. Output takes the form of a single homogeneous good (Y), which is produced using labour (N) and land (T). The relationship between inputs and output is described by an aggregate production function of the form: $Y_t = BT_t^{0.5} N_t^{0.5}$. Assume that the availability of land is fixed, with $T=1$. The dynamics of population (N) is described by the following equation: $\dot{N} = \nu[y - \bar{y}]$. (a) Where ν is a positive parameter, $y=Y/N$ and $\bar{y} = 2$ is the subsistence level of per capita income. Assume initially that $B=18$. (b) Explain the equation that describes the dynamic of population in this economy. (c) Find out the steady state of the model and represent it in a graph. Is this steady-state stable? (d) Suppose now that the discovery of a new fertilizer improves B from 18 to 20. Following this change, will the population expand indefinitely? Why? What happens to the population density, N/T ? € Suppose instead that B was expanding continuously at a rate of 2% per year? Would population expand at 2% per year as well? Why? What if ν was very small?

- 1.2.** Consider an isolated Malthusian economy (Alfa), where $Y_t = BT_t^{0.5} N_t^{0.5}$, $T=4$, $\bar{y} = 10$, and $B=100$ (exogenous). (a) (a1) Assuming that, initially $N=256$, how much will be per capita income in that year? (a2) How will the economy evolve onwards? Explain the theory for the dynamics of population. (a3) Describe the steady state for population and population density. (a4) Represent in a graph. (a5) Discuss the stability of the equilibrium. (b) Consider an economy (Beta), also isolated, identical to Alfa except in that $T=1$. (b1) With $B=100$ and $N=256$, will per capita incomes in the two economies converge? Quantify. (c) Suppose that technology evolved very slowly, as a function of the size of population? Would the two economies converge? Discuss.
- 1.3.** Consider a closed economy devoted to agriculture, where the aggregate production is $Y_t = BT_t^{0.75} N_t^{0.25}$, where initially $T=4$, and $B=8$. The population dynamics can be described by the following equation $N_t - N_{t-1} = 100[y_{t-1} - \bar{y}]$ where $\bar{y} = 1$ is the subsistence level of per capita income and t is a time subscript for centuries. (a) Find out the steady state in this economy, (N^*, y^*) and represent in a graph. Is it a stable steady state? Explain. (b) Suppose that, at moment $t=1$, some swamps were drained, so that the arable land expanded by 4%. How much would be per capita income in that century? And population in the century after? How much would be population, per capita income and population density in the long run? Represent in a graph. (c) Assume that, instead of exogenous, technology was a function of the last century's population $B_t = 0.125N_{t-1}$. (c1) Explain the intuition; (c2) Explain what would happen to technology, per capita income and population in the years that followed the swamp drainage. (c3) would the economy face any Malthusian barrier again? (d) To which extent does this model help explain real world facts?
- 1.4.** Consider a closed economy in the Malthusian regime, where the aggregate production is $Y_t = BT_t^{0.5} N_t^{0.5}$. In this economy, technology is a function of the last century's population $B_t = 0.01N_{t-1}$. The population dynamics can be described by the following equation $N_t - N_{t-1} = 250[y_{t-1} - \bar{y}]$ where $\bar{y} = 1$ is the subsistence level of per capita income and t is a time subscript for centuries. Initially, $T=100$, and $N=100$. Describe what will be the impact of an increase in the availability of land to $T=125$. Complete the following table:

Time	t-1	t	t+1	t+2
Land (T)	100,00	125,00	125,00	125,00
Technology (B)	1,00			
Population (N)	100			
Per capita income (y)				

- 1.5.** Consider the following production function for the World economy: $Y_t = B_t T$, where T refers to a fixed amount of land, and $\dot{B} = bN_t^\theta B_t^\varepsilon$, and $0 \leq \varepsilon \leq 1$, $0 \leq \theta \leq 1$. Further

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

assume that population expands whenever per capita income $y = Z/N$ increases above the subsistence level, $\bar{y}$. Choose one: (a) A small ε means: (a1) a sizeable “stepping on shoes” effect; (a2) a negligible fishing out effect; (a3) a large standing on shoulders effect; (a4) none of the above. (b) This model will display a strong scale effect when: (a1) $\varepsilon = 0$, and $\theta = 1$; (a2) $\varepsilon = 1$ and $\theta = 0$; (a3) $\varepsilon = 1$ and $\theta = 1$ (a4) $\varepsilon = 0$ and $\theta = 0$. (c) In the long term, the model will display exogenous growth when: (a1) $\varepsilon = 0$, and $\theta = 1$; (a2) $\varepsilon = 1$ and $\theta = 0$; (a3) $\varepsilon = 1$ and $\theta = 1$ (a4) $\varepsilon = 0$ and $\theta = 0$.

- 1.6.** Consider an economy where people live two periods. In the first period, people are young, they work, they have children and they support their parents. In the second period, people are too old to work or to have children, so they need assistance from their children to sustain their consumption needs. Each family is only concerned with its lifetime utility function, given by $U = \ln c_t + \ln c_{t+1}$. Further assume that: family income during the working age period is equal to 10 monetary units; the cost of rising children is equal to 1; each child delivers 1 unit of its income to his parents in the old age. There are no social security or capital markets. (a) Formalize the utility maximization problem of a period-1 individual. Write down the intertemporal budget constraint. (b) Describe the welfare gains associated to the fact that people can have children. Use a graph to illustrate your answer. (c) What happens to the optimal choice when the family income increases from 10 to 12? Show in a graph. Explain how this relates to the Malthusian theory. (d) What happens if the cost of rising children increases from 1 to 1.25? € Show that the problem above is equivalent to that of a static optimization in which individuals derive an intrinsic utility from having children. (e) Suppose now that banking services become available, so that households could borrow or lend any amount of money at a zero interest rates. Would children still be a profitable investment?
- 1.7.** The following table illustrates the “demographic momentum”. Initially, the population is stable with a fertility rate equal to 2. The number of fertile women in each generation is equal to half of the new-borns 30 years before, and the death rate is stable at 1/3. In year zero the fertility rate jumps temporarily from 2 to 3. (a) Explain how the birth rate is determined in the model, (b) Explain why the temporary shock in fertility at year 0 produces lasting effects in the birth rate for more than two centuries.

Year	-60	-30	0	30	60	90	120	150	180	210	240
Initial Population	100.0	100.0	100.0	116.7	127.8	135.2	140.1	143.4	145.6	147.1	148.0
Fertile women	16.7	16.7	16.7	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0
Fertility Rate	2	2	3	2	2	2	2	2	2	2	2
Birth Rate	33%	33%	50%	43%	39%	37%	36%	35%	34%	34%	34%
Death Rate	33%	33%	33%	33%	33%	33%	33%	33%	33%	33%	33%
Population growth	0.0%	0.0%	16.7%	9.5%	5.8%	3.7%	2.3%	1.5%	1.0%	0.7%	0.4%

- 1.8.** Consider an economy where two goods, Manufactures (Y) and Agriculture (Z), are produced using labour input, only. For simplicity, assume that the total labour force in the economy is equal to 100 and that both production functions are linear in labour: $Y = BN_Y$; $Z = AN_Z$. (a) Find out the expression for the production possibilities frontier. Display it in a graph, assuming that the productivity parameters are $A=1/2$ and $B=1$. In

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

that case, what will be the opportunity cost of expanding one unit of manufactures output? (b) Assume now that we are dealing with a closed economy and that the utility function is given by $U = \ln(Z - \bar{Z}) + \ln Y$, with $\bar{Z} = 40$. (b1) Interpret the parameter $\bar{Z}$. (b2) Find out an expression relating the equilibrium level of employment in manufactures with the parameter A. (b3) Assume that this country experiments an agricultural revolution, with the productivity parameter A shifting from $A=1/2$ to $A=1$. Explain what happens to employment in manufactures. (c) Suppose that productivity in manufactures evolves according to $B_t = 0.1N_{Y,t-1}$. (c1) Interpret this rule. (c2) What happens to productivity in manufactures after the agricultural revolution? (c3) Does employment in manufactures change at all? Why? (d) Assume now that two countries, say England and India, engage in international trade. Assume that, before openness, England experimented an agrarian revolution as described in b) and the implied transformation, as described in c). India, on the contrary, was still in the first stage ($A=1/2$ and $B=1$): (d1) Has England absolute advantages in agriculture? (d2) Has England comparative advantages in agriculture? (d3) Admitting that both economies open up to international trade, how will employment evolve in both countries? (d4) Taking into account rule (c), are comparative advantages likely to change in the future?

2 The basic Solow model

“A thrifty society will, in the long run, be wealthier than an impatient one, but it will not grow faster” [Robert Lucas Jr.]

Learning Goals:

- Understand the extent to which capital accumulation can overcome the law of diminishing returns
- The mechanics of the Solow growth model.
- Evaluation of the Solow model through the lenses of the Kaldor facts.
- Distinguish the effects of a change in the saving rate from those of an exogenous change in technology
- Discuss the optimality of the saving rate in the context of the Solow model
- Growth accounting

2.1 Introduction

The neoclassical theory of economic growth was pioneered by two independent economists, the American Robert Solow and the Australian Trevor Swan³⁰. The main innovation of the Neoclassical model in respect to the Malthusian Model is the replacement of land by “capital” in the production function. By “capital”, we mean machinery, buildings, and other equipment. This modification is more than a mere change in form: contrary to land, which is available in finite supply, capital can be produced and accumulated. This opens an avenue to overcome diminishing returns on labour: by allowing the capital stock to expand over time, the Solow model avoids the tendency for productivity to decrease with the size of

³⁰ Solow, R., 1956. “A contribution to the theory of economic growth”, *Quarterly Journal of Economics* 50, 65-94. Swan, T., 1956. “Economic growth and capital accumulation”, *Economic Record* 32, 334-61.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

population that plagues the Malthus model. In the Solow model, the long run is not characterized by a low-income equilibrium trap. Still, because capital itself faces diminishing returns, capital accumulation alone cannot generate long-run growth: in the Solow model, a higher investment rate translates into a higher level of per capita income in the long run, but it does deliver sustained growth of per capita income.

This chapter presents the Solow model in its simplest formulation. In Section 2.2, we describe the main assumptions of the model and we characterize the equilibrium. In Section 2.3, we confront the predictions of the model with the main stylized facts of Modern Economic Growth. In Section 2.4 we discuss how the main endogenous variables of the model respond to changes in the exogenous parameters. Section 2.5 addresses the trade-off between current and future incomes underlying saving decisions. Section 2.6 extends the model to the case in which the saving rate is optimally decided by individual agents. Section 2.7 introduces a growth accounting exercise to show what the Solow model cannot explain. Section 2.8 concludes.

2.2 The Solow model

2.2.1 The neo-classical production function

In the Malthus model, with all else constant, an increase in the size of the labour force leads to a decline in per capita income. This is a direct consequence of the Law of Diminishing Returns: as long as the availability of land is unchanged, its higher intensive use will translate into lower labour productivity. The Solow model retains the assumption of diminishing returns to labour, but explores another property of the neoclassical production function: Constant Returns to Scale (see box). Under constant return to scale, it is possible for per capita income to remain constant over time, provide capital and labour expand at the same rate.

In the following, we consider the following specification for the economy' aggregate production function³¹:

$$Y_t = AK_t^\beta N_t^{1-\beta}, \quad 0 < \beta < 1 \quad (2.1)$$

where K denotes for the economy' capital stock, N for the size of the labour force, Y for output and A is Total Factor Productivity. Dividing the production function (2.1) by the size of the workforce (N), one obtains an expression that relates per capita output to capital per worker:

$$y_t = A\left(\frac{K}{N}\right)^\beta = Ak_t^\beta, \quad (2.2)$$

In (2.2), $y=Y/N$ denotes for per capita income and $k=K/N$ is the capital-labour ratio. Equation (2.2) is called the production function in the *intensive form* and stresses the role of the capital-labour ratio as a key driver of per capita income. Thus, as long as capital and labour expand at the same rate, the capital labour ratio will remain constant and so will do per capita income.

The interesting feature of the Solow model is that we do not need to postulate K to grow at the same rate as the labour force. As we will see next, the properties of the model are such that the capital stock, even if endogenous, will end up growing at the same rate as the labour force in the long run, ensuring that per capita income remains constant.

Box 2.1 Constant returns to scale

The key assumption of the neoclassical growth model is that of constant returns to scale (CRS). CRS means that if one increases the use of all inputs by a given proportion, output will rise in the same proportion. For instance, duplicating the use of labour and capital

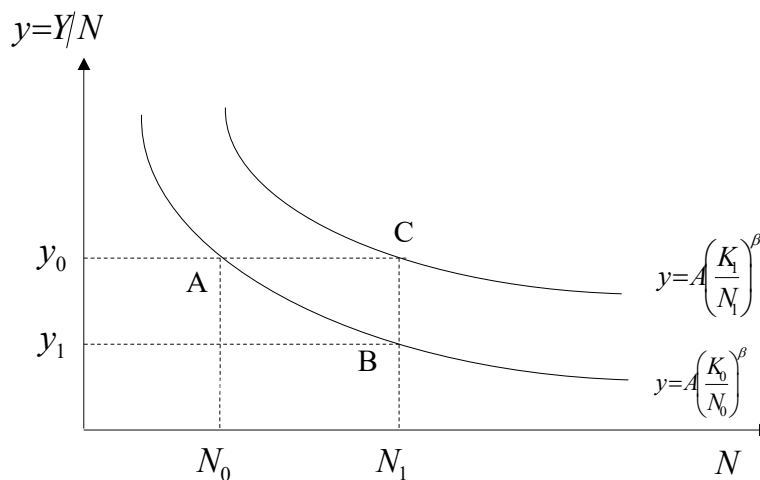
³¹ The Solow model is consistent with more general specifications for the production function. This simple formulation is assumed for mathematical convenience.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

we will obtain the double of output. The CRS property can easily be checked in equation (2.1): for any $q \neq 0$, $A(qK)^{\beta}(qN)^{1-\beta} = qY$.

Note that CRS is not inconsistent with the LDR: the Law of Diminishing Returns states that output will expand less than proportionally than one varying input, when all others are held constant. The Constant Returns to Scale property applies when *all* inputs vary in the same proportion at the same time.

Figure 2.1. Diminishing returns versus constant returns to scale



The Figure displays the average product of labour as a function of the size of the labour force. The curve is negatively sloped because of diminishing returns (from A to B). When the capital stock increases, the schedule shifts to the right. If the capital stock and population increase in the same proportion, the average product of labour remains constant (from A to C).

To illustrate the difference, we refer to Figure 2.1. The figure displays the average product of labour as a function of the labour force. The schedule is negatively sloped because of the Law of Diminishing Returns: all else constant (including the capital stock), an increase in the use of labour from N_0 to N_1 implies a decline in the output per worker from y_0 (point A) to y_1 (point B). This is the main mechanism in the Malthusian model. If, in alternative, when labour expands from N_0 to N_1 , the capital stock also increases, and exactly in the same proportion (such that $K_0/L_0 = K_1/L_1$), then the capital labour ratio remains constant and so will be per capita income, y_0 .

In general, production functions may exhibit decreasing returns to scale (when output grows less than proportionally than all inputs) and increasing returns to scale (when output grows more than proportionally than all inputs). Decreasing returns to scale in its pure form is <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

unlikely: if we managed to increase all inputs in a given proportion, there should be no reason for output to respond less than proportionally. The case with increasing returns is trickier and will be addressed later in this book.

2.2.2 Main assumptions of the Solow model

Consider a closed economy with no government with a large number of small firms producing a single homogeneous good, Y , using two inputs: labour (N), and capital (K). Population and the labour force are the same. Inputs are hired from households, who are also the owners of the firms and the consumers in this economy. Households spend a fraction $(1-s)$ of their income on consumption, and save the remaining fraction, to acquire new capital. The capital stock depreciates at a constant rate, δ . Perfect competition and flexible prices are assumed, so full employment holds each moment in time. Population is instead assumed to expand at some exogenous rate, n ³².

The production function of the representative firm i takes the following form:

$$Y_{it} = A_i K_{it}^\beta N_{it}^{1-\beta}. \quad (2.3)$$

where Y_i , K_i and N_i refer to output, capital and labour employed by each firm. This production function exhibits Constant Returns to Scale in labour and capital and decreasing marginal returns to each of these inputs. The parameter A stands for the level of technology, which is exogenous to each firm. Throughout this chapter it will be assumed that the level of technology is constant over time:

$$A_i = A. \quad (2.4)$$

³² Since in this model there is no tendency for per capita income to approach a low-level subsistence trap, we don't need to worry with a Malthusian barrier limiting population growth. In alternative, we abstract from fertility decisions, assuming that the growth rate of population is exogenous.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Assuming that all firms are equal and face the same relative prices, the aggregate production function in this economy will be (2.1), where $Y = \sum_i Y_i$, $K = \sum_i K_i$ and $N = \sum_i N_i$.

Households use income to consume or to save. In the aggregate, this implies:

$$Y_t = C_t + S_t \quad , \quad (2.5)$$

where, C and S denote for aggregate consumption and aggregate savings, respectively.

The Solow model combines the neoclassical assumptions of perfect competition and flexible prices, with a Keynesian consumption function, whereby a constant fraction of income is allocated to savings. In particular, it is assumed that³³:

$$S_t = sY_t \quad (2.6)$$

Given the constant saving rate, s , the flow equilibrium in this economy is given by:

$$sY_t = I_t \quad (2.7)$$

where I denotes gross investment³⁴. By investment, we mean acquisition of real assets, such as machinery, buildings and other equipment with the aim to achieve higher production in the future³⁵.

³³ Below, we argue that turning the savings rate a choice variable does not alter the properties of the equilibrium.

³⁴ Equation (2.7) implicitly postulates the price of the capital good to be the same as that of output. As an example, think that the only output in this economy was potatoes: potatoes can be either consumed or planted (invested) to grow more potatoes. If however total investment included a plough, in that case a given amount of saved output (potatoes) would translate into more or less capital accumulation (ploughs), depending on how many potatoes would be necessary to buy a plough. The implications of changing the relative price of capital are examined in other chapters.

³⁵ This shall be distinguished from purchases of financial assets, such as bonds or shares, with the aim to obtain future returns. A confusion may arise in that the later is often labelled as *financial investment*.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

With the passage of time, capital wears out or becomes obsolete. This process is called *depreciation*. In this model, it is assumed that the depreciation rate (δ) is exogenous and constant over time. Hence:

$$\dot{K}_t = I_t - \delta K_t \quad (2.8)$$

Equation (2.8) states that the change in the capital stock (net investment) is equal to gross investment minus depreciation. The fact that the capital stock depreciates over time implies that some minimum investment will be needed just to avoid the erosion of the existing capital stock.

Finally, the population growth rate is:

$$n = \dot{N}_t / N_t \quad (2.9)$$

The above equations describe the basic Solow model.

2.2.3 Factor prices and factor income shares

We assume that households are the owners of capital, which they acquire with the product of their savings. There is also a debt market, whereby households can lend to other households at the interest rate r , but since the economy is closed, the assets and liabilities emerging from that trade cancel out in the aggregate. Households rent labour and capital to firms. Labour services are paid at the wage rate w . Capital services must be paid the opportunity cost of holding capital, plus a compensation for the fact that capital erodes over time, δ . The sum $r + \delta$ is the “user cost of capital”.

It is assumed that firms are price-takers both in the product market and in factor markets. When this is so, we know that profit maximization delivers demands for inputs equal to the respective marginal products. Formally, the profit function of each individual firm is given by:

$$\pi_{it} = Y_{it} - (r_t + \delta)K_{it} - w_t N_{it} \quad (2.10)$$

The first order conditions of profit maximization problem are:

$$\frac{\partial \pi_i}{\partial N_i} = (1 - \beta)K_i^\beta N_i^{-\beta} - w_t = (1 - \beta)\frac{Y_{it}}{N_{it}} - w_t = 0$$

$$\frac{\partial \pi_i}{\partial K_i} = \beta K_i^{\beta-1} N_i^{1-\beta} - (r_t + \delta) = \beta \frac{Y_{it}}{K_{it}} - (r_t + \delta) = 0$$

Since firms are all alike, this leads to the following aggregate demand functions for labour and capital, respectively:

$$w_t = (1 - \beta) \frac{Y_t}{N_t} \quad (2.11)$$

$$r_t + \delta = \beta \frac{Y_t}{K_t} \quad (2.12)$$

2.2.4 Factor income shares

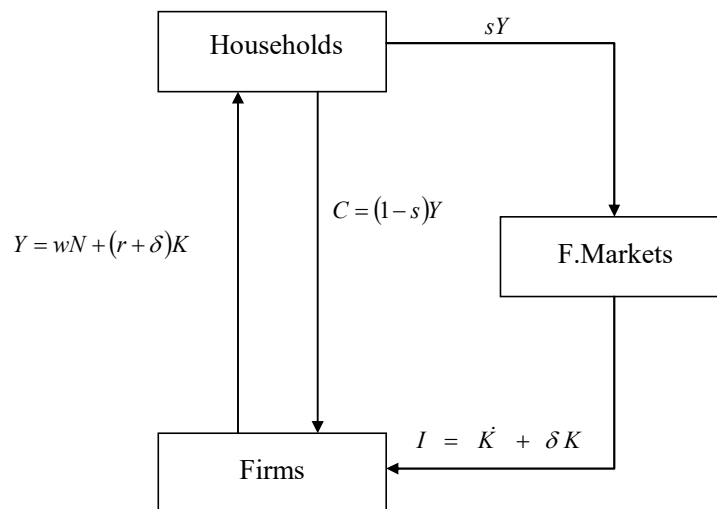
Equations (2.11) and (2.12) imply that the shares of capital and labour on national income, $(r + \delta)K/Y$ and wN/Y , are constant and equal to β and $1 - \beta$, respectively. That is, even though the prices and quantities of capital and labour may vary over time, changes must be such that the shares of national income paid out to each factor of production remain constant. This is a direct implication of assuming perfect competition and a production function with constant returns to scale.

2.2.5 The flow income chart

The flow income chart of this economy is displayed in Figure 2.2. From the expenditure angle, the sum of investment and consumption demands equal to output. From the income angle, output is paid to labour and capital owners in the form of labour income and capital income. Households use their income to consume and to save. Savings are equal to investment.

In the figure, the item “financial markets” is mediating savings and investment. This financial market is frictionless, in the sense that it involves no transactions costs³⁶.

Figure 2.2: The flow income chart in the basic Solow model



The figure displays the flow of payments in this economy. Factor incomes are paid to households, who consume and save, acquiring bonds. The resources raised by sales of bonds are invested by other households to acquire new capital (there are no transactions costs or frictions in the financial market).

2.2.6 The Fundamental Dynamic Equation

To understand how the model works, note that, in the absence of technological progress, the main determinant of per capita output (2.1) is the capital-labour ratio (i.e. the amount of capital available per worker). This variable is pre-determined each moment in time, given the existing stock of capital and the size of population. The capital-labour ratio

³⁶ If each household relied on own savings only to invest, a financial market would not be needed. But one may consider the possibility of different households having different propensities to save or to invest. You may think this model as with investors issuing bonds that pay an interest rate r to finance their purchases of capital. These bonds are acquired by savers. Each household can be simultaneously a saver and an investor, but not necessarily balanced. In the aggregate, net savings of the household sector will be equal to total investment.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

may however change over time, depending on investment, depreciation, and population growth. Formally, let's take the time derivative on k , to obtain:

$$\dot{k} = \left(\frac{\dot{K}N - \dot{N}K}{N^2} \right) = \frac{\dot{K}}{N} - \frac{\dot{N}}{N}k \quad (2.13)$$

After some substitutions using (2.2) and (2.7)-(2.9), we obtain the so-called Fundamental Dynamic Equation of the Solow model:

$$\dot{k}_t = sAk_t^\beta - (n + \delta)k_t \quad (2.14)$$

This equation states that the capital-labour ratio increases with per capita saving ($sy = sAk_t^\beta$) and decreases with the depreciation rate (δ) and the population growth rate (n).

The term $(n + \delta)$ in (2.14) may be interpreted as the rate of depreciation of the capital-labour ratio: on one hand, the depreciation rate reduces k by causing the capital stock K to wear out over time; on the other hand, population growth results in less capital available for each worker (this negative effect of population growth on capital per worker is called *capital dilution*). According to equation (2.14), the change in the capital-labour ratio over time is positive whenever per capita savings (investment) exceed the depreciation of the capital-labour ratio, and conversely.

2.2.7 Graphical illustration

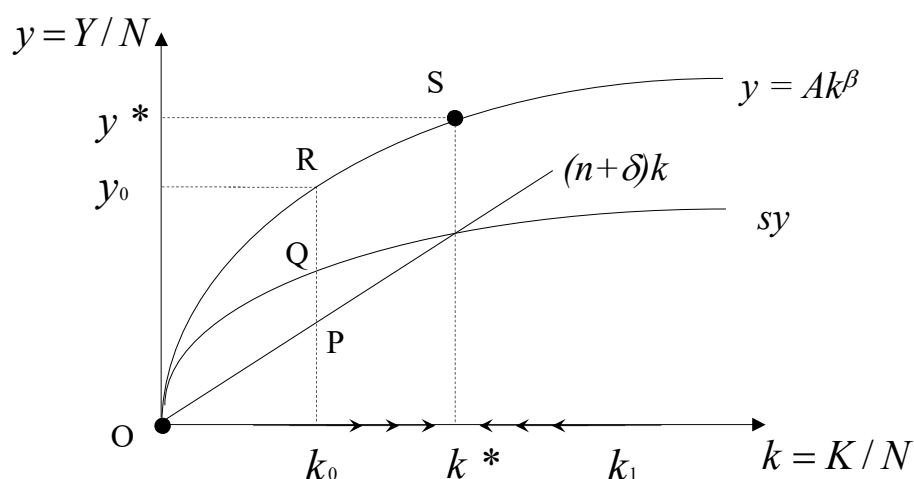
Figure 2.3 describes the dynamics and the equilibria of the model, as implied by equation (2.14). The uppermost curve is the production function in per capita terms (2.2). The figure also depicts the two terms at the right-hand side of (2.14): per capita savings (sy), and the locus $(n + \delta)k$. The latter is known as the *break-even investment line*: it depicts, for each level of capital per worker, the exact amount of gross savings that will be necessary to offset the corresponding capital depreciation and capital dilution.

To see how the capital-labour ratio evolves over time, consider an initial situation where the capital-labour ratio is equal to k_0 . In that case, per capita income will be y_0 , of which QR devoted to consumption and k_0Q devoted to savings. Since per capita savings exceed the “break even investment” (given by k_0P), from equation (2.14) it follows that the change in the capital-labour ratio will be positive. In words, since the economy generates more savings (and hence more investment) than the needed to keep the amount of capital-

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

labour ratio constant, the capital-labour ratio will increase. Then, as k increases, savings per capita increase less than proportionally. The reason is once again diminishing returns: since income per worker grows less than proportionally than the stock of capital per worker, savings cannot grow as fast as depreciation. And a moment will come when the two schedules cross each other: at $k=k^*$ per capita savings are just the needed (but no more) to equip the new entrants into the labour force and to replace the depreciating capital, implying that the capital labour ratio remains constant. This is the *steady state* (equilibrium) of the model³⁷.

Figure 2.3. Dynamics and equilibria in the Solow Model



The figure displays the production function in the intensive form, per capita savings and the break-even investment line. The steady state occurs when the break-even investment line crosses the schedule describing per capita savings. If the economy starts out on the left (right) of the steady state, per capita savings will be greater than (less than) the required to keep the capital labour ratio constant.

2.2.8 The steady state

³⁷ We invite the reader to use a similar reasoning to explain why the capital-labour ratio converges to the steady state departing from k_1 .

Formally, the equilibria of the model are obtained solving (2.14) for $\dot{k} = 0$. This equation has only two solutions, the trivial one ($k=0$) and:

$$k^* = \left(\frac{sA}{n + \delta} \right)^{\frac{1}{1-\beta}}. \quad (2.16)$$

Because the model predicts that the economy converges to the steady state (2.16) from any departing point in its neighbourhood, this equilibrium is said to be *stable*³⁸.

Substituting (2.16) in (2.2), one obtains the steady state level of per capita income:

$$y^* = A^{\frac{1}{1-\beta}} \left(\frac{s}{n + \delta} \right)^{\frac{\beta}{1-\beta}} \quad (2.17)$$

Since parameters A , s , n and δ are all constant, equations (2.16) and (2.17) imply that, in the steady state, capital per worker and per capita income are also constant.

Note that this outcome is in full conformity with the CRS property: if labour and capital are set to grow at rate n , then output will also grow at rate n (implying a constant level of per capita income). This is why the CRS assumption plays a key role in this model.

2.3 The Solow model and the facts of economic growth

2.3.1 The Solow model and the Kaldor acts

Robert Solow developed his famous model with the main purpose being a better understanding of the growth performance of the US economy in the twentieth century. He was particularly interested in explaining the long-run tendency for output and capital to grow at the same rates – a statistical regularity first documented for the U.S. economy by the

³⁸ Formally, the equilibrium described by k^* is *locally* stable because the condition $\partial \dot{k} / \partial k < 0$ holds for any point in its neighbourhood. The reader may verify that the same condition does not hold in the neighbourhood of the trivial steady state, $k=0$. The later is an *unstable* equilibrium.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Russian-American economist, Simon Kuznets. Solow also wrote the model with the so-called “Kaldor’s facts” in mind. These are six “remarkable historical constancies” (empirical regularities) that the British economist Nicholas Kaldor identified as characterizing modern economic growth.

In particular, the Kaldor stylized facts are³⁹:

1. Output per worker grows over time at a sustained rate
2. The capital stock per worker grows over time at a sustained rate
3. The capital-output ratio exhibits no clear trend over time;
4. The real return to capital is relatively constant over time;
5. The shares of labour and of capital on national income are roughly constant over time;
6. There are wide differences in the growth rate of productivity across countries.

Kaldor did not claim that these facts hold each moment in time. For instance, per capita output falls during recessions and the real interest rate fluctuates significantly over the business cycle. Over long periods of time, however, these facts tend to show up in the statistical data. Hence, they shall provide a natural benchmark for models focusing on long run growth to be confronted with.

As we just saw, equations (2.16) and (2.17) imply that, in the long run, per capita income and the capital labour ratio do not grow at all. Hence, the basic Solow model fails to capture the Kaldor Stylized facts 1, 2, and 6.

To check whether Fact 3 is met, let’s divide (2.16) by (2.17), to obtain the output-capital ratio in the steady state:

$$\left(\frac{Y}{K}\right)^* = \frac{n + \delta}{s} \quad (2.18)$$

³⁹ Kaldor, N, 1961. *Capital Accumulation and Economic Growth*. In F.A. Lutz and D.C. Hague (eds.), *The theory of Capital*. New York: St Martin Press. Kaldor, N., 1957. A model of Economic Growth. *The Economic Journal* 67 (268), 591-624.

This ratio is constant because the three parameters on the right-hand side are all constant (in Figure 2.3, this ratio is measured by the slope of the ray OS). We conclude that the model captures the stylized fact number 3: the long run tendency for capital (K) and output (Y) to grow at the same rate. If the average product of capital is constant in the steady state, then the interest rate will also be constant in the steady state (from equation 2.12). Hence, the Kaldor's stylized fact 4 is also implied by the model. Regarding fact number 5, we already saw that it holds in this model (equations 2.11 and 2.12).

Summing up, the basic Solow model captures facts 3, 4,5, but it fails to capture facts 1, 2, and 6.

2.3.2 Savings, population and per capita income in the real world

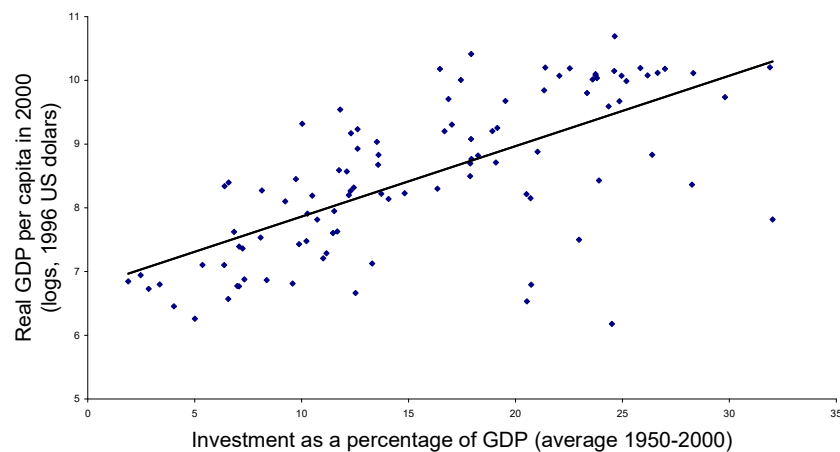
Other implication of the Solow model concerns the relationship between per capita income and saving rates and population growth rates. According to equation (2.17), countries with higher saving rates and slower population growth should enjoy higher standards of living than countries with lower saving rates and fast population expansion.

In figures 2.4 and 2.5, we check how these two predictions of the model go along with the real-world facts. The figures plot the level of GDP per capita in the year 2000 with, respectively (i) gross investment as a share of GDP and (ii) population growth rates. Figure 2.4 reveals a positive correlation between investment rates and per capita incomes. Figure 2.5 reveal a negative correlation between population growth and per capita incomes. Both figures are in broad accordance to the Solow model.

Note, however, that this evidence does not prove that the Solow model is correct. For example, it could be that the low saving rates in the poorest countries were explained by the fact that people living at the margin of subsistence could afford to save. On the other hand, poorest countries may exhibit faster population expansions simply because they still didn't make their demographic transition (see Box 2.2). Thus, while this evidence is in accordance to the Solow model, it does not prove that the Solow model is the right one to capture this evidence. Correlation is not causality.

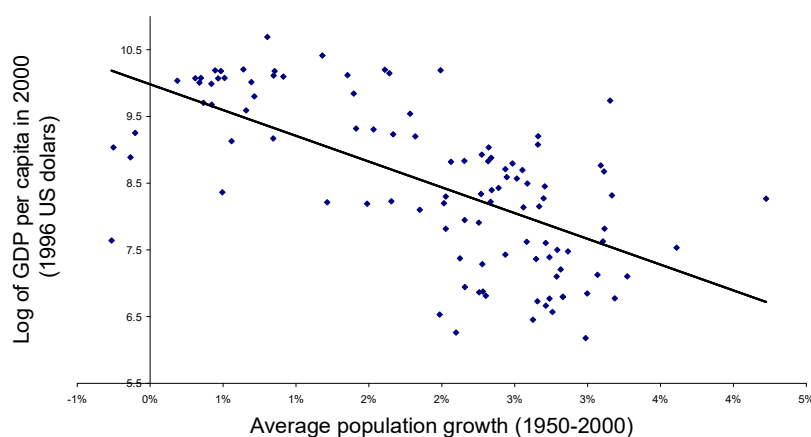
Figure 2.4: Per capita GDP and gross Investment 1950-2000

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>



Source: Summers, Robert and Heston, Alan. (1991). The Penn World Table (Mark 5): an expanded set of international comparisons, 1950-1988, *Quarterly Journal of Economics*, 106(2), May, 327-68. The sample includes 169 countries and average data over the 50 year period from 1950 to 2000.

Figure 2.5: Per capita GDP and population growth rates, 1950-2000



Source: same as Figure 2.4.

Box 2.2 Demographic transitions and poverty traps

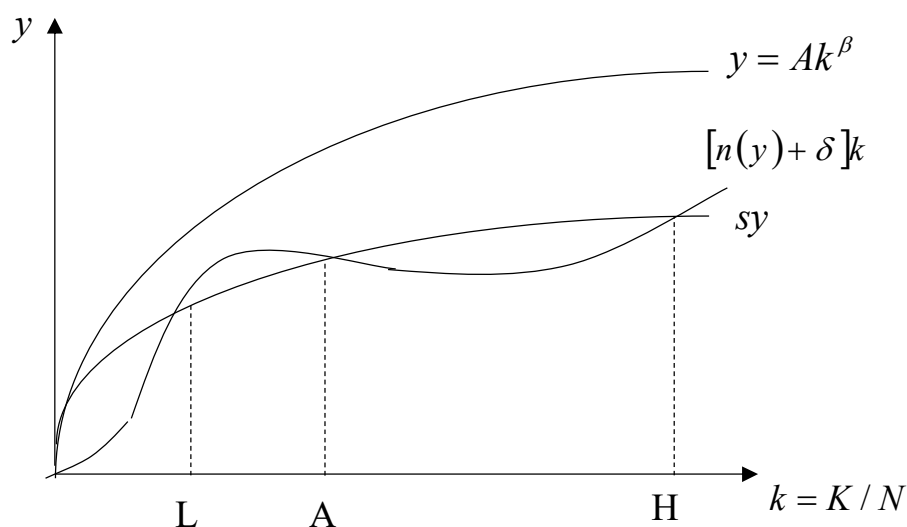
The model outlined above postulates a constant rate of population expansion, n . The departure from the Malthusian assumption looks sensible to address the growth problem of economies that already made their demographic transitions. However, it may be interesting to investigate how the model predictions may change when one allows the rate of population growth to be determined endogenously.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

The stylized fact of demographic transitions is that a country's population is expected to expand at moderate rates both when per capita income is low (high birth rates and high death rates) and when per capita income is high (low birth rates and low death rates). In intermediate stages, the growth rate of population is expected to be high, because death rates are low and birth rates are still high. Adding a non-linear relationship between population growth and per capita income to the Solow model, one obtains a break-even investment curve that is non-linear, as depicted in Figure 2.6. In this case, the model may display multiple equilibria.

In Figure 2.6, there are four possible equilibria: origin, L, A, and H. The equilibrium represented by H corresponds to a low population growth rate and a high level of per capita income. Equilibrium A is an intermediate equilibrium, characterised by fast population growth. Equilibrium L is characterised by a low rate of population growth and a low level of per capita income.

Figure 2.6. A poverty trap in the context of the Solow model



In the figure the basic Solow model is extended, to account for a population growth rate that depends non-linearly on per capita income. In this case, there are three non-trivial equilibria, L and H (stable) and A (unstable). Because equilibrium L is inferior to H, it is called a poverty trap.

The equilibria described by H and L are both stable: like in the baseline model, departing from a neighbour point at their left (right), per capita savings are greater than (are less than) the break even investment, and hence the capital labour ratio will increase (decrease) until reaching the steady state. The equilibrium described by A is unstable: if, <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

departing from this equilibrium, the capital stock decreases (raises) by a small amount, then the saving rate becomes lower (higher) than the break even investment and the capital labour ratio will decrease (increase), until the low (high) income steady state is reached. Because the equilibrium L is stable and dominated by another possible outcome (equilibrium H), it is called a “poverty trap”.

A key question in models with multiple equilibria concerns what equilibrium will prevail. In the case at hand, the stability properties of the model imply that *history* plays a critical role in equilibrium selection: if an economy starts out in the bad equilibrium L, it will remain in the bad equilibrium. If the economy starts out in the good equilibrium, H, it will remain in the good equilibrium.

Moreover, a policy designed to move the economy out of the poverty trap L may fail to do so, unless it is powerful enough to push the economy to the right side of threshold point A. Suppose that the economy was initially at point L and then it received some external aid to improve its capital stock. Unless the amount of aid was large enough, the impact would be merely temporary: the initial rise in output per capita would lead to an acceleration of population expansion, which in turn would require a higher saving rate just to stand still. Since the saving rate is constant, the capital-labour ratio would fall back, driving the economy again to the poverty trap. The “demographic transition” would not take place because the saving rate was too low to induce the required increase in the capital-labour ratio during the intermediate stage. In the end, the extra capital obtained was “spent” in an increasing population, rather than in improving living standards. If, however, the donation was large enough for the capital-labour ratio to expand ahead the critical level A, then the economy would have been able to escape the trap, moving towards the high-income steady state, H. This reasoning suggests that that international assistance to a poor country trapped in a Post-Malthusian regime will be useless unless it is large enough for the economy to overcome the trap.

Note however that a large investment in physical capital is not the only avenue for this economy to escape the trap: a policy of birth control and family planning, if successful in reducing the gap between the birth rate and the death rate in the intermediate stage, could eventually smooth the break-even investment line the enough for the multiple equilibria to disappear.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

2.4 Transitional Dynamics

An important feature of the Solow model is that, if the economy is not in the steady state, it will converge to the steady state. But the economy does not jump instantaneously from one steady state to the other: since capital accumulation is bounded by the availability of savings, there will be an adjustment period, during which the economy *approaches* the new steady state.

This point is very important because in the real-world economies may be found in adjustment processes, rather than in the steady state: e.g. perhaps because the savings rate has risen recently but not yet been reflected fully in higher per capita income. In the light of the Solow model, that economy will be experiencing a transitory growth, until reaching the new steady state. The following sections address specifically the issue of transition dynamics.

2.4.1 What happens if the savings rate increases?

Figure 2.7 shows how the model adjusts to a rise in the saving rate. When the saving rate rises from s_0 to s_1 , the curve measuring per capita savings shifts upwards, moving *the* steady state from k_0^* to k_1^* .

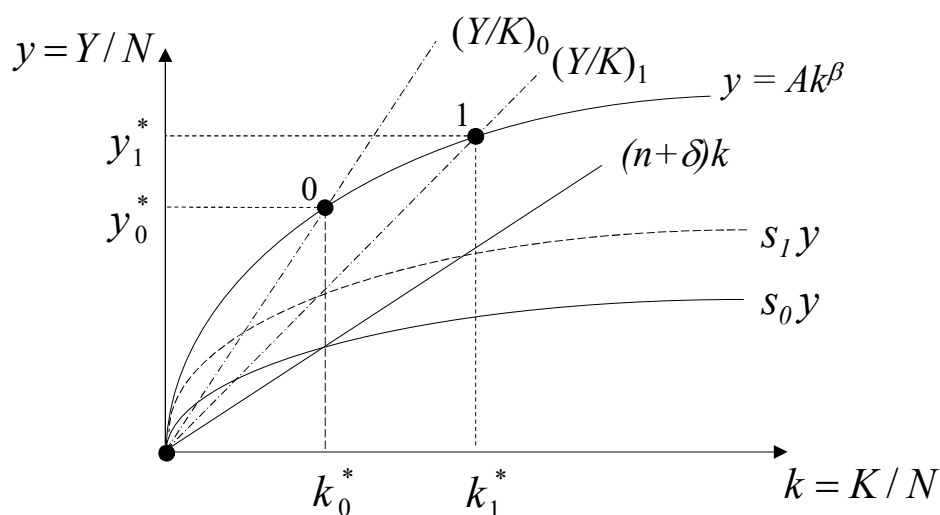
Thus, the economy engages in an expansionary process until the new equilibrium is met. In the long run, the rise in the saving rate has produced a *level effect*: that is, in the new steady state, the economy will enjoy a higher *level* of output per worker than in the steady state before. During the transition from one steady state to the other, per capita output had increased. But, because of diminishing returns, this increase in output per capita was no more than a *temporary* phenomenon.

It is worth mentioning that the average product of capital (Y/K) in the new steady state is lower than in the old steady state. This can be checked by reference to equation (2.18), where the saving rate enters in the denominator. Visually, in Figure 2.7 you see that the slope of the ray that departs from the origin and crosses the production function at point I is lower than the one corresponding to the ray that crosses the production function at point O . Referring to (2.12), this also means that *the interest rate has declined*: intuitively, one consequence of a higher availability of capital per worker in the economy is that using capital becomes relatively cheaper.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

The implication of what we just learned is that, in low-income countries with low savings (say $s=10\%$), a growth surge could eventually be achieved by raising the saving rate. However, once the economy reached the new steady state, per capita income would stagnate again. Thus, in order to achieve further increases in output per worker, one would need to raise the saving rate again, and again. And clearly, there are limits in exploring this avenue: the maximum level of the savings rate observed in countries in the real world is around 30-35%. Rates much higher than this obviously eat into available consumption and so the current standard of living. The conclusion is that it will be impossible to achieve a continuous growth of per capita income by increasing the saving rate on a continuous basis.

Figure 2.7. A higher saving rate raises the steady state level of per capita income but only boosts growth rates temporarily



The figure shows how the steady state in the Solow model changes with an exogenous increase in the saving rate. In the new steady state (point 1), the capital labour ratio and per capita output are higher than before, but the average product of capital (Y/K) is lower, due to diminishing returns. In the long term, per capita income is again constant (the increase in the saving rate produced a “level effect”).

2.4.2 What happens if population growth or the depreciation rate decline?

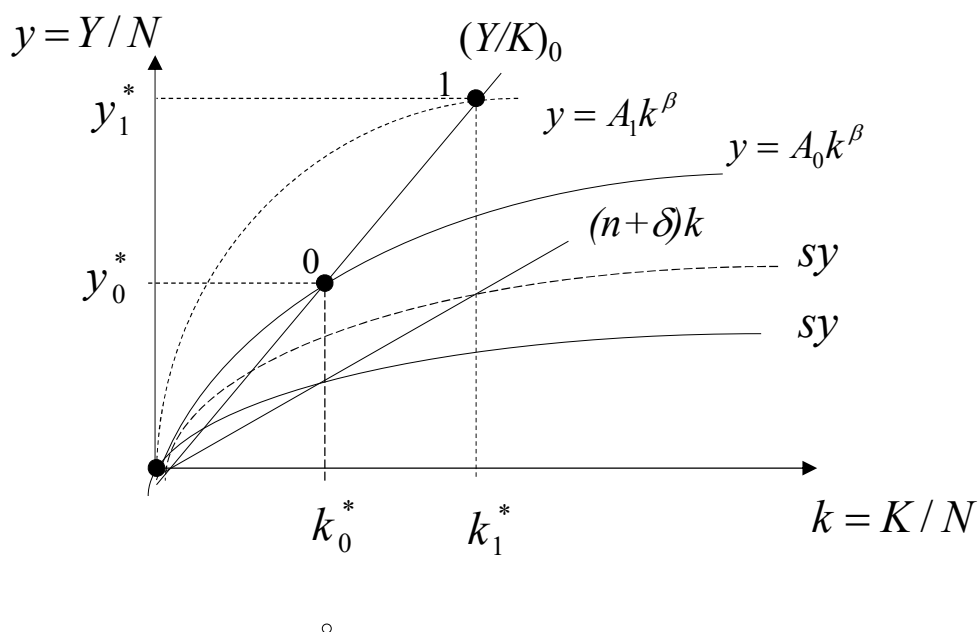
From equation (2.17) we see that a fall in the population growth rate has a similar effect to that of a rise in the savings rate. In terms of the Figure 2.3, the difference is that the change in the steady state will be caused by a downward shift of the break-even investment line. Thus, both output per worker and capital per worker will increase, but this will happen only during the transition from one steady state to the other. In the new steady state, the average product of capital - and interest rate - will be lower than in the initial steady state.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

2.4.3 What happens if the level of technology improves?

Figure 2.8 describes how the model adjusts to a once-and-for-all technological change. In the figure, when TFP changes from A_0 to A_1 , both the production function and the curve measuring per capita savings shift upwards, tilting the steady state from k_0^* to k_1^* .

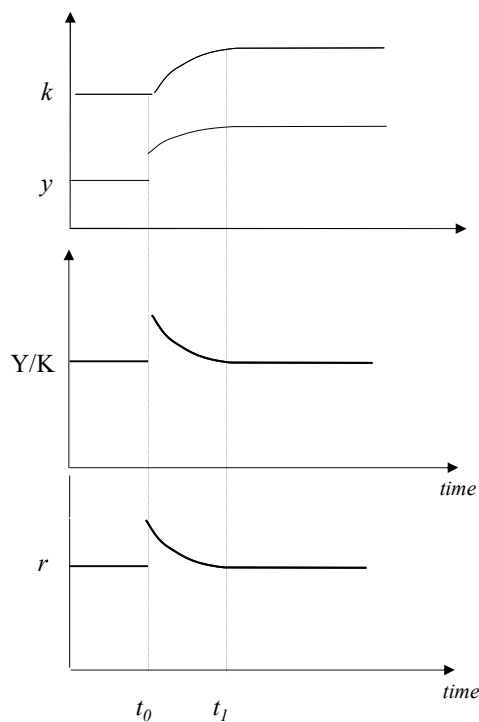
Figure 2.8. An increase in technology raises the steady state level of per capita income but leaves the output-capital ratio unchanged



In contrast to the case of an increase in the saving rate, in this case there is an initial jump in per capita output: the productivity increase means that more production is achieved out of the initial capital labour ratio. The paths of output per worker, the average product of capital, and of the interest rate are described in Figure 2.9. As shown in the figure, at the time of the shock (t_0), all the three variables jump upwards. During the adjustment to the new steady state, the average product of capital declines again and so will do the interest rate. In the new steady state (after t_1), the average product of capital and the interest rate are the same as before the shock (you may confirm this by observing that the long run level of Y/K (equation 2.18), does not depend on the level of technology, A). In the long run, the technological improvement materialized with higher levels of per capita income and capital per worker, but with the same interest rate.

Figure 2.9. The time paths of per capita output, the capital –labour ratio, the output capital-ratio and the interest rate, following an improvement in technology

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>



Following a technological improvement, per capita income first jumps (at a constant k), and then starts growing along with the increase in k , until the new steady state is reached. The average product of capital jumps at the impact, and returns to the initial level in the long run.

2.5 The Golden Rule

2.5.1 The Golden Rule of capital accumulation

In the Solow model, a higher saving rate delivers a higher level of per capita income in the steady state. Does this mean that any increase in the saving rate is desirable? From the household point of view, the answer to this question is negative.

To see this, remember that households care with *consumption*, not with income. Thus, while a higher saving rate brings a higher per capita income in the steady state, a higher *proportion* of the later will be foregone consumption. The final impact on consumption will depend on the balance between these two opposing effects.

Mathematically, the saving rate that maximizes the level of per capita consumption in the steady state can be found in the following manner:

$$\max_k c = y_t - sy_t, \text{ subject to } \dot{k} = 0.$$

where $c=C/N$ denotes for per capita consumption. In the steady state, this is equivalent to chose k so as to maximise⁴⁰:

$$c = Ak_t^\beta - (n + \delta)k. \quad (2.19)$$

The first order condition of this problem leads to:

$$\frac{\partial y}{\partial k} = n + \delta \quad (2.20)$$

This condition is called the "Golden Rule of Capital Accumulation"⁴¹. It states that the steady state level of per capita consumption is maximised when the slope of the production function (i.e. the marginal product of capital) is equal to the slope of the break-even investment line.

Geometrically, this problem is illustrated in Figure 2.10. The Golden Rule is met at point k^G . Whenever the steady state level of capital per worker is below the golden rule (that is, when $k^* < k^G$), the rise in income that results from a possible increase in k^* more than offsets the rise in the amount of savings that is necessary to sustain such equilibrium, implying that more resources are available for consumption. Conversely, whenever the steady-state capital-labour ratio is higher ($k^* > k^G$) the rise in income that would result from any further raise in k^* is less than the required increase in savings, implying that less resources become available for consumption.

An alternative representation of the golden rule is in Figure 2.11. The figure describes the steady state level of per capita consumption as a function of the steady state capital labour ratio, as implied by (2.19). It basically corresponds to the difference between per capita

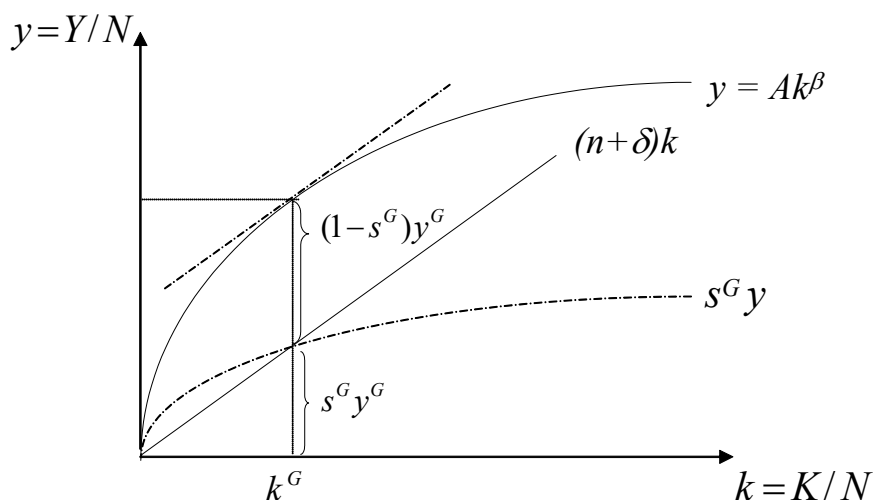
⁴⁰ Note the in the steady state $sy^*=(n+\delta)k^*$. The symbol “*” - which refers to the steady state - is suppressed to simplify the algebra.

⁴¹ Phelps, E., 1961. "The Golden Rule of Accumulation: a Fable of Growth man". *American Economic Review*, 51, 638-643.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

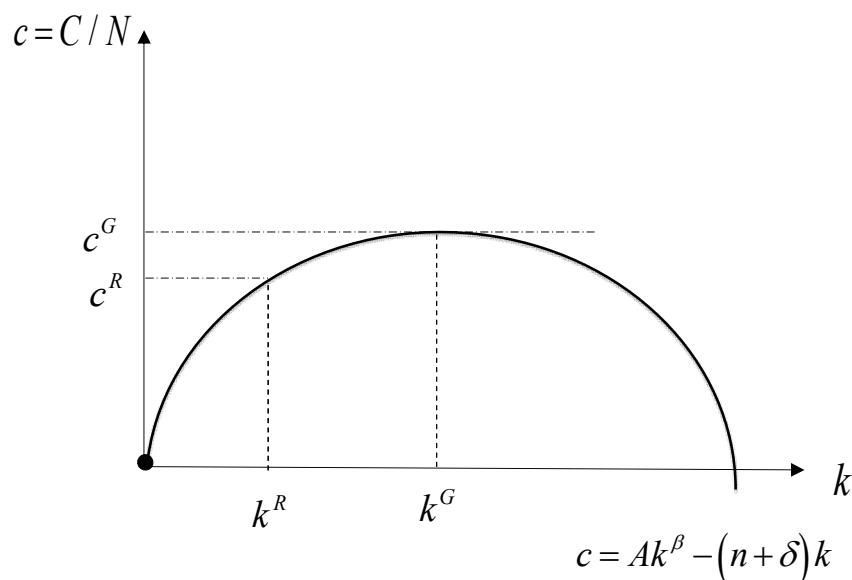
output and break-even investment in Figure 2.10. The maximum of this curve corresponds to the golden rule level of k , $\frac{\partial c}{\partial k} = 0$, given by (2.21).

Figure 2.10: Illustration of the Golden Rule



A steady state in the Solow model must be such that the savings locus crosses the breakeven investment line. Given an economy's breakeven investment line, among all possible saving rates, the golden rule saving rate is the one that determines the larger distance between the breakeven investment line and per capita output. That will be the one where the slopes of the two schedules are the same.

Figure 2.11. Per capita consumption in the steady state



The figure describes the steady state level of per capita consumption as a function of the capital-labour ratio. The golden rule corresponds to the maximum possible consumption level in the steady state. The alternative allocation described by R in the figure corresponds to the “modified golden rule”, that comes out in a model with endogenous savings and impatient consumers (see Section 2.6).

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Algebraically, the golden rule level of k^* is given by:

$$k^G = \left(\frac{\beta A}{n + \delta} \right)^{\frac{1}{1-\beta}}. \quad (2.21)$$

The value of s that turns k^G into a steady state is called the “golden rule” saving rate. Comparing (2.21) with the general solution for steady states (2.16), we conclude that the golden rule saving rate is:

$$s^G = \beta \quad (2.22)$$

That is, the golden rule saving rate is equal to the share of capital in income⁴².

2.5.2 Dynamic efficiency

Suppose that an economy was in a steady state at the left of the golden rule (that is, with $k^* < k^G$), with and per capita consumption below the maximum possible. Would that mean that the government should step in, trying to increase the economy’ saving rate (say, by adjusting taxes and subsidies, as illustrated in Box 2.3)?

As a general principle, if savings are decided by optimizing agents, any policy altering their choices should make them worse off. So, unless there are good reasons to believe that some impediment prevents consumers from optimally decide their savings, or that any market failure is turning individual decisions socially unacceptable, there will be no case for intervention.

As a policymaker, you could argue that the impatience of the current generation (reflected in low saving rates) was impacting adversely on the inter-generational distribution of income, as less capital would be available for future generations. However, forcing the

⁴² An alternative avenue to obtain (22) is replacing (2.17) in the maximization problem (2.19) and take the derivative in order to s .

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

current generation to save more would not be a Pareto improvement. Because of this, any equilibrium at the left of the golden rule is said to be *dynamically efficient*. An equilibrium is said to be dynamic efficient if the path of savings cannot be changed to strictly increase consumption at some point in time without lowering it at any point in time⁴³.

A different case occurs when the initial steady state lies at the right-hand side of the golden rule (that is, if $k^* > k^G$). In that case, by reducing savings today, consumption would increase *both* today and tomorrow. Since a “free lunch” is readily available, this case is labelled as *dynamically inefficient*.

At the first sight, a case of dynamic inefficiency looks at odds with the principle that agents are optimizers: after all, if households were saving too much, they should realize that by reducing savings today, they would be able to consume more, today and in the future.⁴⁴ However, in the real World, individuals do not leave forever. In a world with overlapping generations, it is possible that the current generation, concerned with its income after retirement, engages in a saving effort that reveals excessive from the social point of view. In that case, a policy aiming to reduce current savings today could in principle improve the welfare of both current and future generations. Under this reasoning, a social security mechanism based on inter-generational transfers (pay as you go), could help mitigate the concerns of individuals while young, preventing excessive savings. Whether this theoretical case is realistic or not is a different question. In practice, we observe that the shares of capital in national income vary from 0.3 and 0.4. According to the model formulation, this corresponds to the contribution of capital to output, β . Since real world saving rates are, in general, lower than 30%, one may conclude that “dynamic inefficiency” is not at all a general case.

⁴³ Phelps, E., 1965. Second Essay on the Golden Rule of Accumulation, *American Economic Review* 55, 793-814.

⁴⁴ Of course, this presumes that individuals are free to choose their saving rate. A textbook counterexample is the case of the old Soviet Union: during the communist regime, most consumption goods were rationed, implying that individuals had income to spend but no goods to buy, being forced to save more than desired.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Box 2.3. Using taxes and subsidies to meet the golden rule

Suppose you were a benevolent central planner wanting to maximize the steady state level of per capita consumption of your citizens. How would you achieve this objective? One possibility would be to use taxes and subsidies. To illustrate this, consider a tax on production τ (subsidy if negative) which proceeds are returned back to households after households decided the amount of consumption and savings; if negative, this implies a confiscation of part of the household' desired consumption). The government budget is assumed to be balanced, that is $T = \tau Y$.

Assuming, as before, a constant saving rate s , total savings in this economy will be given by:

$$S = s(1 - \tau)Y = \dot{K} + \delta K.$$

To solve the model in the new version, just note that $s(1 - \tau)$ shall be used instead of s in the fundamental dynamic equation. Proceeding as before, the steady state level of per capita income will now be given by:

$$y^* = A^{\frac{1}{1-\beta}} \left(\frac{s(1-\tau)}{n+\delta} \right)^{\frac{\beta}{1-\beta}}. \quad (2.23)$$

The corresponding steady state level of per capita consumption is:

$$c^* = (1-s)(1-\tau)y^* + \frac{T}{N} = [1-s(1-\tau)]y^* = [1-s(1-\tau)]A^{\frac{1}{1-\beta}} \left(\frac{s(1-\tau)}{n+\delta} \right)^{\frac{\beta}{1-\beta}}. \quad (2.24)$$

Maximizing (2.24) with respect to τ , one obtains:

$$\tau = 1 - \frac{\beta}{s}. \quad (2.25)$$

According to (2.25), the golden rule tax rate on output depends on the gap between the actual saving rate s and the golden rule saving rate β : the optimal tax will be negative (subsidy) if the saving rate falls below the golden rule; it will be positive (tax) if the saving rate is higher than the golden rule; and it will be zero if the saving rate satisfies exactly the golden rule.

2.6 The case with endogenous savings

2.6.1 An optimal consumption rule

In the Solow model, the saving rate is assumed exogenous. The neoclassical model can however be extended, to account for the case in which individuals optimally decide their saving rates. It is not in the scope of this book to solve complicated dynamic optimization models. So, in the following – and throughout the book – we will refer to a very simple formula (for details on how this formula is obtained, see Appendix 2.1).

$$\gamma_t = r_t - \rho \quad (2.26)$$

In (2.26), $\gamma_t = \dot{c}/c$ denotes for the growth rate of per capita consumption, and ρ is the rate of time preference (that is, the rate at which individuals are willing to trade one unit of utility today for utility in the future). The rate of time preference measures the impatience of individuals.

According to (2.26), as long as the interest rate is higher than the rate of time preference, individuals will optimally increase their consumption over time. If however the interest rate falls below the rate of time preference, individuals will optimally reduce consumption over time. When $r_t = \rho$, the optimal level of consumption will be constant.

2.6.2 What happens when the rate of time preference decreases?

To see the implication of replacing an exogenous saving rate by (2.26) in the neoclassical growth model, remember that in the competitive equilibrium the interest rate is determined by the marginal product of capital (equation 2.12). The latter, in turn, is a negative function of the capital-labour ratio (as implied by the Law of Diminishing returns).

Suppose that, departing from an initial steady state, the degree of impatience of individuals decreases. This is reflected into the fact that individuals demand a lower return to postpone consumption. At an unchanged interest rate, savings will increase, and consumption will fall instantaneously at the time of the shock. Since more savings today translate into more investment, the capital labour ratio will then start increasing, inducing a temporary growth of per capita income and consumption.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Then, as the stock of capital per worker increases, the marginal product of capital declines, and so will do the interest rate. At the time the interest rate becomes equal to the rate of time preference again, the desired consumption becomes constant over time (eq. 2.26) and the process of capital accumulation stops. In the new steady state, both consumption and per capita income are higher than before, but both will be constant again.

The main conclusion is that making savings endogenous does not alter the main property of the model: if, for any reason, people become more patient and save more, there will be a permanent increase in the steady state level of per capita income, but no permanent effect on economic growth.

2.6.3 The modified golden rule

A question that arises is how the endogenous saving rate looks like, in terms of our model. Before addressing this question, it is important to note that in the model with endogenous savings, the saving rate (savings divided by income) is not in general constant along the transition to the steady state. When the economy reaches the steady state, however, both per capita consumption and per capita income become constant over time, so the saving rate will be constant as well.

To obtain the steady state saving rate in the model with endogenous savings, let's first substitute the market interest rate (2.12) in (2.26), to obtain the growth rate of per capita consumption each moment in time:

$$\gamma_t = \beta \frac{Y_t}{K_t} - \delta - \rho$$

In the steady state the growth rate of consumption is zero and the capital output ratio is given by (2.18). Using these conditions in the equation above, and solving for the saving rate, one obtains:

$$s = \beta \frac{n + \delta}{\rho + \delta} \quad (2.27)$$

The steady state saving rate (2.27) is the “modified golden rule”. It is less than the golden rule, because the consumer is assumed to be impatient, as reflected in parameter ρ . Intuitively, the impatience reflected in the rate of time preference means that an infinitely

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

lived consumer will, in general, prefer a steady state consumption lower than the maximum feasible⁴⁵.

The steady state level for capital per worker implied by (2.27) can be obtained directly substituting that expression in (2.16). This gives:

$$k^R = \left[\frac{\beta A}{\rho + \delta} \right]^{\frac{1}{1-\beta}}$$

This optimized capital-labour ratio is represented in Figure 2.11. The main notable difference in respect to the Solow model is that a balanced growth path with a capital stock above the golden-rule level is not possible: if the economy was in such a path, households would optimally reduce their saving and take advantage of this opportunity.

2.6.4 Exogenous or endogenous savings?

The consumption rule (2.26) was obtained assuming that all individuals have access to a frictionless financial market, where they can borrow or lend any amount of income at a given interest rate r . Acknowledging this assumption is very important, when thinking how the model applies to the real World. In the real world, financial markets are far from perfect. Many poor households have no access to borrowing, and this is especially true in developing countries. In that case, households are more likely to consume based on current income, rather than on future incomes. This means that a consumption function depending only on current income and with an exogenous saving rate, as assumed in the basic Solow model, may actually be very sensible to describe emerging economies with underdeveloped financial markets.

⁴⁵ Formally, the condition $\rho > n$ is needed for the solution of the maximization problem to be well defined (for an explanation, see for instance, Romer 2001, chapter 2.1 [Romer, D., 2001, Advanced Macroeconomics, McGraw-Hill]).

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

2.7 The Solow Residual

In the sections above, we saw that a country's per capita output can increase both because the capital-labour ratio increases or due to technological change. Empirically, a technique to disentangle the contribution of these two factors to economic growth using observable data is "growth accounting".

Basically, the technique departs from a log-differentiation of the production function (2.1):

$$\frac{\dot{Y}}{Y} = \frac{\dot{A}}{A} + \beta \frac{\dot{K}}{K} + (1 - \beta) \frac{\dot{N}}{N} = \hat{A} + \beta \hat{K} + (1 - \beta)n \quad (2.28)$$

This equation states that the growth rate of output equals the growth rate of total factor productivity ("A") plus a weighted average of the growth rates of physical capital and labour, where the weights are the corresponding elasticities in the production function. In the intensive form, the equivalent is:

$$\hat{y} = \hat{A} + \beta \hat{k} \quad (2.28a)$$

If factor markets are competitive, as assumed in the Solow model, the parameter β shall be equal to the share of capital in national income (equations 2.11 and 2.12). In the real world, it has been observed that the share of capital in national income ranges from 0.3 to 0.4. Thus, based on observed variables one can estimate the unobservable growth rate of the technological parameter, as the difference between the actual growth of output and the growth implied by factor accumulation (hence the label *Solow residual*)⁴⁶:

$$\hat{A} = \hat{Y} - \beta \hat{K} - (1 - \beta)n \quad (2.29)$$

In the intensity form, the Solow residual is obtained as:

$$\hat{A} = \hat{y} - \beta \hat{k} \quad , \quad (2.30)$$

⁴⁶ Solow, R. 1957. "Technical change and the aggregate production function", *Review of Economics and Statistics* 39, 79-82.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

with $\hat{y} = \hat{Y} - n$ and $\hat{k} = \hat{K} - n$.

By construction, the Solow residual measures the part of actual output growth that is not accounted for by factor accumulation. This is more than simple “technological progress”. Technological progress, in narrow sense, refers to changes in the “efficiency” with which the existing technology and inputs are combined. The Solow residual also accounts for unmeasured changes in the quality of inputs (skill increments in the labour force, quality of land, climate) and statistical errors.

In his original paper, Solow (1957, op. cit) found out that the fraction of per capita income growth in The US along 1909-1949 accounted for capital accumulation was very small: only one eighth, comparing to seven-eighths attributed to “technical change”. This finding was corroborated by many other studies for many other economies.

Table 2 displays the results of a recent growth accounting exercise focusing on Britain since the 13th century (references in the table). The figures in the table indicate that capital played little role in economic growth in the pre-industrial era. Until 1830, most economic growth was explained by the growth rate of TFP. During the transition to modern economic growth (1830-1860) capital accumulation by passed technological change as the main source of growth. Still, TFP contributed to 0.43p.p. (28 per cent) of per capita income growth in that period.

This evidence points to a fundamental limitation of the Solow model in its basic formulation: by assuming that A is constant, this model ignores a critical ingredient of economic growth: technological progress.

Table 2.1: Accounting for the growth of British GDP per head, 1340s to 1860s (% per annum)

	Output Growth	Population growth	Output per worker	Contributions:	
				Capital Deepening	TFP Growth
1340s-1400s	-0.73	-1.27	0.54	0.34	0.20
1400s-1450s	-0.21	-0.13	-0.08	-0.13	0.05
1450s-1640s	0.5	0.53	-0.03	-0.01	-0.02
1640s-1690s	0.84	-0.04	0.88	0.22	0.67
1690s-1830s	1.08	0.74	0.34	0.01	0.26
1830s-1860s	2.28	1.17	1.11	0.68	0.43

Notes: Output and capital stock are in constant 1700 prices. Wights for labour and capital are 0,5 and 0,4 respectively. Source: Broadberry, S and A M de Pleijt (2021), "Capital and Economic Growth in Britain, 1270-1870: Preliminary Findings", CEPR Discussion Paper 15889.

2.8 Key ideas in Chapter 2

- The Solow model combines the neoclassical assumptions of perfect competition and flexible prices with a Keynesian consumption function, whereby consumption is a linear function of current income.
- The model explores the assumption of Constant Return to Scale to overcome the limitations imposed by diminishing returns. The properties of the model are such that in the long run the capital stock grows at the same rate as population, implying a constant level of capita income.
- The model accords to the real world facts that the capital-output ratio, the interest rate and the shares of labour and capital in per capita income are roughly constant over time, but it fails to deliver the most basic of the stylized facts of economic growth, namely that output per capita (and real wages) tends to grow over time: in this model, any growth in per capita income is merely temporary, reflecting the adjustment in the economy from one steady state to the other.
- The model offers the credible suggestion that countries with high savings rates and low population growth rates should expect higher levels of per capita income in the long run than countries with low saving rates and rapid population expansion.
- In the context of the Solow model, there is a saving rate that maximizes the level of per capita consumption in the steady state. This is not to say that a change in the saving rate towards that level will be Pareto improving: this will only be the case when the starting situation is of dynamic inefficiency.
- In the context of the Solow model, it is not possible to achieve a continuous growth of per capita income through successive increases in the saving rates. First, because saving rates have a feasible limit. Second, if continuous growth was accounted for increases in the saving rates, the real interest rate should be decreasing over time, and this does not conform with the Kaldor facts.
- Making saving endogenous does not rescue the model from its main limitation, that the long run growth rate of per capita output is zero.
- In this model, an exogenous improvement in technology leads to a higher level of per capita income, without causing any decline in the interest rate. This suggests that the key to overcome the main limitation of the Solow model is to allow for a continuous improvement in technology.
- Growth accounting is the technique of measuring a country's productivity change by the difference between output growth and the "contribution" of inputs. In general, growth accounting exercises reveal that total factor productivity expands over time. This evidence adds to the need of enriching our model, allowing technology to expand over time.

Appendix 2.1: The optimal consumption path in a simple 2-period model

The problem of how much an individual should save was first addressed in an inter-temporal optimizing framework by a mathematician from Cambridge UK called Frank Ramsey. Ramsey died at the age 26 and his seminal contribution remained obscure for long time by the economics profession, because at that time most economists were not familiar to dynamic optimization. His work was re-discovered four decades later, by Cass and Koopmans, who used it to characterize the optimal saving paths in the context of the neoclassical growth model⁴⁷.

In this appendix, we illustrate the optimal consumption rule in a dynamic context referring to a discrete time framework. Consider an individual whose life-time utility function is given by

$$U = \sum_{t=1}^{\infty} \frac{u(c_t)}{(1+\rho)^{t-1}}, \quad (\text{a2.1})$$

where $u(c_t)$ is a concave function, c_t is real consumption in period t and ρ is a given rate of time preference.

Assume that this individual has full access to financial markets, being therefore able to borrow or lend any amount of income at the interest rate r . His problem is to maximize the lifetime utility function, subject to $\sum_{t=1}^{\infty} \frac{c_t}{(1+r)^{t-1}} = \Omega$, where Ω denotes for lifetime wealth.

From the first order conditions of the maximization problem one obtains the so-called *Euler equation* for each two consecutive periods:

$$u'(c_{t+1}) = u'(c_t)(1+\rho)/(1+r).$$

⁴⁷ Ramsey, F., 1928. A mathematical theory of savings. *Economic Journal*, 38, N° 152, 543-559. Cass, D., 1965. Optimum growth in an aggregative model of capital accumulation. *Review of Economic Studies* 32 (3), 233-240. Koopmans, T., 1965. On the concept of optimal growth. In *The Econometric Approach to Development Planning*. Rand-McNally, Chicago.

This equation states that the marginal utility of consumption in the next period must be equal to the marginal utility of consumption in the current period, weighted by the ratio of the rate of time preference to the market discount rate. In other words, this rule implies that the consumption level each period must be such that an extra unit of consumption would make the same contribution to lifetime utility no matter to what period is allocated.

In the main text, we stick with the convenient assumption of logarithmic preferences, that is $u(c_t) = \ln c_t$. In this case, the Euler equation simplifies to:

$$c_{t+1}/c_t = (1+r)/(1+\rho). \quad (\text{a2.2})$$

Denoting by γ the growth rate of per capita consumption, the later expression becomes equal to:

$$1+\gamma = (1+r)/(1+\rho),$$

Which, by approximation gives (2.26).

Problems and Exercises

Key concepts

- *Gross investment vs. net investment. Break even investment line. The Kaldor facts. Poverty trap. Golden rule. Dynamic inefficiency (a la Phelps). The Solow residual*

Essay questions:

- Explain how the steady state in the Solow model relates to the CRS property
- To which extent the basic Solow model is capable of describing real world facts?
- Why can't the Solow model generate a sustained growth of per capita income?
- Is the Golden Rule saving rate an optimal saving rate?

Exercises

- 2.1. Consider an economy where the aggregate production function $Y=AF(K,N)$ exhibits Constant Return to Scale, positive and decreasing marginal productivity and unit elasticity of substitution between factors. Admitting that the saving rate, the population growth rate, technology and the rate of capital depreciation are all constant and exogenous: (a) Describe in a graph the steady state of this economy. Is it a stationary steady state? Why? (b) Describe in a graph the effects of the following changes on the

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

- long run level of per capita output: An increase in the population growth rate; an earthquake that destroys part of the capital stock. (c) Describe the effects of a rise in the saving rate in the time paths of the following variables: capital per worker; per capita income; per capita consumption. (d) Describe the effects of a rise in the level of technology on the time paths of the following variables: per capita output; capital per worker; interest rate. (e) In light of the Solow model, is there a tendency for per capita output levels in different countries to approach each other in the long term? Why?
- 2.2.** Consider an economy where the production function is given by: $Y_t = 20K_t^{1/3}N_t^{2/3}$, where N_t is the number of workers in period t . In this economy, 25% of income is saved, the labour force grows at 2.5% and capital depreciates at 2.5%. We also know that in this economy there is perfect competition, and wages and prices are fully flexible. (a) Compute the steady state values of capital and output per worker. Represent in a graph and describe the stability of the equilibrium. (b) Suppose this economy was affected by a hurricane, which reduced its capital stock. Discuss the subsequent dynamic adjustment of this economy with the help of a graph.
- 2.3.** Consider two economies, A and B, sharing the same technology, given by $Y = K^{0.5}N^{0.5}$. Assume that the saving rates in A and B are, respectively 10% and 20% and that the sum $n+\delta$ is equal to 10% in both countries. (a) Suppose that initially the capital-labour ratio was equal to 2 in both countries. What will be the corresponding initial levels of per capita consumption and per capita income? (b) Starting from the position described in a), compare the evolution of per capita income in both economies as time goes by. Discuss.
- 2.4.** Consider an economy where the production function is given by $Y_t = 0.2K_t^{1/3}N_t^{2/3}$. In that economy, 25% of income is saved, capital depreciation is 5% and population is constant and equal to 1000 inhabitants. (a) Find out the steady state values of per capita income, per capita consumption, real wages and the interest rate. (b) Find out the saving rate that would maximize C/N in *steady state*, where C is consumption. Illustrate with the help of a graph the adjustment dynamics of Y/N and C/N admitting that the saving rate actually changed to that level. (c) Suppose you were a benevolent planner with power to set a tax on production, which revenues were distributed lump sum to consumers. What would be the level of τ if you wanted to maximize the steady state level of per capita consumption? Would such policy be welfare improving?
- 2.5.** Consider an economy where the aggregate production function is given by $Y_t = A_t K_t^{1/2} N_t^{1/2}$. In this economy, the saving rate is 20%, capital depreciates at 5% per year, and population is constant and equal to $N=100$. (a) Assume that $A_t = 1$. (Find out the steady state values for: (a1) per capita income; (a2) interest rate; (a3) capital and labor income shares; (a4) wage rate. (a5) To what extent does this model comply with the Kaldor stylized facts? (a6) Represent the equilibrium in a graph and discuss its stability. (b) Sticking with $A_t = 1$, analyze graphically and quantify, when possible, the short term and long-run effects of a fall in the saving rate to $s=2.5\%$ on: (b1) per capita income; (b2) per capita consumption; (b3) the interest rate. Considering your findings, would the saving rate be a good candidate to explain: (b4) why some countries are much richer than others? (b5) long term growth? Elaborate. (c) Departing from $s=20\%$, analyze graphically and quantify, when possible, the implications of a decrease in A to $A_t = 0.125^{0.5}$, namely on: (c1) per capita income; (c2) the interest rate. Based on your

findings, could A be a good candidate to explain: (c3) why some countries are much richer than others? (c4) long term growth? (c5) What would be the theoretical problems with this explanation? (c6)

- 2.6.** Consider an economy where the labour income share is 75%. What would be the Solow residual, if both output and capital were growing at 3% per year and the labour force was expanding at 1.5%?

3 Exogenous Growth

“The Solow model did not assume that technical progress was exogenous—that is, determined outside the model. Rather, the model made the assumptions necessary to produce a model of an economy with a dynamic equilibrium, a path to which, in the long run, the economy would settle down. The implication of those assumptions was that technical progress had to be exogenous to the model”. [Lant Pritchett]

Learning Goals:

- Understand why the Solow model cannot explain technological change
- Solve the model with exogenous technological progress
- Acknowledge the extent to which the modified model helps explain real world facts
- Explain the model’ implications regarding convergence
- Understand the implications for growth accounting of using alternative definitions of technological progress

3.1 Introduction

As shown in Chapter 2, the basic Solow model does not account for the essential stylized fact of Modern Economic Growth: that output per capita tends to grow over time. This limitation was noted by Robert Solow itself in its original article, where he also provided a brief indication of how technological progress could be incorporated into the model.

This chapter shows how the Solow model can be adapted to account for the possibility of technological progress. As we will see, this modification rescues the model from its main limitation and renders it capable of describing most stylized facts of economic growth. The Chapter is organized as follows: in Section 3.2, we explain why the Solow model cannot account for endogenous technological progress. Section 3.3 presents the extended version of the Solow model with exogenous technological progress. Section 3.4 discusses how the main variables of the model adjust to changes in exogenous parameters. In Section 3.5, we show how this extended version is helpful to understand many facts of modern economic growth. Section 3.6 discusses the implications for growth accounting. Section 3.7 concludes.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

3.2 Perfect technological diffusion

Technology is different from most other goods, in that it is composed by *ideas*, rather than by *objects*. One implication of technology's non-physical nature is that its use is *nonrival*: it can be used by more than one person at the same time without losing its effectiveness. This is in sharp contrast to physical capital: if someone is using an equipment, no one else can use that equipment at the same time. In other words, equipment is "rival" in its use. This is not true for ideas and knowledge, even if they do come packaged up in bits of capital equipment: the fact that a given company uses some software to manage its operations does not preclude other firms from using the same software⁴⁸.

Technology may vary, however, in its degree of *excludability*. Excludability is the degree to which an owner of something can prevent others from using it without consent. Much of the knowledge, because of its nature, is non-excludable: it is difficult to prevent an agent from using a good idea, once he or she becomes aware of it. Still, there are ways of preventing others from using particular pieces of knowledge: for instance, trade secrets, patents and copyrights, are mechanisms through which agents try to keep competitors away from their inventions.

In this chapter, we stick to the assumption of *perfect technological diffusion*: that is, once a new technology becomes available, it becomes equally available to all agents at the same time. The reason to assume that technology spills over at no cost is that we are dealing with a model with perfect competition. Under perfect competition, all information is freely available to all agents in the economy.

The implication of assuming perfect technological diffusion is that technology becomes a *pure* public good: no user will be willing to pay for it and no self-interested agent will engage in a deliberate effort to produce it. In terms of our model, this implication is

⁴⁸ On the no rivalrous nature of knowledge, there is a famous quote of Thomas Jefferson, in a letter to Isaac McPherson, in 1813: "Its peculiar character...is that no one possesses the less, because every other possesses the whole of it. He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me".

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

rather convenient: we don't need to worry with returns to innovation or to model the research activity. The other face of the coin is that technological progress is doomed to enter in the model exogenously: since there is no reward to innovation, any technological progress must take place for non-economic reasons (such as unintended discoveries that come out through the passage of time).

3.3 The extended Solow model

3.3.1 Labour augmenting technological progress

In the model of Chapter 2, the state of technology, A , was assumed constant over time. In this chapter, it is assumed instead that technology expands over time at the constant rate, g :

$$A_t = A e^{gt} \quad (3.1)$$

With such specification, technological progress has the effect of “renumbering” the isoquants of the production function: as time goes by, each isoquant corresponds to a higher level of output than before. The “shape” of the isoquants remains unchanged: for each relative factor price, the optimal proportion in which inputs are used remains unchanged. Technological progress specified this way is labelled “Hicks Neutral”⁴⁹.

3.3.2 The two components of “technology”

Not all productivity changes are related to technology in the narrow sense: two economies may have access to the same body of technical knowledge but may follow different models of social organization. These differences will show up as differences in TFP in a production function with capital and labour. To account for these different aspects of

⁴⁹ Technically, technological progress is said to be “Hicks Neutral” if the Marginal Rate of Technical Substitution remains unchanged, for each given capital-labour ratio.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

“technology”, equation (3.1), includes two components: a constant parameter (A), and a term that expands over time.

The first component, A, shall be interpreted as capturing the influence of factors that affect the *level* of productivity. For instance, a country climate may influence the overall relationship between inputs and output, for each level of technology. We may also interpret this component as capturing aspects of economic “efficiency”, such as those related to taxation, externalities, bureaucracy, and the quality of institutions. These factors influence the effectiveness with which inputs to production *and* a given state of technology are combined to produce output. The second component – which grows continuously over time – is thought to capture the role of technological progress, in the engineering sense.

3.3.3 Other assumptions

The remaining assumptions of the model are the same as in the basic Solow model. For your convenience, we reproduce the main equations here:

$$Y_{it} = A_t K_{it}^{\beta} N_{it}^{1-\beta} . \quad (2.3)$$

$$sY_t = I_t \quad (2.7)$$

$$\dot{K}_t = I_t - \delta K_t \quad (2.8)$$

$$n = \dot{N}_t / N_t \quad (2.9)$$

3.3.4 Labour in efficiency units

Since technological progress (3.1) causes the production function (2.3) to shift upwards continuously over time, solving the model in this new version is not as straightforward as it was in the basic formulation. But with the help of a small trick, we can turn the new formulation similar to the previous one. The trick is to re-write the model in terms of a new variable, L, defined as “labour in efficiency units”.

Substituting (3.1) in (2.3) and aggregating across firms, we can rewrite the aggregate production function in the following convenient form:

$$Y_t = A K_t^{\beta} L_t^{1-\beta} , \quad (3.2)$$

where:

$$L_t = N_t \lambda_t \quad (3.3)$$

$$\lambda_t = e^{\gamma t}, \text{ with } \gamma = g/(1 - \beta). \quad (3.4)$$

In (3.2), the term L measures labour in “efficiency” units (i.e, the number of workers adjusted for their – time varying - efficiency level). The term λ refers to the “effective labour input per worker”.

Under the assumptions above, the “effective labour input per worker” grows at an exogenous rate, γ . That is, as time goes by, the typical worker becomes more efficient because new abilities are costlessly bestowed upon him at the rate γ . The rate γ is labelled the “Harrod neutral” or “Labour Augmenting rate of technological progress”. It is called *Labour Augmenting* because, analytically, it produces the same effect in production of an increase in raw labour, N^{50} .

Note that, with the transformation above, the production function (3.2) gets a form similar to that of (2.1): the main difference is that we replaced N by L . This similarity is the trick we needed to return to the “previous problem”, which we already know how to solve.

3.3.5 The Fundamental Dynamic Equation revisited

To solve the model, we make use of a new variable, $\tilde{y} = Y/L$, which will be labelled as output “per unit of efficiency”. Remember, however, that our variable of interest – the one that measures economic progress – is per capita income, $y = Y/N$. Using (3.3), the relationship between the two variables is:

⁵⁰ Technically, technological progress is said to be “Harrod Neutral” if does not alter the shares of labour and capital on income, for each given capital-labour ratio. When the production function is a Cobb-Douglas, the shares of labour and capital are constant and equal to their elasticities in production, $1-\beta$ and β , respectively. Hence, any Hicks neutral technological progress will also be Harrod neutral. For a given rate of Hicks neutral technological progress (g), the equivalent rate of Harrod neutral technological progress (γ) is larger. The reason is that, in the later case the burden of technological progress is carried by one factor, only..

$$y_t = \tilde{y}_t \lambda_t = \tilde{y}_t e^{\gamma t} \quad (3.6)$$

Dividing all terms in (3.2) by L , one obtains the production function in the *intensive* form:

$$\tilde{y}_t = A \tilde{k}_t^\beta, \quad (3.7)$$

where and $\tilde{k} = K/L$ denotes physical capital “per unit of efficiency”.

Taking time derivatives in $\tilde{k}$ and using (2.7), (2.8), (2.9), and $\dot{L}/L = \dot{\lambda}/\lambda + \dot{N}/N = \gamma + n$, the modified version of the Fundamental Dynamic Equation results as follows:

$$\dot{\tilde{k}}_t = s A \tilde{k}_t^\beta - (n + \delta + \gamma) \tilde{k}_t. \quad (3.8)$$

The two terms in the right-hand side of (3.8) are depicted in Figure 3.1, together with equation (3.7). The first term measures gross investment per unit of efficiency labour. The second term gives the “break-even investment”, that is, the one that would be necessary to compensate for the “depreciation” of $\tilde{k}$. Note that the later includes the depreciation of physical capital and the growth rate of “effective” labour, $n + \gamma$.

Apart from the way the endogenous variable is defined, the interpretation of equation (3.8) is the same as that of the corresponding equation in the basic Solow model, (2.14): in brief, $\tilde{k}$ rises whenever gross investment per unit of efficiency is higher than the break-even investment - as in $\tilde{k}_1$ of Figure 3.1 - and conversely. The model of Section 2 is a particular case of the model developed in this section, with $\gamma=0$.

3.3.6 The steady state

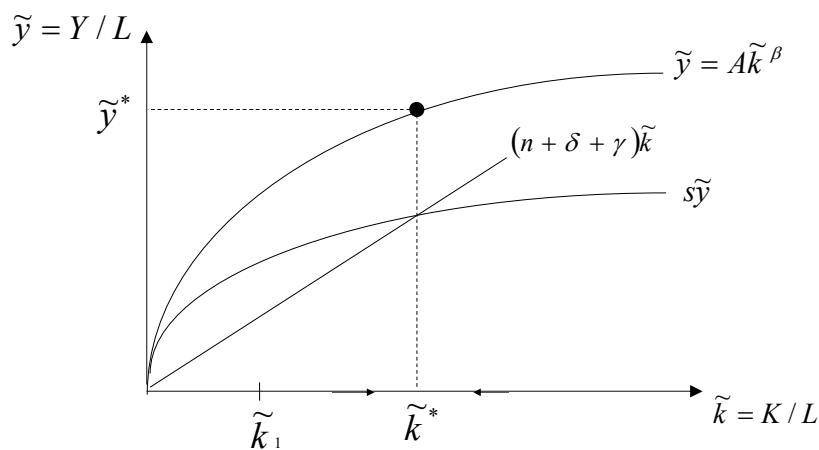
The steady state of the model is obtained setting $\dot{\tilde{k}} = 0$ in (3.8). Using $k_t = \tilde{k}_t e^{\gamma t}$, $y_t = \tilde{y}_t e^{\gamma t}$, we find the paths of capital per worker and per capita income in the steady state:

$$k^* = \left(\frac{sA}{n + \delta + \gamma} \right)^{\frac{1}{1-\beta}} e^{\gamma t} \quad (3.9)$$

$$y_t^* = A^{\frac{1}{1-\beta}} \left(\frac{s}{n + \delta + \gamma} \right)^{\frac{\beta}{1-\beta}} e^{\gamma t} \quad (3.10)$$

Equation (3.10) states that, in the steady state, per capita income grows continuously at the rate γ . How can that be? The labour augmenting technological progress has the effect of neutralising the diminishing returns to capital: by economising progressively the input whose supply cannot be changed – labour – technological progress allows effective labour to increase at the same rate as the number of machines (capital) and at the same time the number of machines per worker increases at rate γ .

Figure 3.1. The Solow model with technological progress



The figure depicts the equilibrium of the Solow model with technological progress. The vertical and horizontal axes measure output and capital per labour in efficiency units. The equilibrium is found when savings per unit of efficiency labour is equal to break-even investment.

3.4 Transitional Dynamics

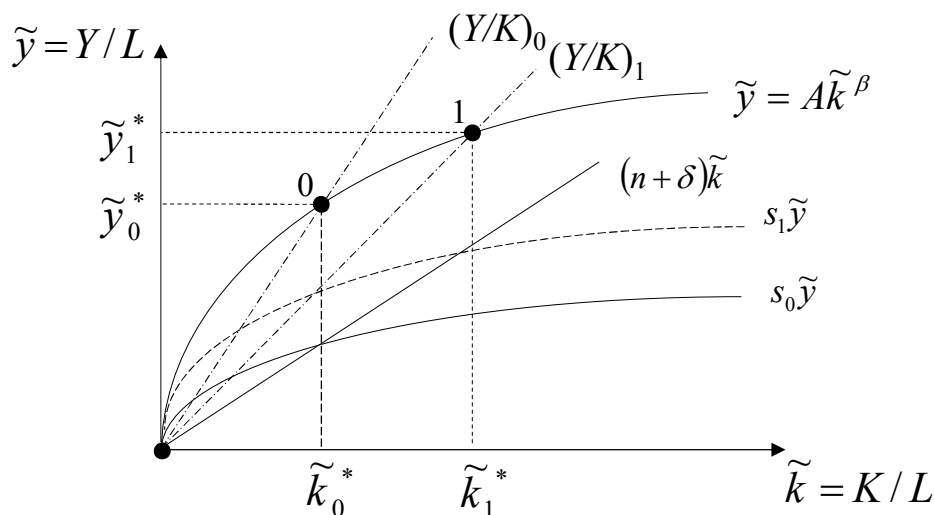
3.4.1 What happens if the Savings Rate increases?

In figure 3.2 we examine the implications of a once-and-for-all increase in the saving rate. The rise in the saving rate shifts the steady state to the right, from point 0 to point 1. In the new steady state the average product of capital, Y/K , is lower than in the initial steady state. This means that the interest rate has declined from one steady state to the other (equation 2.12).

Because capital accumulation is bounded each moment in time by the availability of savings, the economy does not jump immediately from point 0 to point 1: it slowly converges to the new steady state. Figure 3.3 describes the time paths of the main variables of the model, following a rise in the saving rate. The top panel depicts the evolution of output per unit of efficiency labour ($\tilde{y} = Y/L$) and capital per unit of efficiency labour ($\tilde{k} = K/L$). The Middle panel depicts the paths of output per capita ($y=Y/N$) and of per capita consumption ($c=C/N$) – these are displayed in logs, so as to stick with linearity. The bottom panel depicts the paths of the interest rate (r) and of the growth rate of per capita consumption (γ).

Let t_0 be the moment at which the saving rate raises to the new level. Assume that in the moment just before, the economy was in a steady state, with a constant level of output per unit of efficiency and with per capita income growing at the exogenous rate γ (like in point 0 in figure 3.2). Since capital and labour are both pre-determined, at the time of the shock (t_0) per capita income remains initially unchanged. Per capita consumption, however, falls at the impact, because a higher proportion of income is devoted to savings.

Figure 3.2. A higher saving rate raises the steady state level of output per unit of efficiency units



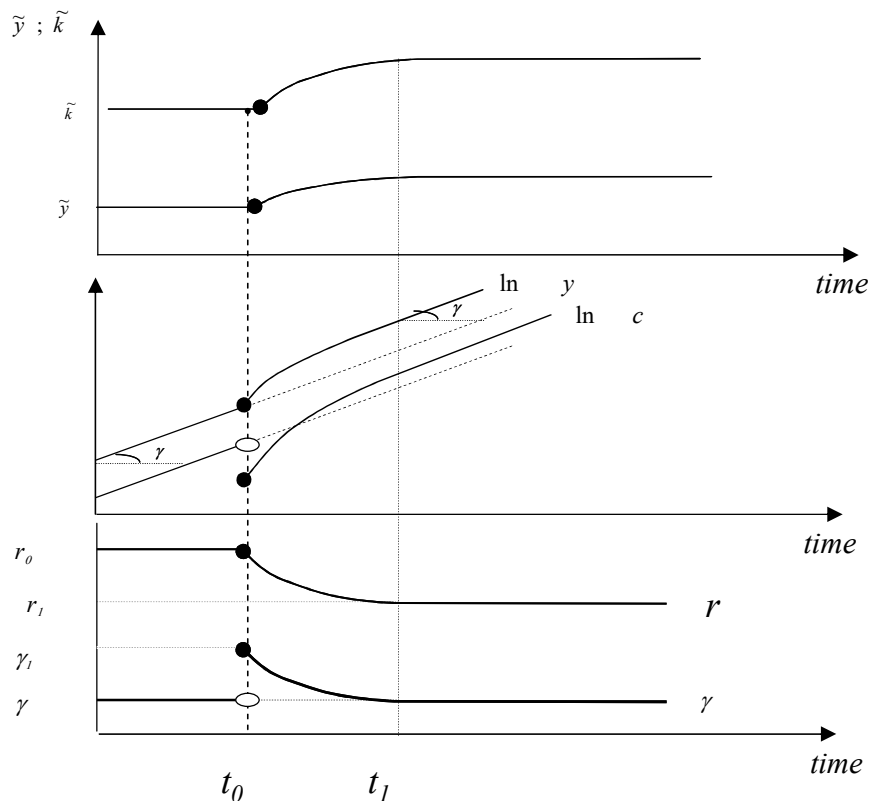
The figure shows how the equilibrium of the model changes with an increase in the saving rate. In the new steady state (point 1), both capital and output per unit of efficiency labour are higher than before, and the average product of capital (Y/K) is lower, due to diminishing returns.

The adjustment process takes place between t_0 and t_1 . Since savings become temporarily greater than the break-even investment, both K/L and Y/L start increasing. This

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

means that the growth rates of output per capita and of per capita consumption jump temporarily ahead of the rate of technological expansion, γ . As the economy approaches the new steady state, diminishing returns show up, implying that the growth rate of output per capita falls back, approaching the exogenous rate, γ . The new steady state occurs when output per capita grows at the same rate as technology, $\gamma_t = \gamma$.

Figure 3.3 – Implications of a rise in the saving rate



The figure describes the time paths of output and capital per efficiency units, per capita income and consumption, the interest rate and the growth rate of per capita output, following a once-and-for all increase in the saving rate. At the impact per capita consumption declines, reflecting the increased savings. After the shock, per capita income accelerates, expanding temporarily faster than technology. As time goes by, the growth rate of per capita income decreases until meeting the rate of technological expansion in the new steady state. From one steady state to the other, the interest rate declined, reflecting the higher abundance of capital relative to the previous path.

As in the case without technological progress, in the new steady state (after t_1), the consumption path may be above or below the original consumption path. Referring to our discussion in Section 2.5, this will depend on how the new and the old saving rates compare to the Golden Rule saving rate. Figure 3.6 depicts the special case, in which the rise in the saving rate leads to a higher level of per capita consumption in the steady state. You may easily verify that the Golden Rule saving rate in this version of the model is the same as <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

before: it corresponds to the share of capital in total income, β . Also remember that an increase in the savings rate towards the golden rule is not necessarily welfare improving.

Box 3.1. Growth effects and level effects

At this stage, it is important to introduce the distinction between *level effects* and *growth effects*:

- A level effect occurs when changing a model's parameter changes the steady state without affecting the growth rate of the economy in the steady state.
- A growth effect occurs when a change in a parameter alters the growth rate of the economy in the steady state.

In the Solow model, changes in parameters like the saving rate and the population growth rate produce level effects, only. A growth effect could only occur if the exogenous rate of technological progress was changed.

3.5 The extended Solow model meeting the real-world facts

3.5.1 Revisiting the Kaldor's facts

As discussed in Chapter 2, the main limitation of the simpler version of the Solow model is that it cannot account for two stylized facts of economic growth, namely that per capita output and capital per worker tend to grow over time (Kaldor's facts 1 and 2). As shown in equations (3.9) and (3.10), the assumption that technology expands continuously brings the model into compliance with these two facts.

As before, the model is consistent with the evidence that the shares of labour and capital on national income tend to be constant over time (Kaldor's fact 5). Note that firms take technology as given, so they maximize profits with respect to K and N , as in the simple Solow model. Hence, conditions (2.11-2.12) also hold in this more sophisticated version of the model. The stylized fact 5 is a direct consequence of assuming CRS and perfect competition.

Dividing (3.10) by (3.9) we obtain the expression for the average product of capital in the steady state:

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

$$\left(\frac{Y}{K}\right)^* = \frac{n + \delta + \gamma}{s} \quad (3.11)$$

Since all parameters at the right-hand side of (3.11) are constant, this means that the equilibrium average product of capital is constant over time. As a by-product – and using equation (2.12) – it follows that the interest rate is constant in the steady state. This is the Kaldor fact number 4.

A novelty with the augmented model is that real wages increase over time: since the wage rate is proportional to per capita output (equation 2.11), in the steady state wages are increasing at rate γ . This is another feature of the augmented model that makes it more compliant with the facts of Modern Growth.

Finally, we turn to fact number 6 (“There are wide differences in the growth rate of productivity across countries”). As long as we stick to the assumption of perfect technological diffusion, the model predicts that *in the long run all countries should be growing at the same rate, γ* . That is, the steady states of the different countries should be characterized by per capita incomes evolving in *parallel* over time.

Note however that such conclusion only holds in the long run: equation (3.10) refers to the steady state, it is expected to hold only for countries that already adjusted fully to changes in their exogenous parameters. For countries that are engaged in a *transitional dynamics*, the current growth rate of per capita income may be higher or lower than γ , depending on whether the starting point is below or above the corresponding steady state: countries that start out below (at the left of) their respective steady states are expected to grow faster than countries that start out above their steady states. Hence, growth rates may differ considerably across countries each moment in time. This is consistent with the Kaldor’ fact number 6.

3.5.2 Absolute convergence

A question that has received considerable attention in the economic profession is whether there is a general tendency for poor countries to grow faster than rich countries. At least since David Hume⁵¹, economists have been arguing that technological spillovers, diminishing returns and capital mobility provide poor countries with an impetus to “catch up”. The hypothesis that poor economies tend to grow faster than rich economies is known as *absolute convergence*.

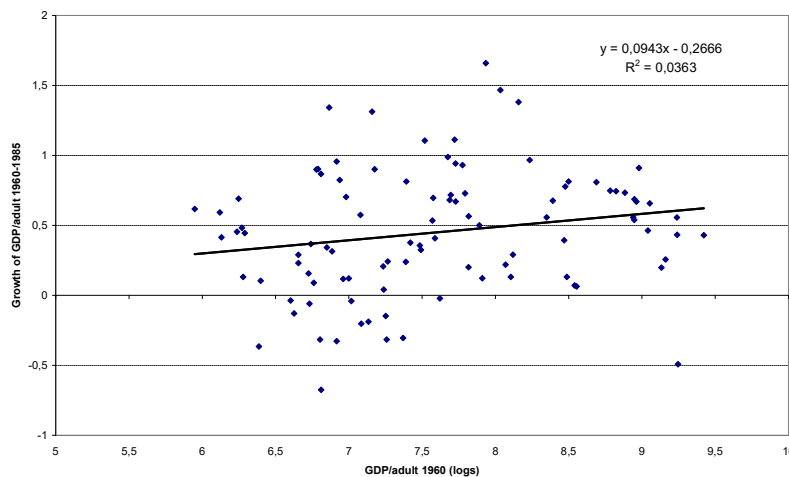
A simple way to investigate the convergence hypothesis using cross-sectional data is by plotting growth rates of per capita incomes against the initial levels of per capita incomes, for different countries along a period of time: if there was a general tendency for per capita incomes to approach each other, then poorer countries should grow faster than richer countries. Figure 3.4 illustrates such an exercise, using a sample of 98 non-oil countries. The figure relates the growth rate of GDP per working age person from 1965 to 1985 (vertical axis) with the corresponding 1965 level. If there was a general tendency for poor countries to grow faster than rich countries, the slope of the regression line should be negative. However, this is not the case. The conclusion is that “absolute convergence” does not hold as a general rule in this sample of countries during this time period.

Note however that this evidence does not contradict the Solow model: *the Solow model does not imply that countries should converge to same level of per capita output*. As stated in equation (3.10), countries differing in terms of the fundamental parameters (those that determine the steady state, such as the saving rate and the population growth rate) are expected to reach different steady states.

Figure 3.4: Evidence of Non-Convergence in 98 Countries

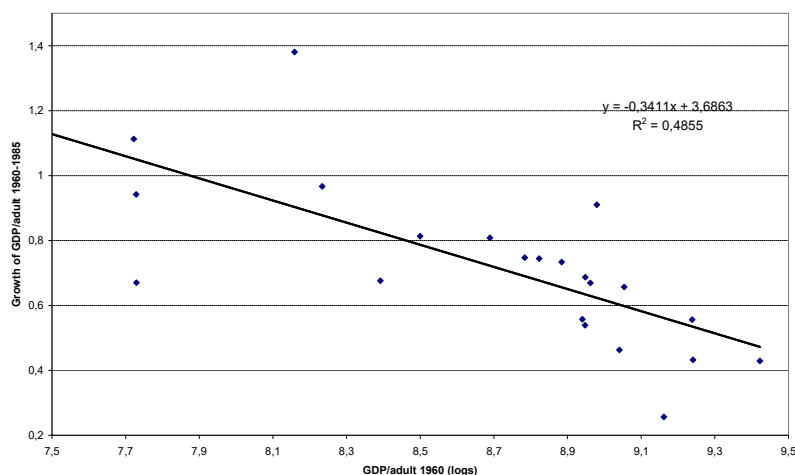
⁵¹ Hume, D., 1758. *Essays and Treaties' on Several Subjects*. London: A. Millar.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>



The figure crosses growth rates of per capita incomes with initial levels of per capita incomes. The fact that no negative correlation is found implies that there is no systematic tendency for initially poorer countries to grow faster than rich countries. Source: Mankiw, G., D. Romer and D. Weil, 1992. "A contribution to the empirics of economic growth", Quarterly Journal of Economics, 107 (2), 407-38

Figure 3.5: Evidence of Absolute Convergence among 22 OECD Countries



The figure crosses growth rates of per capita incomes with initial levels of per capita incomes, restricting the sample to OECD economies. The negative correlation obtained implies that in this particular sub-sample there has been a tendency for initially poorer countries to grow faster than richer countries. Source: Same as figure 3.4

Of course, economies that are similar in terms of the fundamental parameters, such as the saving rate and the population growth rate are expected to approach steady states that are close to each other. Thus, if one restricts the sample to countries that are similar, one may well observe that those with initially lower levels of per capita income grow faster than those with higher per capita incomes. An example of this is displayed in Figure 3.5. This figure restricts the sample of Figure 3.4 to OECD countries, only. In this particular sub-sample, we <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

are able to identify a negative correlation between growth rates and initial per capita incomes. This is not to say that all these countries are converging exactly to the *same* steady state: what happens is that, in this particular sub-sample, departures from steady state (i.e, transitional dynamics) account for a larger share of the cross-country variation of per capita incomes than differences in the steady states, which are arguably small. Note that the data refers to the post-WWII period, during which many European countries were rebuilding their capital stocks⁵².

Summing up, in light of the Solow model, it will be impossible to predict whether a country will grow faster or slower by observing its initial income relative to other countries. In light of the Solow model, it is not the initial income that determines a country' growth rate, but instead its *distance relative to the steady state*: economies with per capita incomes that fall behind their steady states should grow faster than economies with per capita incomes that are above the respective steady states. This property of the model, is known as *conditional convergence* and will be subject to further scrutiny in the next chapter.

Box 3.2. Explaining the convergence test

In figure 3.4, the convergence hypothesis was investigated plotting the growth rates of per capita income against in a period, against the initial levels of per capita incomes. Formally, the hypothesis of “absolute convergence” can be tested empirically by estimating the following regression equation:

$$\ln y_{it} - \ln y_{i0} = a + b \ln y_{i0} + \varepsilon_i \quad (3.13)$$

⁵² Evidence of a negative relationship between per capita income growth and the initial level of per capita income is often found in samples restricted to industrial countries or their regions (for instance, Baumol, 1986, Barro e Sala-i-Martin, 1991), but this evidence reveals fragile to small sample modifications (De Long, 1988). Baumol, W., 1986. “Productivity Growth, Convergence and Welfare: What the Long Run Data Show”, American Economic Review 76 (5), 1072-1085. De Long, J., 1988. “Productivity Growth, Convergence and Welfare: Comment”. American Economic Review 78 (5), 1138-1154. Barro, R. and Sala-i-Martin, X., 1991. “Convergence across states and regions”. Brooking Papers on economic activity 1, 107-158.

where the dependent variable is the growth rate of per capita GDP between period zero and period t in country i and the regressors are: a constant (a) and the initial level of per capita GDP in country i ($\ln y_{i0}$). The term ε_i is a random disturbance. In (3.13), the “absolute” convergence hypothesis is assessed by investigating the sign and significance of b : If $b < 0$, this means there is a general tendency for initially poor economies to grow faster than rich economies.

To see how this test relates to the Solow model, consider the following equation, describing the dynamics of per capita income towards the steady state (see Appendix 3.1 for details):

$$\ln y_t - \ln y_0 = \pi + (1 - e^{-\nu})(\ln \tilde{y}^* - \ln y_0), \quad (3.14)$$

with $\nu = (1 - \beta)[n + \delta + \gamma] > 0$.

Equation (3.14) states that the growth rate of an economy depends on its distance relative to the steady state: if the economy starts out in the steady state, the expected growth rate is γ ; if the economy is below (above) the steady state, its growth rate will be higher (lower) than γ . In general, this equation states that *per capita income converges to a steady state and the speed at which it does so relates inversely to the initial distance to the steady state*. This is what we mean by *conditional convergence*.

The relationship between the parameters of the regression equation (3.13) and those of the structural relationship (3.14) is straightforward:

$$a = \pi + (1 - e^{-\nu}) \ln \tilde{y}^* \quad (3.15)$$

$$b = -(1 - e^{-\nu})$$

Thus, a regression equation of the form (3.13), by postulating the same intercept to all countries, implicitly imposes that steady states are the same. With no surprise, tests for *absolute* convergence perform very poorly in World-wide samples, where dramatically different countries are pooled together in a regression equation.

To overcome this limitation, many studies have allowed the intercept (the steady state) to differ across countries. This is done adding to the regression model variables that are thought to determine the steady state ($\ln \tilde{y}^*$), such as the saving rate, the population growth

rate, and efficiency, λ (equation 3.10). We will discuss these tests of *conditional* convergence in Chapter 4.

Box 3.3. How Long?

The term $\nu = (1 - \beta)[n + \delta + \gamma] > 0$ in equation (3.14), measures the *speed* of adjustment of per capita income to the steady state. To quantify this, let's calibrate this equation with reasonable parameters:

- According to the Solow model, the elasticity of labour in the production function can be assessed by the (observable) share of labour on national incomes. National accounts data for different countries reveal that the labour share in national income varies from 60% to 70%. Thus, a reasonable assumption for the elasticity of capital in the production function is $\beta = 1/3$.
- In the Solow model, the rate of technological progress, γ , is equal to the growth rate of per capita output in the steady state. The long-run evidence for industrialised countries reveals that per capita incomes have evolved at an average rate around $\gamma = 2\%$. A popular assumption in the literature is to consider $\delta + \gamma = 0.05$.
- The population growth rate varies considerably across countries. As an example, consider a population expanding at 1% per year.

Under the above assumptions, the speed of convergence towards the steady state will be $\nu = 0.04$. With such a value, how long it will take for that country to get “halfway” to its balanced growth path? Using equation (3.14), the answer is $e^{-0.04t} = 0.5$, which solves for $t = 17$ years.

3.5.3 Interpreting growth patterns using the Solow model

By now, we have been confronting the Solow model with stylized facts referring to samples of countries. A different question is whether the model is helpful to interpret specific growth patterns of individual countries. To address this question, let's consider Figure 3.6, that depicts the evolution of per capita incomes in some *advanced* economies, namely the United States, France, Japan and Germany, over the period from 1871 to 2001.

We observe that:

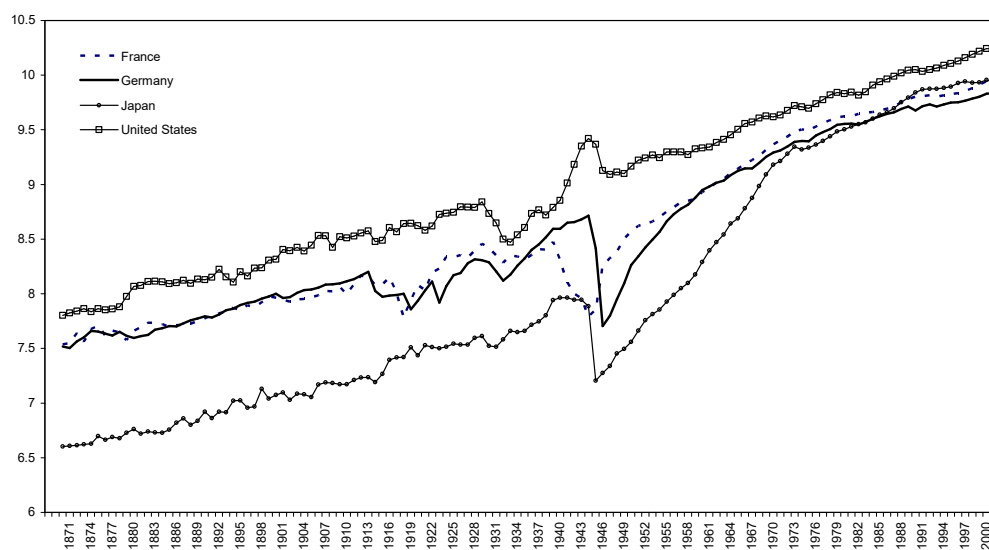
<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

- In all cases, per capita GDP exhibits an upward trend. This is accounted for by the Solow model with technological progress.
- The growth rates of per capita output appear to be stable in the long run and quite similar across countries. This suggests that the draconian assumptions of technology evolving at a constant rate g and with all countries drawing from a *common technological pool* may not be at odds with reality in this very *particular sample*, composed by a group of advanced countries⁵³.
- The long-term paths of per capita incomes are parallel but not coincident. The Solow model does not imply that steady states should be the same for all countries: according to equation (3.10), the level of per capita income in each country depends on the saving rate (s), the population growth rate (n) and efficiency (A).
- During the long period from 1870 to 2011, some major disruptions pushed the US, France and Germany away from their respective long-run paths. These events included the Great Depression in the 1930s, the First World War (1914-1918) and the Second World War (1939-1945). As time went by, per capita incomes look like having return to the earlier paths. According to the Solow model, the steady state levels of per capita output are independent of a country's initial capital endowments. So, if some disaster destroys part of the capital stock, per capita GDP will fall initially, but then it will recover until returning to the initial steady state. In Figure 3.6, we see that this prediction of the model fits quite well the cases of US, Germany and France. For instance, during WWII, per capita output in Germany and France dropped significantly, but this was followed by a fast recovery that brought these economies back to the earlier path.

⁵³ Note that this set of countries is very specific: they share a set of characteristics that make them permeable to technological innovations discovered by each other. Many other countries in the world will hardly draw so easily from this "technological pool".

- In the case of Japan, a “level effect” is likely to have occurred after the Second World War. The path of Japan points indeed to a distinct case from those of, Germany and France: after WWII this country seems to have moved from a lower steady state to a higher one, closer to that of United States. In light of the Solow model, such move could be explained by a change in a fundamental parameter, such as the saving rate, the population growth rate or other country-specific effects, as captured by the efficiency parameter, A . Any change in one of these exogenous parameters implies a change in a country steady state.

Figure 3.6- Per Capita GDP in Japan, France, Germany and US, 1871-2001



The figure displays the evolution of per capita incomes (in logs) in a sample of industrial countries. The fact that per capita incomes evolve basically in parallel is consistent with the idea that these countries share the same body of technological knowledge. This figures also suggest a tendency for these countries to return to respective balanced growth paths after major disruptions. The exception is Japan, that managed to achieve a “level effect” after WWII. Source: Maddison, A., 2001. *The World Economy: a Millenial Perspective*. Development Centre, Paris.

3.6 Growth accounting revisited

3.6.1 Autonomous versus induced contributions

Having in mind the extended version of the Solow model, we now revisit the growth accounting exercise introduced in Section 2.7. As for an illustration, consider the figures of US economic growth. According to the World bank (1991)⁵⁴, during the first half of the twentieth century, the growth rate of GDP was about 3% per annum, on average. Estimates also indicate that in the same period, the capital stock expanded at about 3% per annum, whereas labour (hours worked) expanded at 1% per annum, only. Assuming a labour share in national income of two thirds, the implied Solow residual (using 2.29) is:

$$\hat{A} = 0.03 - \left(\frac{1}{3}\right)0.03 - \left(\frac{2}{3}\right)0.01 = 0.013.$$

This means that labour and capital jointly accounted to about 1.7 p.p. per annum to GDP growth. The residual balance of 1.3 percentage points per annum is accounted for by “technological change”. Using the intensive form (2.30), the conclusion is even more startling: 1/3 of the change in per capita GDP is accounted for the increase in the capital labour ratio, and the remaining 2/3 of the change is accounted for TFP:

$$\hat{A} = 0.02 - \left(\frac{1}{3}\right)0.02 = 0.013.$$

The fact that both capital and output expanded at around 3% is however consistent with the view that this economy evolved along a balanced growth path (with Y/K constant, and hence with a constant saving rate). But if the US economy was indeed evolving along a balanced growth path, why should a growth accounting exercise like (2.29) indicate a contribution of capital to GDP growth equal to 1%?

⁵⁴ World Bank, World Development Report 1991, Washington..

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

To answer this question, note that the Solow model with exogenous growth predicts the capital stock to grow over time in the steady state at the rate $\dot{K}/K = \gamma + n$. This growth, however, is not autonomous (i.e, it is not implied by a change in the saving rate); it is instead an endogenous response to the expansion of the effective labour force: if there was no population growth or technological progress ($\gamma=n=0$), then the capital stock would be constant, unless the saving rate had recently increased.

Taking this into account, we conclude that growth accounting exercises based on equation (2.29) measures the total contribution of capital to growth, without disentangling whether the observed growth in capital is induced by an increase in the saving rate (transitional dynamics), or is a mere response to a growing effective labour. When the economy is in the steady state, for instance, such an exercise will attribute $\beta(\gamma + n)$ to the growth of capital, $(1 - \beta)n$ to the growth of population and $g = \gamma(1 - \beta)$ to technological progress. And yet, these parameters are zero if there is no population growth or technological progress.

3.6.2 An alternative approach

To disentangle whether a country growth process is mainly driven by transitional dynamics or by steady technological change, there is a growth accounting exercises based on the following re-arrangement of the production function, (3.7)⁵⁵:

$$y_t = A_t^{\frac{1}{1-\beta}} \left(\frac{K_t}{Y_t} \right)^{\frac{\beta}{1-\beta}} \quad (3.16)$$

⁵⁵ David, P., 1977. Invention and Accumulation in America's Economic Growth: a Nineteen century parable, Journal of Monetary Economics, Special Supplement VI, 176-228.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

This rearrangement emphasizes the relationship between the capital output ratio and per capita income. In light of the Solow model, in a steady state, this ratio changes when the saving rate changes, but is independent of the technological level, A (equation 3.11).

Log-differentiating this, we get

$$\hat{y} = \frac{1}{1-\beta} \hat{A} + \frac{\beta}{1-\beta} (\hat{K} - \hat{Y}) . \quad (3.17)$$

In (3.17) the contribution of capital to the growth rate of per capita GDP is now evaluated by the extent to which its growth rate exceeds that of output growth. This decomposition actually *expurgates* the growth rate of the capital stock from the part that is induced by the exogenous parameters. Thus, growth accounting based on equation (3.17) will capture the contribution of capital, *only* to the extent that the country is involved in a process of transition dynamics.

Using this new approach and the same figures for the US economy, one obtains:

$$2\% = 2\% + \frac{1}{2}(3\% - 3\%) = \gamma$$

That is, along the first half of the last century, capital accumulation by itself did not account for the US per capita income growth: the only exogenous source of per capita income growth along this period was (Harrod neutral) technological change, which expanded at 2%, on average. The capital stock evolved at a rate of 3% per year, but this was merely a response to the increasing population and TFP growth.

Box 3.4. The TFP controversy in East Asia

There are many examples in the literature of practical applications of growth accounting. A particularly controversial application refers to the so-called Asian Tigers (Hong Kong, Singapore, South Korea and Taiwan). Economists have for many years turned to these successful countries for clues that can explain sustained development and so provide

examples to the rest of the world. In particular, in 1991, the World Bank came down strongly in favour of the idea that the East Asian economies, by operating sounder domestic policies than other developing countries, relied mostly on TFP growth: almost 2 percentage points of its overall growth of almost 7 per cent per annum⁵⁶. By contrast, TFP growth rates in both Africa and Latin America were almost zero in that same period.

In an influential 1995 article, Alwyn Young dissented from this view. The author found that the bulk of the impressive growth achieved in the four East Asia miracle countries in the period 1966-1990 could be attributed much more narrowly to their fast rates of factor accumulation, and not to their exceptional levels of TFP growth. Columns (1), (2) and (3) of Table 3.1 present a summary of the Young estimates. These are obtained using a decomposition similar to (2.29), with the difference that the author aggregated raw labour and education levels into a single measure of labour. The data in column (3) indicate that TFP growth accounts for only a small proportion of GDP growth in these countries. Singapore for example was a particularly bad performer, with TFP growing at a rate close to zero. Young concluded that these growth miracles have achieved very high growth rates because of their ability to achieve high investment rates (in physical capital and in human capital), and a drastic increase in the fraction of population at work (largely via the increased labour force participation of women). Paul Krugman popularise these results contending that the key for success in the Asian Tigers was "transpiration" (factor accumulation) rather than "inspiration" (technological change)⁵⁷.

The implication of this finding for the economic profession was obviously disappointing: if technological catch up played only a minor role in East Asia's growth, this means that traditional factor accumulation was relatively more important than the quest for policies that encourage economic efficiency and innovation.

⁵⁶ World Development Report 1991. Oxford University Press, New York.

⁵⁷ Krugman, P. , 1994. "The myth of Asia's Miracle", *Foreign Affairs* 73(6), Nov-Dec, pp.62-78.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

With no surprise, such controversial conclusion became subject to further scrutiny in the years that followed. One avenue that was explored relates to our discussion in Section 3.9: traditional growth accounting tends to overstate the role of factor accumulation because it does not control for the component of factor accumulation that is merely induced by technological change. Peter Klenow and Rodriguez Clare stressed this point and computed the Harrod neutral rates of technological progress *implied by the Young estimates* of Total Factor Productivity. Their results are reproduced in column (4) of Table 3.1. Clearly, the adjusted measures of TFO growth are higher than those in Column (3).

Table 3.1. - Productivity growth in East Asia

	Young (1995)				Klenow and Rodriguez-Clare (1997)		
	Y (1)	Y/N (2)	TFP (3)	Harrod Neutral (4)	Y/N (5)	Harrod Neutral (6)	Rank (98 countries) (7)
Hong Kong	7.3	4.7	2.3	3.7	5.5	4.4	4
Korea	10.3	4.9	1.7	2.5	5.4	2.5	17
Singapore	8.7	4.2	0.2	0.3	5.1	3.3	6
Taiwan	9.4	4.8	2.6	3.5	5.3	3.0	7

Notes: Columns (1) - (3) display average growth rates for the period 1966-1990 (1966-1991 in Singapore). Columns (4)-(6) are from Klenow and Rodriguez Clare (1997) refer to the period 1960-85. Column (7) displays the rank order of estimates in Column (6) in a sample of 98 countries. Sources: Young, A ., 1995. "The tyranny of numbers: confronting the statistical realities of the East Asian growth experience", Quarterly Journal of Economics CX (3), 641-680. Klenow, P. and Rodriguez-Clare, "The Neo-Classical Revival in Growth Economics: Has it Gone Too Far?", NBER Macroeconomic Annual, 1997.

A second problem relates to data measurement. In general, growth accounting exercises are very sensitive to key assumptions, such as the weights assumed in the production function, and the treatment of human capital. To illustrate how sensitive estimates are to changes in methodology, we display in columns (5) and (6) of Table 3.1 the estimates proposed by Klenow and Rodriguez Clare for output per capita and for (Harrod neutral) productivity growth in these countries. Although the estimates for South Korea and Taiwan

roughly match those of Young, estimates for Singapore and Hong Kong are much higher. In the case of Singapore, the difference is qualitatively important⁵⁸.

The third caveat to the Young results is that the contribution of TFP growth should be evaluated *per se* and not as a proportion of output growth. In light of the neoclassical model, for any two countries enjoying the same rate of technological progress, the one that starts with a lower capital labour ratio should exhibit faster growth. Thus, a lower *proportion* of output growth will be accounted for by improvements in TFP, even though both countries share the same rate of technological progress. Column (7) displays the rank order of the Harrod neutral rate of technological progress estimated by Klenow and Rodrigues-Clare in their 98 countries sample. Clearly, the TFP growth rates in the four East Asian economies are quite respectable. In general, estimates of TFP growth using comparable data for a large set of countries reveal that East Asian productivity growth is relatively high when compared to other regions.

All in all, despite the initial controversy, the evidence points to the case that TFP growth (technology and aggregate efficiency) has indeed played a much greater role in the economic transformation of East Asia than the original Young estimates suggest.

3.7 Discussion: What we have achieved?

In Chapter 2 we saw that, because of diminishing returns, the basic Solow model is not capable of describing the real-world fact that per capita income tend to increase over time in economies engaged in modern growth.

In this chapter, we saw that assuming an exogenous rate of technological progress we turn the model capable of describing most stylised facts of economic growth. In particular, the model conciliates a sustained growth of per capita income with an interest rate that

⁵⁸ The authors attributed the bulk of the difference to different assumptions regarding the capital income share (they used 0.30, instead of 0.48 in Young, 1995) and also to the different data set used for employment growth.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

remains constant over time. With this refinement, the model becomes much more compliant with real world facts. In plus, we saw that the Solow model is consistent with the evidence that poor countries do not tend to grow faster than rich countries. In a word, the Solow model can be fairly said to provide the correct answers to the set of questions it was intended to address.

The main drawback of the model is that it cannot address the causes of technological progress. The model describes how economies evolve over time, and can be extended to account for the role of technological progress in delivering long-run growth. However, the model fails to explain why technological progress takes place at all. The reason for this was already discussed at the beginning of this chapter: by assuming perfect competition, the model implicitly assumes that technology is freely available to everybody, so no profit-making economic agent will find incentives to invent new technologies. Without accounting for the incentives to invent new technologies, the model is forced to assume that technology grows exogenously. Questions such as: "who produces technological progress and why" cannot be addressed by the Solow growth model.

Despite its limitations, the Solow model provides a framework that shall be seen as the centrepiece to describe the mechanics of per capita income growth over time. As such, it became the workhorse of growth models and it provides the basis for many advanced models in macroeconomics.

3.8 Key ideas of Chapter 3

- Because it assumes perfect competition and rules out market failures, the Solow model can only account for exogenous technological progress.
- The exogenous rate of technological progress allows effective labour to expand faster than population. If – as implied by the model – the capital stock expands proportionally to effective labour, then output will also expand at the same rate as effective labour. The implication is that capital per worker and per capita income will both increase at the rate given by the labour augmenting rate of technological progress.
- With this small amendment, the Solow model can account for the fact that most countries see output per capita increasing over time.
- As before, changes in the saving rates can only produce “level effects”: that is, the growth rate of per capita income increases temporarily, but in the long run it falls back to the level given by the rate of technological progress.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

- The Solow model does not imply that in the long run all countries should have the same level of per capita income (“absolute convergence”). According to the Solow model, per capita incomes differ in the steady state depending on country characteristics, such as the saving rate and the population growth rate. The Solow model implies that countries that start out behind their respective steady states should grow faster than countries that are close to their steady states. This property of the model is known as “conditional convergence”.
- The Solow model is capable of describing many stylized facts of economic growth. It fails however to explain economic growth, because the parameter that ultimately determines the rate at which per capita incomes are expanding is exogenous to the model.
- Growth accounting exercises have different interpretation depending whether the technological parameter is specified as Harrod neutral or as Hicks neutral. If one wants to capture the contribution of capital after expurgating the capital accumulation that is induced by technological change, the Harrod neutral measure should be used.

Appendix 3.1 Transition dynamics in the Solow model

The stability properties of the neoclassical growth model ensure that the economy converges to its steady state and that the speed of adjustment depends on how far the economy is from the steady state.

Formally, the speed of convergence may be assessed using a first-order Taylor-series approximation of (3.8) around the steady state (3.9). This gives:

$$\begin{aligned}\dot{\tilde{k}} &= \frac{\partial \dot{\tilde{k}}}{\partial \tilde{k}} \bigg|_{\tilde{k} = \tilde{k}^*} (\tilde{k} - \tilde{k}^*) = \left[\beta s \frac{Y}{K} - (n + \delta + \gamma) \right] (\tilde{k} - \tilde{k}^*) = -(1 - \beta)(n + \delta + \gamma) (\tilde{k} - \tilde{k}^*) \\ &= -\nu (\tilde{k} - \tilde{k}^*) \text{ with } \nu = (1 - \beta)(n + \delta + \gamma) > 0,\end{aligned}$$

where we used equation (3.11) to eliminate Y/K . The last equation is a first-order, non-homogeneous differential equation, whose solution is given by:

$$\tilde{k}_t - \tilde{k}^* = e^{-\nu t} (\tilde{k}_0 - \tilde{k}^*)$$

This equation states that the change in $\tilde{k}$ each moment in time declines as the economy approaches its steady state. When the economy is in the steady state, the second term on the right-hand side is zero, implying a constant level of $\tilde{k}$. When the economy is below the steady state, the second term on the right-hand side is positive, implying that $\tilde{k}$ is rising. This means the model exhibits *local stability*.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Since y is a continuous function of k , as a linear approximation, y approaches the steady state at the same rate as k . Then, it can also be shown that the dynamics of y in natural logarithms is given by⁵⁹:

$$\ln \tilde{y}_t - \ln \tilde{y}^* = e^{-\nu} [\ln \tilde{y}_0 - \ln \tilde{y}^*]$$

Thus, the speed of adjustment of per capita income to the steady state is given by $\nu = (1 - \beta)[n + \delta + \gamma] > 0$. This equation suggests a natural regression to study the rate of convergence in the context of the Solow model: subtracting $\tilde{y}_0$ in both sides, rearranging and using the identity $\ln y_t = \ln \tilde{y}_t + \gamma t$ to eliminate $\tilde{y}_t$ and $\tilde{y}_0$, one obtains (3.14).

Problems and Exercises

Key concepts

- *Perfect technological diffusion. Labour in efficiency units. Harrod neutral vs Hicks neutral technological progress. Level effect vs. growth Effect. Absolute Convergence. Conditional Convergence.*

Essay questions:

- Comment: “In the Solow model, technology *has to* be assumed exogenous”.
- Comment: “The Solow model does not imply that poor countries should grow faster than rich countries”

Exercises

- 3.1.** Consider an economy where the production function is given by: $Y_t = A_t K_t^{1/3} N^{2/3}$, where $A_t = 16e^{0,02t}$ describes the technology and N is the (constant) number of workers. In this economy, 25% of income is saved the capital depreciation rate is 1%. (a) Describe the main equations of the model and find out the fundamental dynamic

⁵⁹ The student will thank us for skipping the tedious mathematical derivation.

- equation for K/L , where L is labour in efficiency units. (b) Find out the equilibrium values of K/L , Y/L and K/Y . (c) Describe the time-paths of per capita income (Y/N), the wage rate, the interest rate and the factor income shares in the steady state. Are these paths in accordance to the real-world facts? (d) Suppose that a war destroyed part of the stock of capital of that economy. Describe the subsequent evolution of per capita income (Y/N). (e) How did the growth rate of per capita income and the interest rate evolved during the transition path? Explain.
- 3.2.** In economy W, the aggregate production function is given by $Y_t = A_t K_t^{1/4} N_t^{3/4}$, where N refers to population. In this economy, $s=24\%$, $n=1\%$, $\delta=0$, and $A_t = 8e^{0.015t}$. (a) Find out the steady state values of K/L and Y/L , where L is labour in efficiency units. Represent the equilibrium in a graph and explain why it is stable. (b) Compute the interest rate, the capital and labour income shares, and explain the steady state patterns of wages and per capita income. Are these results in accordance to the Kaldor facts? (c) Assume that a benevolent planner managed to increase the saving rate in this economy to $s=27.78\%$. (f1) Would the growth rate of per capita income change? (f2) What about the interest rate? (f3) Consumers would be better off? Explain, quantifying when possible.
- 3.3.** Consider an economy (Oldland) where the production function is given by $Y = A_t K_t^{1/3} N_t^{2/3}$, where N measures the number of workers. It is known that, in this country 25% of income is saved, population is expanding at 0.5% per year, the capital stock depreciates at 3% and $A_t = 20e^{0.01t}$. (a) Find out the equilibrium levels of K/L , Y/L and K/Y of this economy, where L represents labour in efficiency units. Discuss the stability of the equilibrium and represent it in a graph. (b) Describe the short and long run effects of a rise of the saving rate in the following variables: per-capita income, growth rate of per-capita income, per-capita consumption and interest rate. (c) Admit that in Oldland per-capita income was ten times higher than in Newland. In what conditions could you state that Newland was growing faster than Oldland? (d) Knowing that technology was the same in both countries, find out what the interest rate in Newland should be. Would the two economies converge? Discuss.
- 3.4.** Consider an economy composed by a large number of small and identical firms. The available technology for each of them is given by $Y_i = 0.5 K_i^{0.5} L_i^{0.5}$, where $L = \lambda N$ measures labour in efficiency units. Population doesn't grow, the depreciation rate is 3.5% and the saving rate is 20%. Admit also that $\lambda = e^{0.015t}$. (a) Find out the equilibrium values of K/L , Y/L and K/Y of this economy. (b) Describe the long run behaviour of per-capita output, wages and the interest rate. (c) Assume now that saving rate was determined according to the following rule $\gamma = \dot{c}/c = r_t - \rho$, where ρ is the rate of time preference and r the interest rate. Explain this rule. Find out the value of ρ that is consistent with $s=20\%$. (d) Describe the adjustment of the model following a decrease in ρ to 0.05. What will be the implied saving rate in the steady state?
- 3.5.** Consider an economy where capital and output are growing at 3% per year and the labour force is expanding at 1%. Assuming that the elasticity of capital in the production function (β) was one third, compute the contribution of capital to output growth, using: (a) conventional growth accounting; (b) growth accounting based on the following re-parameterisation of the production function: $\hat{y} = A^{1/(1-\beta)} (K/Y)^{\beta/(1-\beta)}$. Interpret the differences. Repeat the exercise assuming that output growth was 4.5%.

4 The Neoclassical model with Human Capital

Why doesn't capital flow from rich countries to poor countries?

[Robert Lucas Jr.]

Learning Goals:

- Understand why the Solow model cannot account for large cross-country differences in per capita income
- Understand the role of unspecified ingredients to production, hidden in the TFP term.
- Acknowledge that augmenting the model with human capital makes the model more compliant with the real-world facts.
- Development accounting
- Acknowledge the main achievements as well as the main limitations of the neoclassical growth model.

4.1 Introduction

The Solow model is capable of describing the main real-world facts of economic growth. Not surprisingly, the Solow model became the workhorse model in the theory of economic growth and it still is. During the last 30 years, however, various researchers have subjected the model to a demanding empirical scrutiny. One direction explored in this further investigation relies on the fact that the model specification implies not only the expected *signs* of certain parameters but also their approximate *magnitudes*. In particular, the model implies broad orders of magnitude for the coefficients linking per capita income to the savings rate and the population growth rate. It happens however that these magnitudes do not conform too well with the empirical evidence regarding cross-country income inequality.

The key parameter in this further investigation is the elasticity of output in respect to physical capital, β . If, according to the model, there is perfect competition and factors of production are paid their marginal products, then the elasticity of output in respect to physical

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

capital should correspond to the share of capital in national income. As we will see, the empirical observation that the latter stands at around 30% to 40% makes the model incapable of describing the large differences in per capita incomes that we observe in the real world. The conciliation would require, for example, differences in savings rates as between countries much larger than those that we actually observe in reality.

This chapter explains these inconsistencies and highlights the fact that the term that we have been labelling as “technology” (A) cannot be equal across countries. The chapter then focuses on an ingredient that has been hidden inside A : Human Capital. As we will see, augmenting the production function to account for cross-country endowments in human capital makes the neo-classical model more compliant with the real-world facts.

Section 4.2 reviews the above-mentioned limitations of the Solow model. Section 4.3 introduces the concept of Human Capital and proposes an augmented production function, including human capital. Section 4.4 shows how the neoclassical growth model can easily be adapted to account for Human Capital. Section 4.5 turns to the empirical evidence to evaluate how far the extended version can go in accounting for cross-country income differences. The conclusion is that augmenting the production function with human capital improves the predictability of the model but it does not allow it to fully explain the existing per capita income differences. The chapter concludes on the need to go again inside A to identify more country-specific factors that prevent per capita income levels to approach each other.

4.2 The Lucas’ 1990 Paradox

4.2.1 Explaining cross-country income differences using the Solow model

In the Solow model, the steady-state level of per capita income is given by:

$$y^* = A^{\frac{1}{1-\beta}} \left(\frac{s}{n + \delta + \gamma} \right)^{\frac{\beta}{1-\beta}} e^{\eta} \quad (4.1)$$

Since the model assumes perfect technological diffusion, advances in technology $A_t = A e^{g t}$, are equally shared across economies. Hence, cross-country differences in per capita incomes shall be accounted by differences in saving rates and in population growth rates.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

In light of (4.1) higher saving rates and lower population growth rates should be associated with higher levels of per capita income. As shown in figures 2.5 and 2.5 this prediction of the model confronts well with the real world facts. A question remains however regarding magnitudes: how much should saving rates and population growth rates differ across-countries for the model to capture large differences in per capita income as we observe in reality?

To examine this question consider two countries; say a Rich Country (R), and a Developing Country (L). Assuming that advances in technology are equally shared across countries, equation (4.1) implies that the ratio of per capita incomes in the *steady state* must be:

$$\frac{y_R}{y_L} = \left(\frac{s_R}{s_L} \right)^{\frac{\beta}{1-\beta}} \left(\frac{n_L + \delta + \gamma}{n_R + \delta + \gamma} \right)^{\frac{\beta}{1-\beta}} \quad (4.1a)$$

In order to predict the difference in per capita incomes between two countries given the saving rates and the population growth rates, one need an estimate for β .

In light of the new-classical model, the term β should be equal to the share of capital in national income: because the model assumes perfect competition and rules out market failures (such as externalities), it predicts that factors are paid according to their marginal products. This is reflected in equations (2.11) and (2.12), which state that the shares of capital and of labour in national income shall correspond to the respective elasticities in the production function, β and $1 - \beta$.

In the real world, we observe that the share of labour in national income stands around 60%, depending on the country. Given this, one may reasonably calibrate the Solow model setting the parameter β equal to $1/3$. With such calibration, the elasticity of income in respect to the savings rate, $\beta/(1 - \beta)$ in equation (4.1a), becomes equal to 0.5: that is, a 1 percentage point (p.p.) increase in the saving rate implies a 0.5p.p increase in the steady state per capita income. The inconsistency with the empirical facts arises from the observation that this elasticity is too small to account for the large differences in per capita incomes that we observe in reality.

To get a sense of this problem with the model, assume that $\beta=1/3$, $\delta=3\%$ and $\gamma=2\%$ (the last corresponding to the trend growth rate of per capita GDP in the U.S.). Then, assume

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

that the two countries differ in the respective saving rates and population growth rates: $s_R = 0.2$, $s_L = 0.24$, $n_L = 0.01$ and $n_L = 0.03$. These values match roughly the average rates observed in US and Tanzania between 1960 and 2000. With these parameter values, equation (4.1a) would predict per capita incomes in the rich country and in the poor country differing only by a factor of 1.05 (that is, the difference between the average income of a US citizen and that of a Tanzanian citizen should be 5% only). Obviously, this is too small: by 2000, per capita income in the US was 32 times higher than that of Tanzania.

Now, as an extreme case, assume that the poor country had instead a saving rate as low as $s_L = 0.015$ and a population growth rate as high as $n_L = 0.05$. Clearly, these assumptions are extreme, even for developing countries: no country has sustained a population growing at 5% per year and few save as little as 1.5% of the respective income. Still, these figures would imply a ratio of per capita incomes of 4.7, only. This is still too small to account for the observed differences in per capita incomes we observe in reality⁶⁰.

The conclusion is that, even using dramatic assumptions concerning the key parameters of the model, in no way we are capable of predicting cross country income disparities of the magnitude that are observed in reality.

4.2.2 Using the production function

By making use of equation (4.1), exercise (4.1a) implicitly postulates that economies are in the respective steady states. But it could be that the poor country was still on its way to the steady state. In that case, part of the poor country's income gap vis-a-vis the rich country could be due to the fact that its income potential (as determined by n and s) was not yet fully materialized.

⁶⁰ If one used instead $\beta=0.4$, which is also a reasonable assumption, the ratio of per capita incomes in the extreme scenario would rise to 7.9. Still too small.

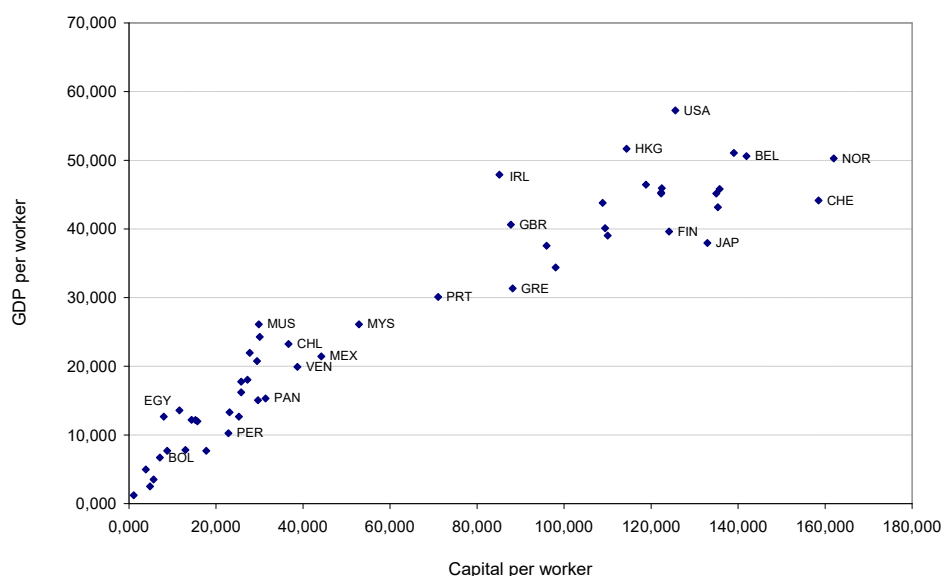
To gain a further insight, we now give away the assumption that economies are in the steady state. We simply focus on the production function:

$$y_t = A_t k^\beta \quad (4.2)$$

If the production function (4.2) was the same for all countries, then the factor explaining cross-country income inequalities should be capital per worker, k . Note that this conclusion does not depend on whether a country is in the steady state or engaged in transition dynamics.

Evidence that rich countries have more capital per worker than poor countries is provided in Figure 4.1 (see also Box 4.1). The figure crosses data on per capita income and on capital per worker for 53 countries. The figure confirms that capital per worker varies considerably across countries (by a factor of 100:1) and that richer countries tend to enjoy higher levels of capital per worker. The non-linearity in the relationship between per capita income and capital per worker is also suggestive of diminishing returns, as assumed in the Solow model. The question is whether the relationship between the capital per worker and per capita output in that figure is consistent with our guess for the capital-output elasticity, β .

Figure 4.1 – GDP per worker and capital per worker, 2000



Source: Caselli and Feyrer (2007). The Figure crosses capital per worker and GDP per worker for 53 countries in the year 2000. [Caselli, F., Feyrer, J. 2007. The marginal product of capital. Quarterly Journal of Economics vol. 122(2), 535-568].

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

From Figure 4.1, let's consider a developing country, say Peru. By 2000, per capita income in Peru was roughly 17.9% of the corresponding level in the United States. Could such difference be fully explained by a difference in the capital labour ratio? To answer this question, let's assume again that β is equal to $1/3$, and solve (4.2) for k , to find out the implied level of capital per worker in Peru. The answer is $k = 0.179^3 = 0.0057$. That is, for the per capita income difference between Peru and the US to be explained by the capital labour ratio only, one would need a capital labour ratio in Peru equivalent to 0.57% of that in the United States. This is clearly unrealistic: by 2000, the capital labour ratio in Peru stood at about 18% of the corresponding level in the US.

4.2.3 Why doesn't capital flow from rich to poor counties?

A third way of looking at the same problem was proposed by the Nobel Laureate Robert Lucas Jr: if, cross country differences in per capita incomes were only related to differences in the ratio of capital per worker, then poor countries should have much higher interest rates than rich countries⁶¹.

How much higher? To answer this question, Lucas solved equation (4.2) for k , and used (2.12), to obtain:

$$r + \delta = \beta A_t^{1/\beta} y^{(\beta-1)/\beta}. \quad (4.3)$$

Now consider a Rich country, (R), and a Less Developed Country, (L), using the same technology (4.2). In that case, the ratio of capital returns in the two countries should be equal to:

$$\frac{(r + \delta)_L}{(r + \delta)_R} = \left(\frac{y_t^R}{y_t^L} \right)^{\frac{1-\beta}{\beta}} \quad (4.3a)$$

⁶¹ Lucas, R., 1990, "Why doesn't capital flow from rich countries to poor countries?", American Economic Review 80, 92-96.

As an example, Lucas considered the cases of USA and of India, which per capita incomes in the year 2000 differed by a factor of 15:1⁶². Assuming, for instance, $\beta=0.4$, the exponential term in equation (4.3a) equals 1.5. Thus, to explain a 15:1 gap in per capita incomes, the ratio of capital returns should be $15^{1.5}=58$. In plain language, if capital were generating a real return of around 4% in the USA, then the corresponding average return in India, after adjusting for risk-premia, should be $58 \times 4\% = 233\%$. Clearly, this is wholly unrealistic⁶³.

Lucas pointed out that such high interest differentials cannot hold in a world with capital mobility. In face of such high differentials, capital goods should be flowing from rich countries to poor countries. Moreover, one should observe almost no investment in capital-abundant countries until capital-labour ratios - and hence interest rates - were more or less equal across the globe.

If that was so, per capita incomes would be doomed to converge in absolute terms: if the problem in the developing World was lack of capital and if that was reflected in higher returns to investment, then capital should flow from rich countries to poor countries until all per capita incomes were equalized.

Clearly this story does not conform to today's realities: capital does not flow in huge amounts to poor countries, interest differentials are nowhere near as large as 58 times and there is no systematic tendency for poor countries to grow faster than rich countries.⁶⁴

⁶² According to Maddison (2001), in 2000, per capita incomes in USA and in India in comparable units (PPP) were, respectively, \$28.129 and \$1.910. [Maddison, A., 2001. The World Economy: a Millennial Perspective. Development Centre, Paris].

⁶³ With $\beta=1/3$ the required difference in interest rates would jump to 225.

⁶⁴ Lucas observed that during the XVIII and XIX centuries, at a time when production functions were better described in terms of labour and land, *labour* actually moved from labour abundant countries (Europe) to labour scarce countries (New World). That movement favoured economic convergence. In the XX century, capital replaced land as the main factor in the production function and became the potentially mobile factors, at a time when strong restrictions on labour mobility were erected. The potential for capital to move did not materialize, however, in per capita income convergence.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Box 4.1 The return to physical capital in poor countries

The question as to whether the marginal product of capital is higher or lower in poor countries than in rich countries has important policy implications: if one concludes that the marginal product of capital differs substantially across countries, this means that impediments to capital mobility are preventing physical capital from being more efficiently allocated across the globe. In that case, there would be scope for expanding the world output by promoting a better allocation of capital between rich countries and poor countries (for instance, through external aid). Caselli and Feyrer investigated the extent to which the low levels of capital per worker in poor countries are indeed associated with higher marginal products of capital⁶⁵.

In their calibration exercise, the authors isolated the parameter A from the influence of land and other natural resources, specifying a production function with constant returns to scale on three inputs: labour, capital and “natural capital”. Accounting for this third input implies that the share of physical capital in national income (β) cannot not be estimated as one minus the share of labour on national income. In the case of United Kingdom, for instance, the authors estimated a share of labour equal to 75% and a share of physical capital equal to 18%, only. The 7% difference corresponded to the share of natural capital. In Bolivia, the share of labour was estimated to be 67% while the share of physical capital was only 8%, which implies 25% for natural capital. Since the share of income rewarding natural capital tends to be high in countries specialized in agriculture and in mineral exports, ignoring that influence would imply an overestimation of the contribution of capital to output in poor countries.

A second caveat pointed out by Caselli and Feyrer relates to the relative price of capital. If the price of output and the price of physical capital were the same – as assumed in the Solow model – then in a world with perfect capital mobility one should observe an

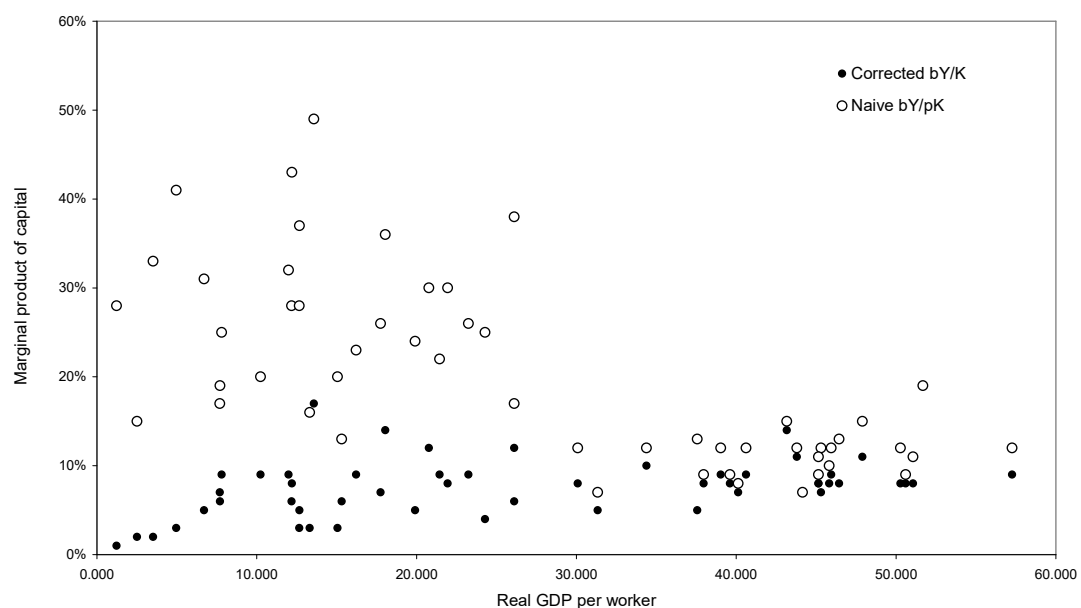
⁶⁵ Caselli, F., Feyrer, J. 2007. The marginal product of capital. Quarterly Journal of Economics vol. 122(2), 535-568.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

equalization of marginal products of capital across countries. However, in the real world, one unit of capital does not cost the same as one unit of output. In general, the relative price of capital is higher in poor countries than in rich countries, due to tariffs, transport costs and other distortions. Since in poor countries, firms must spend more units of output to buy one unit of capital, they need to achieve a higher return per unit of invested capital, all else constant.

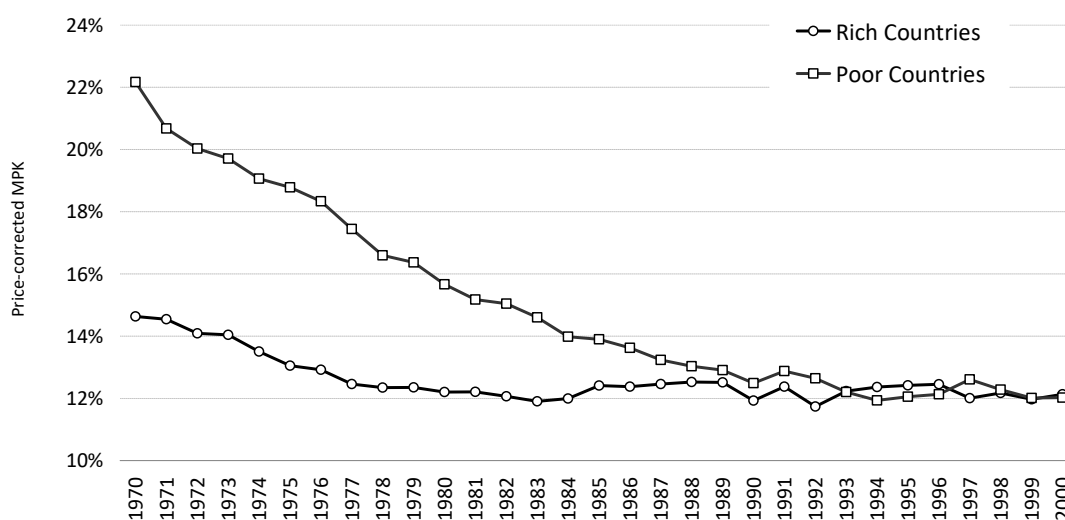
Formally, let p_K be the relative price of capital goods (in units of Y). In that case, when one unit of output is saved, the investor will be able buy $1/p_K$ units of physical capital, only. With a marginal product of physical capital equal to $\partial Y/\partial K = \beta(Y/K)$, the change in output obtained by saving one unit of output will be $\beta(Y/K)(1/p_K)$. Taking this into account, Caselli and Feyrer corrected the conventional measure of the Marginal Product of Capital, dividing it by the relative price of capital.

Figure 4.2 – The marginal product of capital (naïve and corrected estimates)



Source: Caselli and Feyrer (2007), op.cit. The naïve estimates correspond to the marginal product of capital computed assuming a production function with capital and labour, only, and ignoring differences between the price of output and the price of capital. The corrected estimate accounts for the role of natural capital and for the existing cross-country divergences between prices of output and prices of capital.

Figure 4.3 – The price-corrected marginal product of capital in rich countries and in poor countries



Source: Mello (2008).

Figure 4.2 compares two alternative estimates of the marginal product of capital: the “corrected estimates”, and the “naïve estimates” (that is without price correction and without accounting for the share of natural capital). As the figure reveals, the naïve estimates point to high and variable marginal products of capital in poor countries, whereas in rich countries the marginal product of capital tends to be lower. That being the truth, it would support the idea that capital has not flowing the enough from rich countries to poor countries, eventually reflecting impediments to capital mobility. Removing these impediments should deliver economic convergence and a more efficient allocation of capital across the globe.

After the corrections, however, the story is different. As shown in Figure 4.2, when corrections for the natural capital and for the relative prices are made, the marginal product of capital becomes unrelated to the level of per capita income. In other words, from the investor’s point of view, the return to investment is not higher in poor countries than in rich countries. This evidence suggests that the large cross-country differences in capital per worker that we observe in reality cannot be attributed to lack of capital mobility: a reallocation of world capital so as to exactly equal the marginal products of capital across

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

rich and poor countries would not bring much difference in terms of cross-country income inequality relative to what we observe today.

A different question regards the differences in the marginal product of capital across time. Extending the work of Caselli and Feyrer for the period 1970-2000, Mello (2008) showed that the equalization of marginal products of capital across the world is a rather recent phenomenon⁶⁶. The author's evidence is summarized in Figure 4.3. The figure displays the price-corrected marginal product of capital for a group of rich and a group of poor countries, along the period from 1970 to 2000 (the author did not correct for natural capital). As shown in the figure, initially the marginal product of capital differed substantially between rich and poor countries (5.52p.p. in 1970): at that time, large efficiency gains could have been achieved enhancing cross-country capital mobility. An indeed, along the decades that followed restrictions to capital flows were progressively waved. As shown in the figure, this move towards financial openness delivered a progressive convergence of capital returns between rich countries and poor countries. By 1990, the return differential was already as low as 0.37p.p, confirming the conclusion of Caselli and Feyrer: in today's world, the potential gains from a further integration of world capital markets would be relatively small.

4.2.4 So, how can we explain these inconsistencies?

If cross-country differences in per capita income were accounted for by differences in the availability of capital per worker only, then the marginal product of capital should be very high in poor countries and physical capital should flow from rich countries to poor countries. However, this does not happen in our days: capital is not flowing massively from rich countries to poor countries and there is no evidence that the marginal product of capital is substantially higher in poor countries than in rich countries.

⁶⁶ Mello, M., 2009. Estimates of the Marginal Product of Capital, 1970-2000. The B.E. Journal of Macroeconomics 9(1), article 16.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

The conclusion is that our calibration of the production function (4.2) is missing something. Returning to (4.2a), there are two obvious possible avenues to explore: one is to consider that the elasticity of physical capital in production (β) is larger than that implied by the capital income share. That would be the case if capital was not correctly priced in the market, because of externalities. This avenue will be addressed in Chapter 6. Alternatively, one may concede that the parameter A, that we have so far labelled as “technology” cannot be equal across countries. Along the second avenue, consider for instance the cases of Hong Kong and of Japan. As you may observe in Figure 4.1, Hong Kong enjoys a higher output per worker than Japan, with a lower level of physical capital per worker. This is suggestive that other factors apart from capital per worker (hidden inside A) are driving cross-country income differences.

Admitting that the parameter A may differ across countries implies however recognising that some of the ingredients that are hidden inside this broad “total factor productivity” term are not suitable to spill over instantaneously from one country to the other. “General knowledge”, that comes out in newspapers, scientific journals, or through the internet, diffuses easily across the world, and hence is not a good candidate to explain why the parameter A may differ across countries. But many other ingredients are country-specific. This is the case of natural resources, differences in climate that affect agricultural productivity, differences in the education and health of residents, the quality of government policies, culture, and so on.

To a large extent, looking to what is inside A and evaluating the potential of some of the components in it to diffuse across the space constitutes our agenda for the entire course. For the moment, however, we are particularly interested in disentangling the role of a particular input to production: *Human Capital*. Human capital is distinct from knowledge in that it is “embodied” in people, and hence is not mobile across countries, unless people migrate.

4.3 Human capital

4.3.1 What is human capital?

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Human capital is the term used to coin the stock of knowledge and health embodied in workers. The notion of human capital comes from the observation that people *invest* in knowledge and in health, through schooling, on-the-job training, exercise and healthcare, with the aim to obtain a return, just like people invest in physical capital.

Like physical capital, human capital depreciates over time. As an example think on when you adapt to a new release of your computer operating system: you have to learn how to operate with the new version, throwing away part of the time invested in learning how to operate with the old version. A similar reasoning holds for health: health depreciates over time, for instance, with age. Consequently, continuous investment in health is required (exercise, safe nutrition, health care), so as to prevent health capital from eroding too fast.

In the real world, we observe that the two components of human capital tend to be correlated to each other: more educated people tend to be more aware of the advantages of a healthy nutrition and of exercise, and tend to be more healthy too; by the same token, healthier individuals, with longer life expectancies, are likely to invest more in education, because they have longer payback periods. Health and education tend to move along.

Human capital is different from physical capital in that it is *embodied in individuals* and hence cannot be sold or rented for use by a third person, as an equipment. Also, because human capital is embodied in individuals, it cannot diffuse instantaneously across the globe like knowledge in general. Human capital can potentially move from one country to another through labour mobility, but in our days this avenue is considerably restricted by immigration laws.

4.3.2 Human capital as an input to production

There are several reasons to believe that human capital should be considered as an input to production, different from raw labour measured, say, in workdays: first, more knowledgeable workers will be able to accomplish more complex tasks with a minimum

outlay of time; second, healthy and well-nourished workers are expected to have more physical and mental energy to learn and undertake their tasks than sick workers⁶⁷; third, health capital increases the amount of healthy time available for work, by reducing incapacity, disability and absenteeism; fourth, individuals with longer life expectancy have incentive to work harder, because they need to save more for their longer retirement period. All in all, these various effects imply that a higher level of human capital should impact positively on output and, by then, on the average product of labour⁶⁸.

In order to account for the role of human capital in production, let's consider a production function where *both* “raw labour” and “human capital” enter as inputs to production:

$$Y = A_t K_t^\beta H_t^\alpha N_t^{1-\alpha-\beta}, \quad \text{with } \alpha+\beta < 1 \quad (4.4)$$

In (4.4), the new variable “H” stands for Human Capital (knowledge and health) and all other variables are defined as before. In light of this specification, production needs “bodies” and “brains” and it is not possible to substitute completely “bodies” for “brains” or “brains” for “bodies”. All are essential to production.

This production function exhibits CRS in all these inputs: that is, one would be able to duplicate Y if one could simultaneously duplicate the use of the three factors, K , H and N . As in the basic Solow model, we are constrained by the fact that one input (N) evolves at

⁶⁷ According to some studies, in the 18th century France, individuals in the bottom 10 percentile of consumption had daily caloric intakes that were less than the energy needed to work. In the centuries that followed, improvements in nutrition resulted in a higher workers' effort and on a higher fraction of working-age population being able to work). [Fogel, R., 1991. The conquest of high mortality and hunger in Europe and America: timing and mechanisms. In: Lander, D., Higgonet, C., Rosovsky, H. (eds.), *Favourite of Fortunes: Technology, Growth and Economic Development Since the Industrial Revolution*. Cambridge MA: Harvard University Press].

⁶⁸ In addition to the direct effect of human capital on production, human capital impacts indirectly, through the permeability of a country to technological change. Skilled and educated workers are more likely to adopt new technologies and to contribute themselves to technological change than less educated people. These effects run from H to A and then to Y . For expositional convenience, we are abstracting from any cross-country differences in technology, and consequently for the indirect impact of human capital via technological change.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

exogenous rate, n . Thus, there will be diminishing returns on reproducible factors altogether, K and H . This property makes the model of the same category of the basic Solow model.

In the intensive form, the production function becomes:

$$y_t = A_t k_t^\beta h_t^\alpha, \quad (4.5)$$

where $h = H/N$ is defined and human capital per worker.

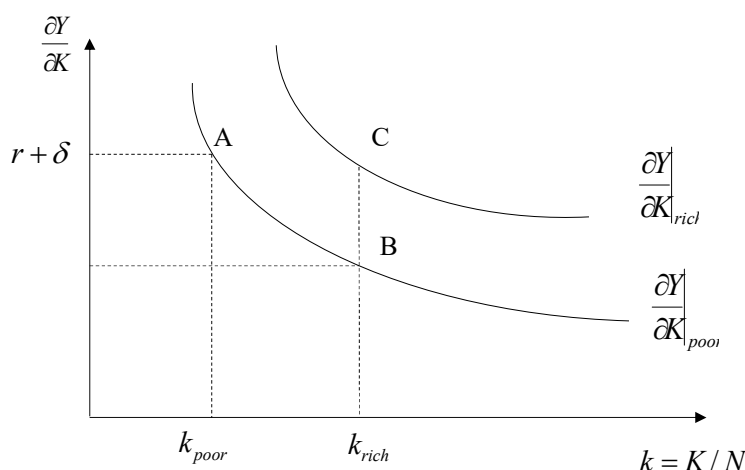
4.3.3 Human capital and the productivity of physical capital

When two inputs are complementary in production, employing an additional unit of one input makes the other more productive.

To see this in terms of (4.4), let's take the partial derivative in respect to physical capital:

$$\frac{\partial Y}{\partial K} = \beta \frac{Y}{K} = \beta [A_t k_t^{\beta-1} h_t^\alpha] \quad (4.6)$$

Figure 4.4 – What happens to the marginal product of physical capital when human capital per worker increases?



The figure displays the marginal product of capital as a function of capital per worker. This curve is parametric in the level of human capital.

In (4.6), the marginal product of physical capital is a positive function of the endowment of human capital. Figure 4.4 illustrates this. The figure displays the marginal product of capital (4.6) as a function of capital per worker in two different countries that

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

differ in terms of human capital per worker: a “rich” country and a “poor country”. Both schedules are downward sloping due to diminishing returns: more capital per worker translates into a lower productivity of capital (and hence, a lower interest rate), everything else constant. However, these curves are parametric in the endowment of human capital per worker. Thus, if a rich country has simultaneously more physical capital and more human capital than the poor country, then the required difference in interest rates might not be that high.

4.4 The model of Mankiw, Romer and Weil (MRW)

In this section, we show how the steady state of the Solow model differs once we allow for a role for human capital in production. This extension is due to Mankiw, Romer, and Weil (1992)⁶⁹.

4.4.1 Main assumptions

As in the Solow model, the authors assumed that the *common* technological parameter in (4.4) increases continuously, at an exogenous rate g :

$$A_t = A e^{gt} \quad (4.7)$$

Rearranging, the production function (4.4) in terms of “efficiency labour”, we get:

$$Y = A K_t^\beta H_t^\alpha L_t^{1-\alpha-\beta}, \quad (4.8)$$

where

$$L_t = N_t e^{\gamma t} \text{ and } \gamma = g/(1 - \alpha - \beta). \quad (4.9)$$

⁶⁹ Mankiw, G., D. Romer and D. Weil, 1992. “A contribution to the empirics of economic growth”, Quarterly Journal of Economics, 107 (2), 407-38.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Note that in this model, workers become more productive both because of labour augmenting technological progress and because of investment in human capital. The critical distinction between these two factors is that the former is strictly exogenous, while the later is produced and accumulated, just like physical capital.

As in the Solow model, it is assumed that one unit of output can be transformed at no cost into either one unit of human capital or into one unit of physical capital. The resource constraint of the economy is given by:

$$Y_t = C_t + I_t + I_t^H, \quad (4.10)$$

where I_t^H refers to investment in Human capital and all other variables are defined as in the Solow model.

It is also assumed that people invest constant fractions of their incomes in human capital. Let s_H be the fraction of income invested in human capital and s the fraction of income invested in physical capital. The dynamics of K and H are given, respectively, by:

$$sY_t = I_t = \dot{K}_t + \delta K \quad (4.11)$$

$$s_H Y_t = I_t^H = \dot{H}_t + \delta H. \quad (4.12)$$

For simplicity, the model assumes that the stocks of physical and of human capital depreciate at the same rate, δ . The depreciation of human capital may be interpreted as the erosion of knowledge (obsolescence, forgiveness) net of the benefits from experience.

4.4.2 The steady state in the MRW model

To solve the model in the simplest possible manner, there is a helpful clue: in the steady state, Physical Capital and Human Capital must grow at the same rate. At this stage, the reason should be intuitive: like in the simple Solow model, diminishing returns imply that it does not worth having one input growing faster than the others: if, for instance, physical capital was growing faster than human capital, then the return to physical capital would decrease relative to that of human capital, creating the incentives for people to accumulate less physical capital.

Imposing $\dot{K}/K = \dot{H}/H$ in the two equations above, one obtains a critical condition:

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

$$\frac{H}{K} = \frac{s_H}{s} \quad (4.13)$$

This condition states that the ratio of human to physical capital in the steady state shall be proportional to the corresponding investment rates. Solving for H, and substituting in the production function (4.4), one obtains⁷⁰:

$$Y = BK_t^\theta L_t^{1-\theta} \quad (4.14)$$

where

$$B = A \left(\frac{s_H}{s} \right)^\alpha, \text{ and} \quad (4.15)$$

$$\theta = \alpha + \beta \quad (4.16)$$

Equation (4.14) holds in the steady state only. Given the similarities between (4.14) and the standard production function in the Solow model, you may guess that the steady states should be similar. To some extent, this happens to be true. Adapting equation (4.1) for the parameters in (4.14), we get:

$$y_t^* = B^{\frac{1}{1-\theta}} \left(\frac{s}{n + \delta + \gamma} \right)^{\frac{\theta}{1-\theta}} e^{\gamma t} \quad (4.17)$$

Substituting back using (4.15) and (4.16), one obtains the steady state level of per capita income in this augmented model:

$$y_t^* = \left[\frac{A s_H^\alpha s^\beta}{(n + \delta + \gamma)^{\alpha+\beta}} \right]^{\frac{1}{1-\alpha-\beta}} e^{\gamma t} \quad (4.18)$$

⁷⁰ In a footnote to their paper (footnote 12), Mankiw et al. (1992) noted that the properties of the model change dramatically in the case in which $\theta=1$. In that case, equation (4.14) becomes linear in K , delivering “endogenous growth”. This alternative case will be addressed in the next chapter.

where $\gamma = g/(1 - \alpha - \beta)$. Note that (4.1) can be interpreted as a particular case of (4.18), with $\alpha=0$.

Equation (4.18) states that the steady state level of income per capita depends positively on the saving rate and negatively on the rate of population growth (as before). It also states that per capita income rises in the long run at a constant rate γ (as before). The novelty here is that the level of per capita income also depends on the share of income devoted to Human Capital Accumulation, s_H : an increase in s_H gives rise to a *level effect* (output per capita will expand temporarily until the new steady state is reached). The same happens with investment in physical capital.

As in the basic Solow model, the golden rule for capital accumulation can be obtained maximizing the steady state level of per capita consumption: $\max_{s, s_R} c_t^* = (1 - s - s_R) y_t^*$, where y_t^* is given by (4.18). The solution to this problem is as expected: $s = \beta$ and $s_R = \alpha$.

4.4.3 Factor income shares in the MRW model

The production function (4.4) applies to the economy as a whole. Assuming that all firms in this economy are identical, profit maximization leads to the following aggregate demands for physical capital, human capital and raw labour:

$$\frac{\partial Y}{\partial K} = \beta \frac{Y}{K} = r + \delta \quad (4.19)$$

$$\frac{\partial Y}{\partial H} = \alpha \frac{Y}{H} = r_H + \delta \quad (4.20)$$

$$\frac{\partial Y}{\partial N} = (1 - \alpha - \beta) \frac{Y}{N} = w \quad (4.21)$$

In equation (4.20), we use r_H to describe the return of human capital. We do not set $r_H = r$, to account for the possibility of market failures that prevent the returns of the two types of capital to be equal.

As expected, equations (4.19)-(4.21) imply that factor income shares are equal to the corresponding elasticities in the production function: the share of physical capital on national income should be equal to β , the share of human capital should be α , and the share of raw

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

labour should be $1 - \alpha - \beta$ (remember that this is a direct consequence of profit maximization under CRS).

In practice, however, national accounts do not disentangle what is the share of human capital from what is the share of raw labour: workers are paid a salary that reward both components at the same time. National accounts therefore only provide a figure for $(1 - \alpha - \beta) + \alpha = 1 - \beta$, the total share of labour in national income, irrespectively as to whether this corresponds to human capital or raw labour. As we already know, this figure varies across countries, standing at around 50% to 70% of national incomes. Estimates of α and β using regression analysis are shown below, in tables 4.1 and 4.2.

4.4.4 Cross-country income differences revisited

A question that naturally arises is whether adding human capital to the production function turns the model more compliant with the real world facts. In particular, one wants to understand the extent to which the augmented model accounts for the large cross-country income differences, as we observe in reality.

Proceeding as before, let consider a “rich country” and a “poor country”. Sticking with the assumption that advances in technology are equally shared across the globe, from (4.18), the ratio of per capita incomes in the steady state becomes:

$$\frac{y_R}{y_L} = \left(\frac{s_R}{s_L} \right)^{\frac{\beta}{1-\alpha-\beta}} \left(\frac{s_{HR}}{s_{HL}} \right)^{\frac{\alpha}{1-\alpha-\beta}} \left(\frac{n_L + \delta + \gamma}{n_R + \delta + \gamma} \right)^{\frac{\alpha+\beta}{1-\alpha-\beta}} \quad (4.18a)$$

Comparing to (4.1a), we see that now we have two investment rates (in physical and in human capital) affecting the per capita income gap.

To calibrate (4.18), we need to assume some values for the parameters, α and β . For simplicity, consider a case with $\alpha=\beta=1/3$, $\delta=3\%$ and $\gamma=2\%$. This gives:

$$\frac{y_R}{y_L} = \left(\frac{s_R}{s_L} \right) \left(\frac{s_{HR}}{s_{HL}} \right) \left(\frac{n_L + 0.05}{n_R + 0.05} \right)^2$$

Now, consider two countries: a rich country with $s_R = s_{HR} = 0.2$, $n_R = 0.01$, and a poor country with $s_L = s_{HL} = 0.10$ and $n_L = 0.03$. Such parameterization is reasonable, give

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

the world evidence. Replacing these figures in the equation above, one obtains a per capita output difference of 7 to 1. This is much more than in the Solow model with equivalent assumptions (around 2 to 1). Still, it is half-way relative to what we need!

In general, explaining real world cross-country income differences using equation (4.18) delivers better results than using equation (4.1) (see Box 4.2). But the empirical evidence also reveals that a significant part of the cross-country variation of per capita incomes remains to be explained. In other words: accounting for the role of Human Capital improves the explanatory power of the model and helps mitigate the Lucas (1990) paradox, but it is not enough: this suggests that one need to go again inside the technology parameter A , to disentangle other ingredients that can explain why per capita incomes are so different across countries and regions.

Box 4.2. Empirical test on the MRW model – steady state

Instead of calibrating equation (4.18) to compare pairs of countries, Mankiw, Romer and Weil used regression analysis to investigate how well the model accounts in general for observed cross-country income disparities in the real world. In particular, the authors estimated a log-linear version of (4.18), given by:

$$\ln y_i^* = a + \frac{\beta}{1 - \alpha - \beta} [\ln s_i - \ln(n_i + \delta + \gamma)] + \frac{\alpha}{1 - \alpha - \beta} [\ln s_{Hi} - \ln(n_i + \delta + \gamma)] + u_i \quad (4.18b)$$

In this equation, the parameter $a = \eta + \ln A / (1 - \alpha - \beta)$ refers to the “state of technology” at moment t , and is assumed common to all countries. Equation (4.18b) implicitly assumes that all countries are in their steady states or, more generally, that deviations from the steady state are random. The random disturbance u_i captures these deviations, as well as country specific effects determining differences in the level of A , such

as differences in resource endowments or in climate⁷¹. The Solow model corresponds to the particular case with $\alpha=0$.

The authors used a sample of 98 countries for the year 1985. The proxies used were: the growth rate of the working-age population for n ; the ratio of real investment to GDP for s (period averages); the fraction of the eligible population enrolled in secondary education for s_H ; the real GDP per working age person for y ; the authors also postulated $\delta + \gamma = 0.05$ for all countries.

The estimates of the Solow model (equation 4.18b imposing $\alpha=0$) are displayed in the first column of table 4.1 (standard errors in italics). According to these estimates, the Solow model accounts for 59% of cross-country variation of per capita incomes. The estimated coefficients have the predicted signs and are significant. Nonetheless, the implied value for β is as high as 0.60 (using $1.48 = \beta/(1 - \beta)$). This is another incarnation of the Lucas' 1990 puzzle: in order for the observed differences in per capita incomes to be accounted for by differences in physical capital per worker alone, one would need a contribution of physical capital to production much larger than the corresponding observed income share. In light of the current model, the coefficient on physical capital is biased upwards because human capital is being omitted from the regression equation.

The results using the augmented model (equation 4.18) are displayed in the second column of table 4.1. With no question, the augmented model performs better empirically than the basic version. All coefficients have the expected signs and are highly significant. The new estimate for β (0.31) is now closer to the real-world data regarding the share of capital on national income. Notably, this model explains 78% of the cross-country variation in per

⁷¹ In practice, the possibility of country specific effects leads to a potential econometric problem: if negative country specific effects (e.g., poor resource endowments) discourage capital accumulation (as is likely), the error term will be correlated to the saving rate and therefore estimates will be biased. This is one of the main objections to the empirical implementation of this model. To get around this problem, some authors proposed panel data estimation, which allows for the control of country ("fixed") effects (Caselli et al, 1996). [Caselli, F., Esquivel, G. and Lefort, F. 1996. Reopening the convergence debate: a new look at cross-country growth empirics. *Journal of Economic Growth*, 40, pp 363-389].

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

capita incomes. This is a quite challenging result, as it suggests that three quarters per capita income differences across countries can be explained by differences in physical and human capital accumulation, while technological differences are assumed random.

Table 4.1. Estimation results for the steady state: the Solow model versus MRW

Dependent variable: log GDP per working age person in 1985		
Sample	Non-oil countries	
Observations	98	
constant	6,87 <i>0,12</i>	7,86 <i>0,14</i>
$\ln s - \ln(n + 0,05)$	1,48 <i>0,12</i>	0,73 <i>0,12</i>
$\ln s_H - \ln(n + 0,05)$		0,67 <i>0,07</i>
Implied beta	0,60 <i>0,02</i>	0,31 <i>0,04</i>
Implied alfa		0,28 <i>0,03</i>
R- Squared	0,59	0,78
S.E.E.	0,69	0,51

Source: Mankiw et al (1992), Tables I and II. Notes: Standard errors are in italics. The table displays the regression results of an equation describing the steady state level of per capital income, in light of the Solow model (first column) and the MRW model (second column).

4.4.5 Conditional convergence again

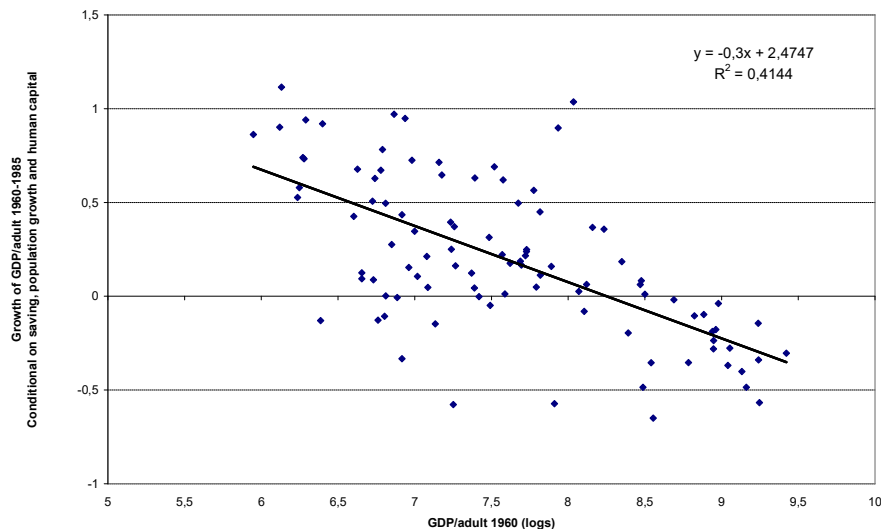
In Figure 3.4, we showed that there is no general tendency for poor countries to grow faster than rich countries. In Section 3.5, it was argued that this finding is fully consistent with the neoclassical growth model: the model allows countries to reach different steady states. The model also implies that countries grow faster the farther they are from their steady states. This property of the neoclassical model is labelled *conditional convergence*.

Empirically, the conditional convergence hypothesis can be tested by observing the relationship between the growth of per capita income along a period of time and the initial level of per capita income, after the variables determining the steady state are controlled for -

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

remember equation (3.14): if that relationship is significant and negative, this is taken as evidence of conditional convergence.

Figure 4.5 Evidence of Conditional Convergence (98 countries)



Source: Mankiw et al, 1992. The figure displays the relationship between the growth rate of per capita GDP and the initial level of per capita GDP, after controlling for the effects of variables determining the steady state (saving rate, investment rate in human capital, and population growth rate). The negative slope is suggesting of diminishing returns and conditional convergence.

Figure 4.5 shows the results of a test on conditional convergence performed by MRW using their theoretical model - details in Box 4.3. The vertical axes measures the growth rate of per capita income, after controlling for the effects of: the saving rate; investment in human capital; and the population growth rate. The horizontal axes measures the initial per capita income.

Clearly, the figure reveals a strong negative association between growth and initial incomes, after controlling for the different steady states. This evidence strongly contrasts to the finding in Figure 3.4, which uses the same data, but without controlling for differences in the steady state. The conclusion is that the worldwide evidence is not supportive of *absolute convergence*, but it is supportive of the neoclassical proposition of *conditional convergence*.

Box 4.3. Testing for Conditional Convergence

Mankiw et al (1992) implemented an exercise similar to that replicated in table (4.2) to a subsample of countries consisting in OECD economies, only. The results were however

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

disappointing: under the MRW specification the R-squared was only 0.28, and the coefficients on investment and population growth were not significant. A natural explanation is that OECD countries were not evolving around the respective steady states: the Second World War caused significant departures from steady states, and by 1985 these countries were eventually still engaged in a transition dynamics towards the respective steady states. If that was the case, a simple estimation of equation (4.18b) - that applies to the steady state - is not likely to deliver reasonable results. In general the same criticism may apply to any of the 98 countries in the larger sample: since the regression model (4.18b) does not account for transition dynamics, it cannot capture the possibility of some countries being engaged in temporary adjustment path towards the respective steady states, following some change in a fundamental parameter.

To overcome this limitation, Mankiw et al (1992) proposed an alternative estimation method. To understand their test, we refer to equation (3.14), which describes the transition dynamics in the Solow model. This equation also holds in the MRW, after adapting the parameter measuring the “speed of convergence” to the existence of human capital: $\nu = (1 - \beta - \alpha)(n + \delta + \gamma)$. Note that ν is now lower, implying a slower convergence than in the original Solow model.

The test for conditional convergence consists therefore in investigating the sign and significance of parameter b in an equation like (3.14), after *controlling for the determinants of the steady state*: that is, replacing y^* by 4.18b. Formally, the empirical model is⁷²:

$$\ln y_t - \ln y_0 = \mu - \kappa \ln y_0 + \frac{\kappa\beta}{1 - \alpha - \beta} [\ln s - \ln(n_i + \delta + \gamma)] + \frac{\kappa\alpha}{1 - \alpha - \beta} [\ln s_H - \ln(n_i + \delta + \gamma)]$$

(4.23)

where $\mu = \gamma + \kappa \ln A / (1 - \alpha - \beta)$ and $\kappa = (1 - e^{-\nu t})$, $\nu = (1 - \beta - \alpha)(n + \delta + \gamma)$.

⁷² It is worth noting that imposing a common convergence parameter ν (while population growth rates differ across countries), is a potential source of bias (a discussion in Lee et. al., 1997). [Lee, K., Pesaran, M., Smith, R., 1997. Growth and convergence in a multi-country empirical stochastic Solow model. *Journal of applied econometrics* 12, 357-392]

Table 4.2 describes the results obtained by Mankiw et al. (1992), for a worldwide sample of 98 countries and for 22 OECD economies. The third column of Table 4.2 also shows the results of a similar estimation, for a sample of 37 African countries, implemented by Murthy and Ukpolo (1999). These authors used the same dataset as Mankiw, Romer e Weil (1992) except in that their measure of investment in human capital also included both primary and secondary school attainment, instead of secondary school attainment only.

Table 4.2 – Tests for conditional convergence

Dependent variable: log difference GDP per working age person 1960-1985			
Sample	Non-oil	OECD	Africa
Observations	98	22	37
<i>constant</i>	2,46 <i>0,48</i>	3,55 <i>0,63</i>	2,16 <i>0,65</i>
$\ln y_0$	-0,30 <i>0,06</i>	-0,40 <i>0,07</i>	-0,35 <i>0,10</i>
$\ln s - \ln(n + 0,05)$	0,50 <i>0,08</i>	0,40 <i>0,15</i>	0,44 <i>0,16</i>
$\ln s_H - \ln(n + 0,05)$	0,24 <i>0,06</i>	0,24 <i>0,14</i>	0,31 <i>0,17</i>
Implied beta	0,48 <i>0,07</i>	0,38 <i>0,13</i>	0,40
Implied alfa	0,23 <i>0,05</i>	0,23 <i>0,11</i>	0,28
Implied speed of convergence	0,014 <i>0,00</i>	0,021 <i>0,00</i>	0,017
R- Squared	0,46	0,66	0,41
S.E.E.	0,33	0,15	

Source: Non-oil countries and OECD: Mankiw et al (1992), Table VI. African countries: Murthy and Ukpolo (1999), Table 1. [Murthy, N. and Ukpolo, V. (1999), A test of the conditional convergence hypothesis: econometric evidence from African countries, *Economics Letters* 65, 249-253]. Notes: Significant levels in italic. The regression equation is (4.23), describing the transition dynamics of the MRW model.

As shown in the table, in the three samples, all coefficients have signs in accordance to (4.23) and are statistically significant. Since the coefficients of $\ln y_0$ are negative and significant, this means that the *conditional convergence* hypothesis holds in the three samples. The partial association between the growth rate of per capita GDP and the initial level of per capita GDP implied by the first column of Table 4.2 is displayed in Figure 4.5.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

The R squared for OECD countries is now 66%, revealing that accounting for the transition dynamics makes a difference in this subsample.

In respect to the parameter estimates, Table 4.2 point to some differences as between the OECD and African countries. In particular, the estimated average speed of conditional convergence (ν) is 1.7% for African countries and 2.1%, for OECD countries. This means that the time required to eliminate half of the initial gap from their steady states in Africa is about 42 years, which compares to 35 years in the OECD sample⁷³.

4.5 Productivity matters!

The model proposed by MRW follows the basic Solow model, emphasizing the role of factor accumulation as an explanation for the observed cross-country differences in per capita incomes⁷⁴. Empirically, the results seem to validate the view that technology differences have a limited role when compared to that of investment in physical and human capital. This finding is in line with the claim of Alwin Young (discussed in Box 3.4), that the East Asian growth miracles were mostly fuelled by old-style factor accumulation, rather than by technological change. If that was true, it would imply that one should focus on investment and in education as the proximate causes of growth, rather than, for instance, on technological change.

With no surprise, these conclusions were subject to a further scrutiny in the literature. In a seminal contribution, Klenow and Rodriguez Claire (1997) questioned the idea that cross-country differences in per capita income are mainly explained by factor accumulation. Implementing alternative calibrations of the production function (4.8), the authors showed

⁷³ To assess how long it takes an economy to get halfway to its balanced growth path, the reader is referred again to equation (3.15): the answer is $e^{-\nu t} = 0.5$, which solves for $t = -\ln(0.5)/\nu$.

⁷⁴ “This paper takes Robert Solow seriously”, wrote the authors in the first paragraph of their seminal article (Mankiw, Romer and Weil, 1992, pp. 407).

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

that productivity differences play a dominant role in explaining cross-country differences in per capita income and also in growth rates⁷⁵.

Klenow and Rodriguez Claire considered a production function equal to (4.8). This production function can be solved for per capita income as follows:

$$y_t = A_t^{\frac{1}{1-\alpha-\beta}} \left(\frac{K_t}{Y_t} \right)^{\frac{\beta}{1-\alpha-\beta}} \left(\frac{H_t}{Y_t} \right)^{\frac{\alpha}{1-\alpha-\beta}} \quad (4.24)$$

As explained in Section 3.6, re-writing the production function in such a way has the advantage of abstracting from the capital accumulation that is induced by technological change. Accordingly, the first term in the right-hand side measures the Harrod Neutral level of technology.

Klenow and Rodriguez-Clare calibrated equation (4.24) for 98 countries over the period 1960-85, assuming $\alpha=0.28$ and $\beta=0.30$. Then, they expressed all country' variables as percentage of the corresponding values in the United States (that is, US=1.00).

Table 4.3 displays some of their results. In the table, take Tanzania, for instance. In 1985, per capita income in this country was only 3% of that in the United States. How much of that difference could be explained by physical capital alone? In 1985, the K/Y ratio in Tanzania was 59% of that in the United States. If human capital played no role in the production function (i.e, if $\alpha=0$), then the contribution of the capital labour ratio would be $(K/Y)^{\frac{0.3}{1-0.3}} = 0.79$. That is, with equal productivity and no role for human capital, Tanzania should have a per capita income level corresponding to 79% of that in the US! Now, let's include human capital in the production function. In the table, we see that the ratio H/Y in Tanzania was 37% of the corresponding variable in the United State. Hence, the joint contribution of physical and human capital to relative per capita income would be equal to:

⁷⁵ Klenow, P. and Rodriguez-Clare, "The Neo-Classical Revival in Growth Economics: Has it Gone Too Far?", NBER Macroeconomic Annual, 1997

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

$$X = (0.59)^{\frac{0.3}{1-0.28-0.30}} (0.37)^{\frac{0.28}{1-0.28-0.30}} = 0.35. \quad (4.25)$$

We conclude that adding human capital improves the estimate, but we are still very far from reality. In light of (4.24), the remaining difference can only be accounted for by the productivity parameter:

$$\lambda = A^{\frac{1}{1-\alpha-\beta}} = 0.03/0.35 = 0.08. \quad (4.26)$$

Thus, in the case of Tanzania, the main reason why per capita income is so low relative to that of the United States is that total factor productivity is lower than in the United States. In general, Klenow and Rodriguez-Clare concluded that “richer economies tend to have higher K/Y, higher H/Y and higher A, with a dominant role for A, a large role for K/Y and a modest-to-large role for H/Y” (pp. 89).

Table 4.3 – Development Accounting (1985, US=1.00)

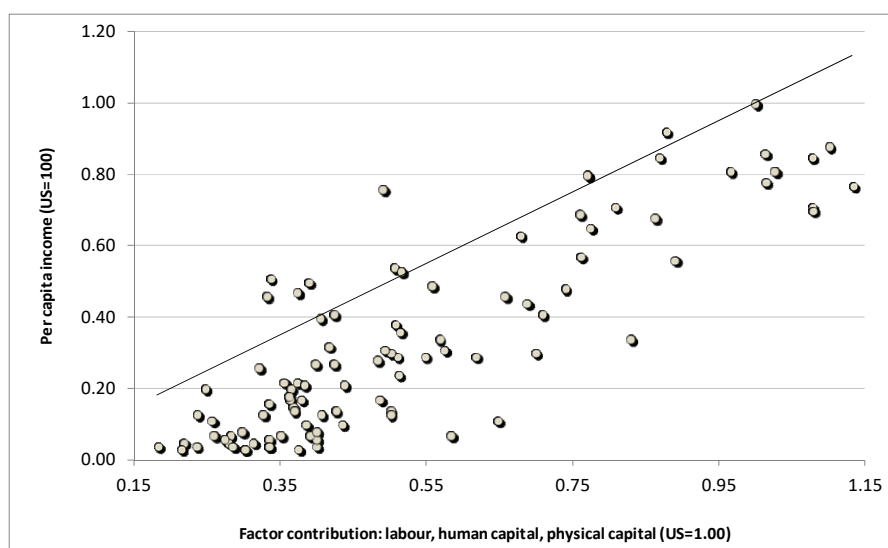
	Output per Working Age Person Y/L	Factor Contribution			Productivity (Harrod Neutral) λ
		K/Y	H/Y	X	
SOUTH AFRICA	0.29	0.84	0.45	0.52	0.57
TANZANIA	0.03	0.59	0.37	0.35	0.08
BRAZIL	0.32	0.70	0.40	0.42	0.77
INDIA	0.08	0.71	0.38	0.41	0.20
INDONESIA	0.13	0.59	0.45	0.40	0.32
MALAYSIA	0.31	0.68	0.51	0.48	0.63
FRANCE	0.80	1.47	0.45	0.77	1.04
NETHERLANDS	0.85	1.28	0.61	0.86	0.98
PORTUGAL	0.34	1.21	0.34	0.56	0.60
TURKEY	0.21	0.79	0.37	0.44	0.48
U.K.	0.68	1.23	0.64	0.86	0.79
PAPUA N.GUINEA	0.10	1.08	0.26	0.43	0.23

Source: Klenow and Rodriguez-Clare (1997). The table displays the decomposition (4.24) for each country, as a percentage of the corresponding variables in the United States, assuming $\alpha=0.28$ and $\beta=0.30$.

A systematic view of the role played by factor accumulation versus factor productivity in their sample of 98 countries is provided in Figure 4.6. The figure crosses the combined contribution of physical and human capital (X) with per capita incomes, with all variables being measured as a percentage of the corresponding values in the United States. The 45° line corresponds to cases where per capita income gaps are fully accounted for factor accumulation gaps. Vertical distances relative to the 45^a line therefore measure the productivity gaps. As the figure reveals, most observations fall below the 45 degrees line. <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

This means that technological gaps play a key role in explaining cross-country differences in per capita incomes.

Figure 4.6 – Development Accounting: contributions of factor accumulation versus per capita income (US=1.00)



Source: Klenow and Rodriguez-Clare (1997). The figure compares the observed per capita output differences with those predicted by equation (4.13) assuming equal TFP. Vertical distances relative to the 45° line measure the importance of TFP.

Box 4.3. Development accounting

The technique of calibrating a production function to measure productivity differences vis-a-vis a reference country is known as “development accounting”. Development Accounting shares with growth accounting the feature that it uses national accounts data to calibrate a production function and disentangle the relative contributions of inputs and TFP to economic performance. Instead of measuring growth rates, however, the method focuses on variables in levels, relative to a reference country (in this case, US=1.00).

In general, “development accounting” exercises reveal that a sizeable proportion of cross-country income disparities cannot be explained by factor accumulation, human capital included. Productivity differences matter.

4.6 Discussion: where are we standing

The basic formulation of the Solow model stresses the role of physical capital in explaining cross-country income differences. Although we observe that richer countries tend to exhibit higher levels of capital per worker than poor countries, a further investigation reveals that the weight given to physical capital in the neoclassical production function is too low to account for the existing per capita income disparities. Furthermore, if capital per worker was really the main explanation for cross-country income differences, capital should flow massively from rich countries to poor countries, which does not happen in reality.

A possible solution to this problem is to allow the parameter A to differ across countries. The parameter A , or total factor productivity, accounts for the influence of all other factors apart from those specified in the production function. For instance, natural resources, human capital, government policies, institutions and advances in technology are thought to impact on the level of A . To the extent that these ingredients are different across countries, one might be able to explain the Lucas paradox.

The avenue followed in this chapter consisted in specifying human capital as an argument of the production function (or, in other words, releasing human capital from A in the original model). Such avenue was proposed by Mankiw, Romer and Weil, in their extension of the Solow model. Like physical capital, human capital can be accumulated through investment: human capital belongs to the general category of reproducible inputs.

With no question, augmenting the model so as to increase the weight of reproducible inputs in the production function improves its explanatory power and its compliance with the real-world facts. However, further empirical scrutiny of the augmented model reveals that a significant proportion of per capita income differences remains to be explained. This suggests that one needs to turn again to the parameter that we label as “technology”, to search for reasons why it can differ across countries. Also, a greater understanding of the growth question will imply learning why technology is expanding over time, instead of simply presuming that it is expanding exogenously at some rate γ .

4.7 Key ideas of Chapter 4

- The Solow model correctly predicts that countries that save more should enjoy higher levels of per capita income, but it fails in predicting how much. The reason is that the

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

fundamental parameter through which cross-country differences in saving rates are mediated – the capital-output elasticity – is, according to the model's assumptions, too small.

- This limitation of the model was put in a simple way by the Robert Lucas Jr: If differences in per capita income were basically explained by differences in the availability of capital per worker, then returns to capital should be very high in capital scarce countries, and capital should flow from rich countries to poor countries. This does not happen in reality.
- In order to solve these inconsistencies, different avenues have been proposed in the literature: the one analysed in this chapter consists in expurgating the “technology parameter” from the influence of an input that is hidden in the Solow specification: human capital.
- Human capital includes both education and wealth. Like physical capital human capital can be accumulated through investment, and it depreciates over time.
- When the neoclassical production function is augmented with human capital, the marginal product of physical capital becomes parametric in human capital. If richer countries have not only more physical capital but also more human capital, the required difference in interest rates does not need to be so large.
- Still, empirical assessments using the extended model reveal that physical capital and human capital together are not enough to account fully for the existing cross-country per capita income differences.
- This means that, in order to have a complete picture of why some countries are richer than others, one should turn again to parameter A and disentangle other factors that influence the productivity of nations.

Problems and Exercises

Key concepts

- *Human Capital. The Lucas (1990) paradox. Conditional convergence. Development accounting*

Essay questions

- Comment: “Because poor countries have less physical capital per worker, returns to capital should be higher there”.

Exercises

- 4.1. Consider an economy (W) where the production function is given by $Y = A_t K_t^{0.5} N_t^{0.5}$, where N measures the number of workers. It is known that, in this country 16% of income is saved, population is expanding at 0.5% per year, the capital stock depreciates

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

- at 2.5% and $A_t = 4e^{0.005t}$. (a) Find out the equilibrium levels of K/L , Y/L and K/Y in this economy, where L represents labour in efficiency units. Discuss the stability of the equilibrium and represent it in a graph. (b) Describe the time-paths of per capita income (Y/N), the wage rate, the interest rate and the factor income shares in the steady state. To which extent this model complies with the Kaldor facts? (c) Consider a second economy (V) which per capita income is 1/16 of per capita income in W . In the absence of further information, in which conditions could you state that per capita income was expanding faster in V than in W ? (d) If the only difference between V and W was the saving rate, how much should that be in V ? what would be the corresponding interest rate? (e) If the only difference between V and W was parameter A , how much should that be in V ? what would be the corresponding interest rate? (f) Given your findings in d) and e), discuss the limitations of the Solow model in explaining cross-country income differences.
- 4.2.** In Micronia the production function of each individual producer is given by: $Y_i = A_i K_i^{1/3} N_i^{2/3}$, where N is the number of workers. In this country the saving rate is 25%, population is constant and the depreciation of the capital stock is 3% a year. Assume that $A_t = 5e^{0.04t}$. (a) Find out the equilibrium values of K/L , Y/L and K/Y , where L is work per efficiency units. Plot it in a graph and explain the stability of the equilibrium. (b) Compute the interest rate and N/Y . Are these values consistent with the empiric evidence? (c) Consider other economy with the same structure of the above except for the saving rate. What would be the saving rate of that economy for per capita income to be 1/10 of the level found in a(i)? Explain the result. (d) What are the advantages of the specification of Mankiw, Romer and Weil (1992) compared to the model of a(i)?
- 4.3.** Consider economy P , where the production function takes the following form, $Y_t = AK^\beta N^{1-\beta}$, and where the share of labour on national income is 2/3. It is also known that per capita income in this economy is about 20% of the corresponding level in the economy R . (a) If the only difference between P and R was the level of capital per worker, how much should be k in P , as compared to R ? (b) In that case, how much should be the interest rate in P as compared to R ? Could such difference be explained by risk premiums? (c) Suppose the data revealed k in P to be roughly 51.2% of the corresponding level in R . If that was the case: (c1) which parameter in the model should capture the remaining difference in per capita income? How much should that parameter differ in the two countries? (c2) would the marginal products of capital still differ in the countries? By how much? Could that difference be explained by risk premiums and other impediments to capital mobility?
- 4.4.** Consider the following production function $Y_t = 3K^{1/3}N^{1/3}H^{1/3}$. Assume that $N=1$ and $H=8$. Compute the marginal product of physical capital and display it a graph. Explain what happens to the marginal product of capital when H increases to $H=27$.
- 4.5.** In Myanmar, per capita income in 1985 was 4% of the level observed in the United States. In both countries, the share of labour on national income was about 60%. In the following, assume that technology was the same in both countries and that there are no
- <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

market failures. (a) Suppose the production function was given by $Y_t = A_t K_t^\beta N_t^{1-\beta}$ in both countries. In that case, $r + \delta$ in Myanmar should be equal to (i) 125 times that in the US; (ii) 8 times that in the US; (iii) 25 times that of the the US; (iv) none of the above. (b) Suppose the production function was given by $Y_t = A_t K_t^{0.4} H_t^{0.3} N_t^{0.3}$, and human capital per worker in Myanmar was equal to 2.56% of the level in the US. In that case, $r + \delta$ in Myanmar should be equal to (i) 125 times that in the US; (ii) 8 times that in the US; (iii) 25 times that of the the US; (iv) we can't compute this value.

- 4.6.** Consider two economies, a rich country, and a poor country, with per capita incomes differing from 10 to 1. Assume that the two economies have the same production function $Y_t = K_t^\beta H_t^\alpha N_t^{1-\alpha-\beta}$. Also assume that economies are open to capital flows, implying the equalization of interest rates (and by then, the marginal product of physical capital). Would returns on human capital be equalized? Discuss.

5 The AK model

“...a level effect can appear as a growth effect for long periods of time, since adjustments in real economies may take place over decades”. [Sachs and Warner].

Learning Goals:

- Understand why getting rid of diminishing returns one can obtain unceasing growth via factor accumulation.
- Distinguish the case with endogenous savings.
- Review different models where simple factor accumulation can generate endogenous growth.
- Acknowledge the empirical challenges raised by the abandonment of diminishing returns.

5.1 Introduction

Along the previous chapters, we learned that, under diminishing returns, factor accumulation cannot, by itself, explain the tendency for per capita income to grow over time. For this reason, a sustained growth of per capita income can only be achieved in the neoclassical model by postulating an exogenous rate of technological progress.

In this chapter it is shown that, by getting rid of diminishing returns on capital accumulation, one can obtain continuous growth of per capita income *without* the need to postulate an exogenous rate of technological progress. This result is shown in terms of the so-called AK model. The AK model differs critically from the Solow model in that it relies on a production function that is linear in the stock of capital. In this model, per capita income grows continuously in the equilibrium, without any tendency to stabilize. In that model, a rise in the saving rate produces a permanent increase in the growth rate of per capita income. This contrasts with the Solow model, where a rise in the saving rate only delivers only a “level effect”.

The pitfall of the AK model is that the assumption of diminishing returns plays a very central role in economic thinking. Hence, it cannot be abandoned without a well motivated

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

story. Some of the sections below describe alternative theories and models that were proposed to motivate the departure from the assumption of diminishing returns on capital. Although none the models described in this chapter shall be seen as *the true model*, they all offer alternative avenues to think the various factor that might contribute to economic development.

Sections 5.2 describes the AK model in its simpler formulation. Section 5.3 extends the AK model to the case of endogenous savings. Section 5.4 reviews alternative models that emulate the AK model. Section 5.5 addresses the empirical evidence on the relationship between savings on economic growth. Section 5.6 concludes.

5.2 The basic AK model

5.2.1 Getting rid of diminishing returns

Consider a closed economy where the population growth rate, the savings rate and the depreciation rate are all constant over time. The novelty relative to the Solow model is that the production is linear in K :

$$Y_t = AK_t, \quad A > 0 \quad (5.1)$$

In (5.1), the parameter A stands for the level of technology (or *aggregate efficiency*), and is assumed constant.

In light of (5.1), production depends only on capital and there are no diminishing returns. The reader may get suspicious about this formulation: after all, does it make sense to model production without labour? In fact, we don't need to assume that labour has no role: in the following sections, we will review alternative models that, while accounting for the role of labour in production, emulate versions of production function (5.1). For the moment, however just stick with this simple formulation.

Dividing (5.1) by N , one obtains a linear relationship between per capita income and capital per worker:

$$y_t = Ak_t \quad (5.2)$$

The remaining equations of the model are the same as in the basic Solow model:

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

$$Y_t = C_t + S_t \quad (2.5)$$

$$sY_t = I_t \quad (2.7)$$

$$\dot{K}_t = I_t - \delta K_t \quad (2.8)$$

$$n = \dot{N}_t / N_t \quad (2.9)$$

From (2.7), (2.8), (2.9) and (5.2), we obtain the dynamics of the capital labour ratio:

$$\dot{k}_t = sAk_t - (n + \delta)k_t \quad (5.3)$$

This equation is similar to (2.14), with the only difference that now $\beta=1$. This small difference has an important implication: since both terms on the right-hand side of (5.3) are linear in k , only by an exceptional coincidence of parameters would this expression be equal to zero. Hence, the general case in the AK model is that there is no steady state.

Dividing (5.3) by k , one obtains the equation that describes the growth rate of capital per worker in this economy:

$$\frac{\dot{k}}{k} = sA - (n + \delta)$$

Since output is linear in K , the growth rate of capital per worker is also the growth rate of per capita income. That is:

$$\gamma = sA - (n + \delta) \quad (5.4)$$

This equation states that the growth rate of per capita income rises with total factor productivity (A) and the saving rate (s) and declines with the depreciation of the capital-labour ratio (n and δ). As long as $sA > (n + \delta)$, per capita income will expand forever, at a constant growth rate. Note that this conformity with the real-world facts is achieved without the need to postulate any exogenous technological progress.

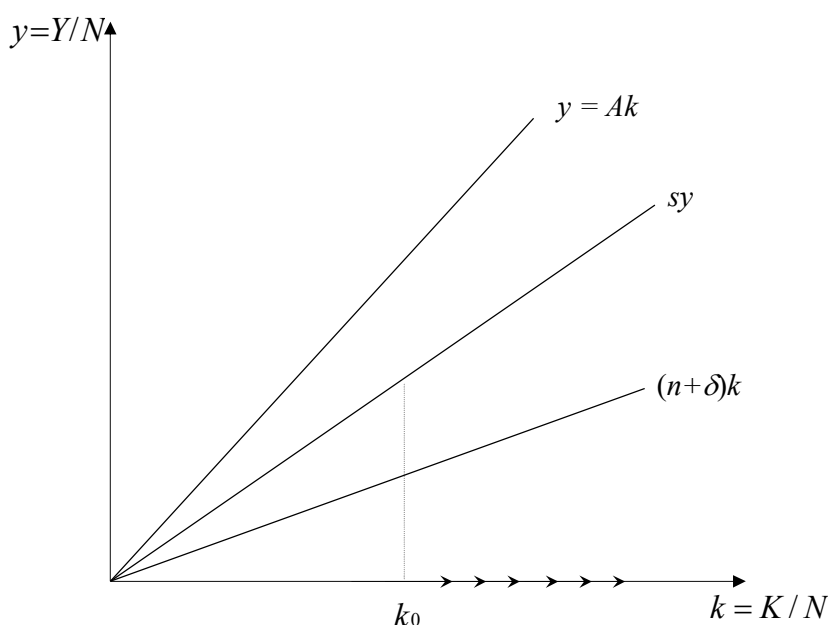
Because the growth rate of per capita income in (5.4) is influenced by the other parameters, instead as being given, the model is categorized as of *endogenous growth*.

5.2.2 A Graphical Illustration

Figure 5.1 describes the dynamics of the AK model. The horizontal axis measures the capital labour ratio (k). The vertical axis measures output per capita (y). The top line shows the production function in the intensity form, (5.2); the middle line corresponds to gross savings per capita (the first term in the right hand side of 5.3); the lower of the three lines is the break-even investment line (the second term in the right hand side of 5.3).

Since the production function is now linear in k , the locus representing gross savings never crosses the break-even investment line (compare with Figure 2.3). This means that, as long as $sA > n + \delta$, per capita output will grow forever, without any tendency to approach an equilibrium level.

Figure 5.1. The AK model



When the production function is linear, the curve describing per capita savings never crosses the break-even investment line. Hence, the capital-labour ratio and per capita output will expand without limits.

5.2.3 What happens when the saving rate increase?

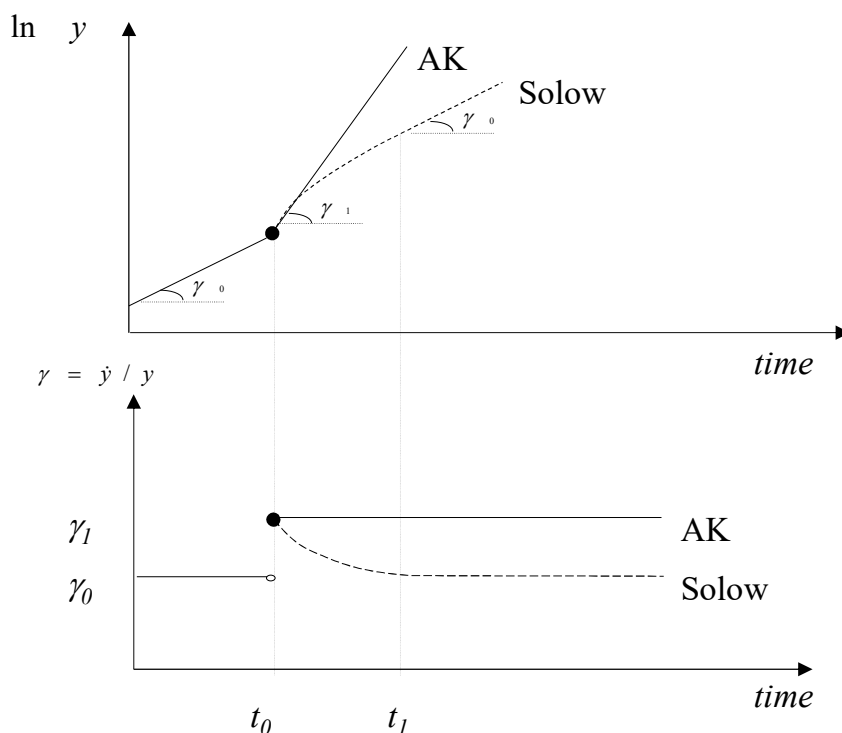
The AK model differs drastically from the Solow model, in that changes in the exogenous parameters alter the long run growth rate of per capita income, rather than the level of per capita income.

Figure 5.2 compares the paths of per capita income in the AK model and in the Solow model following a once-and-for-all increase in the saving rate at time t_0 (the case with the

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Solow model was already discussed in detail in Figure 3.3). The top part of the diagram shows levels and the bottom part shows growth rates.

Figure 5.2: The AK model and the Solow model compared for a rise in the saving rate



The figure compares the response of per capita income to an exogenous increase in the saving rate in light of the AK model versus the Solow model. While in the Solow model this gives rise to a level effect, in the AK model, there is a growth effect. The figure also suggests that, in the short term, the Solow model and the AK model produce similar predictions.

In the Solow model, the rate of growth of per capita income jumps initially to a higher level, but then it declines slowly over time, until returning to the previous - *exogenous* - rate of technological progress. Because of diminishing returns, the long run growth rate of per capita income is independent of the saving rate. In the AK model, the rise in the saving rate has a permanent effect on growth: there is no tendency for the growth rate of per capita income to decline as time goes by. The growth rate of per capita output is proportional to the saving rate.

5.2.4 The Harrod-Domar equation

A useful comparison between the AK model and the Solow model regards their respective behaviours in the long run. Using (5.1) and (5.4) we get:

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

$$\gamma = s \frac{Y}{K} - (n + \delta) \quad (5.5)$$

This equation is known as the Harrod-Domar equation. The difference between (5.5) and (3.11) (that holds in the Solow model in the long run) refers to the variables that are exogenous and endogenous in this equation. In both models, s , n and δ are exogenous. But the two models differ in respect to the exogeneity of δ and Y/K : In the AK model, Y/K is exogenous and γ is endogenous. By contrast, in the Solow model, γ is exogenous and Y/K is endogenous.

Hence: In the Solow model, a rise in the saving rate leads to a lower average productivity of capital in the steady state. That is, Y/K declines from one steady state to the other (Figure 3.2). In the AK model, Y/K is constant (equal to A). Hence, a rise in the saving rate can only be accommodated in the model by an increase in the growth rate of per capita income, γ .

Because the AK model predicts that changes in A or in the saving rate produce growth effects, it goes far beyond the neoclassical model in stressing the relationship between economic policies and economic growth: government policies, such as taxes and subsidies, that affect economic efficiency and consumption-saving decisions may alter the long run' rate of economic *growth*, rather than simply altering the *level* of per capita income.

5.2.5 No convergence

The AK model does not predict convergence of per capita incomes, even among similar economies. According to (5.4), two economies having the same technology and savings rates will enjoy the same growth rate of per capita income, regardless of their starting position. In that case, the respective per capita incomes will evolve in parallel without any tendency to approach each other. This contrasts to the Solow model, where countries with similar parameters should approach the same per capita income level in the steady state.

Moreover, since changes in technology (A) and in the saving rate (s) affect growth rates permanently, countries with different parameters should exhibit different growth rates of per capita income. In a world where policies differ substantially across countries, the rule should be that of divergence of per capita incomes, rather than of convergence.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

5.3 Incarnations of the AK model

5.3.1 The Harrod-Domar model

The true predecessor of the AK model was developed independently by two economists, Roy Harrod, and Evsey Domar⁷⁶. The Harrod Domar model preceded that of Solow by several years and obviously it was not motivated by any explicit intention to improve on the Solow model. The HD model was developed in the aftermath of the Great Depression, as a dynamic extension of Keynes' general theory. The aim was to discuss the business cycle in the U.S. economy. Since at that time, unemployment was very high, the focus of the model was on the relationship between investment in physical capital and output growth.

The main assumption of the Harrod-Domar model is that capital and labour are pure complements, meaning that they cannot substitute for each other in production. The underlying production function is Leontief:

$$Y_t = \min \{AK_t, BN_t\}, \quad (5.6)$$

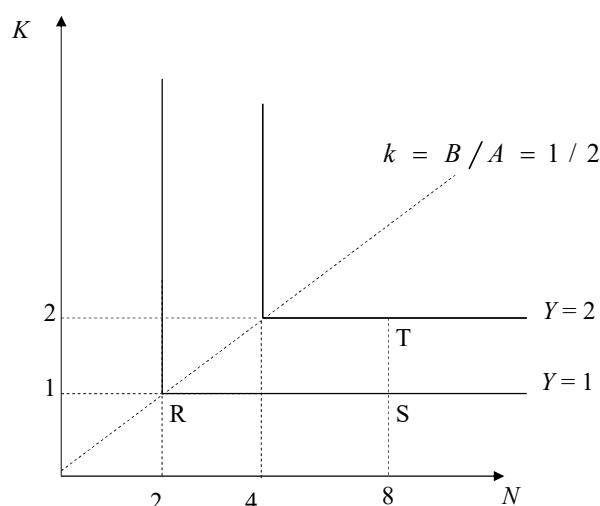
where A and B are positive constants.

The Leontief production function contrasts with the Cob-Douglas production function in that inputs cannot substitute for each other. As an illustration, suppose that your output (Y) was a meal consisting in a "steak with two eggs". To produce any amount of this output you would need steaks (K) and eggs (N) in a proportion of two eggs per one steak. Figure 5.3 illustrates this, by displaying the isoquants corresponding to $A=1$ and $B=0.5$. Thus, to produce one meal ($Y=1$), you need at least one steak and two eggs (point R). If you employed one steak and 8 eggs, your maximum production would still be equal to one meal (point S).

Figure 5.3: The Leontief production function

⁷⁶ Harrod, R. , 1939. "An essay in dynamic theory". Economic Journal 49, 14-33. Domar, E., 1946. "Capital expansion, rate of growth and unemployment", Econometrica 154, 137-47..

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>



The figure describes two isoquants where inputs to production are complementar. The straight line A/B corresponds to the efficient combinations. If the economy's endowment point lies on the right-hand side (left hand side) of this line, there will be unemployment of labour (capital).

Now think that this production function applied to the economy as a whole and that K and N referred to capital and labour. If the economy's endowments were $K=1$ and $N=8$, the economy would be producing $Y=1$ only, wasting 6 unit of labour (point S). From that point, expanding the quantity of labour would not deliver higher output, because labour cannot substitute for capital. The only way to expand production will be increasing the stock of physical capital. If one managed to increase the stock of capital to $K=2$, the output level would jump to $Y=2$ (point T), and unemployment would be reduced to 4 units of labour. Raising production by incrementing the stock of capital (K) in an economy with surplus labour (N) is basically how the Harrod-Domar model works.

Mathematically, a situation of employment surplus occurs when $K/N < B/A$. In that case the relevant branch of the production function (5.6) is the first, implying a linear relationship between output and K , $Y=AK$. This is basically the AK model. Then, given the exogenous saving rate and the population growth rate, from (2.5)-(2.9), you'll obtain the growth rate of per capita income as described by (5.4).

The main limitation of the Harrod-Domar is that factor prices play no role in driving the economy towards full employment of labour and capital. This contrasts with the Solow model, where real wages and the real interest rate adjust to ensure full employment each moment in time. Thus, even if output is expanding over time at the rate $\dot{Y}/Y = sA - \delta$, this may not be enough for per capita income to increase: in case $sA - \delta < n$, the economy does

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

not save the enough to keep the capital labour ratio unchanged. Population will be expanding faster than output, surplus labour will be increasing (chronic underproduction) and per capita income will be declining.

If, on the contrary, the capital stock grows faster than population ($\gamma = sA - \delta - n > 0$), then per capita income will be increasing over time, and eventually the surplus labour will be eliminated in the long run. Still, the mechanics of the model is such that per capita income cannot growth indefinitely. The reason is that at the time the full employment locus is crossed, the relevant segment of the production function in (5.6) shifts to , and the binding constraint in production becomes the availability of labour: beyond this point output is be bound to expand at the same rate as population, implying a constant level of per capita income thereafter⁷⁷.

Box 5.1: The ghost of financing gap

One of the reasons why the Harrod Domar equation (5.5) became so popular is that it offers a simple and appealing formula to forecast economic growth. This formula was also extensively used by international organizations, such as the World Bank, to calculate a country' financing needs.

If equation (5.5) was true, one could easily forecast a country' economic growth, using the saving rate, the depreciation rate and an estimate for the average product of capital, A. Since the later is not readily available in national accounts, a possible proxy would be the ratio of net investment to the change in real GDP over two consecutive years:

$$ICOR = \frac{\Delta K}{\Delta Y} = \frac{\text{net investment}}{\text{change in GDP}}$$

⁷⁷ Alternatively, one may assume that, after unemployment is eliminated, wages start increasing and the economy enters in "modern economic growth". A seminal contribution along this avenue is due to Arthur Lewis (1954). [Lewis, W., 1954. "Economic Development with Unlimited Supplies of Labour". The Manchester School of Economic and Social Studies 22, 139-191].

This is the known as the "Incremental Capital-Output Ratio", ICOR.

As an example, consider a poor economy where the $ICOR = 3$ and the observed investment ratio (s) is 15%. Assuming a depreciation rate equal to 4%, equation (5.4) implies that output will grow at $0,15/3 - 0,04 = 0,01$. Now suppose you were a consultant for that economy, advising on poverty alleviation. You could well conclude that the saving rate in this economy was too low. If, for instance, population was growing at 2%, that would imply a fall in per capita income...

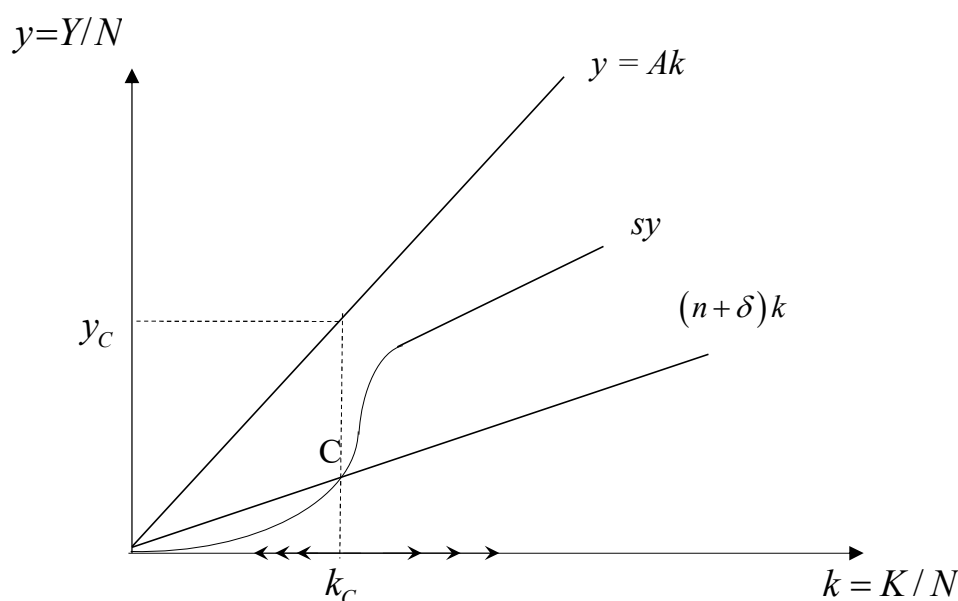
You could, then, use the HD equation the other way around: how much should the investment rate in this country, for per capita income to increase at some desired rate? Suppose you wanted income per capita to expand at 2% per year. With the population growing at 2% and a measured ICOR equal to 3, following (5.5), you would need a net investment amounting to 24% of GDP. Since domestic savings were only 15%, you could request the international donors to fill the "financing gap", equal to 9% of GDP.

Economists in international institutions, such as the World Bank, the IMF, the Inter-American Development Bank, the European Bank for Reconstruction and Development used models based on the HD equation to estimate the amount of savings (and/or aid) necessary for poor countries to achieve some desired rate of economic growth. This philosophy was supported by the understanding that people living near the subsistence level cannot save the same as rich people.

To illustrate the argument, we refer to figure 5.4. This figure is similar to 5.1, the difference being that the saving rate starts out very low and then increases with per capita income. In that case, a bifurcated growth pattern emerges, whereby per capita income increases forever or decreases forever, depending on the initial level of capital per worker. Now suppose that a poor country was trapped in situation with insufficient savings. Say, because a natural disaster caused the capital stock to fall below the critical level, or because a demographic explosion tilted the break-even investment line upwards, turning the existing capital stock insufficient. In that case, per capita income would start decreasing over time. Arguably, foreign aid could fix this: if the external aid succeeded in raising savings above the critical level, it could be that the country engaged in a self-sustained growth path. In this case, foreign aid would need only to be temporary.

Recent proponents of this idea include Sachs (2005) and the United Nations Millennium Development Goals Project. Sachs (2005): “(...) if the foreign assistance is substantial enough, and lasts long enough, the capital stock rises sufficiently to lift households above subsistence...growth becomes self-sustained (...)” (p. 246). United Nations (2005, p. 19): “The key to escape the trap is to raise the economy’s capital stock to the point where the downwards spiral ends and self-sustained economic growth takes over” (p. 19)⁷⁸.

Figure 5.4: The AK model with a non-linear saving rate



The figure describes a case where, at low levels of income, the saving rate is an increasing function of income. In this case, the model displays one unstable equilibrium, where per capita income is stagnant. At the left of this equilibrium, savings are insufficient, and per capita income decreases over time; at the right-hand-side, per capita income increases over time.

The view that external aid would be the panacea to help poor countries to escape poverty traps was seriously criticised by William Easterly in controversial paper called “The

⁷⁸ [Sachs, J., 2005. The end of poverty: economic possibilities for our times. New York: Penguin Press. United Nations, 2005. Millenium Development, Project Report, United Nations, New York].

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

ghost of financing gap”⁷⁹. The author noted that, over the past four decades, large amounts of international financial assistance to the developing world did not translate into faster economic growth. Using a sample of 146 countries along the period from 1950 to 1992, the author failed to find a robust positive linear relationship between aid and economic growth.

Does this mean that the HD equation is wrong? Not necessarily. But probably one should not trust too much historical values of the average product of capital (the ICOR) to guess the marginal impact of new investments: if for instance, part of the external aid is diverted into unproductive uses (frivolous expenses, corruption fees), then much of the higher saving rate will be offset by a lower A . Arguably, the impact of external aid on growth shall depend on the quality of policies and institutions of the recipient country (see discussion in Box 5.3). Note however, by now, models have been silent regarding the role of policies and institutions.

5.3.2 *A one sector model with Physical and Human Capital*

Another way of accounting for the role of labour in production and obtain an AK type of model is by considering two types of capital, physical and human capital, and assuming that constant returns apply to the *broad* concept of capital⁸⁰.

To see this formally, consider the following production function:

$$Y = AK^{\beta}H^{1-\beta} \quad (5.7)$$

In this production function, there are diminishing returns to physical capital and to human capital in isolation, but there are constant returns to scale in reproducible factors. This contrasts to the Solow and the MRW models, where returns to reproducible factors are decreasing due to the presence of a non-reproducible factor, labour.

⁷⁹ Easterly, W., 1999. "The ghost of financing gap: testing the growth model of the international financial institutions". *Journal of Development Economics* 60 (2), 423-438.

⁸⁰ Rebelo, S., 1991. Long run policy analysis and long run growth, *Journal of Political Economy*, 99, 500-521.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Also note that this production function does not necessarily exclude raw labour from production. Indeed, one may think human capital, H , as measuring *quality adjusted* labour, that is, the number of workers, N , multiplied by the human capital of the typical worker (h):

$$H = hN . \quad (5.8)$$

The implied assumption in (5.8) is that the quantity of workers, N , and the quality of workers, h , are substitutes. With such a specification, raw labour needs no longer to be a source of diminishing returns: multiplying h and K by a given constant implies that production Y will expand proportionally, even if N remains constant. The CRS property ensures the linearity between production and reproducible factors, and this is what we need to generate economic growth. This linearity does not apply to MRW model, because the non-reproducible factor (N) cannot be perfectly substituted by human capital, implying diminishing return on broad capital (K plus H).

To see this how the model works, let's return to the MRW assumptions that people save a constant fraction of their incomes in the accumulation of human capital, just like they do for physical capital⁸¹:

$$\dot{K}_t = sY_t - \delta K_t \quad (5.9)$$

$$\dot{H}_t = s_H Y_t - \delta H_t \quad (5.10)$$

Because of diminishing returns to each type of capital, it doesn't make sense for people to accumulate one type of capital faster than other. Hence, the two types of capital will be expanding at the same rate $\dot{K}/K = \dot{H}/H$. Using (5.9) and (5.10) this implies

$$\frac{\dot{H}}{H} = \frac{s_H}{s} \quad (5.11)$$

⁸¹ Equation (5.10) describes the change in H . In term of, human capital per worker, the corresponding dynamics will be $\dot{h}_t = s_H y_t - (n + \delta) h_t$.

Substituting (5.11) in (5.7), we obtain a variant of the AK production function:

$$Y = A \left(\frac{s_H}{s} \right)^{1-\beta} K \quad (5.12)$$

Comparing to (5.1) we see that now the average product of capital embodies the propensity to invest in human relative to physical capital (that is, you can look at A in equation 5.1 as including this effect).

Using this in (5.5), the growth rate of per capita income in this variant of the AK model is:

$$\gamma = s^\beta s_H^{1-\beta} A - n - \rho. \quad (5.13)$$

This equation shows that it is perfectly possible to have diminishing returns to physical capital alone and yet having sustained growth of per capita income. What we need is to have constant returns to all types of capital (or *reproducible* inputs) when considered together. Comparing to (5.4), what matters for growth is not the investment rate in physical capital alone, but rather the average investment rate in physical and human capital.

5.3.3 A two-sector model of endogenous growth

Another incarnation of the AK model was proposed by Usawa, as early as in 1965⁸². The author extended the Solow model, considering two sectors of production: one producing final goods that can be either consumed or accumulated as capital stock (productive sector) and other pooling together various activities that contribute to the efficiency of labour, such as health, education, and research (“educational” or “research” sector). The production sector employs labour and capital. The educational sector employs labour only.

⁸² Usawa, H., 1965. Optimum technical change in an aggregative model of economic growth”, International Economic Review, 6, 19-31.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

In the model, it is assumed that workers devote a fraction $1 - \mu$ of their working time to production of goods and the remaining μ to the improve labour efficiency, λ . The production function for final goods is given by:

$$Y = AK^\beta [(1 - \mu)\lambda N]^{1-\beta} \quad (5.14)$$

The change in technology is a positive function of the fraction of labour allocated to the educational sector:

$$\dot{\lambda} = b\mu\lambda \quad (5.15)$$

The parameter b shall be interpreted as the *productivity in the research sector*. In this model, the linearity that is needed to generate endogenous growth arises from the fact that the production function for technology (5.15) depends linearly on the level of technology, through the *standing on shoulders effect*: a constant fraction of working time devoted to research produces a constant growth rate of technology that is independent on the existing level of technology. In other words, there are no diminishing returns to technology on technology creation). With such an assumption, a policy change that successfully increases the proportion of time allocated to research (μ) or that improves the productivity in that sector (b) will impact positively and permanently on the growth rate of per capita income⁸³.

As for physical capital accumulation, the assumption (5.9) is retained. This model can be solved in the same manner as the Solow model. For mathematical convenience, let's rewrite the production function (5.14) as follows:

$$\tilde{y} = A(1 - \mu)^{1-\beta} \tilde{k}^\beta \quad (5.16)$$

Where $\tilde{y} = Y/L$, $\tilde{k} = K/L$, and $L = \lambda N$. Proceeding as usual, the fundamental dynamic equation becomes:

⁸³ If, in alternative technological progress had diminishing returns of technology (reflecting a “fishing out effect”, leading to $\dot{\lambda} = b\mu\lambda^\varepsilon$ with $\varepsilon < 1$), then the growth rate of technology would tend to zero, no matter how much effort was allocated to the educational sector. Sustained growth could not be achieved.

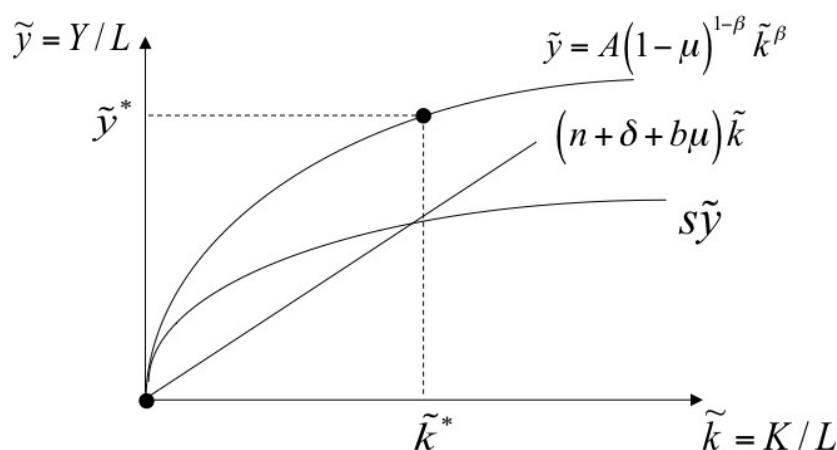
<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

$$\dot{\tilde{k}} = sA(1-\mu)^{1-\beta}\tilde{k}^\beta - [n + \delta + b\mu]\tilde{k} \quad (5.17)$$

Comparing with (3.8) you can verify how similar this model is with the Solow model. The main difference is that the parameter determining the effectiveness of labour, rather than growing exogenously, is now dependent of other parameters in the model.

This model is hybrid, in the sense that it shares characteristics with the AK model and with the Solow model. It shares with the neoclassical model the feature that it has a stable steady state. To find the steady state, we just need to solve for $\dot{\tilde{k}} = 0$. Figure 5.5 illustrates the steady state of the model. Like in the Solow model, changes in s produce “level effects”, causing the steady state level of output per unit of efficiency to increase. Because both $\tilde{y}$ and $\tilde{k}$ are constant in the steady state, the output-capital ratio is constant and so will do the interest rate. In the steady state: $\hat{Y} = \hat{K} = \gamma + n$ and $\hat{y} = \hat{Y} - n = b\mu$.

Figure 5.5: the steady state in the two-sector model



The dynamics of the Usawa model is similar to that of the Solow model. There is a stable steady state when the break-even investment line crosses the savings locus. The main difference is that both the production function and the break-even investment line are tilted when the fraction of time devoted to research changes.

Contrasting to the Solow model, the long run growth rate of per capita income ($b\mu$) is related to other parameters in the model. It depends on the proportion of working time that people allocate to the educational sector and on the effectiveness of that time, b . Thus, for instance, a policy that is successful in inducing an increase in the proportion of time allocated to research raises the growth rate of the economy on a permanent basis (growth effect). The

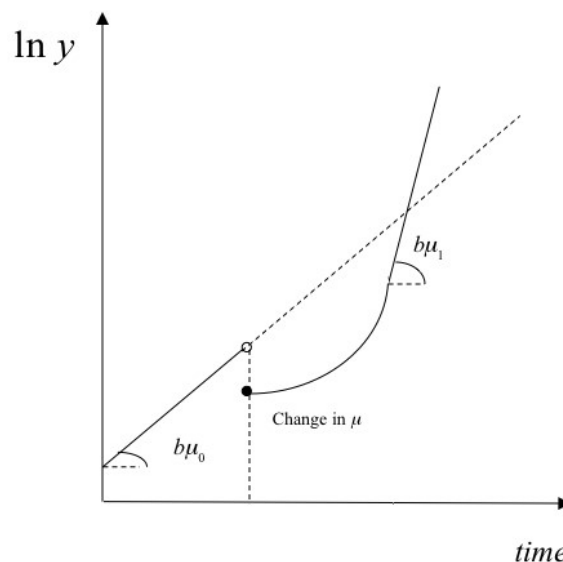
model is capable of generating sustained growth of per capita income without the need to assume exogenous shifts in the production function.

Technically, sustained growth is obtained in this model because the production function of technology is free of diminishing returns. In other words, the model overcomes diminishing returns to physical capital by postulating a linear production function for technology. Physical capital can then be accumulated without seeing its productivity declining, because technology is expanding. Rewriting the production function (5.14), we see that:

$$Y = A \left(\frac{1-\mu}{\tilde{k}} \right)^{1-\beta} K \quad (5.18)$$

Since in the steady state $\tilde{k}$ is constant, the long run version of the model is no more than another incarnation of the AK model. In the short run, however, $\tilde{k}$ is not in general constant, so the model also displays transition dynamics.

Figure 5.6 – The path of output per capita following an increase in μ



When the fraction of time employed in research increases, output in the production sector declines, causing per capita income to fall. However, there is a growth effect that, in the long run, more than offsets the negative level effect. Immediately after the shock, the growth rate of per capita income is low, reflecting the fact that the capital per unit of efficiency labour (horizontal axes in figure 5.5) is approaching a new – lower – level.

Figure 5.6 describes the path of per capita income in this economy following an increase in the time allocated to the educational sector: at the impact, there is a negative effect on per capita income, because less time is devoted to production. As the times go by, however, the growth rate of per capita output accelerates, due to the faster rate of technological expansion. Note that in Figure 5.5 the production function and the break-even investment line shift in opposite directions. Hence, $\tilde{k}$ starts decreasing until meeting the new steady state. This implies that, during the transition period the growth rate of per capita output is less than the corresponding level in the steady state. As the economy approaches the new steady state, the decline in $\tilde{k}$ decelerates, implying a convergence of per capita income growth to the new steady state growth rate, $b\mu_1$.

A question that naturally arises is how people decide the optimal level of μ . Intuitively, the time allocated to the educational sector versus final good production shall be determined by an arbitrage conditions, stating that the payoff of the two activities at the margin should be the same. Thus, the optimal allocation of time shall balance a variety of factors, such as the productivity of research (b), the opportunity cost of research time (the wage rate), the level of impatience of people, and so on. However, with the present formulation, we cannot explore these ideas formally.

5.4 The AK model with endogenous savings

Thus far, the saving rate in the AK model has been assumed exogenous. In this section we show that, when the model is extended to allow individuals to optimally trade consumption today for consumption in the future, a second channel linking efficiency to growth is opened up.

5.4.1 Adding the optimal consumption rule to the AK model

In what follows, let's recall the simplest possible optimal consumption rule, introduced in appendix 2.1:

$$\gamma = r - \rho \quad (5.19)$$

This equation states that, as long as the interest rate is higher than the rate of time preference, there will be an incentive for households to increase consumption over time. This, <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

in turn, is achieved through a higher saving rate. Note that in this model (because there is no transitional dynamics), consumption and income evolve in parallel each moment in time. Hence, (5.19) can be seen as describing simultaneously the growth rate of per capita consumption and the growth rate of per capita income.

To find out how the growth rate of per capita income relates with the remaining parameters of the AK model, one needs an interest rate. As before, we assume that firms are perfectly competitive and maximize profits. In this case, capital will be paid its marginal product, A . That is:

$$r + \delta = A \quad (5.20)$$

Substituting (5.20) in (5.19) and rearranging, one obtains:

$$\gamma = A - \delta - \rho \quad (5.21)$$

This equation describes the growth rate of per capita income in a version of the AK model where consumers are allowed to trade consumption today for consumption in the future at interest rate r . Comparing to (5.4), you see that now it is the rate of time preference, instead of the saving rate, that determines the rate of economic growth.

5.4.2 Transpiration responds to inspiration!

From the qualitative point of view, equation (5.21) brings no novelty relative to the case with exogenous savings, (5.4): a lower rate of time preference (that is, a change in consumer preferences towards more consumption in the future and less consumption today), by raising the saving rate, leads to a higher rate of capital accumulation and a higher growth rate of per capita income.

However, comparing (5.21) to (5.4), we observe that the impact of A on growth is now much larger than in the case with exogenous savings. For instance, with a saving rate equal to 20%, the impact of a unitary change in A on growth in light of (5.5) is 0.2. In light of (5.21), however, the impact of A on growth is one to one. That is: five times more.

What makes the assumption of endogenous savings so powerful that it can alter dramatically the relationship between the efficiency parameter and growth? The reason is that, when A increases, there are two effects: On one hand, when A increases for a given s , the growth rate of per capita income increases, just like in (5.4); on the other hand, when A

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

increases, there is an additional impact mediated by the interest rate, r : a higher marginal productivity of capital translates into a higher return on capital and this, in turn, induces a higher saving rate, for each rate of time preference. Then, a higher saving rate, impacts positively on growth.

Formally, the double impact of A on growth under endogenous savings can be assessed substituting (5.21) in (5.4) and solving for s , to obtain the (endogenous) saving rate: $s = 1 - (\rho - n)/A$. For the problem to be well behaved, $\rho > n$ must hold⁸⁴. Taking the partial derivative in respect to A , we verify that the impact of a change in A on the saving rate is $\partial s / \partial A = (\rho - n) / A^2$. The total impact of a change in A is the sum of the direct impact of A on γ with the indirect impact, through s : $d\gamma / dA = \partial\gamma / \partial A + (\partial\gamma / \partial s)(\partial s / \partial A) = s + (\rho - n) / A = 1$.

This finding is of the upmost importance to understand the mechanics of many endogenous growth models. A typical assessment based on equation (5.4) is that a country may either grow through “inspiration” (A) or through “transpiration” (s). But, we just saw that “transpiration” responds to “inspiration”: a more efficient resource allocation, leading to a higher marginal product of capital, implies a higher return on investment. Thus, agents will be willing to forego a higher proportion of their consumption to save more.

With this finding, one may rewrite equation (5.4) in the following form:

$$\gamma = s \left(\overset{-}{\rho}, \overset{+}{A} \right) A - (n + \delta) \quad (5.22)$$

The fact that capital accumulation is itself endogenous in respect to the level of productivity has important policy implications: how can governments improve the level of efficiency in an economy? Is just a question of implementing the right policies and imitate best practices? Or there are institutional and cultural elements that prevent governments from adopting the best policies? Or is it shall be the case that, irrespectively the good intentions of governments, that are natural factors such as climate and location that prevent productivity in

⁸⁴ We skip the proof of this.

different geographical locations to approach each other, dooming countries facing adverse geographical conditions to permanent lower growth?

5.4.3 Proximate versus fundamental causes of economic growth

According to equation (5.4), a low rate of economic growth can be explained either because a country does not invest enough (s) or because it has low productivity of capital (A). Dealing with the development question at a deeper level, however, one may ask why some countries save and invest more than others and why some countries reach higher levels of efficiency than others. In other words, one would like to consider as endogenous some of the parameters that the model takes as exogenous.

To some extent, equation (5.22) is a step in that direction: according to this equation, individuals will save more in countries where the productivity of capital is higher⁸⁵. This gives parameter A a key role in this model.

The following chapters will be devoted to a better understanding of what is behind parameter A . In this quest, we will relate the level of A to the quality of economic policies, among other factors. We will argue that countries with sound economic policies are expected to achieve higher efficiency levels than countries with poor economic policies. But another question will immediately arise: why do some countries implement better policies than others? To answer this question, we need to address the incentives of policymakers. These, in turn depend on the quality of political institutions. These, in turn, are embedded in even deeper factors underlying human societies, such as social norms, culture and geography.

In a word, as one deepens the analysis, we move from the *proximate* causes of economic growth (the exogenous parameters in equation 5.4), to the *fundamental causes* of

⁸⁵ Equation (5.22) stresses the causality from “inspiration” to “transpiration”. However, the reversal may also be true. In Chapter 6 we’ll address precisely some theories according to which the level of A is enhanced by capital accumulation. The possibility of mutual causation implies that savings and efficiency may reinforce each other, both positively and negatively, raising the possibility of multiple equilibria and poverty traps.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

economic growth, which ultimately determine why in a given country the parameter values are what they are. These fundamental causes are essential to understand why some societies make choices that translate into the adoption of better policies and more modern technologies.

This is not to say that simple models like (5.4) are useless. On the contrary, they are essential to understand *the mechanics of economic growth*. In particular, the role of investment and technology as *mediators* between country characteristics and economic performances. But dealing with the growth question at a deeper level, one may want to understand what is behind the parameters that this model takes as exogenous.

Box 5.2. Developing accounting by Hall and Jones

In a famous article, Hall and Jones (1999) proposed an exercise of development accounting based on a production function of the form⁸⁶:

$$Y = K^{1/3}(\lambda H)^{2/3} \text{ with } H = hN. \quad (5.22)$$

In (5.22), K denotes for physical capital, H for human capital, h for human capital per worker, N the number of workers and λ is a measure of productivity.

They then proposed the following re-specification:

$$y = \left[\left(\frac{K}{Y} \right)^{\frac{1}{2}} h \right] \lambda = X\lambda \quad (5.23)$$

This equation breaks down differences in output per capita into differences in the capital-output ratio, differences in human capital per worker and differences in productivity. The first two terms (in the square brackets) account for the contribution of inputs (*transpiration*, X) and the last term measures TFP (*inspiration*, λ). As in standard growth

⁸⁶ Hall, R., and C. Jones, 1999. "Why do some countries produce so much more output per worker than others?", *The Quarterly Journal of Economics* 114 (1), 83-116.

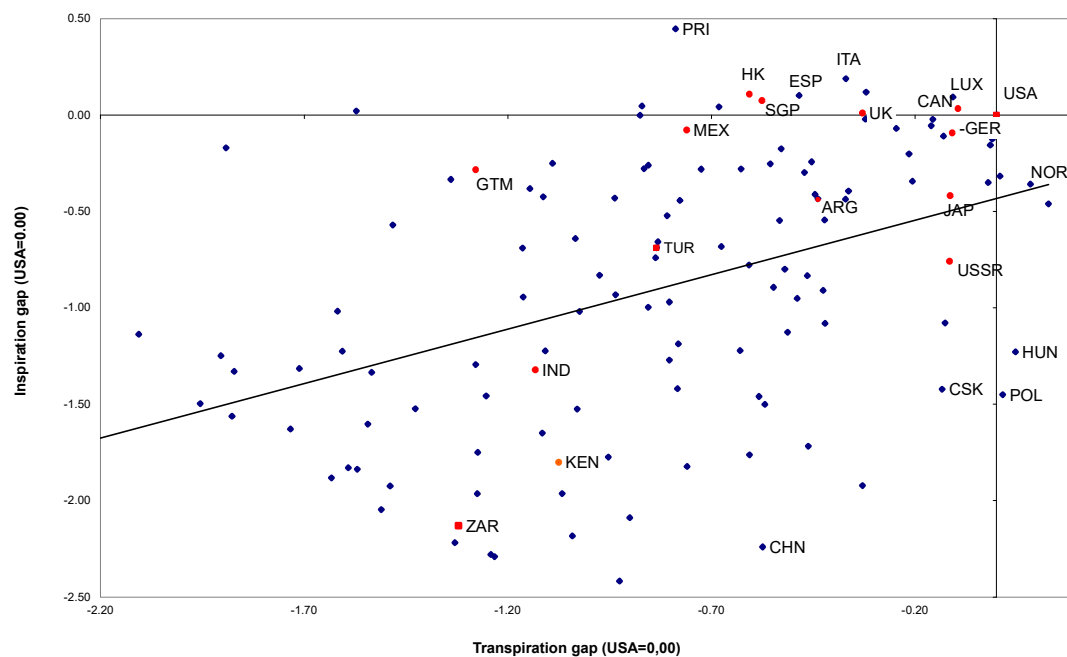
<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

accounting, the later is obtained as a residual. Hall and Jones implemented this decomposition for the year of 1988. In their calculations, all variables are measured relative to the United States. Human capital per worker h was computed as a function of the (observed) average years of schooling.

Figure 13.7 summarises the authors calculations of X and λ . For each country, the vertical axes measures the estimated value of λ relative to the corresponding value in the US, in logs. For instance, according to the figure, Italy has a positive *inspiration gap* vis-à-vis the world leader, while USSR has negative *inspiration gap*. The horizontal axes measures the extent to which a country accumulated more or less human and physical capital (as captured by X) than the US. For instance, according to the figure, Norway has positive *transpiration gap* while Argentina has a negative *transpiration gap*.

In the figure, the positively sloped line is a 45° line, describing the points for which the two gaps are of the same size. For example, Argentina and Turkey exhibit patterns that are “balanced” in terms of inspiration gap and transpiration gap. On the contrary, the USSR had a large inspiration gap and very low transpiration gap, reflecting a massive investment in physical and human capital in a poor innovative environment and low ability to implement foreign technology. An interesting pattern in Figure 5.7 is that countries with large inspiration gaps also tend to exhibit large transpiration gaps. In other words, the two variables are largely correlated across countries.

Figure 5.7: Inspiration gap versus transpiration gap for 127 countries (US=0.00)



Source: Hall and Jones (1999).

As we know, correlation is not causality. Hall and Jones (1999) argued that this correlation was capturing a causality from λ to X . The authors conjectured that cross-country differences in physical and human capital accumulation and productivity are fundamentally related to cross-country differences in “social infrastructure”, that is, “the institutions and government policies that determine the economic environment within which individuals accumulate skills, and firms accumulate capital and produce output” (p.84). A social infrastructure will be favourable to economic development if it provides an environment that is supportive of investment and innovation, rather than promoting theft, corruption or confiscatory taxation.

To test this proposition, the authors constructed a measure of “social infrastructure”, as a combination of two indexes. The first is a measure of government anti-diversion policies (defined as an average of five indexes: law and order, bureaucratic quality, corruption, risk of expropriation and government repudiation of contracts). The second is a measure of openness to trade. Controlling for feed-back effects, the authors found that countries with a good social infrastructure tend to have high capital intensities, high human capital per worker and high output per worker.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

The Hall and Jones (1999) results point to a distinction between the proximate causes of growth (as captured by capital accumulation) and the deep causes, which ultimately determine the individual's willingness to produce and invest. According to this interpretation, policy choices such as the size of the government, the rate of inflation and innovation are all best thought as outcomes (that is, proximate causes) rather than as determinants. This work points to the critical role of institutions as drivers of long run economic performance.

5.5 Empirical controversies

5.5.1 Level effects or growth effects?

The AK model differs dramatically from the exogenous growth model, in terms of the relationship it establishes between the investment rate and economic growth. In the Solow model, a change in investment rate alters the steady state level of per capita income; in the AK model, a change in investment rate alters the growth rate of per capita income. This prediction suggests a natural way of assessing what is the model that better complies with the real world facts: if, on average, countries that invest more achieve higher rates of per capita income, eventually that will be supportive of the AK model.

Table 5.1 displays some evidence provided by Klenow and Rodriguez-Clare (2005)⁸⁷. The table shows the estimates of nine regression equations, implemented for three samples: a large sample with 96 countries, a subsample consisting in OECD economies, and a subsample consisting in non- OECD economies. In each panel, two dependent variables are experimented: the growth rate of per capita income along 1960-2000 ("growth"), and the log of per capita in 2000 ("level"). In the left panel, the explanatory variable is the investment rate in physical capital; in the right-panel, the independent variable is the investment rate in human capital.

⁸⁷ Klenow, P. and Rodriguez-Clare, A., 2005. "Externalities and Growth". In Aghion, P., and Durlauf, S. (eds), Handbook of Economic Growth, North Holland, Amsterdam, Chapter 11, 817-866.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Table 5.1. Investment rates correlate more with levels than with growth rates

	Independent variable: S			Independent variable: SH		
	Dependent variable: Y/N		number of countries	Dependent variable: Y/N		number of countries
	Growth	Level		Growth	Level	
All countries						
beta	0.111	1.25	96	0.21	0.313	74
Std.dev	(0.017)	(0.13)		(0.060)	(0.026)	
R-sq	0.32	0.48		0.15	0.67	
OECD						
beta	0.02	0.76	23	-0.259	0.119	21
Std.dev	(0.047)	(0.358)		(0.078)	(0.024)	
R-sq	0.01	0.18		0.37	0.56	
Non-OECD						
beta	0.124	0.842	73	0.367	0.314	53
Std.dev	(0.023)	(0.162)		(0.095)	(0.043)	
R-sq	0.29	0.28		0.22	0.51	

Source: Klenow and Rodriguez-Clare (2005). Notes: Y/N is GDP pr worker; S is the physical capital investment rate; SH is years of schooling (for the 25+ population) divided by 60 years (working live). These variables are averages over 1960-2000.

As shown in the table, investment rates in physical capital and in human capital are both positively correlated with growth in the sample of 96 countries and in the non-OECD sub-sample. In the OECD sample, investment in physical is not significant, and investment in human capital gets the wrong sign in explaining economic growth. In contrast, we observe that both investment variables are significant and with the right sign when predicting per capita income levels. In general, the correlation coefficient is higher when the dependent variable is per capita income in levels than when the dependent variable is the growth rate of per capita income. Broadly speaking, this evidence is more supportive of the new-classical growth model than the AK model. But – still – a positive relationship between investment and growth cannot be discarded.

A pitfall in the discussion above is that it relied on the steady state prediction of the Solow model, that investment rates determine the level of per capita income. without taking into account the transition dynamics: remember that the Solow model predicts a temporary relationship between investment and growth, while the economy is moving from one steady state to the other. Therefore, in order to clearly reject the AK model, one would need evidence showing that permanent changes in investment rates produce *temporary* but not *permanent* effects on per capita GDP growth. Along this avenue, Charles Jones (1995)

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

challenged the validity of the AK type of model⁸⁸. Using time series evidence from 15 OECD countries, the author first observed that, along 1950-1988, investment rates, literacy rates and school enrolment rates have increased significantly in most countries, and this didn't materialize in faster economic growth. This evidence was later questioned by Li (2002), who argued that the Jones' case against the AK model was fragile to sample modifications and also to the adoption of broader measures of capital, which are more in line with the AK model⁸⁹.

In practice, a major difficulty in disentangling whether the "true model" is the Solow model or the AK is that the two models are observationally equivalent for a period of time. Since this transition period can be quite long (in the MRW formulation, for instance, half of the transition dynamics takes as long as 35 years) and because most reliable datasets with comparable data start after 1950, it is not easy to assess using time series data whether changes in the investment rate have long run *level effects* or long run *growth effects*.

5.5.2 Tests on conditional convergence

Another key difference between the neo-classical model and the AK model refers to convergence. The neo-classical model predicts conditional convergence. This hypothesis states that countries converge to the respective steady states, and that the speed of convergence varies in direct proportion to the distance to the steady state. This property of the neoclassical model contrasts to the AK model, where changes in parameters impact once-and-for all on the growth rate of per capita income, without any tendency for per capita income to return to a previous path. Hence, another possible avenue to disentangle what is the "true" model is to test whether conditional convergence holds in reality.

⁸⁸ Jones, C., 1995. "Time series tests of endogenous growth models", Quarterly Journal of Economics, 110, 495-525

⁸⁹ Li, D. , 2002. "Is the AK model still alive? The long run relation between growth and investment re-examined", Canadian Journal of Economics 35(1), 92-114.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

The workhorse of empirical research in tests for conditional convergence are the so-called *cross-country growth regressions*⁹⁰. Basically, the method consists in estimating the growth rate of per capita income as a function of a range of variables that control for differences in steady states, and also for the initial level of per capita GDP, that controls for transition dynamics. Conditional convergence is assessed by the significance of the coefficient on the initial level of per capita income.

An example of cross-country growth regression was already introduced in Box 4.2. In the MRW formulation, differences in steady states are accounted for by differences in investment rates in human and in physical capital. As for parameter A , the authors simply assumed it to vary randomly across economies, reflecting differences in climate, natural resources, institutions and other, without disentangling their differential contributions.

More generally, one may extend the model so as to include a range of possible determinants of A . In general, the equation to be estimated in an extended neoclassical framework will be:

$$\ln y_t - \ln y_0 = a - b \ln y_0 + \psi X + \chi Z + u_t, \quad (5.24)$$

where X is a vector of variables capturing factor accumulation that are present in the MRW model (propensities to invest in physical and human capital, and the population growth rates) and Z is a vector of other variables determining the level of A . The conditional convergence hypothesis is assessed investigating the significance of b . In case this parameter is positive and significant this means that growth rates are proportional to the distance to steady states, which accords to the idea of conditional convergence. In case the assumption $b=0$ is not rejected, then changes in the explanatory variables impact on the growth rate permanently, supporting the endogenous growth model (5.5).

⁹⁰ Cross country growth regressions were popularized by Barro (1991). The pioneers of this approach were however Robison (1971) and Kormendi and Meguire (1985). [Barro, R. J. , 1991. "Economic growth in a cross-section of countries", *Quarterly Journal of Economics* 106:2, 407-43. Robinson, S., (1971). Source of growth in Less Developed Countries: a cross-section study, *Quarterly Journal of Economics* 85 (3), 391-408. Kormendi, R., Meguire, P., 1985. Macroeconomic determinants of growth: cross country evidence, *Journal of Monetary Economics*, 16 (2), 141-63].

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

The advantage of cross-country growth regressions relative to simple growth accounting is that, rather than estimating A as a residual, they try to identify the policies and other factors that underlie the cross-country differences in A . Variables that have been found significant in empirical work include proxies for the quality of policies and institutions (such as trade openness, the rule law, political risk, inflation, financial depth) and of geographical conditions. The discussion in Box 5.3 illustrates this.

In general, the evidence with cross-country growth regressions using large samples of countries has been favourable to the conditional convergence hypothesis (an illustration in table 5.2). That is, the coefficient on the initial level of per capita income has been found to be, in general, negative and significant in cross-country growth regressions. This fact inspired Robert Barro to state: *“It is surely an irony that one of the lasting contributions of endogenous growth theory is that it stimulated empirical work that demonstrated the explanatory power of the neoclassical growth model”* (Barro (1997), p. x).

Box 5.3. The external aid controversy

The question as to whether external aid helps or not a country achieve faster economic growth is obviously very important from the policy point of view. With no surprise, this question has been subject to empirical scrutiny.

A branch in the literature has investigated the possibility of the impact of aid being conditional on the recipient country's characteristics. A particularly influential study was a background paper to the 1998 World Bank *Assessing Aid* report, by Burnside and Dollar⁹¹. The authors run some regressions trying to explain the growth rates of per capita income along the period from 1970 to 1993, using a sample of 56 developing countries. The original results of Burnside and Dollar are reproduced in columns (1) and (2) of Table 5.2. In equation (1), the growth rate of per capita GDP is correlated with: the logarithm of initial per capita

⁹¹ Burnside, C. and Dollar, D., 2000. “Aid, policies and growth”, American Economic Review 90 (4), 847-868.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

GDP (capturing conditional convergence); the degree of ethnic fractionalisation, the rate of political assassinations and the product of these two variables (to capture political instability); an index of institutional quality; the ratio of money to GDP (to capture financial development); two regional dummies, for sub-Saharan Africa and East Asia; a “policy index” (compounding the government budget surplus, inflation and openness to international trade, to capture the quality of domestic policies); and external aid as a percentage of GDP.

In column (1), we see that the t-ratio on AID/GDP is too low (0.28 in column 1). The authors then concluded that aid, by itself, does not explain growth. Column (2) differs from column (1) by adding an interaction term, given by the product of the variables AID/GDP and the Policy Index. Because this last variable was found to be significant while AID/GDP alone was not, the authors concluded that aid only leads to more growth in a sound policy framework⁹².

Table 5.2. Growth regressions explaining the growth rate of per capita GDP in 56 developing countries

⁹² The authors also tested the possibility of aid to be detrimental to policies. However, no significant relationship was found between the amount of aid received and the quality of the domestic policies.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

	(1)	(2)	(3)
Estimation method	OLS	OLS	OLS
Initial GDP	-0.61 (1.09)	-0.60 (1.05)	-0.54 (0.96)
Ethnic fractionalization	-0.54 (0.75)	-0.42 (0.58)	0.12 (0.16)
Assassinations	-0.44* (1.69)	-0.45* (1.73)	-0.38 (1.55)
Ethnic fractionalization * Assassinations	0.82* (1.86)	0.79* (1.80)	0.70 (1.63)
Institutional quality	0.64** (3.76)	0.69** (4.06)	0.69** (4.02)
M2/GDP (lagged)	0.014 (1.08)	0.012 (0.86)	-0.02 (1.54)
Sub-Saharan Africa	-1.60** (2.19)	-1.87** (2.49)	-1.58** (2.04)
East Asia	0.91* (1.69)	1.31** (2.26)	1.57** (2.63)
Burnside-Dollar policy index	1.00** (7.14)	0.71** (3.74)	0.78** (4.05)
Aid/GDP	0.034 (0.28)	-0.021 (0.13)	1.49** (3.92)
(Aid/GDP) * policy index		0.19** (2.71)	0.09 (1.34)
Fraction of land in tropics			-0.70 (1.32)
(Aid/GDP) * fract. of land in tropics			-1.52** (4.02)
Observations	275	270	270
Countries	56	56	56
R ²	0.36	0.36	

Sources: Burnside and Dollar (2002) for columns (1) and (2) (regressions (3) and (4) in the original paper; Dalggaard et al., (2004) for column (3) (column 5 in the original)

Notes: The dependent variable is real per capita GDP growth. All regressions include time dummies. Robust t-statistics in parentheses. significant at 10%; ** significant at 5%.

These results caused a significant reaction in the economic profession, as it implied that foreign aid is useless in countries pursuing bad policies. Not surprisingly, they were subject to an intense scrutiny by other researchers. In general, this further investigation revealed sensitivity of the Burnside and Dollar results in respect to alternative specifications

of the regression model or of the sample period⁹³. Burnside and Dollar then shifted their focus from the quality of policies to the quality of institutions. Using a cross section of 124 countries over the 1990s, they found that, while aid alone is not significant related to growth, aid interacted with the degree of institutional quality is significant⁹⁴.

This was not however a final word. In column 3 of table 5.1 we reproduce the results of another study, by Daalgard et al. (2004)⁹⁵. The authors stressed the role of geography in explaining growth and accordingly they included the fraction of a country's land located in the tropics as explanatory variable. In column (3), we see that the policy-aid interaction becomes insignificant, while aid and aid interacted with the climate became significant. These results suggest that aid has a positive impact on growth, but the impact decreases for countries located in the tropics. This last result is, of course, disappointing because it points to a critical role of geography - which cannot be changed by human actions - rather than of policy, which people can change.

In the last few years, many other studies have investigated the extent to which the impact of aid on growth is conditional on third variables. The main conclusion is that there is no definitive answer regarding the variable that better interacts with aid: some studies suggests it is policy, others point to the critical role of institutions, and some others to geography. This disparity of results suggests that the inter-play between aid, local circumstances and growth is eventually too complex to be captured by a simple regression equation. It also points to the difficulty in establishing empirically the right determinants of economic growth, when the candidate variables are so many, as compared to the limited number of observations (see discussion in Box 5.4).

⁹³ Easterly, W., Levine, R. Roodman, D. 2004. "Aid, Policies, and Growth: Comment," *American Economic Review*, vol. 94(3), pages 774-780, June. Islam, N., 2005. Regime changes, economic policies and the effects of aid on growth. *The Journal of Development Studies* 41 (8), 1467-1492.

⁹⁴ Burnside, C. and Dollar, D., 2004b. "Aid, Policies, and Growth: Revisiting the evidence", World Bank Policy Research Working Paper 3251, March.

⁹⁵ Daalgard, C., Hansen, H., Tarp, F., 2004. On the empirics of foreign aid and growth. *Economic Journal* 114(496): F191-F216.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Box 5.4. Pitfalls of cross-country growth regressions

The Empirical analysis using cross country growth regressions face a number of limitations⁹⁶.

First, because the theory does not provide an unambiguous guide to the choice of elements of Z , there is a lot of uncertainty regarding the right model specification. In practice researchers have proposed more and more variables to complement the baseline specification, each one stressing a causal relationship between a particular variable and growth. This, in turn, brings a familiar econometric problem: because explanatory variables tend to be correlated to each other (countries performing badly in a given indicator also tend to perform badly in other indicators), there is a large scope for multicollinearity: the significance of each variable in the equation is influenced by the particular combination of variables included in the regression. This problem makes very difficult to assess empirically which variable is more correlated to growth and how much (e.g, if inflation rates, exchange rate volatility and political instability go wrong together, how one can disentangle the various contributions to growth?).

Second, there is a problem of endogeneity: although it may appear natural that the parameter estimates (ψ and χ in equation 5.25) contain information of causal effects on economic growth, this is not necessarily true. Some right-hand-side variables may be econometrically endogenous in the sense that they are jointly determined with the rate of economic growth: for instance, the same factors that make a country invest more in physical capital may also have a direct effect on a country growth rate. In that case, the estimated parameter will be biased and will provide little information regarding the direction of causality.

Third, even if all variables on the right-hand side were exogenous, many of them could be “symptoms”, rather than “syndromes”. For instance, consider the measurement of

⁹⁶ For a discussion, see Durlauf, S., Johnson, P., 1995. Multiple regimes and cross-country growth behaviour. *Journal of Applied Econometrics*, 10, 365-384.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

human capital. Shall we choose the secondary school enrolment or the primary school enrolment? Since these tend to be correlated to each other, they render one another insignificant when both are included in the regression equation. So which one should we choose? Moreover, a given symptom may be interpreted as capturing different syndromes. For example, a negative correlation between inflation and growth means bad macroeconomic management or a large tax evasion that forces the government to rely on revenues from money creation?

Fourth, there is a problem of parameter heterogeneity: parameter values estimated with cross section exercises that pool together very different countries may fail to accurately capture any of each. As once stated by Arnold Harberger: “What do Thailand, the Dominican Republic, Zimbabwe, Greece and Bolivia have in common that merits their being put in the same regression analysis?.

Fifth, the lack of a structural model stating how much the parameter A depends on each policy variables makes it difficult to go beyond general statements on observed correlations and to provide a convincing interpretation of the results.

Other problems of cross-country-growth regressions include: the presence of outliers, measurement errors, and model linearity. Despite the extensive econometric improvements that have been adopted to overcome these limitations, the results of cross-country growth regressions have still to be taken with caution.

5.6 Discussion

The AK model stresses the relationship between policies and economic growth. This contrasts with the Solow model, whereby changes in the key parameters only produce level effects. The empirical evidence has not been, however, very favourable to the simpler version of the AK model. In general, country characteristics, such as the saving rate and aggregate efficiency are found to influence the *levels* of per capita income, rather than *growth rates*. This view is supported by an extensive empirical literature favourable to the conditional convergence hypothesis.

Does this mean that we shall abandon the AK model? The answer is no.

First, remember that the important link between efficiency and growth is also present in the neoclassical model: the difference is that in the later the growth effect will be <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

transitory. That is, you may interpret the AK model as a *short-run* version of the neoclassical growth model. With half of the transition period between steady states in the neoclassical model taking as long as 35 years, whatever the true model is, we are doomed to accept that policy actions may influence economic growth for a *considerable* period of time⁹⁷.

Second, the AK model is much easier to solve than the Solow model. Because of this, from the expositional point of view, it is often more convenient to study the impact of particular policies in the context of the AK model than in the context of the Solow model, especially when the math becomes too complex. Of course, in doing so, one shall take into account that any conclusion regarding the impact of the policy on growth would be spelled out in term of level effects, if adapted to the context of the neo-classical model. In some of the upcoming chapters, we will follow this approach.

Last, but not at all the least, the basic AK model illustrates how linearity avoids the basic problem of diminishing returns, generating long-term growth. Linearity is a basic feature of most endogenous growth models focusing on technological change. The Usawa model provides an illustration of this: in that model, linearity in production of technology is what is needed to generate long term growth. The AK model can therefore be interpreted as a toy version of more complex endogenous growth models, whereby knowledge expands through investments in R&D. Knowledge shares with capital the characteristic that it can be built over time by sacrificing some of today's consumptions. Interpreting investment as foregone consumption in a broad sense (that is, including physical assets, human capital and R&D), is the what we need to rescue the AK model as a very valid framework to think the mechanics of economic growth.

⁹⁷ Easterly (2005) calibrated a simple neoclassical growth model with a share of total capital equal to 2/3 (which accords to MRW) and with other reasonable values for the remaining parameters. He found that a tax decrease from 30% to zero raises per capita income by a factor of 2.25 times. The author also showed that immediately after the change in policy, the growth rate of the economy jumps up by almost 8 percentage points relative to its steady state. Only in the very long run (more than 5 decades after), the growth effect wears off and the growth rate returns to its long run level. The author concluded that policies have significant effects in the neoclassical model, too (pp. 1024-1026).

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

5.7 Key ideas of Chapter 5

- The AK model reveals in a simple manner that getting rid of diminishing returns, factor accumulating alone can generate continuous growth of per capita income. In the context of the AK model, changes in the saving rate produce “growth effects” rather than “level effects”.
- Extending the AK model to the case with endogenous savings, the direct effect of aggregate efficiency on growth is reinforced by an indirect effect via a higher return on savings. The implication is that, wherever financial markets are more developed, the impact of policy changes on growth is more dramatic.
- The model with endogenous savings appeals to the distinction between proximate causes of growth, like the savings rate and aggregate efficiency, and fundamental causes of growth, that determine why some countries have higher investment rates and better efficiency than others.
- Extending the model in a variety of ways, we found that one can interpret K in a broad sense, including other reproducible inputs to production and technology.
- In its simpler formulation, the AK models displays no transitional dynamics. There are some hybrid models, however, like the Usawa model and an extended version of the neo-classical model that at the same time have transition dynamics and display unceasing growth.
- The Harrod Domar equation inspired the idea that complementing low domestic savings in poor countries by foreign aid would be a key to generate economic growth. In practice, however, the impact of external aid on the growth varied significantly across countries, depending on the quality of domestic policies, institutions, and geography.
- The empirical evidence of conditional convergence has been more favourable to the neo-classical model than to the simpler versions of the AK model whereby higher saving rates generate faster economic growth.
- The AK model can be seen as a toy version of more complex models of endogenous growth based on technological change.

Problems and Exercises

Key concepts

- *Harrod Domar equation. Level effect versus growth effect. Proximate versus fundamental causes of economic growth. New-classical model of endogenous growth. Cross-country growth regressions.*

Essay questions:

- Referring to the Harrod-Domar equation, compare the AK model and the Solow model in respect to the variables that are exogenous and endogenous. In particular, examine the impact of an increase in the saving rate in light of the two models.
- Comment: “Poor countries, with underdeveloped financial markets, are more likely to tolerate bad policies than rich countries with developed capital markets”.
- Explain why the Usawa model is hybrid. In the context of this model, which policies could influence the rate of economic growth?

Exercises

- 5.1. [AK Simple]** Consider an economy where the production function is given by $Y=AK$. In this economy, the saving rate is s , the population grows at rate n and the capital depreciation rate is δ . (a) Does this production function satisfy the usual neoclassical properties? Why? (b) Describe analytically and graphically the dynamics of per capita income in this economy. Is there any stable equilibrium? (c) Does this model predict convergence of per capita incomes across economies? (d) Describe, comparing with the Solow model, the impact of: (i) a fall in the population growth rate; (ii) An increase in A .
- 5.2. [AK endogenous savings]** Consider an economy, where the production function is given by $Y=0.2K$, the population grows at 2% per year, the capital depreciates at 3% and the saving rate is 25%. (a) Find out the growth rate of per capita income in this economy. (b) What will be the effect of A increasing to 0.25? (c) What would be the effect of a simultaneous increase of A to 0.25 and s to 0.4? (d) Now assume that the saving rate was endogenous, as implied by the following optimal consumption rule: $\gamma_t = r_t - 0.17$. Analyse in this case the implications of an increase in efficiency from 0.2 to 0.25. (e) Comparing the two models, find out the expression that relates the saving rate to efficiency (A). Explain why a change in the efficiency parameter (A) impacts more on growth when savings are endogenous.
- 5.3. [Harrod-Domar]** Consider a closed economy without government, where population is equal to one thousand inhabitants, and constant over time. In this economy, the relevant production function is given by $Y = 0.5K$, capital deteriorates at the rate of $\delta = 0.04$ per year, and consumption is a linear function of income, according to $C = 8000 + 0.84Y$. (a) Suppose the initial capital stock in this economy was exactly $K = 200.000$. In this case, per capital income: (i) will be growing over time; (ii) will be decreasing over time; (iii) will be stagnant. (b) Suppose that, due to some external support, the capital stock in this economy jumped temporarily to $K = 250.000$. In this case, the long run growth rate of per capita income will approach: (i) 0; (ii) 0.008; (iii) 0.04. (c) Suppose this economy started out with a capital stock equal to $K = 250.000$, but the production function was actually given by $Y = \min[0.5K, 150N]$. In this case, the long run growth rate of per capita income will approach: (i) 0; (ii) 0.008; (iii)

0.04.(d) In the conditions set out in (c), the long run value of the capital-labour ratio will be: (i) decreasing over time; (ii) $k=300$; (iii) $k=400$; (iv) increasing over time.

- 5.4. [Rebelo]** Consider an economy where the aggregate production function is given by $Y = K^{0.5} H^{0.5}$, where $H=hN$, N is the number of workers and h measures the quality of labour. We also know that in this country 20% of income is saved and invested in physical capital, the population is constant, and the physical capital stock depreciates at an annual rate of 4%. (a) Assume for the moment that the quality of labour evolves exogenously according to $\dot{h}/h = 0.01$. (a1) Discuss that assumption and its implications for the model; ((a2) Determine the equilibrium values of $k = K/N$ and $y = Y/N$. Describe the equilibrium of the model on a graph and discuss its stability; (a3) How well does this model explain the stylized facts of economic growth? (b) Now assume that $\dot{H} = s_H Y - \delta H = 0.0125Y - 0.04H$, where s_H refers to the investment rate in human capital. (b1) Does this version of the model still verify the neo-classical properties? Why? (b2) Compute the growth rate of per capita income in this economy and explain its dynamics with the help of a graph. Describe the income per capita dynamics of this model. Represent it on a graph. (c) Compare, in the light of both models: (c1) the short run and the long run effects of a rise in the saving rate. (c2) the convergence hypothesis.

- 5.5. [Usawa]** Consider an economy where $Y_t = K_t^{1/2} [(1-\mu)(N_t \lambda_t)]^{1/2}$, the saving rate is 20%, population is constant, and capital depreciates at the rate $\delta = 0.04$. Technology evolves according to $\lambda_t = e^{\gamma t}$, $\gamma = b\mu$, and $b=0.05$. (a) Assume for a moment that $\mu = 0$. (a1) Find out the steady state level of per capita income. (a2) Represent in a graph. (a3) In the steady state, per capita income and the interest rate evolve according to the Kaldor stylized facts? (b) (b1) Explain the equation describing the technological change. Examine the implications of an increase in μ to $\mu = 0.2$. In particular: (b2) Describe the change in equilibrium with the help of a graph. (b3) Compute the new equilibrium values of $\tilde{y}$ and $\tilde{k}$. (b4) Describe the time path of per capita income before and after the change. (b5) In the steady state, per capita income and the interest rate evolve according to the Kaldor stylized facts? (b6) was the change in μ welfare improving? (c) Compare the impacts of an increase in the saving rate in cases (a) and (b). (d) Does this model display conditional convergence?

- 5.6. [Usawa]** Consider an economy where the production function is given by $Y_t = (1-\mu)^{1/2} K_t^{1/2} (N \lambda_t)^{1/2}$, where $\mu = 0.75$ is the fraction of time devoted to production. In this economy, the saving rate is 15%, the population is constant and capital does not depreciate. The productivity of labour accumulates at the rate $\gamma = \dot{\lambda}/\lambda = b\mu$, where $b=0.02$. (a) Explain the equation describing technological progress. (b) Using the equation describing the change in $\tilde{k} = K/L$ (the fundamental dynamic equation), find out the steady state values of $\tilde{y}$ and $\tilde{k}$. (c) Examine the implications of an increase in the saving rate from $s=15\%$ to $s=18\%$. In particular, compute the new equilibrium values of $\tilde{y}$ and $\tilde{k}$. Describe the change in a graph and

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

explain what will happen to the interest rate. (d) Returning to the initial figures, examine the implication of an increase in b to $b=0.04$. In particular, compute the new equilibrium values of $\tilde{y}$ and $\tilde{k}$. Describe the change in a graph and explain what will happen to the interest rate. (e) Compare the effects on the path of per capita income, $y=Y/N$, of the changes described in c) and in (d).

- 5.7.** Consider the following production function and law of motion of per capita consumption: $Y_t = K_t^\beta N_t^\alpha H_t^{1-\beta-\alpha}$, with $\alpha, \beta \leq 1$, and $\gamma = r - \rho$. Assume that the depreciation rate is identical for the two capital types and that population does not grow over time. (a) Suppose that $\alpha=0, \beta<1$. Explain if it is possible to obtain sustained growth in the long-run through factor accumulation. Would conditional convergence hold in that case? What would be the impact of an increase in ρ on the interest rate and on per capita income? (b) Suppose that $\alpha+\beta<1, \alpha, \beta \neq 0$. Discuss the advantages of this parameterization comparing them to the results obtain in (a). Would conditional convergence hold in this case? And sustained growth of per capita income? (c) Suppose that $\alpha+\beta=1, \alpha, \beta \neq 0$. Would conditional convergence hold in that case? And sustained growth of per capita income? What would be the impact of an increase in ρ on the interest rate and on per capita income?

Part II – Technology and its diffusion

6 External economies and learning by doing

“One for all! All for one!” [Alexandre Dumas]

Learning Goals:

- Distinction between internal and external economies of scale
- Acknowledge the different types of external economies related to capital accumulation
- Distinguish the implications of non-decreasing versus increasing returns to scale for economic growth and convergence
- Explain why increasing returns are a source of cumulative causation
- Discuss the role of external economies in shaping comparative advantages and the pattern of international trade

6.1 Introduction

The main limitation of the Solow model is that technological progress is assumed exogenous. As already explained in section 3.1, this is a natural implication of the model's assumptions: since there is perfect competition and technology is assumed to diffuse instantaneously at no cost, no single user will be willing to pay for it and no selfish agent will engage in a deliberate effort to invent a new technology. In the Solow model, all income is distributed to the owners of capital and labour, and no profit is left to reward successful research.

It is however possible to construct a model sticking with perfect competition and perfect technological diffusion where the level of technology is endogenous. The trick is to assume that the level of technology evolves over time as an *unintended* consequence of investment decisions by individual firms. As long as technological improvements arise unintentionally, there will be no need for reward. There is, however, an externality, in that each firm benefits from each other firm's investments. Each firm will consider the level of technology as exogenous, and hence it will not take into account the impact of its investment decisions in the aggregate. As firms invest, they will be contributing to the common technological change. With externalities, it is possible to stick with the assumptions of

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

diminishing returns and perfect competition at the firm level with endogenous technological change. This idea was first developed by Alfred Marshall (1890), to explain the tendency for some industries to concentrate in few areas⁹⁸.

In the growth literature, this avenue was first explored in the early 1960s by Marvin Frankel (1962) and Kenneth Arrow (1962)⁹⁹. The purpose of Frankel was to reconcile the convenient properties of the new-classical production function regarding factor allocation and income distribution, with the hypothesis of constant returns to capital that delivers unceasing growth in the Harrod-Domar model. The purpose of Arrow was to model technological change as an unintended outcome of cumulative working experience, a phenomenon labelled “learning by doing”. These ideas were rediscovered two decades later by Paul Romer (1986) and Robert Lucas Jr. (1988)¹⁰⁰. Romer extended the Arrow model of learning by doing to the case with non-diminishing returns in reproducible factors, obtaining unceasing growth. Lucas emphasized the role of externalities associated to human capital as a source of non-diminishing returns. The contributions of Romer and Lucas marked the first wave of the so-called *new growth theory*.

This chapter shows how externalities associated to capital accumulation may overcome the limitation imposed by diminishing returns, delivering a production function for the economy as a whole with the required properties to generate unceasing growth, without the need to depart from perfect competition. Section 6.2 describes the general model introduced by Marvin Frankel, a particular specification of which delivers an AK production function in the aggregate. Section 6.3 explains why the competitive equilibrium with externalities is no longer efficient, and discusses the possible role of the government in addressing the market failure. Section 6.4 focus on specifications of the model with

⁹⁸ Marshall, A., 1890. *Principles of Economics*, London: Macmillan.

⁹⁹ Frankel, M., (1962). The production function in allocation and growth: a synthesis. *The American Economic Review* LII(5), 995-1022. Arrow, K., 1962. “The economic implications of Learning by Doing”, *Review of Economic Studies* 29: 155-173

¹⁰⁰ Romer, P. 1986. “Increasing returns and long run growth”. *Journal of Political Economy* 94, 1002-37. Lucas, R., 1988. “On the mechanics of economic development”. *Journal of Monetary Economics* 22, 3-42.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

externalities that deliver increasing returns in the aggregate, allowing the model to display both unceasing growth and agglomeration effects. Section 6.5 discusses the implications of external economies of scale for comparative advantages and for the benefits of international trade. Section 6.6 concludes.

6.2 Externalities on capital accumulation

6.2.1 A “development modifier”

In his 1962 paper, Frankel first observed that the neo-classical production function used in the Solow model is capable of describing factor allocation and income distribution but is not capable of generating sustained growth of per capita income. In turn, the linear production function used in Harrod-Domar model, is capable of generating long-run growth, but it does not offer a satisfactory theory for factor allocation and income distribution. Frankel then proposed a method to conciliate the two production functions, so that the desirable properties of each but none of the limitations are retained: the key for such conciliation was to assume a production externality whereby the “overall level of development of a region” impacts positively on the productivity of each private firm.

The author observed that firms in developed economies are able to produce more with given inputs of capital and labour than firms in underdeveloped regions. Frankel related this to “various external effects”, such as “improvements in the organization and the quality of labour, technical change, external economies of scale, and better social overhead facilities in the form of transport and communication networks” (p 1001). These external effects are *bounded in space*, in the sense that they are specific to a given economy, acting therefore as a local public goods.

To model these externalities, Frankel assumed that each individual firm faces a Cobb-Douglas production function, where TFP is a positive function of the economy-wide capital stock. Formally, let the production function for each individual firm i be given by:

$$Y_i = BK_i^\beta N_i^{1-\beta} \quad , \quad (6.1)$$

where Y_i , K_i and N_i denote, respectively, for output, capital and labour employed by firm i . The technological parameter, B - the “development modifier”, as coined by Frankel - was

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

assumed to depend positively on the aggregate levels of capital and labour. Its more general specification was as follows:

$$B = A \frac{K^\varepsilon}{N^{\varepsilon'}}, \text{ with } \varepsilon \geq 1 - \beta \text{ and } 0 \leq \varepsilon' \leq \varepsilon, \quad (6.2)$$

where $K = \sum_i K_i$ and $N = \sum_i N_i$ stand, respectively for the *aggregate* levels of capital (human, physical) and labour in the economy.

According to (6.2), an increase in the aggregate stock of capital impacts positively on the productivity of each firm. Thus, whenever a firm accumulates capital for private reasons, it will be “indirectly” contributing to the productivity of all other firms. The productivity term (6.2) also accounts for a negative externality on aggregate labour, in case $\varepsilon' > 0$. This effect captures the possibility of the positive externality related to capital accumulation being partially or totally *diluted* by the size of the labour force. When, for instance, $\varepsilon' = \varepsilon$, the firm productivity depends on the aggregate stock of capital per worker, rather than with the aggregate stock of capital in absolute terms. When instead $\varepsilon' = 0$, what matters is the absolute level of capital in the aggregate, not the capital labour ratio. In the following, we’ll consider alternative cases regarding the relative weights of these external effects.

Production externalities specified in this manner are labelled “Marshallian” or “Technological” externalities”. Comparing to the Solow model, the model with externalities retains the assumption that technology spills over instantaneously across firms at zero cost. Hence, like in the Solow model, there are no private incentives to improve the level of technology. But the level technology is now endogenous: even though there are no purposeful efforts to develop new technologies, the level of technology improves over time as a by-product of capital accumulation, which is driven by economic decisions.

6.2.2 Factor prices in the competitive equilibrium

Because each firm is small relative to the economy, its decisions will have a negligible impact on the aggregate. Hence, each firm will take parameter B as exogenous, and independent of its investment decisions. Each individual firm will perceive its own production function in the form (6.1), with the standard neoclassical properties of constant returns to scale and diminishing returns to capital.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Profit maximization by individual price taking firms will therefore deliver the usual conditions stating that firms employ labour and capital until their marginal products equal the respective factor prices:

$$r + \delta = \frac{\partial Y_i}{\partial K_i} = \beta \frac{Y_i}{K_i}, \quad (6.3)$$

and

$$w_i = \frac{\partial Y_i}{\partial N_i} = (1 - \beta) \frac{Y_i}{N_i}. \quad (6.4)$$

Because all firms are equal, we have $Y_i/K_i = Y/K$ and $Y_i/N_i = Y/N$.

Thus, in the competitive equilibrium, the shares of capital and of labour on domestic income are given, respectively, by β and $1-\beta$. This is the very convenient result, as it implies that factor shares remain in accordance to the Kaldor stylized fact, regardless the presence of externalities.

6.2.3 The aggregate production function

In the presence of externalities, the aggregate production function differs from the individual production functions, even if all firms are alike. The aggregate production function is obtained substituting (6.2) in (6.1) and summing up across all firms. This gives:

$$Y = AK^{\varepsilon+\beta}N^{1-\beta-\varepsilon'} \quad , \quad (6.5)$$

with $Y = \sum_i Y_i$. The *aggregate production function* (6.5) may deviate from the neoclassical assumptions of constant returns to scale and diminishing marginal returns. For instance, when $\varepsilon + \beta \geq 1$, returns to capital are non-decreasing. As we already know, this is the condition we need for a model to display *unceasing growth* through capital accumulation. On the other hand, whenever $\varepsilon > \varepsilon'$, the aggregate production function will exhibit *increasing returns to scale*: that is, rising capital and labour by a given proportion causes output to increase more than proportionally. As we will discuss next, this scale effect makes larger economies more attractive for location, acting therefore as a mechanism of economic divergence.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Note that at the individual level, production functions retain the neoclassical properties of constant returns and diminishing returns to capital. The aggregate production function departs from these properties because of the externality, which individual firms – being small - do not take into account.

6.2.4 The AK model again

Equation (6.5) is general enough to account for external effects of different magnitudes. Frankel was however concerned with the particular case where externalities are such that the aggregate production function becomes exactly linear in the economy' capital stock, to mimic the Harrod-Domar model. Therefore, the author focused on the case with $\varepsilon = \varepsilon'$ and $\varepsilon + \beta = 1$ ¹⁰¹. In this case, the positive externality in K is exactly enough to overwhelm the normal process of diminishing returns to capital, and – at the same time - the negative externality on labour exactly matches the externality on capital, implying that returns to scale remain constant. The aggregate production function (6.5) becomes exactly linear in K:

$$Y_t = AK_t . \quad (6.5a)$$

The aggregate production function (6.5a) takes the AK form, but firm level production functions (6.1) retains the neoclassical properties. Each firm perceives its production function as with diminishing returns to capital, so it will employ capital and labour according to (6.3) and (6.4). In the aggregate, the production function is linear in K, so the marginal product of capital will never decline and the economy will never stop growing. In this model, policies influencing the rate of capital accumulation will indirectly influence the level of technology, and by then the rate of economic growth.

¹⁰¹ In case $\varepsilon = \varepsilon'$ and $\beta + \varepsilon < 1$, the aggregate production function exhibits constant returns to scale and diminishing returns to capital. As you may easily check, in that case the steady state growth rate of output is equal to the growth rate of population, just like in the basic Solow model. The only difference is that, because of the externality, private returns to capital in laissez faire will be too low.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

The advantage of this model when compared to the simple AK model, is that it does not rely on the peculiar assumption that labour plays no role in production. Like in the Solow model, both factors are used in production and factor income shares are in accordance to reality. Note that the model also accords to the other Kaldor stylized facts: the wage rate and per capita income grow steadily over time, the user cost of capital is constant and equal to $r + \delta = \beta A$ (equation 6.3), and growth rates may be different in different economies.

6.3 The market failure and optimal intervention

6.3.1 Growth accounting revisited

Conventional growth accounting (as exemplified in Section 2.7) typically uses the share of capital on national income as the proxy for the contribution of capital to production. The discussion in Chapter 4 revealed, however, that this procedure delivers an estimate for the contribution of capital (β), that is too small to account for the observed differences in per capita incomes across countries. In order to account for such large differences, one would need a contribution of capital to production much larger than that implied by the observed income shares.

The Frankel model offers an explanation for this puzzle: a larger contribution of capital to production than that implied by the observed shares in national incomes might be accounted for by externalities. Formally, equation (6.5) reveals that, as long as the externality parameter ε is positive, the *actual* contribution of capital to production ($\beta + \varepsilon$) is larger than that implied by its share in income (β). Log-differentiating (6.5), one obtains:

$$\hat{Y} = (\varepsilon + \beta)\hat{K} + (1 - \beta - \varepsilon)n \quad (6.6)$$

In (6.6), input changes have two effects, a private one and an external effect. The external effect may amplify or diminish the private effect, depending on the sign and magnitude of the respective parameter. For instance, when $\varepsilon = 1 - \beta$, a one-percentage point increase in the capital stock will result in a one-percentage point increase in output, a result that conforms with the AK model (and that Frankel argued to conform as well to the U.S. data).

Equation (6.6) suggests that conventional growth accounting, by underestimating the effective contribution of capital to production, overestimates the Solow residual.

6.3.2 The social return of capital

With externalities, the competitive equilibrium is not efficient. Each firm, being small relative to the economy, decides its capital stock taking into account its own profits, only. The positive contribution of its investment decisions to the overall capital stock are considered negligible and hence ignored. Still, investments made by all firms together impact positively on the profit of each individual firm. Thus, the competitive equilibrium delivers a suboptimal level of investment.

Formally, the marginal contribution of aggregate capital to aggregate production (i.e, taking into account the externalities) as stated in (6.5) is:

$$\left(\frac{\partial Y}{\partial K} \right)_{social} = (\beta + \varepsilon) \frac{Y}{K} \quad (6.7)$$

In its profit maximization problem, the firm considers only the narrow private returns to capital (equation 6.3). As long as $\varepsilon \neq 0$, there will be a wedge between private returns and social returns. The wedge between private returns and social returns to capital implies that *incentives are misaligned*: in the decentralized economy, investment will be too low.

6.3.3 Optimal intervention

The wedge between social returns and private returns to capital constitutes a market distortion: investment in physical capital is too low relative to the efficient allocation. Given this sort of diagnosis, a benevolent planner may find appropriate to subsidize capital accumulation.

To find out the optimal subsidy, let's rewrite the individual firm profit function, but allowing for subsidy $\tau_K < 0$ on capital incomes:

$$\pi_i = BK_i^\beta N_i^{1-\beta} - (r + \delta)(1 + \tau_K)K_i - wN_i \quad (6.8)$$

In light of this specification, the cost of one unit of capital –the cost to firms - is $(r + \delta)(1 + \tau_K)$, while capital owners receive as net income $(r + \delta) > (r + \delta)(1 + \tau_K)$. Regarding

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

how this subsidy will be financed, assume that lump sum taxation is available, so that the policy will not imply further distortions.

From the first order conditions of profit maximization, one obtains (instead of 6.3):

$$\frac{\partial Y_i}{\partial K_i} = \beta \frac{Y}{K} = (r + \delta)(1 + \tau_K) \quad (6.9)$$

To get the incentives right, the government must set the subsidy such that the (net) rental price of capital, $(r + \delta)$, fully reflects the social return of capital (6.7). That is, τ_K should be set such that:

$$r + \delta = \beta \frac{Y/K}{(1 + \tau_K)} = (\beta + \varepsilon) \frac{Y}{K} \quad (6.10)$$

Solving for τ_K , the optimal (first best) tax rate will be:

$$\tau_K = \frac{-\varepsilon}{\varepsilon + \beta} \leq 0 \quad (6.11)$$

This result is intuitive: if the contribution of capital to production is given by (6.7) and private firms only perceive it to be equal to (6.3), then a subsidy filling the gap will achieve the aim of getting private returns aligned with the social interest.

In the particular case in which $\varepsilon = 1 - \beta$ (the AK model), the optimal subsidy will be such that $\tau_K = \beta - 1$ (note however, that in this extreme case *all* income in the economy would be devoted to capital owners and nothing would be left to raw labour; this would be only possible if K referred to a broad concept of capital, including human capital).

6.3.4 Growth effects

In this model, removing the distortion leads to a greater efficiency and, by then, to faster economic growth. To see this, consider again the optimal consumption rule $\gamma = r - \rho$ and let's focus in the particular case in which $\varepsilon = \varepsilon' = 1 - \beta$ (the AK model).

In the competitive equilibrium, the interest rate is determined according to (6.3). Substituting r in the optimal consumption rule, one obtains the growth rate of per capita income under laissez faire:

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

$$\gamma = \beta \frac{Y}{K} - \rho - \delta = \beta A - \rho - \delta \quad (6.12)$$

With intervention, the user cost of capital would become equal to (6.7) and the growth rate of the economy will be:

$$\gamma^* = (\beta + \varepsilon) \frac{Y}{K} - \rho - \delta = A - \rho - \delta \quad (6.13)$$

Comparing, we see that the growth rate of this economy with an appropriate intervention will be higher than in the laissez faire.

This example illustrates how judicious government intervention might be used to establish the “right” prices and thereby stimulate growth. Note however that such a “perfect” intervention requires a high level of confidence by the government regarding the size of the external effect, as well as availability of non-distortionary taxation. Whenever these conditions are not met, it may well be the case that the government may fail to do better than the market.

6.4 The case with increasing returns

6.4.1 External economies of scale

So far, the analysis focused on the case with $\varepsilon = \varepsilon'$. This is however a very special case. It requires the positive effect arising from a larger stock of physical capital to be exactly offset by the “dilution” effect resulting from a larger number of workers. In this version of the model the aggregate production function exhibits constant returns to scale, even though returns to capital are non-decreasing.

A quite distinct case occurs when $\varepsilon > \varepsilon'$. In that case, expanding the use of capital and labour by a given proportion has a more than proportional impact on output: the aggregate production function exhibits *increasing returns to scale*. Remember that increasing returns do not arise at the individual firm level, but instead at the aggregate level. Because of this, increasing returns are said to be *external* to the firm (Box 6.1).

When the aggregate production function displays increasing returns, there will be a tendency for the region to become larger and larger. To see this, just note that the average product of labour in (6.5) becomes equal to:

$$y = Y/N = Ak^{\varepsilon+\beta} N^{\varepsilon-\varepsilon'} \quad (6.14)$$

This means that that, in a competitive equilibrium, the wage rate – determined according to (6.4) - will also be an increasing function of the size of the workforce.

$$w = (1 - \beta) Ak^{\varepsilon+\beta} N^{\varepsilon-\varepsilon'} \quad (6.4a)$$

Thus, a larger region will be a more attractive place to work than a smaller region. This will generate a tendency for employment to move to the larger region, further expanding the larger region and depressing the smaller region.

This illustrates why increasing returns are a source of divergence: if for whatever reason, a region starts out bigger, increasing returns will assure that it will become a more attractive place to work and invest. With free factor movements, labour will move from depressed areas to the more dynamic region and the later will get bigger and richer, absorbing resources from the rest of the world.

The idea that development brings more development in a virtuous cycle was coined “cumulative causation” by Thorstein Veblen, in 1898. This concept was popularised by Gunnar Myrdal in the 1950s¹⁰². This author contended that labour migration, capital movements and trade may lead to cumulative expansion of the favoured regions and retard the development of backward regions, leading to persistent or even divergent cross-country differences in per capita income.

¹⁰²Veblen, T., 1898, Why is economics not an evolutionary science? *Quarterly Journal of Economics*, 12, 373-97. Myrdal, G., 1957. *Economic theory and underdeveloped regions*, Duckworth, London.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Box 6.1. Internal and external economies of scale

The distinction between “internal” and “external” economies of scale dates back from Scitovsky (1954)¹⁰³. “Internal” economies of scale refer to the case in which a single firm faces a downward sloping average cost curve when increasing *its own* level of output. In this case, there is a tendency for the larger firm to outprice its competitors, becoming larger and larger, until becoming monopolist in the market. Internal economies of scale are inherently linked to imperfect competition.

The concept of “external” economies of scale refers to the case in which scale economies arise at the aggregate (spatial or industry) level. In that case, average costs for the individual firm decline with aggregate output, but not with the individual firm output. “External economies of scale” in the aggregate may co-exist with constant returns to scale and declining marginal productivities at the firm level. Hence, one does not need to abandon the assumption of perfect competition.

6.4.2 Alfred Marshall and the theory of economic geography

The theory of external economies of scale was pioneered by one of the founders of modern economics, the British economist Alfred Marshall. In his book “Principles of Economics” (1920, first published in 1890), Marshall was concerned with the question as to why there is a tendency for some industries to concentrate in few areas within a country (“industrial districts”). Examples at that time included cutlery manufactures in Sheffield, and hosiery firms in Northampton. In our days, similar examples include the Silicon Valley, Hollywood and Las Vegas. This type of spatial concentration of economic activities cannot be explained by proximity to natural resources.

¹⁰³ Scitovsky, T., 1954. Two concepts of external economies. *Journal of political economy*, 62, 143-151.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

To explain the tendency for firms of the same industry to cluster together, Marshal conjectured that the productivity of each firm in a given location may depend positively on the general progress of the corresponding industry *in the same location*, via three types of external effects: availability of specialized suppliers, labour market pooling, and knowledge spillovers.

First, the availability of specialized suppliers: in many industries, production requires the use of specialized inputs, such as intermediate products and specific supporting activities, that cannot be acquired at distance because of high transportation costs. For instance, the production of a motion picture requires a variety of services, such as casting services, sound effects, costume design, choreography, catering, etc. Many of these services are better purchased to specialized firms, because specialized firms can split the fixed costs of investing in technology through different costumers. If, in a given region, there is only one film producer, it will not pay for upstream suppliers to locate in that region. Instead, they will prefer locations where there are already many moviemakers, ensuring that the market is large enough to break even. By the same token, moviemakers will find it more profitable to join locations where other moviemakers are already located, because this will imply a higher market for – and hence a higher availability of - specialized services, competing with each other.

Second, labour market pooling: when many firms and specialized workers locate in the same region, both sides of the market will be less exposed to events affecting a small number of firms or workers. For instance, the failure of one firm will be less problematic for a specialized worker located in a region with many firms than if located in a region with one employer only. The same holds for firms. By clustering together, both firms and workers will benefit from the law of large numbers, being therefore less exposed to specific shocks affecting individual agents.

Third, technological spillovers: workers engaged in a given production process have incentive to observe what similar others are doing, so as to imitate the best practices. Arguably this process of learning takes place more effectively when various firms of the same industry are concentrated in a given location, so that workers belonging to different firms have the opportunity to meet together and discuss technical problems, face-to-face. Personal contacts are essential for the diffusion of Tacit Knowledge (see box 6.2). Similarly,

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

the knowledge embodied in skills may diffuse easily through workers mobility across neighbourhood firms.

All in all, these three types of external effects (often called “Marshallian externalities”) act as a *local* public good, implying that each firm becomes more productive, the more firms of the same industry are located nearby. The local nature of these externalities is essential to explain why economic activities tend to cluster together.

Formally, this theory is modelled assuming that the technological parameter of an individual firm’ production function depends positively on an variable measuring the size of the industry in that location (as done in equation 6.2). In that case, the aggregate production function may display increasing returns to scale, creating the incentive for firms to cluster together in particular regions.

The Marshall model of technological externalities launched the theory of Economic Geography. This theory is concerned with the question of how economic activities are located across the space. The theory had however to be refined to account for the fact that there is no tendency for all activities to be concentrated in a single location. That would be wholly unrealistic. In fact, location decisions are also influenced by *centrifugal* (dispersion) forces, such as congestion effects, whereby the cost of a firm adopting a given location increases with the number of firms that are already in that location. For instance, the concentration of activities in a small area leads to higher land prices, high commuting costs, pollution and other sociological factors. Another dispersion force arises due to transport costs: to the extent that some activities have to be undertaken in the periphery (for instance, agriculture, exploitation of natural resources), being close to the centre implies higher transport costs in transacting with the periphery. The location of economic activities must therefore obey to a balance between centripetal forces and centrifugal forces. Since the weights of these two

forces differ from industry to industry, there is scope for diversity in the spatial organisation of economic activities¹⁰⁴.

Box 6.2. Tacit knowledge

Knowledge is not all alike. Some knowledge is suitable for codification, for instance in manuals, or in textbooks. When this is so, it has the potential to be transmitted at distance. Not all knowledge, however, is suitable for codification. In many technologies, only the broad guidelines are codified. The remainder pieces are embodied in the skills of practitioners. This component of knowledge is dubbed “tacit knowledge”.

The main feature of tacit knowledge is that it can only be transmitted through face-to-face contacts. This concept was first proposed by a Doctor of Medicine and of Physical Science, Michael Polanyi. In his words, “Tacit knowledge can be passed only by example from master to apprentice”(p.53)¹⁰⁵. Since the diffusion of tacit knowledge is mediated through personal contacts, it does not propagate easily across the space.

A natural mechanism through which tacit knowledge may diffuse across the space is via labour mobility. Tacit knowledge is embodied with people and can migrate with people. Similarly, tacit knowledge can be transmitted at distance, via training actions. Multinational firms typically spend considerable amounts of resources in organizing meetings, workshops, demonstrations and seminars for their employees, to overcome the problem of communicating knowledge at distance.

¹⁰⁴ Classical contributions accounting for these centrifugal forces in the context of “Marshallian externalities” include Henderson (1974) and Fujita and Ogawa (1982). The theory was later refined with the works of Paul Krugman and Anthony Venables [Henderson, 1974. “The sizes and types of cities”. *American Economic Review* 64, 640-656. Fujita, M., Ogawa, H., 1982. “Multiple equilibrium and structural transition on non-monocentric urban configurations”. *Regional Science and Urban Economics* 12, 161-196. Krugman, P., 1991. “Increasing returns and economic geography”, *Journal of Political Economy* 99, 483-499. Krugman, P. and Venables, A., 1995. “Globalization and the Inequality of Nations”. *Quarterly Journal of Economics* 60, 857-880].

¹⁰⁵ Polanyi, M., 1958. *Personal knowledge: towards a post-critical philosophy*. Chicago: U. Chicago Press.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

6.4.3 The Arrow model

Kenneth Arrow proposed a model of endogenous technological change, whereby investment by individual firms increase the firm's stock of knowledge through learning by doing (see Box 6.3). Arrow modelled learning-by-doing at the individual firm level, assuming that investment in physical capital impacts in the firm's stock of knowledge¹⁰⁶. This assumption was then combined with the assumption that knowledge leaks. Arguably, firms tend to imitate the improvements achieved by fellow firms, so they all end up benefiting from the accumulated experience of each other. Thus, when one firm invests in new capital, it adds to its own stock of knowledge and at the same time to the common stock of knowledge. In the Arrow's model, the externality on capital accumulation arises because the learning acquired through investment by each one firm leaks out to other firms, giving rise to increasing returns to scale in the aggregate.

Like in the Frankel model, total factor productivity at the firm level is an increasing function of the economy-wide accumulated stock of capital. The main difference in is that in the Arrow model there is no negative externality associated to the number of workers. In terms of equation (6.2), Arrow restricted attention to the case with $\varepsilon' = 0$. This assumption captures the non-rival nature of knowledge: many workers and firms can use the same piece of knowledge without reducing its effectiveness. The implication is that the learning by doing model unequivocally displays increasing returns. With all firms identical, the aggregate production function becomes:

$$Y = AK^{\varepsilon+\beta} N^{1-\beta} \quad (6.5b)$$

Thus, as long as $\varepsilon > 0$, the production function will exhibit increasing returns to scale on capital and labour altogether, giving rise to scale effects and agglomeration economies.

¹⁰⁶ Arrow (1962), p. 157: "each new machine produced and put into use is capable of changing the environment in which production takes place, so that learning takes place with continuous new stimuli".

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

In this model, the assumption of knowledge spillovers is critical for the model to be consistent with perfect competition: if the knowledge created did not leak out, the individual firm accumulating capital would become more productive than its competitors; its returns would be higher and higher and the conditions would exist for this firm to grow alone and capture the entire market, becoming monopolist.

With the assumption of perfect technological diffusion, the model follows in an intuitive manner: each firm, perceiving its production function as a CRS, buys new capital until the *private* marginal product of capital equals the user cost of capital (eq. 6.3). Buying the state-of-the-art capital, the firm inadvertently increases its own stock of knowledge, but this effect is small. Since knowledge leaks, the acquisition of physical capital by each one firm adds to the common stock of knowledge, which impacts positively on the productivity of all firms. Thus, each firm will be more productive, the higher the productive experience (measured by the stock of capital) in the economy as a whole (eq. 6.2). Box 6.4 presents a well-known case-study where learning by doing and knowledge spillovers triggered a process of unceasing growth.

Note that increasing returns on capital and labour altogether is not a sufficient condition for the model to display *endogenous* growth. For this, one would need to assume as well that returns to capital are non-diminishing, $\beta + \varepsilon \geq 1$. The problem with that case is that the model would display a problematic (strong) scale effect, whereby the growth rate of per capita income would be a positive function of the size of population: that is, a larger economy should grow faster than a smaller economy (see Box 6.5). Arrow ruled out that possibility, sticking to the case with $\beta + \varepsilon < 1$.

In the Arrow model, there are increasing returns to capital and labour altogether (and hence agglomeration effects), but diminishing returns to capital alone, ensuring that the capital-output ratio converges to a constant in the steady state, just like in the Solow model. To see this, let's log-differentiate (6.5b), obtaining:

$$\hat{Y} = [\beta + \varepsilon] \hat{K} + (1 - \beta)n \quad (6.15)$$

In the steady state, capital and output grow at the same rate. Imposing this restriction in the equation above, one obtains $\hat{Y}(1 - \beta - \varepsilon) = (1 - \beta)n$. Now using $\gamma = \hat{Y} - n$, we obtain the growth rate of per capita income in the Arrow model:

$$\gamma = \left[\frac{\varepsilon}{1 - \beta - \varepsilon} \right] n. \quad (6.16)$$

Equation (6.16) shows that, as long as population growth is positive, per capita income in this economy will grow over time, without the need to assume an exogenous rate of technological progress. Since the growth rate of per capita income in the steady state is determined by the growth rate of population, which is an exogenous parameter, this model displays *exogenous growth*: changes in policy influencing the saving rate or the efficiency parameter A will alter the steady state level of per capita income, but not its growth rate. Only *level effects* will be achieved.

The model displays a kind of scale effect, in that the long run growth of per capita income depends on how fast population is growing. This is a direct implication of the non-rival nature of knowledge: since sharing knowledge does not involve loss of its effectiveness, the larger the population being served with each given piece of knowledge, the better. Since this “scale effect” is of a second order, it is categorized as “weak”.

Box 6.3. Learning by doing

A possible link between investment in physical capital and total factor productivity is that new investments force users to adapt and learn. When a firm buys a new capital good, it is also acquiring a new production technique. Learning how to operate with the new equipment and organizing the production process so as to better benefit from the opportunity opened up constitutes a technological improvement. The full materialization of this learning process may take time. Workers benefit from experience. With the passage of time, workers get more accustomed to the new capital good. By undertaking similar actions repeatedly, workers perfect their routines and learn how to solve minor problems, becoming more productive as time goes by. This benefit occurs through practice - hence the label “learning by doing”.

The learning by doing effect is often summarized by a “learning curve”, that relates the average cost of producing a given good with and the cumulative installed capacity to produce that good. The fall in production costs associated with each doubling of experience is labelled the “learning rate”.

The “learning curve” was first proposed by a German psychologist Herman Ebbinghaus (1850-1909), in a series of tests on humans consisting in memorizing nonsense syllables. In engineering, the relationship between experience and costs of production was first documented in aeronautics by Theodore Wright (1936)¹⁰⁷. In our days, learning is playing a key role in the transition towards a low carbon economy. Today, fossil fuels account for more than 80% of the world carbon emissions, threatening the global climate and the biosphere. The good news is that renewable energy technologies are still young and hence they have much to benefit with a growing cumulative experience. According to some estimates, the *learning rate* of solar electricity has been around 36%: this means that each doubling of the installed solar capacity, the price of solar electricity declined by 36%¹⁰⁸. By contrast, production of electricity with fossil fuel sources have not benefit from further experience. The implication is that the relative price difference between renewables and fossil fuels is decreasing over time, favouring the transition towards a cleaner world.

Box 6.4. Desh Garment Ltd

In 1980, a major world textile producer from South Korea, Daewoo Corporation, launched a joint venture with a local partner in Bangladesh, the Desh Garment Ltd, to produce garments. At that time, Bangladesh had no productive experience in garment production. And yet this investment triggered a process of learning by doing and technological diffusion that transformed the economy of Bangladesh into a major exporter of textiles.

At the outset, Desh Garment sent 130 Bangladeshi workers to Korea, for training. This training allowed young workers to acquire tacit knowledge, becoming familiar with the process of producing garments. This knowledge was not exactly a rocket science, but simply

¹⁰⁷ Ebbinghaus, H., 1885. Memory: A contribution to experimental psychology. Wright, T., 1936. “Factors affecting the cost of airplanes”. Journal of the Aeronautical Science 3, 122-128.

¹⁰⁸ Roser, Max, 2020. Why did renewable energies become so cheap so fast? And what can we do to use this global opportunity for green growth? Our World in Data, Oxford Martin Programme on Global Development, University of Oxford. December.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

it had not been transmitted before, because there was no garment production in Bangladesh. After operations started, however, the knowledge acquired in practice started leaking, opening up an enormous potential. With no surprise, in the years that followed, 115 workers trained by Daewoo left the company to start their own businesses. The new firms not only produced garments, but also gloves, coats and trousers. Rapidly, an entire exporting sector emerged, through learning by doing and knowledge spillovers¹⁰⁹, and Bangladesh became a world player in textiles and related industries.

Box 6.5. Strong scale effects in the LBD model

In the Arrow model there are diminishing returns to capital, so the ratio Y/K tends to a constant in the steady state. This ensures a constant interest rate and compliance with the Kaldor facts. A different case occurs when $\beta + \varepsilon \geq 1$: when returns to capital are increasing, the marginal product of capital becomes a positive function of the capital-labour ratio. This means that the interest rate will be itself increases over time.

To see this, let's substituting (6.5b) in (6.3) to obtain an expression for the user cost of capital:

$$r + \delta = \beta A k^{\varepsilon + \beta - 1} N^{\varepsilon} \quad (6.17)$$

From (6.12), the endogenous growth rate of per capita income will be:

$$\gamma = \beta A k^{\varepsilon + \beta - 1} N^{\varepsilon} - \rho - \delta \quad (6.18)$$

In case $\beta + \varepsilon = 1$, the growth rate of per capita income is a positive function of the size of the labour force, N , displaying a “strong” scale effect: if the workforce grows at a constant rate, n , the interest rate will be ever-increasing interest rate, and so will do the rate of economic growth. In case $\beta + \varepsilon > 1$, the growth rate of per capita income will be ever-accelerating even if population remains constant. These predictions do not square well with

¹⁰⁹ Rhee, Y., W., 1990. The catalyst model of development: lessons from Bangladesh' success in garment exports. *World Development*, 18, 333-346.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

the real world facts: in general, there is no systematic tendency for large countries to grow faster than smaller countries, nor for growth rates to be increasing over time¹¹⁰.

6.4.4 Lucas: externalities on Human Capital

An alternative way of modelling externalities in capital accumulation is focusing on human capital. This avenue was explored by Robert Lucas Jr. in its 1988' article. The main argument is that people who get educated benefit more in a society where people are educated than in a society with low education levels.

To understand this, ask yourself why the best graduate economists prefer to work at the City of London or at Wall Street – where economics graduates are plentiful – rather than in, say, Mongolia where they are in very short supply. The economist working at City earn his high income in part because of the manner in which its own knowledge is enhanced by those of fellow well-educated economists. This happens because individuals benefit from *interacting* with each other. Exchange of ideas with other professionals enhances individual capabilities. Thus, just like in an assembly line, where the value of each worker's effort depends on the other worker's efforts, complementarities in human capital create the incentive for the best workers to match up with each other: if the best economists are *assembled* together, they will have better ideas and will get a higher payoff from their skills. If, instead, they are partnered with lazy or incompetent economists, they will have a lower reward for any effort that they might individually provide.

Note that this is exactly the opposite of the LDR: with diminishing returns, skills substitute for each other, so they become more valuable where they are scarcer – in Mongolia and not at City. Under diminishing returns, skilled labour would move from rich countries to poor countries. By contrast when externalities in human capital are strong enough to

¹¹⁰ Romer (1986) proposes a model to overcome this problem where the growth rate of knowledge is assumed bounded up, due to diminishing returns in knowledge accumulation. With such assumption, the interest rate becomes bounded up too, delivering a constant growth rate for per capita income.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

overwhelm diminishing returns, then skilled labour will be more valuable where it is more abundant: returns to skills for each individual are an increasing function of the existing skill level in the society. This mechanism helps explain why we observe skilled workers migrating from poor countries to wealthy ones, and not the other way around¹¹¹. It also explains why wages for similar skills and education levels are higher in cities than in rural areas: skilled workers tend to move to where skilled workers are, because staying close to each other makes each one of them more productive.

Like in the case with physical capital, complementarities in human capital imply cumulative causation and vicious cycles: in a nation where skill levels are already deep and well established, people in that nation will find strong incentives to invest in their human skills. But in poorer economies where the skill base is thin, the incentive for individuals to invest in human skills is low. Thus, a country will be rich if it started out rich, a country will be poor if it started out poor.

These vicious and virtuous cycles in human capital accumulation also apply to ethnic groups within societies. If any sort of social segregation delivers people of the same ethnic group a higher probability to match and work with each other than with people from other ethnic groups, then there will be a tendency for education levels to converge within each group: people belonging to the low education ethnic group will invest less in education because working with people with low education implies a low return to education. On the contrary, people belonging to the highly educated ethnic group will have an incentive to invest in education, because the chances of being matched with well-educated people are high.

The same mechanism applies to the other component of human capital, health: an healthy society impacts positively on individual health through lower contagion of diseases.

¹¹¹ Carrington and Detragiache (1998) provide evidence that the stock of immigrant workers in the U.S. is, on average, better educated than the average worker in the home countries. The authors also showed that, for most developing countries, the highest migration rates are observed in the group of individuals with tertiary education [Carrington, W. Detragiache, E. 1998. "How Big Is the Brain Drain?" IMF Working Paper 98/102 (Washington)].

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Thus, an individual's health will be a positive function of the average health in the society. This, in turn, impacts positively on individual productivity, because healthy individuals are likely to be more productive.

6.4.5 Localized versus global technological spillovers

Along this section, we argued that technological externalities favour the spatial concentration of economic activities, acting as a source of economic divergence. This is in sharp contrast with a basic assumption of the Solow model, that knowledge has the potential to diffuse *globally*, acting as a source of (conditional) convergence. A question therefore arises on how to conciliate these two perspectives.

The idea that knowledge travels well across borders linking the growth rates of interdependent economies in the long-run has to be taken seriously. However, it also has to be qualified. A well-documented fact in our days is that technological levels are not uniform across countries. Despite all progresses in telecommunications and the internet, we are far from the neoclassical assumption that knowledge spills over instantaneously at any distance at no cost.

On one hand, the empirical evidence gives supports to the idea that proximity matters for technological diffusion (see Box 6.6). On the other hand – and most importantly – one must recognize that other factors apart from geographical distance influence the pace of technological diffusion. Arguably, the same piece of knowledge will diffuse at different speeds across different countries, depending on the recipient country infrastructure, human capital, culture, as well as economic and political circumstances. Thus, while keeping an eye on the idea that technology has the potential to diffuse across the space, one must take into account that technology diffuses via specific mechanisms of human interaction, and that this diffusion may be retarded by *human-devised barriers*. These barriers give rise to persistent disparities in per capita incomes and in growth rates. This discussion suggests that, one needs to deepen our understanding on the mechanics of knowledge diffusion and on the role of economic policies in overcoming the existing barriers, so as to increase a country's permeability to the innovations occurring elsewhere.

Box 6.6. Proximity and technological diffusion: the evidence

The question as to whether knowledge spillovers tend to be bounded in space or global is of crucial importance for our understanding of the forces that promote economic growth and convergence: if most knowledge spillovers are localized, companies operating nearby benefit more from each other innovations than companies located elsewhere. In this case, there will be an incentive for firms to operate in the same location, giving rise to cumulative causation and divergence. If, in contrast, knowledge spillovers are mostly global, there will be a tendency for laggard economies to catch up and to converge with the more advanced economies.

A strand in the literature has examined technological diffusion in its geographical dimension. An interesting study by Wolfgang Keller using intra-industry data, found that with every additional 1200 kilometres distance, there is a 50-percent drop in technological diffusion, irrespectively of country borders. Other researchers found that technological diffusion tends to be stronger within countries than across countries, that agglomeration advantages are more prevalent in R&D than in operations, and that the benefits of proximity might have declined with improvements in telecommunications. Still, the main conclusion is that literature is proximity still matters for knowledge diffusion (see Keller, 2004, for a survey)¹¹².

6.5 Learning by doing and international trade**6.5.1 Comparative advantages locked in**

External economies and learning by doing weigh on comparative advantages. If cumulative experience makes workers progressively more productive and regions

¹¹² Keller, W., 2004. "International Technological Diffusion". *Journal of Economic Literature* 42: 752-782.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

progressively more attractive to industries they are already specialized in, then a process of cumulative causation arises whereby comparative advantages are reinforced over time. Home firms become progressively more productive in each of the goods initially produced at home, while foreign firms become progressively more productive in each of the goods initially produced abroad. Thus, once a pattern of specialization is established, changes in relative productivity will act to further lock the pattern in.

To illustrate this, consider a World with two goods, say agriculture (Z) and manufactures (Y). Both goods are produced using labour, only, and the home country is small relative to the rest of the World. Wages are assumed flexible. The home country production functions are:

$$Y = AN_Y \quad (6.19)$$

$$Z = BN_Z \quad (6.20)$$

So far, this model is similar to the one analysed in Section 1.5. But instead of assuming that the productivity parameters A and B are exogenous, we now assume that technology in each sector evolves over time as a positive function of the country cumulative experience in that sector:

$$\dot{A} = \varepsilon_Y AN_Y \quad \text{with } \varepsilon_Y \geq 0 \quad (6.21)$$

$$\dot{B} = \varepsilon_Z BN_Z \quad \text{with } \varepsilon_Z \geq 0 \quad (6.22)$$

To abstract from scale effects, it is assumed that the total workforce in the economy is equal to 1, $N_Y + N_Z = 1$. To stick perfect competition, we retain the assumption that learning-by-doing takes place as a pure external effect: each producer ignores the effect of its decisions in the aggregate.

Now, let p be the relative price of the agriculture good in terms of manufactures in the world economy. If, at the time of trade openness, $p > A/B$ (that is, if the opportunity cost of producing the agriculture good at home is lower than the relative price of agriculture goods in the world economy), then the home country has comparative advantage in agriculture. Under free trade, the home country will specialize in agriculture.

Since the production possibilities frontier is linear, eventually the home economy will become fully specialized in agriculture. In that case, the only relevant production function

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

after openness will be (6.20), with knowledge accumulating according to (6.22). This mechanism will then intensify the existing comparative advantages: home firms will be progressively more productive in agriculture goods, while domestic productivity in manufactures is stagnant due to lack of productive experience. The initial specialization pattern, once established, becomes entrenched over time.

6.5.2 *Static versus dynamic benefits of trade*

Whether the specialization pattern achieved under free trade is good or bad for the home economy, it depends on how fast technology in agriculture evolves in that sector, compared to manufactures. If the learning potential of both industries was the same (i.e, if $\varepsilon_Y = \varepsilon_Z$), then the growth rate of per capita incomes at home and abroad would be independent of the specialization pattern. Comparative advantages would still be reinforced with accumulated experience, but there would be no consequences for the international distribution of per capita income.

If however the two industries differ in terms of learning opportunities, growth rates will be dependent on which good the country specializes in the first place. As an extreme example, assume that there is no learning-by-doing in agriculture, that is $\varepsilon_Z = 0$. In that case, openness to trade dooms the home country to a constant level of output, while the foreign economy is expanding over time. If in alternative the country remained in autarky, there would be production in both sectors, and some growth would be achieved through learning by doing in manufactures.

This model suggests that the pattern of specialization implied by comparative advantages is not necessarily the one that delivers faster economic growth. As long as

different goods differ in terms of their learning potential, trade openness may involve a trade-off between static efficiency and dynamic efficiency.¹¹³

Under this reasoning, many economists along time have been arguing that nations, instead of engaging in free trade *tout court*, should instead rely on temporary import protection in particular sectors, so as to develop economies of scale before opening up to the world markets. This idea, known as the *infant industry* argument, was first codified by the U.S. Secretary Alexandre Hamilton (1755-1784), who in turn was inspired by the mercantilist thinking of the 17th century¹¹⁴.

In the context of the learning by doing model, Paul Krugman showed that, in a world with many goods, a policy of selective restriction of imports, protecting first an industry with high growth potential until that industry becomes strong enough to survive in the open market, and then move protection to another industry with high learning potential, could alter permanently the pattern of comparative advantages in the protecting country' favour¹¹⁵.

Box 6.7. Learning by doing and the European fears of globalization

In light of the conventional theory of international trade, countries should specialize according to their comparative advantages. In the real world, however, many economists and think tanks believe that policymakers should not overlook the learning potential generated by a significant productive experience in manufactures. This argument gained importance in the last decades, as trade openness and globalization have generated a worldwide reallocation of

¹¹³ The idea that, in the presence of Marshallian externalities, the static gains from trade and the dynamic gains from trade may not go along was first formulated by Graham (1923) [Graham, F., 1923, Some aspects of protection further considered. *Quarterly Journal of Economics* 37, 199-227].

¹¹⁴ Hamilton, A., 1791. Report on manufactures. Reprinted in Syrett H et al (eds.) The Papers of Alexander Hamilton. New York and London: Columbia University Press, 1961.

¹¹⁵ Krugman (1987) argued that such strategy was followed with success by Japan, during its industrialization process [Krugman, P., 1987. "The narrow moving band, the Dutch disease and the competitive consequences of Mrs. Thatcher", *Journal of Development Economics* 27, 41-55].

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

manufactures away from industrial countries towards emerging economies, namely in Southeast Asia.

In the European Union, the European Commissioner, Jacques Barrot was one of the opposers to open globalization. He argued that giving up the European industrial base, allowing the low skill labour intensive components of the production chain to migrate to emerging economies, taking opportunity of the lower labour costs there, may benefit the European consumer in the short run, but will imply sooner or later the loss of the accumulated knowledge, making it impossible for Europe to explore the potential synergies between universities, research centres and firms, as envisaged by the European leaders: “it is not possible to maintain the knowledge accumulated through learning by doing if not supported by a production activity”, he claimed¹¹⁶.

6.5.3 Terms of trade effects

The model so far has assumed that world prices remain unchanged. However, if even if each country alone is small, it will be affected by shifts in terms of trade, as determined by the world demand and world supply of agriculture and manufacture goods.

To see how terms of trade effects may change our earlier conclusions, assume that the World is divided in two regions, say, North and South. Assume that the respective populations are constant. The North is initially specialized in manufactures and the South is specialized in agriculture. If learning by doing opportunities only occur in manufactures, then manufactures production will be expanding over time, while production in agriculture remains stagnant.

If both goods are normal in consumption, the increasing supply of manufactures will determine a fall in its relative price. This means that one unit of agricultural good will be traded by more and more units of the manufactured good as time goes by. Thus, even if the

¹¹⁶ Barrot, J., 2008. Les illusions d’une Europe sans industries”, Les Echos, 28/4/2008.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

South is unable to expand its production, it will benefit by an increasing purchasing power in terms of manufacture goods. For consumption and utility, what matters is real income expressed in terms of both goods, and this will be increasing through a terms of trade effect. By the same token, the North will be growing faster in terms of its own good (with output measured at domestic prices), but will face an adverse terms of trade effect weighing negatively on real income. On balance, what will happen to real incomes in both regions, it depends on preferences.

To examine this question, let's stick with our example of agriculture versus manufactures. As we already know, the demand for agriculture goods tends to increase less than proportionally than income, due to the Engel Law holds (this reasoning was already explored in chapter 1.5). If that is so, then as the world output increases induced by an expansion in manufactures, one will expect the relative demand to be progressively tilted towards manufactures, implying that terms of trade do not improve enough in the South. In the South, real income in terms of the two goods will be increasing, but less than in the North, and there will be economic divergence.

More generally, for any two goods, whether the terms of trade effect is enough to compensate for the diverging production, it depends on the elasticity of substitution between the two goods in consumption¹¹⁷. When the elasticity of substitution is equal to one (as with Cobb-Douglas preferences), the terms of trade effect exactly offset the differential productivity growth, and relative real incomes remain unchanged. If, in alternative, the two goods were high substitutes, the fall in manufactures prices would lead to a more than proportional increase in the world desired demand for manufactures, and terms of trade in the South would not be improving fast enough. Finally, if the substitutability between the two goods was lower than one, then the terms of trade effect would dominate the learning by doing effect and real income in the North would grow at slower pace, despite its faster technological progress. That would be an (unlikely) case of immiserating growth in the North.

¹¹⁷ This point was made in Lucas (1988).

6.6 Key ideas of Chapter 6

- An avenue that some authors have followed to make technological change endogenous without departing from perfect competition is to assume that technological change arises unintentionally, as a by-product of investments in physical or human capital. When this is so, technological change arises as an externality.
- In the presence of positive externalities on capital, there will be a larger role for capital in production than that implied by the share of capital in national incomes. In that case, the conventional growth accounting overestimates the Solow residual.
- With externalities, the competitive equilibrium is no longer a social optimum. There is scope for government intervention.
- If externalities are large enough to generate non-decreasing returns to capital, the model will display unceasing growth, even with a constant population.
- Aggregate externalities may be a source of cumulative causation. In that case, there will be a tendency for agglomeration of economic activities and to the self-reinforcement of economic disparities, with the richer economies becoming more attractive and getting richer, and the poorer economies remaining poor. To some extent, this was what happened across the World after the Industrial Revolution.
- In reality, economic activities do not move all to a single point because there are also centrifugal forces.
- Knowledge spillovers may act as a source of convergence or as a source of divergence depending on whether they are global or bounded in space. While some knowledge travels well, some other remains mostly embodied in people, requiring face to face contacts to be transmitted. There are also human-devised barriers to technological diffusion that prevent poorer economies to catch up. Economic policies may have a role in mitigating these barriers.
- Learning by doing reinforces comparative advantages, locking-in trade patterns. To the extent that industries differ in terms of learning opportunities, a trade-off may arise between the static gains from trade and the dynamic gains. Along this reasoning, it has been argued that countries should impose temporary protection in selected sectors, so as to achieve a critical mass before opening.

6.7 Problems and Exercises

Key concepts

- *External vs. internal economies of scale . Technological externalities. Tacit knowledge. Agglomeration effects. Cumulative causation. Learning by doing. Infant industry.*

Essay questions:

- Comment: “Perfect competition and increasing returns cannot hold together”
- Explain why the competitive equilibrium fails to deliver the first best allocation in the presence of externalities.
- To which extent do external economies help explain why capital doesn’t flow from rich countries to poor countries?
- What are the implications of knowledge spillovers being localized or global in scope?
- Explain why with learning by doing, the static gains from trade may not go along with the dynamic gains.

Exercises

- 6.1.** Consider an economy which production function is given by $Y = A_t K_t^{1/3} N_t^{2/3}$. In this economy, 16% of income is saved, the population is constant and capital does not depreciate. (a) Consider for a moment that $A_t = 16e^{(0.04/3)t}$. (a1) Describe the behaviour of per-capita income, Y/N , wages and of the interest rate in the steady state, as well as the capital and labour shares on national income. (a2) Are these results consistent with the empirical evidence? (b) Assume now that $A_t = 0.125(K/N)^{2/3}$. (b1) Explain the theory that fits in this specification. (b2) Does this aggregated production function verify the neoclassical properties? Explain. (b3) Find out the dynamics of per capita income in this model and represent it in a graph. (b4) will this economy display conditional convergence?
- 6.2.** Consider an economy where the production function of the representative firm is given by $Y_t = B_t K_t^{1/2} N_t^{1/2}$. In this economy, 16% of income is saved, the population is constant and capital does not depreciate. (a) Assuming perfect competition, find out the capital share on income, as well as the expression for the interest rate as a function of the average product of capital. (b) Assume first that $B_t = 0.1(K/N)^{0.5}$. b1) Explain the theory underlying this specification. b2) Find out the expression of the economy-wide production function. b3) Find out the dynamics of per capita income in this model and represent it in a graph. b4) How much will be the interest rate in this economy? Is this interest rate socially optimal? Why? (c) Consider now the case in which $B_t = 0.1K^{0.5}$. c1) Explain the theory underlying this specification. c2) Find out the expression of the economy-wide production function. c3) How much will be the interest rate in this economy? c4) Discuss the implications of this specification in terms of per capita income convergence.
- 6.3.** Consider an economy where the production function of the representative firm is given by $Y_t = B_t K_t^{1/3} N_t^{2/3}$. In this economy, savings amount to 16% of income, population is constant and capital does not depreciate. (a) Assuming perfect competition, what will be interest rate as a function of the capital-output ratio? (b) Assume that $B_t = 0.15K^{2/3}$. b1) Explain the theory. b2) Find out the aggregate production function in this economy.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

b2) Describe, quantifying, the dynamics of per capita income in this economy. (c) Find out the interest rate in this economy. What are the implications of this model in terms of per capita income convergence?

- 6.4.** Consider a World where the production function of each individual firm is given by: $Y_i = AK_i^{0.4}L_i^{0.6}$, where $A=0.2$ and $L_i = \lambda N_i$ measures labour in efficiency units. In this world, capital depreciates 4% each year and the population in each country is time invariant. In that world, consider a particular country where the saving rate is $s=25\%$ and population is $N=1$. (a) Assume for a moment that $\lambda = (K/N)$. In this case: (i) per capita income will expand at 1% per year; (ii) the interest rate is socially optimal; (iii) there will be conditional convergence (iv) all the above. (b) In the model above, an increase in the efficiency parameter to $A=0.25$ will deliver: (i) a temporary expansion of per capita output; (ii) the same interest rate in the long run; (iii) the same level of per capita output in the long run; (iv) none of the above. (c) Assume now that $\lambda = K$. In this case: (i) the long run real interest rate will be 4%; (ii) the externality is non-rival; (iii) a larger economy should grow faster than a smaller economy; (iv) all the above. (d) Finally, consider the case in which $\lambda = K^{1/3}$. In this case: (i) there will be sustained growth of per capita income; (ii) the long run real interest rate will be 4%; (iii) capital will flow out to the largest economy; (iv); all the above.

7 Research and development

“A system – any system economic or other – that at every given point in time utilizes its possibilities to the best advantage may yet in the long run be inferior to a system that does so at no given point in time, because the latter’s failure to do so may be a condition for the level of speed of long-run performance”. Joseph Schumpeter.

Learning Goals:

- Understand why making knowledge excludable is critical for the existence of private incentives to R&D
- Identify the main sources of knowledge excludability in the real world
- Identify the factors that influence the market value of a discovery
- Acknowledge the different market failures involving innovation

7.1 Introduction

The Solow model, by assuming perfect technological diffusion, cannot capture the incentives of economic agents to engage in research and development (R&D): since knowledge is assumed to diffuse instantaneously at no cost, no agent would be able to reap a return on any eventual invention. In the Learning by Doing model, this limitation of the perfect competition model is circumvented by assuming that technological progress arises as an unintended by-product of investment decisions. In that model, technological change materializes endogenously, though without reward. In this section, we analyse an alternative model of technological change, whereby innovation arises in result of purposeful efforts to develop new technologies. In this model, research is explicitly modelled as an economic activity, with a payoff. In doing so, we depart from the paradigm of perfect competition.

In today’s world, much competition between firms takes the form of firms trying to develop new and better products or less costly methods of producing existing products. Selfish economic agents would not be willing to devote valuable time and resources to R&D, unless they expected a reward in case of success. In many markets, that reward takes the form of an economic rent, arising from the inventor’s exclusive access to the innovation, at least for a period of time. Excludability over the new technology can be acquired via different <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

mechanisms, including trade secrets, lead time, and patents. These mechanisms prevent innovations from leaking out instantaneously, allowing innovators to acquire a temporary market power over their inventions and by then to reap a return on the research efforts. The view that technological progress is driven by the prospect of economic rents is on the basis of the so-called Schumpeterian paradigm of economic growth. In light of this theory, entrepreneurs engage in R&D with the aim to obtain profits. The R&D model therefore departs from the frictionless economy with perfect competition, to assume that knowledge does not diffuse instantaneously. This allows innovators to explore a temporary monopoly power.

In this chapter, we model R&D as an economic activity, with the aim to analyse the economic incentives underlying the invention of new technologies. Section 7.2 introduces some basic concepts. Section 7.3 describes the basic model, with a final good sector and a sector producing intermediate inputs. Section 7.4 analyses the incentives to innovate, focusing on the case in which innovations consist in the introduction of new products. Section 7.5 addresses the case in which innovations arise in the form of more efficient technologies to produce existing goods. Section 7.6 addresses the ex ante incentives for an entrepreneur to engage in R&D. Section 7.7 describes alternative mechanisms in which real world' firms rely, to preserve ownership on their inventions. Finally, in Section 7.8 we discuss the problems underlying the finance of valuable R&D and alternative mechanisms that have been developed. Section 7.9 concludes.

7.2 R&D Taxonomy

7.2.1 Basic research versus R&D

Research and Development activities may be categorized in different types: Basic Research, Applied Research, and Development. *Basic research* relates to studies that aim to improve fundamental knowledge for its own sake, in a manner that may be subsequently helpful across a range of activities. Since the knowledge thereby created discoveries is typically released to become publicly available, it generates no economic rents. Hence, most basic research is carried out in universities and non-profit institutions or with the support of government grants.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Applied research is aimed at generating specific uses for existing knowledge. Private firms are primarily engaged in applied research, with the aim of using knowledge for commercial purposes. In general, inventions result in prototypes not ready for consumer use. The process of further improving the invention and its production process so as to make it marketable is called *development*.

In this chapter, the incentives to R&D are mostly discussed in the context of applied research: that is, with firms gathering from general knowledge ideas that can be mastered and adapted to produce marketable goods with the aim to obtain profits. At the end of the chapter, we tackle the case of basic research, where the supportive role of the government becomes essential.

7.2.2 Horizontal and vertical innovations

In considering the output of R&D, there is a distinction between “horizontal innovations” and “vertical innovations”. *Horizontal innovations* are those that expand the range of available goods. For instance, the bicycle, the automobile, the train and the airplane are all horizontal innovations. Although they all address the same basic problem (e.g., transportation), the fact they do not solve this problem exactly in the same manner implies that consumers will tend to use each newly invented product alongside with the previous ones.

Vertical innovations are those that make existing goods or varieties obsolete. For example, the personal computer has displaced the typewriter as a text processing tool. Also more modern automobiles, computers and software tend to displace older vintages of automobiles, computers and software. So when a vertical invention is achieved, consumers tend to replace the old vintages by the new vintages.

A special category of horizontal innovations are inventions that have efficiency-enhancing effects across many sectors. For instance, breakthrough invention such as the wheel, the alphabet, the steam engine, the telephone, the railroad, the electric power, the micro-ship and the internet changed the organization of production across many sectors, and triggered a chain of complementary innovations that transformed the entire nature of the economy at their time. These are called “General Purpose Technologies”.

7.2.3 Process innovations and product innovations

A different categorization relates to where in the production chain the innovation materializes. Innovations can take the form of firms trying to develop new and better products, or instead less costly methods of producing existing products. Innovations that lead to the introduction of new products are labelled *product innovations*. Innovations that lead to the introduction of less costly methods of producing existing products are labelled *process innovations*.

Both product innovations and process innovations can be achieved either through horizontal innovations or through vertical innovations: for instance, an improvement in operations management in a factory producing shirts (process innovation) can be achieved either by introducing higher-quality versions of existing inputs (for instance, better software - vertical innovation) or by expanding the pool of intermediate inputs (for instance, inventing a new algorithm to solve a challenge in operations).

7.3 The basic R&D model

In this section we describe the basic framework where the economic incentives to R&D are to be analysed. In this framework, there is a final good sector where an homogeneous good is produced using a range of intermediate inputs, which in turn are produced using labour only. The final good sector operates under perfect competition, while intermediate inputs may be produced under imperfect competition.

7.3.1 A production function for final goods

Suppose that aggregate output (Y) is an homogenous good, assembled with m intermediate inputs, according to the following production function:

$$Y = B \sum_{j=1}^m x_j^{1-\beta} \quad (7.1)$$

Each intermediate input x_j is assumed to depreciate fully after use. In light of (7.1), the larger the number of intermediate inputs, m , the higher the final good production¹¹⁸. Hence, an expansion in the number of intermediate inputs m can be seen as form of technological progress, that we label *horizontal*. B is a parameter capturing the role of other inputs (e.g, capital, land), the size of the market, and country-specific factors, such as the quality of domestic policies and institutions.

7.3.2 Production function for intermediate goods

We assume that, once invented, intermediate inputs are produced using labour, only. Labour is homogeneous. Let N_j denote the amount of labour used in the production of intermediate good j . The production function of each intermediate input is assumed linear on labour:

$$x_j = \lambda_j N_j \quad (7.2)$$

The parameter λ_j measures the state of technology in the activity of producing the intermediate input j . Increases in λ_j come along with lower production costs, turning older technologies obsolete. Technological improvements leading to increases in parameter λ_j are labelled *vertical innovations*.

The total number of workers engaged in the production of intermediate inputs is N_Y :

¹¹⁸ Since intermediate inputs enter additively in the production function, the marginal product of intermediate product j is independent of the marginal product of intermediate product $j+1$. Still, firms may substitute one input for another while keeping the level of output constant. In this formulation, intermediate inputs are neither “direct” substitutes nor complement to each other. Technically, they are said to be *additively separable*.

$$N_Y = \sum_{j=1}^m N_j \quad (7.3)$$

Since labour is homogeneous, the allocation of workers across different activities shall obey to an arbitrage condition stating the wage rate must be the same.

7.3.3 Two sources of technological change

Although this model can account for heterogeneity in intermediate sectors, in most of our discussion we don't need this. Hence, let's assume that all sectors are alike, employing exactly the same technology and hiring a number of workers equal to the economy's average (alternatively, you can interpret λ as referring to the "average technology"):

$$\lambda_j = \lambda, \forall j \quad (7.4)$$

$$N_j = \frac{N_Y}{m} \quad (7.5)$$

Using (7.5) and (7.4), in (7.2) and (7.1), total output in the economy becomes:

$$Y = Bm \left(\lambda \frac{N_Y}{m} \right)^{1-\beta} = Bm^\beta (\lambda N_Y)^{1-\beta} \quad (7.6)$$

The last term in (7.6) shows the two possible sources of technological change that expand aggregate output in this model: horizontal (process) innovations (increases in m): expanding the pool of varieties for use in production; vertical (process) innovations (increases in λ): efficiency enhancements along a product line, allowing each variety to be produced at lower cost.

Box 7.1. The division of labour effect

In our model, an increase in the number of varieties m causes output to expand via greater availability of intermediate inputs (equation, 7.1). However, an increase in the number of varieties also gives rise to a *dilution effect*, whereby a given number of workers N_Y is divided by a larger number of varieties (equation 7.5) causing production of each variety to decrease (equation 7.2). On balance, equation (7.6) tells us that the net effect of an increase in m is positive. Why is this so?

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

The reason is that intermediate inputs enter in final good production with diminishing returns ($\beta > 0$): when a new intermediate input becomes available, workers are reallocated away from the production of old varieties to start producing a new variety. This reallocation causes the marginal product of existing varieties to increase, and the marginal product of the new variety to decrease, until they are all equalized. In the end, marginal products are higher across all product lines than before the innovation¹¹⁹. Note that if there were no diminishing returns ($\beta = 0$), aggregate output would not be impacted by horizontal innovations.

This model therefore captures the benefit of splitting production processes into different – and eventually more specialized - sub-tasks allowing workers to become more efficient in each subtask. This mechanism was coined by Adam Smith as “division of labour”.

7.3.4 The trade-off between production and R&D

In the R&D model, technological progress does not come for free. Inventions are assumed to be the output of a research sector employing resources that could otherwise be employed in other uses. These resources could be working time, labs, and equipment.

In the context of our model, the deviation of resources away from production to R&D is captured splitting the total labour force in the economy (N) in two groups: those workers engaged in the production of intermediate inputs (N_Y) and those workers engaged in R&D ($N - N_Y$). We denote by μ the fraction of the labour force devoted to R&D:

$$N_Y = (1 - \mu)N \quad (7.7)$$

Using (7.7) in (7.4), and dividing by N , one obtains an expression for per capita income:

¹¹⁹ See exercise 7.1 for an illustrative numerical example.

$$y = \frac{Y}{N} = (1 - \mu)^{1-\beta} B \left(\frac{m}{N} \right)^\beta \lambda^{1-\beta} \quad (7.8)$$

Equation (7.8) illustrates the trade-off underlying the allocation of working time to R&D versus production of intermediate inputs: if the economy commits a larger share of the labour force to R&D (μ rises), there will be a negative impact on per capita output, because less working time is devoted to production. At the same time, a higher research effort will allow output to expand faster over time, via faster technological change, that may arise either in the form of more intermediate inputs per capita (m/N) or in the form of more efficient ways of producing these inputs (λ). At the macro-level, the reasoning is similar to that in the Usawa' formulation.

7.4 Market structure and operating profits

To examine the microeconomic incentives to R&D, consider the case of an entrepreneur that invested some time in research and managed to discover a new technology. In this section we focus on the case in which new inventions arise in the form of intermediate inputs (horizontal innovations). The case where innovation take the form of more efficient ways of producing existing product (vertical innovation) is discussed in the section after.

7.4.1 Demand for intermediate inputs

We assume that the final good sector operates under perfect competition. There are a large number of identical firms that maximize profits taking the price of each intermediate input p_j as given. Under perfect competition, the total demand for each intermediate input is such that its price p_j equals the marginal product, $\partial Y / \partial x_j$. From (7.1), this gives:

$$p_j = (1 - \beta) B x_j^{-\beta} \quad (7.9)$$

The demand for an intermediate product j is described in Figure 7.1 by the downward sloping curve crossing points M and C.

7.4.2 The case with an horizontal innovation

Consider the case of entrepreneur that discovered a new intermediate input (say j). Because this innovation resulted in the expansion of the number of inputs available to production, it is a *horizontal innovation*. With no question, the fact that a new intermediate input is available constitutes an improvement for the economy as a whole: as explained in Box 7.1, a larger pool of inputs to be used in production allows the economy to take opportunity of the division of labour effect, improving aggregate efficiency. A different question is whether the invention comes along with a gain to the inventor himself. Whether the entrepreneur will gain or not, it depends on the profits obtained in producing (or selling the rights to produce) the new design – *operating profits* - compared to the fixed cost related to R&D.

To analyse the entrepreneur' problem, we refer to Figure 7.1. The downward sloping curve crossing M and C is the demand for the input j by the final good sector - equation (7.9). The marginal cost of producing this intermediate input with technology (7.2) is represented in the figure by the horizontal line crossing T and C (it is assumed that the innovator is price-taker in the labour market, so the wage rate w is given).

Given the marginal cost and the selling price, the operating profits are defined as:

$$\pi_j = p_j x_j - \frac{w}{\lambda_j} x_j \quad (7.10)$$

Whether these profits are positive or nil, it depends on the market structure.

First, consider the case were the new technology becomes freely available to all agents in the economy. In that case, the new variety will be produced by a large number of price-takers facing a marginal cost equal to w/λ_j . Profit maximization when the price is given delivers $p_j = w/\lambda_j$ and zero profits for all firms. This case is represented in Figure 7.1 by point C, where, $p_j = w/\lambda_j$. Of course, since in this case the innovator has no profits, he will not be able to recover the (sunk) cost F involved in the previous research activity.

In alternative, consider the case in which the inventor is the only one authorized to produce the new variety, becoming a price-maker. Substituting p_j for (7.9) in (7.10), his problem will be to choose x_j so as to maximize:

$$\pi = (1 - \beta) B x_j^{1-\beta} - \frac{w}{\lambda_j} x_j. \quad (7.11)$$

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

The solution of the maximization problem is the well-known rule stating that the monopolist' optimal price is a mark-up over the marginal cost:

$$p_j = \frac{1}{1-\beta} \left(\frac{w}{\lambda_j} \right). \quad (7.12)$$

In (7.12), the optimal “mark-up” depends negatively on the price elasticity of the demand curve, $1/\beta$: the lower the elasticity, the higher the mark-up.

The optimal production is obtained substituting (7.12) in (7.9):

$$x_j^M = \left[B(1-\beta)^2 \frac{\lambda_j}{w} \right]^{\frac{1}{\beta}}, \quad (7.13)$$

Using (7.2), the demand for labour by sector j becomes:

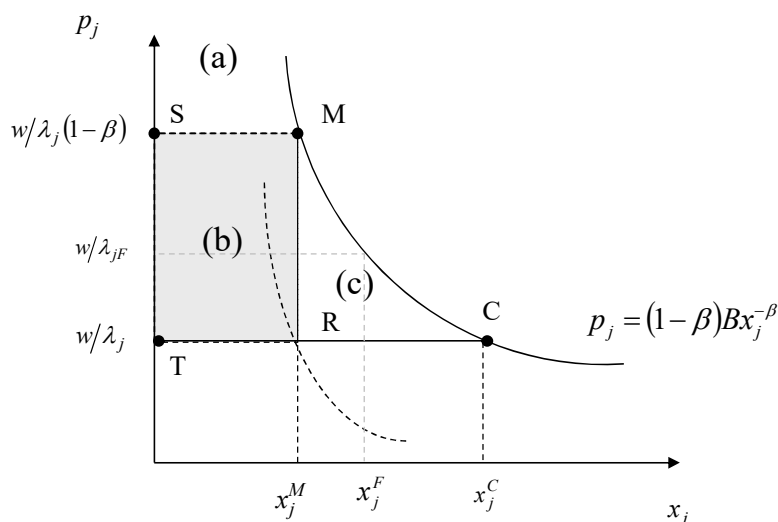
$$N_j = \lambda_j^{\frac{1-\beta}{\beta}} \left[\frac{B(1-\beta)^2}{w} \right]^{\frac{1}{\beta}} \quad (7.14)$$

Finally, substituting (7.12) in (7.11) one obtains the monopolist *operating* profits:

$$\pi_j = \frac{\beta}{1-\beta} w N_j, \quad (7.15)$$

where N_j is defined as in (7.14). In Figure 7.1, the monopoly case is represented by points R and M, which correspond to the intersection of marginal costs with marginal revenues - the dashed curve. The shaded area (b) measures the firm' operating profits.

Figure 7.1. Ex-post monopoly profits in the case of an horizontal innovation



Under perfect competition, the price will be equal to the marginal cost (point C). Under monopoly, the price is such that the marginal cost equals marginal revenues (point M), unless a competitive fringe forces the entrepreneur to set the limit price (w/λ_{jF}).

7.4.3 Static efficiency versus dynamic efficiency

The basic microeconomic theory tells us that monopolies are a source of inefficiency. This can be illustrated in terms of figure 7.1, comparing the welfare gains of the innovation under monopoly and under perfect competition.

The welfare gain of the innovation in the case with monopoly is given by the area (a)+(b), corresponding to the change in consumer surplus and the monopolist' profits, respectively. Under perfect competition, in alternative, consumer prices fall to w/λ_j , and the welfare gain of the innovation is the area (a)+(b)+(c), all accruing to consumers. Hence, the monopoly involves a transfer from consumers to the innovating firm (b) and a deadweight loss to the economy as a whole equal to (c).

The other side of the coin is that ex post monopoly profits are necessary to reward the research effort, without which there would be no consumer gain at all. As with many other problems in economics, there is a trade-off here: some excludability is inefficient from the static point of view, but it may provide the incentives for private agents to develop more ideas, which is good for all. This reasoning led one of the pioneers of modern development economics, Joseph A. Schumpeter (1883-1950) to claim that “static” efficiency and “dynamic” efficiency do not necessarily go along (see the quote at the beginning of this chapter).

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

7.4.4 The equilibrium wage rate and operating profits

In the discussion above, we analysed how the discovery of a new product impact on the entrepreneur profits, assuming all else equal. This is the right assumption to analyse the microeconomic decision of an agent that is small in respect to the economy. Collectively, however, technological change impacts on aggregate income, and by then on the wage rate, affecting individual profits.

To investigate this, let's assume that all m sectors are run by incumbents with full monopoly power, as described by equations (7.12)-(7.15). Since labour is homogeneous, the wage rate must be such that demand for labour in the production sector equals the supply of labour in the production sector. At any given moment in time, that will correspond to solving equation (7.3), given the optimal labour demand in each sector, (7.14). Since all sectors are equal, equation (7.5) holds. Hence, the equilibrium in the labour market becomes¹²⁰:

$$N_Y = m\lambda_j^{\frac{1-\beta}{\beta}} \left[\frac{B(1-\beta)^2}{w} \right]^{\frac{1}{\beta}} \quad (7.16)$$

Solving for w , one obtains $w = (1-\beta)^2 B(m/N_Y)^{\beta} \lambda^{1-\beta}$. This equation shows that, given the labour supply, N_Y , the wage rate increases with technological change. This is valid not only for technological change arising as an expansion in the number of varieties, but also for technological changes arising in the form of more efficient ways of producing the existing varieties, λ . Replacing (7.6) in this equation, the equilibrium wage rate becomes:

$$w = (1-\beta)^2 \frac{Y}{N_Y} \quad (7.17)$$

Using (7.17) and (7.5) in (7.15), and then (7.6), we obtain two alternative expressions to describe the monopoly profits in each sector:

¹²⁰ See Appendix 7.1 for the case in which λ_j differs across sectors.

$$\pi_j = \beta(1-\beta)\frac{Y}{m} = \beta(1-\beta)B\left[\frac{\lambda N_y}{m}\right]^{1-\beta} . \quad (7.18)$$

The first component in (7.18) shows that monopoly profits in each sector depend negatively on the market share ($1/m$), and positively on the size of the market (Y). All else equal, when the number of intermediate inputs increases, there is a dilution effect whereby the demand for each input declines, impacting negatively on monopoly rents. However, a higher number of intermediate inputs causes the size of the market to increase, via (7.6). The last term in equation (7.18) shows that, all else equal, the combined effect is negative: technological progress arising from the expansion in the number of varieties exerts a negative externality on incumbents' profits¹²¹.

7.5 Competition through innovation

The section above explored the case in which innovations consisted in the introduction of new varieties. We now turn to the case of technological improvements along existing product lines – that is, increases in λ .

When innovations are vertical, incumbents are challenged by the possible entry of new competitors offering exactly the same product. Such competition may either force incumbents to reduce prices, to share the market, or even do abandon the market, in case the new technology is more efficient than the older one.

7.5.1 Limit pricing

A first case of competition along a product line occurs when incumbent monopolists face the threat of market entry by less efficient suppliers of the same product. These less

¹²¹ A feature of many models with monopolistic competition is that the number of varieties m is set to increase proportionally to the size of the labour force, N . In that case, the ratio N/m remains constant and the dilution effect on profits is eliminated. The model in appendix 7.2 goes along with this avenue.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

efficient competitors can appear by imitation, knowledge leakages, or through the development of lower quality designs of the same product. Even if these suppliers do not actually enter in the market, they may force the incumbent to set a *limit price*, to prevent entry.

To examine this case, let's return to Figure 7.1. Suppose that the incumbent in the market for j is challenged by a large number of imitators (*competitive fringe*) that cannot exactly replicate the incumbent technology, but are able to produce the same product at some higher cost (lower labour productivity), $\lambda_{jF} < \lambda_j$. In case the imitators' disadvantage is not too large – as illustrated in the figure – then the best the incumbent can do to remain monopolist is to set the price just marginally below w/λ_{jF} (the *limit price*). Setting the limit price, the incumbent is able to undercut its rivals and preserve the monopoly position. However, operating profits, $\pi_{jF} = (w/\lambda_{jF} - w/\lambda_j)x_j$ will be lower than in the unconstrained case.

In sum, potential competitors may constrain the pricing behaviour of the incumbent, even if they don't actually operate. When this is so, it is the competitive fringe that (indirectly) sets the market price, not the monopolist. Consumers are of course better off under limit pricing: prices are lower than in the full monopoly case, and the quantity supplied (x_j^F) is higher. But the incumbent will get a lower return on his research effort.

7.5.2 The case with a more efficient vertical innovation

A different case occurs when the technology developed by a newcomer is more efficient than the existing one. In that case, the opportunity arises for the newcomer to outprice the previous competitors, driving them out of the market.

To analyse this case, assume that prior to innovation the market for product j was perfectly competitive: that is, a large number of firms were producing j with a given technology λ_0 (the suffix j is omitted to simplify the notation). In Figure 7.2, the equilibrium prior to the innovation is described by point C_0 , where the price is equal to the marginal cost (w/λ_0), profits of each firm in the fringe are zero, and the total demand for this variety is x_0^C . Departing from C_0 , suppose that an entrepreneur found a more efficient way

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

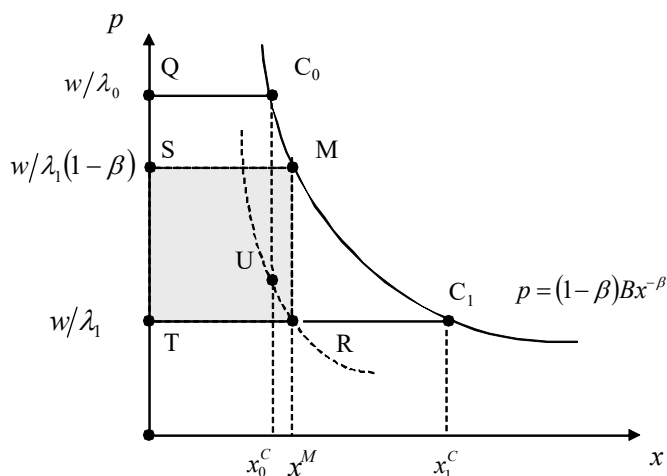
of producing the same product. In Figure 7.2, the vertical innovation is described by the fall in the (horizontal) marginal costs curve from w/λ_0 to w/λ_1 .

As in the case with the horizontal innovation, the innovation may translate into an effective competitive advantage to the innovating firm or not, depending on the innovator's ability to maintain exclusive control over the technology created: If competitors had immediate access to the new design, the market price would fall to w/λ_1 and the total demand for the good would increase from x_0^C to x_1^C . In that case, there would be no monopoly profits and consequently no reward to the time spent in R&D.

If, in alternative, the innovating firm had exclusive access to the new design, it could charge a price lower than the previous competitive price, driving all competitors out of business and become monopolist in this particular sector. In this case, it will be possible for the firm to generate profits to reward the previous research effort.

A previously competitive firm that beats its competitors through a vertical innovation and achieves a monopolist position in the market is said to have *escaped competition*.

Figure 7.2. Ex-post monopoly profits in the case of a drastic vertical innovation



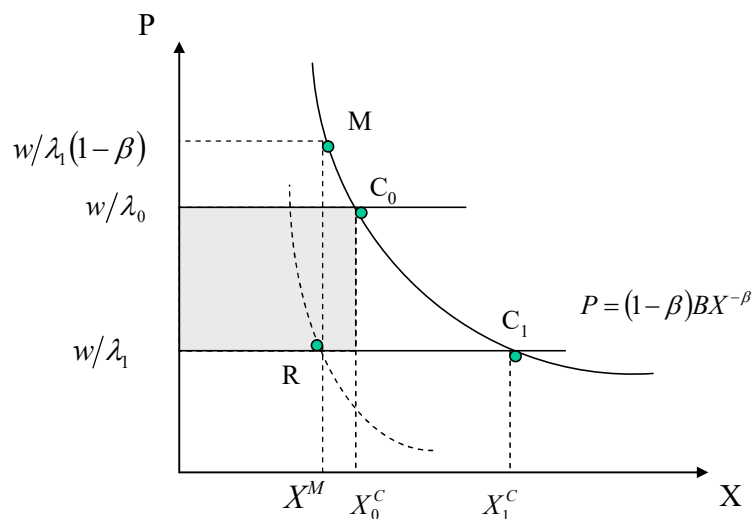
When the vertical innovation is drastic, the innovating firm is able to set the full monopoly price, because the intersection of the marginal cost and marginal revenue in point R implies a price that is lower than the pre-innovation competitive price.

7.5.3 Drastic versus non-drastic vertical innovations

The firm that escapes competition and becomes monopolist does not always set the full monopoly price (7.12). That will be possible only in case the resulting price does not exceed the previous market price. Otherwise, the best the innovator can do is to set a limit price.

These two alternative scenarios are illustrated in figures 7.2 and 7.3. In Figure 7.2, the innovating firm is able to set the full monopoly price, because the intersection of the marginal cost schedule (w/λ_1) with the marginal revenue in point R implies a monopoly price (point M) that is lower than the original competitive price (w/λ_0). This case is known as a *drastic innovation*. The implied operating profits corresponds to the shadow area in the figure.

Figure 7.3. The case with a non-drastic vertical innovation



In the case of a non-drastic innovation, the equality between the new marginal costs curve and marginal revenues (point R) implies a monopoly price (point M) exceeding the competitive price. Hence, the best the innovating firm can do is to set the price just marginally below the competitive price undercutting its rivals and capture the entire market.

Figure 7.3 illustrates the alternative case, of a *non-drastic innovation*. In the figure, the equality between the new marginal costs curve (w/λ_1) and marginal revenues (point R) implies a monopoly price (point M) exceeding the original competitive price (w/λ_0). Hence, the best the innovating firm can do is to set the price just marginally below the previous competitive price (w/λ_0). In doing so, it will be able to undercut its rivals and capture the entire market, pocketing the difference between this price and the new marginal cost, w/λ_1 .

Summing up, a drastic innovation corresponds to a sufficiently large improvement in technology so that the innovator becomes full monopolist. In the case of a non-drastic innovation, previous producers constrain the pricing behaviour of the entrepreneur, even if they don't actually operate¹²². In the case of a drastic innovation, the consumer price falls and quantity increases, so consumers are better off. This contrast to the case of a non-drastic innovation, where quantities remain unchanged and the only source of social gain is the increase in the producer surplus.

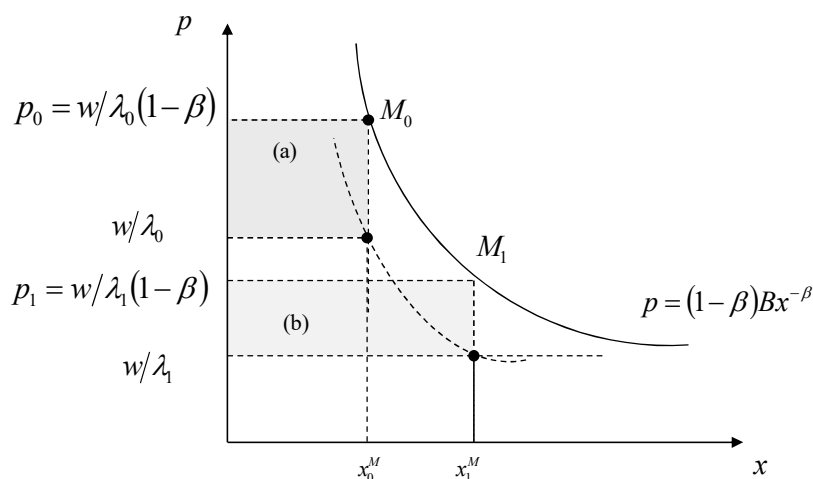
7.5.4 Creative destruction

In Figures 7.2 and 7.3 it is assumed that prior to the innovation the market was under perfect competition. In alternative, one may consider the case where the market was run by an incumbent monopolist. In that case, the vertical innovation comes along with the destruction of an existing economic rent.

In figure 7.4, we illustrate this, referring to a drastic innovation. The equilibrium prior to innovation is described by point M_0 . This equilibrium corresponds to the intersection of the incumbent's marginal costs curve (w/λ_0) with the locus of marginal revenues (the dashed curve), implying a price equal to $p_0 = w/\lambda_0(1 - \beta)$ and a total demand equal to x_0^M (the suffix j is omitted to save algebra). The incumbent's operational profits (7.15) corresponds to the area (a).

Figure 7.4. Creative Destruction

¹²² Formally, you may verify that the innovation will be drastic if: $\lambda_1/\lambda_0 > 1/(1 - \beta)$. Intuitively, when the demand is rigid, the full monopoly price is very high relative to marginal costs and therefore is more likely to exceed the competitive price.



The figure describes a case where a vertical innovation is large enough for the newcomer to become full monopolist driving the previous monopolist out the market. Since previous to innovation the market was monopolized, the arrival of the new technology came along with the destruction of existing rents.

Now assume that an entrepreneur invents a new technology allowing marginal costs to fall to w/λ_1 . As represented in the figure, the innovation is drastic because it allows the entrepreneur to set the full monopoly price and still undercut its rival. The new monopoly price falls to $p_1 = w/\lambda_1(1 - \beta)$ and production increases from x_0^M to x_1^M . With the innovation, the entrepreneur achieves operational profits equal to area (b), and the old rent (a) is destroyed at the benefit of consumers.

The view that firms bringing new technologies enter in the market destroying existing rents is on the basis of the Schumpeterian paradigm of economic growth¹²³. Joseph Schumpeter (1883-1950) theorized that the introduction of new products, new production processes and new forms of industrial organization by innovating firms undermine the marketability and the value of existing designs and production techniques. Innovating firms therefore obtain rents that come along with the destruction of their rivals' rents. The newly generated rents allow inventors to reap a temporary return on their research efforts. But

¹²³ Schumpeter, J., 1912. *Theorie der Wirtschaftlichen Entwicklung*. Leipzig: Dunker & Humblot. Schumpeter, J., 1950. *Capitalism, Socialism and Democracy*. New York: Harpe.

innovation rents do not last forever: sooner or later other firms will come up with new and better designs and production techniques, causing the incumbents' rents to erode.

The process through which technological change leads to the disappearance of old activities and firms, and the reallocation of resources to newer and more promising areas was labelled by Schumpeter as "creative destruction". Along this process, there are winners and losers. Firms that fail do adapt, experiment losses and are forced out of business. Surviving firms are forced to continuously revise their plans and production techniques, in process of permanent adaptation. In light of the Schumpeterian view, creative destruction allows the market economy to incessantly revitalize itself, in a process that resembles the Charles Darwin' theory of natural selection (see Box 7.2).

Box 7.2. The theory of natural selection

In its primitive form, the pea plant evolved a gene that makes its pods explode when peas are ready for germination. This mechanism allows peas to be scattered on the ground, ensuring the survival of the species. In each generation of pea plants, however, a number of mutants grow by accident lacking this key genetic ingredient: pods of mutant peas fail to pop up. In the wild, mutant peas die entombed in their pods. The natural selection therefore ensures that only the healthy pods pass on their genes.

When the man invented agriculture, the direction of natural selection was changed. Humans were not interested in the primitive version of the pea plant, because it is much more convenient to gather pods with peas enclosed directly from the plant, than to search for peas scattered on the ground, one by one. Thus, once the man became a farmer, it started growing the mutant version that fails to explode. Today, the pea plant we see in our fields is the mutant version, not the primitive. Farmers reversed the direction of natural selection: the formerly successful gene became lethal and the formerly lethal mutant became successful.

This example, described by Jared Diamond in his famous book *Guns, Germs and Steel*¹²⁴, illustrate the Darwin's concept of "natural selection": in the nature, each new generation of a species produces a number of mutants. Because in general mutants are not endowed with the same genetic information that their ancestral developed for thousands of years, they are in principle more vulnerable to environmental challenges. The natural processes of differential survival and reproduction does the selection. In critical junctures, however, the mutant "competencies" may turn out to become an advantage instead of a threat: changes in the natural environment may cause a mutant variety to become naturally selected. In these cases, the population undergoes an evolutionary change.

Like living species, economic agents adapt to changes in the economic environment. Agents tend to follow strategies that proved successful in the past. Successful strategies emerge as the outcome of a learning process, in the interaction game between individual competences and the economic environment. Occasionally, agents experiment new strategies. This is innovation. When the new strategy fails, agents retreat to the old strategies. Whenever the new strategy succeeds, the innovator acquires a competitive advantage. This advantage will render previous strategies obsolete. As time goes by, other agents start copying the more effective strategy, until it becomes dominant. This is Creative Destruction.

7.5.5 Neck-and-neck competition

The discussion so far has stressed the idea that product market competition, by eroding the rents that reward successful innovations too soon, discourage R&D. This idea captures the Schumpeterian argument that less market competition is good for growth. There is however another reasoning pointing in the opposite direction: when incumbents face the threat of their rents being eroded by new entrants, they will have incentive to escape

¹²⁴ Diamond, J., 1998. *Guns, Germs and Steel: a short history of everybody for the last 13,000 years*. Vintage, Surrey, UK.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

competition by innovating further. When, in contrast, incumbent monopolists are protected with high barriers to entry, they will have little incentives to keep innovating¹²⁵.

In this section, we complete the analysis on the relationship between competition and economic growth, considering a form of *dynamic competition*, according to which incumbents, facing the threat of their rents being eroded by competing innovations, try to “escape competition” innovating faster. In light of this reasoning, more competitions is good for R&D¹²⁶.

Consider the market of a given sector j , where innovations arise as improvements in labour productivity (λ). Instead of assuming that outsiders always undercut incumbents (figures 7.2-7.4), we now account for the possibility of imitators to exactly catch up with the frontier technology, forcing the incumbent to *share* profits. At any point in time, there will be *two* possible market structures in the industry: “neck-and-neck”, in which more than one firm compete using the frontier technology; and “unequal”, in which only one firm holds the frontier technology and supplies the entire market.

Referring to figure 7.5, assume that the market is initially “unequal”. The incumbent holds technology λ_1 (the “frontier technology”) but its advantage relative to the technology at the fringe (λ_0) is non-drastic. Hence, the best the incumbent can do is to set the price just marginally below w/λ_0 , capturing all the market, and pocketing the difference between the limit price and the marginal cost w/λ_1 . The incumbent profits are equal to the shaded area in the figure (π). All potential competitors are priced out, so their profits are zero.

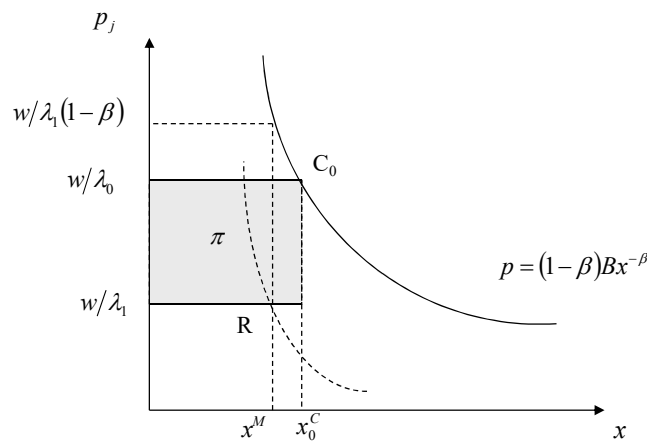
¹²⁵ Arrow (1962) showed that the monopolist’s incentives to innovate are reduced by a “replacement effect”, whereby the rents made possible with the new technology are just replacing rents that already existed under the previous technology [Arrow, K (1962), “Economic Welfare and the Allocation of Resources for Invention”, in Universities-National Bureau Committee for Economic Research (ed.) *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Princeton, NJ: Princeton University Press. pp. 609-626].

¹²⁶ Aghion, P., Harris, C., Howitt, P., Vickers, J., 2001. Competition, imitation and growth with step-by-step innovation. *Review of Economic Studies* 68, 467-492. Aghion, P., Harris, C., Vickers, J., 1997. Competition and growth with step-by-step innovations: an example. *European Economic Review, Papers and Proceedings*, 771-782.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Departing from the “unlevel” case, suppose that one entrepreneur from the fringe successfully innovates and joins the frontier technology, λ_1 . This means that, from now on, two firms will be operating in this market, competing “neck-and-neck”. The profits earned by each firm will depend on how far they will compete with each other: at one extreme, if they engage in open price competition, the equilibrium price will fall to w/λ_1 , resulting in zero profits for both; at the other extreme, if they collude, they can hold the price at w/λ_0 and share equally the profits, obtaining $\pi/2$ each (in this case, the newcomer is said to have “stolen” part of the leader business – see Box 7.3).

Figure 7.5. Neck-and-neck competition



Under “neck-and-neck” competition, the monopoly rent (shaded area) is divided by the number of players using the frontier technology. An increase in the number of players at the frontier reduces the incentives for firms in the competitive fringe to innovate and join the frontier (Schumpeterian effect) but it increases the incentives for firms at the frontier to innovate further and become monopolist of the new technology (“escape competition effect”).

If more firms catch up to the frontier, the share of π obtained by each declines further and the collusive solution becomes more difficult to maintain. Thus, for laggard firms, the higher the degree of competition at the frontier, the lower the incentives to catch up technologically and join the incumbents in the neck-and-neck state. This captures the conventional “Schumpeterian effect”, according to which increased competition discourages innovation.

For firms already in the neck-and-neck state, however, there will be more incentive to innovate the higher the level of competition at the frontier. The more competition at the front,

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

the lower the firm's profits there, and hence the higher the benefit of escaping competition innovating further. In case a firm at the front manages to discover a superior technology, it will be able to undercut its rivals, becoming monopolist in a new "unlevel state". Through this "escape competition effect", there will be a positive relationship between product market competition and innovation.

In sum, once we account for the possibility of neck-and-neck competition, the relationship between product market competition and incentives to innovate becomes ambiguous: on one hand, the higher the intensity of competition in a given market, the lower the incentive for outsiders to innovate and join that market; on the other hand, the larger the number of firms competing neck-and-neck in a given market, the bigger the incentive for one of these firms to "escape competition" innovating further and achieving a monopolist position. In light of this reasoning, R&D intensity should be higher in "unlevel" industries characterized by low competition (where the Schumpeterian effect dominates), and in "neck-and-neck" industries with high competition (where the "escape competition" dominates), but not much in intermediate states¹²⁷.

Box 7.3. The Business Stealing Effect

When an entrepreneur from the fringe successfully joins the leader in neck-and-neck competition, there is a partial deviation of rents from the leader to the newcomer. In this case, the innovator is said to *steal business* from the incumbent. The business stealing effect implies that the rents earned by the imitator correspond to losses by the previous monopolist, without delivering a net gain from the social point of view. On the contrary, the possibility exists for R&D efforts in this case to be welfare reducing.

¹²⁷ Aghion et al. (2005), found an inverted U relationship between product market competition and R&D that is supportive of this view [Aghion, P., Bloom, B., Blundell, R., Griffith, R., Howitt, P., 2005. "Competition and Innovation: An inverted-U relationship". Quarterly Journal of Economics 120, 701-728].

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Referring to figure 7.5, consider again the case of an entrepreneur from the fringe that invests in R&D to imitate the incumbent technology at the frontier, λ_1 . If the two firms collude and share the market equally, then the “business stealing effect” will correspond to half of the shaded area describing the profits. In that case, all the return reaped by the innovating firm will be a mere transfer from the incumbent, and consumers will see no gain at all. As long as the imitation involved a fixed cost, there will be a net loss for the society as a whole, even if the imitator itself had a private gain. From the society point of view, the imitator effort was a mere “stepping on shoes”.

7.6 R&D and an investment decision

In the discussion above, we have focused on the operating profits an innovator after the innovation was achieved. The sum of these operating profits along the economic lifetime of the invention correspond to the reward of the innovation effort. In practice, it may happen that operating profits reveal too low - or end up too soon - to cover the initial investment in R&D. *Ex post*, however, such assessment is useless: once incurred, investment in R&D becomes a *sunk cost*, and hence irrelevant for further decisions. In the production phase, the best the entrepreneur can do is to maximize the eventual operating profits given the market constraints, as explained above (or sell the licence for an equivalent amount), irrespectively of the amount previously invested in R&D.

A different question relates to the decision of engaging in R&D in the first place. Such decision involves an *ex ante* assessment of costs and benefits, taking into account the fixed costs involved in R&D, the uncertainty regarding the outcome of the research activity, and the future flow of operating profits during the life-time of the innovation. In this section, we focus on *ex ante* problem.

7.6.1 The market value of an innovation

The reward of a successful innovation is measured by the discounted sum of operating profits during the lifetime of the innovation. Using a discrete time formulation and ruling out uncertainty, that will be:

$$V = \sum_{t=1}^T \frac{\pi_t}{(1+r)^t} \quad , \quad (7.19)$$

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

where r refers to the time value of resources employed¹²⁸, and T refers to the lifetime of economic rents (in case the innovation is patented, T denotes for the length of the patent period). The discounted sum of operating profits (V) can be interpreted as the “market value of the innovation”: if the innovator decided to sell today the rights to produce with the new technology, and if the interest rate r was the opportunity cost of capital to investors, then the higher bid for the license in an auction would be precisely (7.19).

A particular case of (7.19) is when profits are eternal and constant over time. In that case, the perpetuity formula is obtained:

$$\lim_{T \rightarrow \infty} V = \sum_{t=1}^T \frac{\pi}{(1+r)^t} = \frac{\pi}{r} \quad (7.19a)$$

The assumption that profits are constant over time is not the more realistic one. Arguably, profits may erode over time, reflecting the arrival of new products and competitors, that reduce one's market share. A simple way to capture this is to assume that monopoly profits erode over time at a constant rate q , that is $\pi_t = \pi_1 (1-q)^t$. Replacing this in (7.19) and taking the limit as T approaches infinite, the market value of the innovation becomes:

$$V = \frac{\pi_1}{r+q} \quad (7.19b)$$

In this formulation, the denominator of (7.19b) can be interpreted as the opportunity cost of capital plus a premium to compensate for the time erosion of the cash flow.

A third alternative consists in allowing for uncertainty. Assume that the innovator obtains full monopoly power in the first period, enjoying the profits described by (7.15), but is unsure about how long the monopoly power will last for (T is uncertain). More

¹²⁸ The interest rate could be made endogenous, extending (7.1) so as to account for the role of physical capital. For such an extension, see Aghion and Howitt (1998), chapter 3. [Aghion, P. and Howitt, P., 1998. Endogenous growth theory. Cambridge, MA: MIT Press].

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

specifically, there is a probability q of a superior technology (vertical innovation) being discovered by someone else, turning this technology obsolete. If investors are risk neutral, the value of the license shall obey to a non-arbitrage condition, whereby the expected reward of carrying the license for one period, $\pi_1 - qV$, is equal to the return of investing the same amount of resources at the opportunity cost of capital r , that is: $\pi_1 - qV = rV$. Solving for V , the value of the innovation becomes exactly (7.19b). The denominator of (7.19b) shall be interpreted as an “obsolescence-adjusted interest rate”, capturing the risk of the current technology being displaced by a superior one: in case of no threat ($q=0$), the value of the license will be given by the perpetuity’ formula, (7.19a); in case $q=1$, then profits only materialize for one period, implying $V = \pi_1/(1+r)$.

7.6.2 The net present value of investment in R&D

The decision regarding investment in R&D involves an *ex ante* assessment on what will be the future cash flows (the value of the innovation, V) compared to the initial research costs. This assessment is complicated by the fact that at the time resources are allocated to a research project, investors do not know for sure whether this investment will deliver a marketable invention.

To model this, suppose that the investment in R&D consists in a fixed cost F incurred in period $t=0$. Such (certain) investment may deliver an innovation with value V at $t=1$ with probability b , or nothing with probability $1-b$. In that case, the expected net present value of the research project will be:

$$E[NPV] = bV - F \quad (7.20)$$

A risk neutral entrepreneur will engage in R&D whenever the expected Net Present Value of the project is positive.

Equations (7.19) and (7.20) summarize the key variables underlying the decision to invest in R&D. These include the fixed cost of R&D (F), the probability of success (b), future operating profits (π), *how long* excludability will last (T), and the discount rate (r). In a more restrictive setup, one may replace (7.19) by (7.19b), to account for the possibility of profits to erode over time, or to be destroyed by a competing innovation.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

This formulation may also apply to a deterministic setup. The only difference is that b shall be interpreted as the “productivity” of the research effort: that is, a larger b will correspond to a higher innovation value per unit of investment today.

7.6.3 The break-even value of the innovation

As long as economic agents are free to enter in the research activity, it is natural to assume that competition will drive down the expected NPV in the research sector to zero. In that case, an arbitrage conditions shall hold, stating that the expected value of the research outcome must be equal to the fixed cost in R&D:

$$V = F/b \quad (7.21)$$

In light of (7.21), the larger the required investment in R&D adjusted for the probability of success (F/b), the bigger must be the promised prize (V) so as to keep investors interested. A higher prize, in turn, requires a larger market, a lower elasticity of demand (β), or a longer monopoly lifetime (T).

This reasoning suggests that the nature of R&D in each industry has a word to say in respect to the corresponding market structure: in an industry where R&D costs are high and the probability of innovating is low, innovations will only spring if the market is sufficiently protected, with fewer firms and limited competition, to guarantee that profits are large and last for long. This helps explain the high concentration and the high patent-dependence in industries like the pharmaceutical, where research is very specific and costly, and the risk of failure is high. By contrast, the computer-games industry, where new games may be developed with relatively low investment, exhibits a much more open and competitive structure.

7.6.4 The equilibrium level of R&D (partial equilibrium)

Moving one step further, one may want to determine the equilibrium level of R&D in the context of our basic model. As it is however, the model cannot answer the question. That will require other assumptions, such as whether innovations will be horizontal or vertical, how technological improvements relate to the research effort, how long monopoly profits will last for, and so on. In the growth literature, some models focus on horizontal innovations with

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

perpetual rents (in the line of Jones, 1990)¹²⁹, and other models focus on vertical innovations with rents that disappear randomly due to creative destruction (in line of Aghion and Howitt, 1992)¹³⁰. In Appendix 7.1, we explore a possible solution along the second avenue. In this section, we sketch a possible partial equilibrium solution, considering an individual sector that is small in respect to the entire economy.

In our baseline model, investment in R&D is carried out by workers deviated away from production (equation 7.7). Hence, the term F basically refers to the cost of using research labour. Since labour is homogeneous and freely mobile across sectors, the cost of allocating workers to research can be measured by the foregone wages that otherwise could be earned in final good production. Since each sector is small, it takes the wage rate as given. With this ingredient, the model develops in an intuitive manner: labour is deviated away from production with the aim to obtain rents. Depending on how expected rents compare to the wage rate, workers allocate their time to R&D or to output production. At the margin, a worker must be indifferent between allocating one unit of time to output production or to research.

The key feature of the Schumpeterian model is that each new vertical innovation is fated to become obsolete at a given point in the future, when a superior technology is discovered by a competitor. Thus, when choosing its research effort, the entrepreneur must balance the potential gain of acquiring market power for some time against the cost of seeing profits disappearing because of the arrival of a superior technology. In this model, individual researchers are discouraged by the efforts of other researchers.

In equation (7.19b) the threat of a superior technology being discovered is captured by the probability q . Arguably, this variable shall depend on the amount of time dedicated to R&D in the same industry. Yet whether the relationship between R&D effort and innovation

¹²⁹ Romer, P., 1990. "Endogenous technological change". *Journal of Political Economy* 98, s71-s102.

¹³⁰ Aghion, P. and Howitt, P., 1992, "A model of growth through creative destruction". *Econometrica*, 323-51.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

outcome is linear, quadratic, or something else, has no clear answer. In the following, we follow an ad hoc assumption, postulating that the probability of a new vintage being discovered in a given sector is proportional to the *intensity* of R&D in that sector:

$$q = b\mu \quad (7.22)$$

This equation states that the probability of a technology to be displaced by a competing innovation depends on the research intensity in that sector, μ , times the probability of success, b . Replacing (7.22) in (7.19b), and then in (7.21), the equilibrium level of R&D in that sector must be such that:

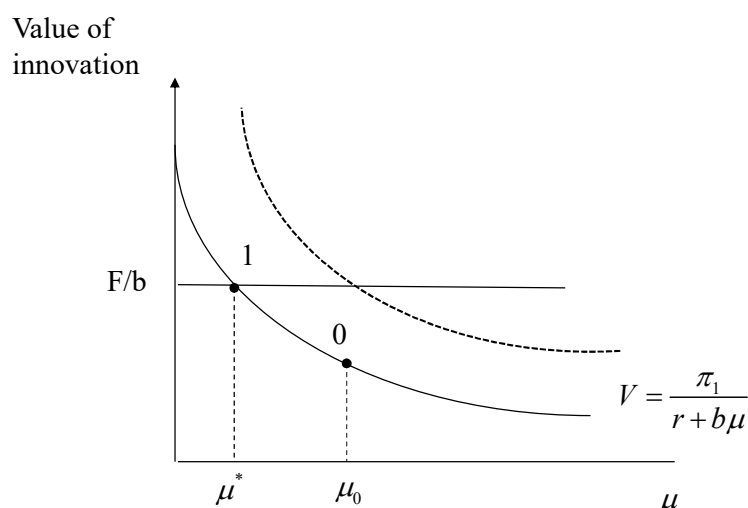
$$\frac{\pi_1}{r + b\mu} = \frac{F}{b} \quad (7.23)$$

Where the fixed cost corresponds to the market value of one unit of labour (i.e, $F=w$).

The equilibrium level of R&D is illustrated in Figure 7.6. The figure displays the two sides of (7.23) as a function of research intensity. The left-hand side is the value of a successful future innovation: it is a negative function of μ due to the “creative destruction effect”: the greater the R&D intensity in the sector, the more likely a competing innovation will destroy one’s rents. At the intercept, $\mu = 0$, the value of the successful innovation is the perpetuity formula (7.19a). The right-hand side of (7.23) is the fixed cost ($F=w$) adjusted for the probability of success.

To see how the equilibrium is reached, assume that initially the research intensity is $\mu = \mu_0$ in Figure 7.6. Point 0 is not an equilibrium because the value of a successful innovation is less than the adjusted fixed cost of R&D, implying a negative expected NPV. Thus, workers in that industry will reallocate time away from research towards production. As the research intensity in the industry decreases, the likelihood of a successful innovations being outpriced declines, implying a higher value of a successful innovation (movement along the curve to point 1). In equilibrium (μ^*) the expected return of the research activity bF must be equal to the fixed cost of R&D, F .

Figure 7.6. Equilibrium intensity of R&D (partial equilibrium)



The figure displays the value of a successful innovation in a given sector as a negative function of the research intensity in that sector, reflecting the creative destruction effect. Free entry in research activity implies that, at the margin, the value of the innovation must equal the fixed R&D costs, adjusted for the probability of success. All else equal, when operating profits increase, the equilibrium research intensity increases.

In light of this model, one can analyse the implications of an increase in the value of successful innovation. That could reflect, for instance, an enlargement in the extent of the market due to openness to international trade: for any given F/b characterizing an industry, a larger market size (B) will imply higher operating profits, and hence a shorter payback period¹³¹. In terms of figure 7.6, the curve describing the value of a successful innovation shifts to the right, increasing the return to R&D. As long as there is free entry in the research activity, the enlargement of the market will come along with a higher research intensity in this sector, and thereby with a faster rate of technological progress: innovators will break even faster because the market is larger and profits are larger, but rents will also be destroyed faster, at the benefit of the consumer. There will be more creative destruction and faster economic growth.

¹³¹ There is a famous quote by Matthew Boulton, a XVIII century British manufacturer and partner of the inventor of the steam engine James Watt, saying: “It is not worth my while to manufacture your engine for three countries only, but I find it very well worth my while to make it for all the world”.

Consider now the impact of an increase in the productivity of R&D, as captured by parameter b . Such a change could reflect an improvement in the organization of the R&D sector that turned the research efforts more successful. In equation (7.23) an increase in b has two effects: on one hand, it improves the probability of innovation reducing the adjusted fixed cost of R&D (the horizontal schedule shifts downwards); on the other hand, it increases the likelihood of creative destruction, reducing the value of a successful innovation (the downward sloping curve shifts to the left). It is easy to check that the former effect turns out to dominate, so when b increases, there will be a higher research intensity and a faster pace of technological progress.

In figure 7.6, the equilibrium is an interior solution. However, this is not a general case: the model does not necessarily imply that the equilibrium level of R&D in any given sector is positive: when the fixed cost is so high that the two curves fail to cross each other (that is, $\pi_1/r < F/b$), then the equilibrium level of R&D will be at $\mu = 0$. That will be the case, for instance, of pieces of knowledge that are impossible to hide or to make excludable, implying $\pi_1 = 0$. In that case, the market mechanism will fail to deliver any innovation in the sector.

Also note that the curve describing the value of a successful innovation refers to the market value, only, reflecting *private returns*. This curve tells us nothing in respect to the *social value* of the innovation. In any given sector, the social value of the innovation can exceed or fall short the market value, depending on the consumer surplus and on a range of external effects, that can be positive or negative. This means that the research intensity under *laissez faire* can deviate significantly from the social optimum, giving scope for government intervention.

7.7 Making knowledge excludable

7.7.1 Excludability sources

The discussion above illustrates the key role of excludability in providing market incentives for R&D. Technology is non-rival, but economic rents are rival. When competitors have instantaneous access to the knowledge created and the right to use the new technology, the innovating firm will not be able to raise the required operational profits to reward its

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

initial research effort. In that case, entrepreneurs will prefer to free ride on the other's research efforts, and henceforth there will be no R&D at all. When, in alternative, any mechanism prevents other firms from using the new design - at least during a certain period of time - then the innovating firm will be able to raise operational profits to reward its R&D effort. The less the technology diffuses, the higher the net present value, and the higher the incentives to R&D.

In the real world, there are different mechanisms in which entrepreneurs can rely, to secure some of the gains of their inventions, before knowledge leaks out completely to competitors¹³².

The first and most obvious mechanism of knowledge excludability is the *trade secret*. By not disclosing the details of an invention, its owner may manage to keep its competitors away from business. This has been the case, for instance, of the famous formula of Coca-Cola, for more than one hundred years. In alternative, the innovator may devote specific efforts to further design the product so as to make very hard for competitors to replicate it. An example of this is encrypting CDs to prevent unauthorized copies.

A common problem in innovative industries is that key ideas are embodied in workers hired and trained by the firm. Thus, there is an obvious risk of workers leaving the innovating firm to join rival firms or to start a competing business independently. In order to avoid this, firms may design compensation schemes that give key employees an incentive stay together (for instance, by sharing profits). In addition, they can introduce non-disclosure and non-compete clauses in employment contracts. In some cases, fellow firms working in a given location set agreements limiting the exchange of skilled workers between them¹³³.

¹³² This discussion presumes that the technology created is useful for competitors: if the invention was so specific that it only served the innovating firm, its diffusion would not be a problem.

¹³³ Whether this is socially good or bad is a different question. For instance, it has been argued that the weak enforcement of non-compete covenants in California may have contributed to the success of Silicon Valley [Gilson, R., 1999. The legal infrastructure of high technology industrial districts: Silicon Valley, Rout 128, and covenants not to compete. New York University Law Review 74, 575-629].

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Not all inventions, however, are suitable to be protected by trade secrets, encryption, or contract clauses in labour contracts: some ideas are so simple that are very easy to replicate (think, for instance, in the wheel or the “post-it”). In general, the passage of time makes even complex ideas very difficult to hide.

In many industries, the most relevant source of excludability is simply *lead-time*. Knowledge leaks only gradually. So in many industries the problem of competitors free-riding on ones’ ideas is circumvented by achieving a faster rate of technological change: innovating firms try to keep the lead continuously developing new sources of differentiation against their competitors. The time length that competing firms take to assimilate new ideas and incorporate them into their own business provides the innovating firm with a first-mover-advantage.

Other advantages for first movers include the time to build up customer loyalty, reputation, and the benefits of experience. Many industries (notably, shipbuilding, aircraft manufacturing, semiconductors) are characterized by a steep learning curve, whereby accumulated experience gives incumbents a significant cost advantage over competitors. This cost advantage does not leak out instantaneously.

Whenever the mechanisms above are not enough to provide the required protection for socially valuable inventions, inventors still have the option to buy legal protection, registering their property rights.

7.7.2 Patents, copyrights and trade marks

Patents are a legal mechanism that establishes private claims on intellectual property rights, permitting innovators to restrict unauthorized use of their ideas. To buy a patent, an inventor must demonstrate that the invention is novel and non-obvious.

A patent grants the inventor exclusive right to its discovery for a definite time length (20 years in Europe and in the US; from 14 to 15 years in the UK). During this period, no producer can use the invention without permission of the patent holder. Patent holders may however license (permit) others to use their invention in exchange for a payment called *royalty*. When the patent expires, other firms are allowed to enter the industry (note that this will not happen with a well maintained trade secret).

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

When applying for a patent, the inventor must disclose the details of its invention. The knowledge revealed is protected in the sense that only its owner can use it to produce the patented output. Yet the *information* in the patent (the technical details of the invention) can be used freely by other firms, to improve their own research projects. New inventions that do not compete directly with the patented output, even when built on the patent information, are in general considered legal.

An instrument related to patents is copyrights. Copyrights apply to art-works and works of authorship when these are attached to a tangible medium, such as a book or a CD. This contrasts to patents, which apply to products, processes, designs and substances. An important distinction between patents and copyrights is that the latter protects the particular expression of an idea, whereas patents protect any tangible embodiment of the idea itself. Therefore, patents allow greater exclusivity than copyrights. In compensation, the society sets copyright terms longer than patents (in the United States copyrights to business last 75 years and copyrights to individuals last for life plus 50 years).

A third category of legal protection of intellectual property are trademarks. The possibility of registering a trade mark encourages firms to develop customer loyalty and reputation. Unlike copyrights and patents, trademarks last forever.

Although legally distinct, patents, copyrights and trademarks can all be viewed as serving the same purpose: they all provide mechanisms of intellectual property protection, preventing others from using an existing idea. The aim is to allow innovators to reap a return on their research efforts.

7.7.3 The economics of patents

The enforcement of legal monopolies by a patent system is not free of controversy. Under monopoly, the firm produces too little and charges too high, imposing a loss on consumers, relative to the perfect competition case (see Box 7.1). Moreover, when ideas are at stake, a further reasoning applies: since the social cost of allowing more users to share any given idea is zero, does it make sense to exclude other people from using that idea? The problem is that, in many markets, valuable ideas would not emerge at all if there were no legal protection of property rights.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

The regulation of property rights has therefore the difficult task of striking a reasonable balance between the static cost of creating legally enforced monopolies with the dynamic gain of providing adequate incentives for researchers. In that problem, there are two key dimensions. The first is the *patent length*: for how long should the patent apply? The second is the *breath* of patent protection: to what range of products should the patent apply?

The optimal length of the patent shall obey to a balance between the need to provide adequate ex ante incentives to researchers and the benefits that will accrue to consumers once the patent expires. The longer the duration of the patent (T), the more time the innovator earns monopoly profits (area b in Figure 7.1), and hence the greater will be the incentive to engage in costly R&D (in figure 7.6, the curve describing the value of a successful innovation shifts to the right). However, a long patent length also implies a long lasting monopoly power, which comes along with a static deadweight loss (area c in Figure 7.1). If the life of the patent is too short, the innovating firm may not be able to generate enough profits to reward the research effort; if the life of the patent is too long, there are more incentives to innovate, but consumers will have to wait too long for open competition. The 20 year patent period is intended to strike a balance between static efficiency and the long run objective of stimulating research and innovation.

A similar trade-off applies to the *breath* of patent protection. If an inventor comes up with a product that is similar to one already patented, shall a patent be given to the new variant? If yes, the first inventor will reap less of the returns of her invention. Excessive coverage, on the other hand, will limit competition through innovation in the neighbourhood of the protected idea: other firms will see their returns to further developing the idea squeezed by the royalties they must pay to the original inventor. The optimal choice involves a balance between the need to stimulate R&D and competition through innovation.

In practice, the breath of patent protection is a matter of dispute in the patent office, with later entrants claiming the right to introduce slightly different innovations or new applications of the original idea without paying the royalties. Because litigation results are not always as desired by established firms, the later often protect the invention against other firms “inventing around”, by establishing property rights on related ideas, even if never used (“sleeping patents”).

Some authors argue that the optimal patent breath and the optimal patent length are not independent. For instance, it has been argued that, because imitators can often get around <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

the patent protection, engaging in a socially costly free ride, and because the incentives to do so increase with the patent length (if the patent duration is short, imitators will find cheaper to wait for the patent to expire), a “short and fat” patent system may be preferable to a “long and thin one”. In principle, a short and fat system will come along with faster creative destruction. This optimal patent breath and length remains however controversial topic¹³⁴.

7.7.4 The case against patents

Many historians have emphasized the role of institutions governing intellectual property rights as a main driver of economic growth. According to this view, the Industrial Revolution was only possible after governments established a proper regulation and enforced property rights, granting inventors with the necessary ex ante incentives to stimulate research and development¹³⁵.

Other authors have argued that the monopoly distortions imposed by the patent and copyright systems are too costly for what they achieve. According to this view, purely private excludability mechanisms, such as first-mover-advantages, lead time, secrecy and imitation delays should provide enough protection for innovation and deliver a better allocation of

¹³⁴ Gallini, N., 1992. Patent policy and the costly imitation. *Rand Journal of Economics* 23, 52-63.
Denicolò, V., 1996. Patent races and optimal patent breath and length. *Journal of industrial economics* 44, March, 249-65.

¹³⁵ Douglas North (1981), P. 164: “The failure to develop systematic property rights in innovation up until fairly modern times was a major source of the slow pace of technological change” [North, D. 1981. *Structure and Change in Economic History*. New York: Norton Other]. See also Landes, D., 1998. *The Wealth and poverty of Nations*. Abacus. Mokyr, J., 2002. *The gifts of Athena: historical origins of the knowledge economy*. Princeton university press, Princeton.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

resources than patent and copyright systems¹³⁶. These authors propose restricting patents and copyrights severely or even eliminating them altogether.

An argument that has been put forward is that patents are less efficient than trade secrets. At the first sight, the opposite looks true: when the inventor buys the patent, he must reveal the “secret”, allowing other inventors to build on it for other uses. On the other hand, consumers will have the opportunity to enjoy the benefits of the innovation when the patent expires. However, the inventor will only prefer to buy patent protection if he foresees that it will be impossible to keep the trade secret. If the trade secret could be maintained for more than 20 years, he would never buy the patent. Thus, secrets that without patent would be doomed to leak after a short period of time will, with the patent, be maintained for 20 years.

A mechanism that helps reduce the inefficiencies generated by patent protection is licensing: patent holders can permit others to use their invention in exchange for a payment called *royalty*. Licensing is welfare enhancing for two motives: on one hand, it prevents competitive innovation and imitation efforts, which are socially costly. On the other hand, because knowledge is non-rival, sharing it, even at a positive cost, is socially better than not to share it at all. Furthermore, from the innovating firm point of view, licensing allows the idea to be used in markets in which the inventor might not have competitive advantage (for instance, in a foreign country). In the real world, licensing is a primary way of transferring know-how across country borders.

The social cost of a patent may be considerable in the case of critical medicines, that could otherwise be sold cheaper and save lives. Because of this, critics of the patent systems have argued that the government (tax payers) should purchase patents for particular

¹³⁶ Merges and Nelson, for instance claim that patents make the entry of creative and energetic newcomers difficult, and argue that “there are many cases in which technical advance has been very rapid under a regime where intellectual property rights were weak or not strongly enforced.” [Merges, R., Nelson, R., 1994. On limiting or encouraging rivalry in technological progress: the effect of patent-scope decisions]. *Journal of Economic Behaviour and Organization* 25, 1-24. See also Boldrin, M. and Levine, D., 2002, “The case against intellectual property. *The American Economic Review* (Papers and Proceedings) 92, 209-212. Kremer, M. 1998. “Patent buyouts: a mechanism for encouraging innovation”. *Quarterly Journal of Economics*, 1998, pp. 1137-1167, November.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

innovations and release them to the public. This would eliminate the ex-post distortion and keep the incentives right. However, implementing such policy would involve a practical problem: if the innovation could be produced in third countries, there would be an incentive for foreign tax payers to free ride on the first country's effort.

In practice, patents are not equally necessary across industries. While some inventions only become available with an enforced patent system, many others become available just as quickly without a patent system. In other words, some inventions are "patent dependent" and others are not. In many industries, sufficient economic incentives for invention and innovation result from secrecy, and first-mover advantages (see Box 7.4). In these cases, the patent system is inefficient, in the sense that the same innovation could be obtained without the cost of granting monopoly power.

Box 7.4 How effective are patents?

A famous research on the effectiveness of patents, using a survey of 650 R&D managers representing 130 lines of business is presented in Table 7.1. The executives were asked to rate the effectiveness of patents as well as of other mechanisms, in protecting their competitive advantages, on a scale from 1 ("not at all effective") to 7 ("very effective").

Interesting enough, for process innovations, patent protection was considered the least effective method of protection. In the case of product innovations, the average score obtained by patents was slightly higher. Still, only secrecy was rated lower than patents. At the industry level, the authors found that only in pharmaceuticals— and for the particular case of product innovations - did the majority of the respondents rate patents as strictly more effective than other means of appropriation.

Another famous study analysed a sample of 100 firms, from twelve broadly defined industries¹³⁷. The author surveyed the firms' R&D directors to ascertain what proportion of

¹³⁷ Mansfield, E., 1986. Patents and Innovation, Management Science, 1986, February.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

their companies inventions were “patent dependent”. The results reveal a quite unimportant role for patents (proportions of 1% or less) in six out of twelve industries, and a moderate role in other three (from 11% to 17%). The three industries reported as more patent dependent where “petroleum” (25%), “Other chemicals” (38%) and “pharmaceuticals” (60%).

Table 7.1. Effectiveness of alternative means of protecting innovations

Effectiveness of alternative means of protecting advantages of new or improved processes and products		
Method of Appropriation	Sample means	
	Processes	Products
Patents to prevent duplication	3,52	4,33
Patents to secure royalty income	3,31	3,75
Secrecy	4,31	3,57
Lead time	5,11	5,41
Moving quickly down the learning curve	5,02	5,09
Sales or services effort	4,55	5,59

Note: Range: 1= not at all effective; 7= very effective. Source: Levin, R., Klevorick, A., Nelson, R., Winter, S., 1987 “Appropriating the returns from industrial research and development. *Brooking Papers on Economic Activity* 3, 783-820.

More recently, authors analysing survey responses from 1478 R&D labs in the U.S manufacturing sector, found that in most industries patents were the least emphasized mechanism of protection¹³⁸. In the pharmaceutical industry, however, patents were considered an effective protection mechanism for more than 50% of all product innovations. In the case of chemicals, the authors also indicate an important role of patents as a mechanism to deter the patenting of close substitutes by rivals (patent blocking).

¹³⁸ Cohen, W., Nelson, R., Walsh, J., 2000. Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not), NBER Working Paper 7552.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

7.8 Financing research and development

7.8.1 Financing constraints

The existence of market incentives for a firm to engage in R&D is not a sufficient for investment in R&D to materialize: even when expected profits are high, valuable research projects may fail to be implemented, due to lack of finance. In the real world, many firms face difficulties in raising capital to finance their R&D projects.

There are many reasons underlying the market failure related to the finance of R&D. First, research projects may fail: either because nothing much of relevance is discovered or because the invention may be beaten in the last minute by a competing firm in the patent office. Research projects involve a significant probability of ex post returns being insufficient to cover the initial loan, raising the likelihood of involuntary default. Second, financial transactions are characterized by a problem of asymmetric information, that is more pervasive in R&D projects: either because of the technical complexity of the project or simply because researchers do not want to disclose the technical details, investors do not in general fully understand what is envisaged by the researcher. This rises a typical problem of moral hazard: because it is difficult to monitor the true effort and the quality offered by the researchers, there is ample scope for low levels of commitment and to the hiding of relevant aspects of new ideas in the event of success. Third, R&D projects do not in general provide valuable collaterals: in the case of a mortgage loan, if the borrower defaults, the bank can seize the real asset. In contrast, when the bank lends for R&D and the research project fails, the bank may end up with nothing. This problem may be, of course, bypassed by the borrower offering other asset as collateral, but this mechanism may not be available for many start-ups, especially in poor countries.

Because capital markets are not efficient, R&D intensive firms tend to use own capital to finance their research efforts. This is not a big issue for large companies already

established in the market. Established firms can raise capital for new R&D out of their profits on past R&D. But for new entrants, especially small firms, lack of financing may constitute a significant barrier to entry and a source of market imperfection¹³⁹.

A mechanism that overcomes the problems in loan contracts is *venture capital*. Specialized financial institutions such as investment banks may invest in promising R&D projects or in start-up companies, demanding in compensation for the risk taking an ownership stake in the project. Since this comes along with the right to assign a manager in the firm, venture capital not only ensures financing, but also quality management and overcomes the problem of moral-hazard that plagues conventional loan contracts. Still, venture capital rarely meets the existing needs, especially at the smaller end of the market, where transaction costs are high relative to expected returns.

In general, financial development is favourable to R&D: as new and more complex securities become available, risk-spreading opportunities for investors increase, with the consequence of increasing the availability of funds to risky projects. In developing countries, where bank loans dominate as source of finance, entrepreneurs tend to choose inferior but safer strategies, such as relying more on imitation than on invention¹⁴⁰.

Box 7.5. Positive externalities on R&D

Along this chapter we have stressed the role of knowledge excludability in providing market incentives to innovate. A different question is whether knowledge excludability – even when fully achieved – will provide the enough incentives for entrepreneurs to engage in the socially optimal R&D. In general, because social returns to innovation do in general exceed the private returns, the level of R&D in the competitive equilibrium will be in general

¹³⁹ The fact that a significant fraction of R&D investments are self-financed lead Schumpeter to defend that large firms have an advantage because they have more resources to invest and setup expensive laboratories. Schumpeter also conjectured that large firms are more likely to engage in R&D because they can explore economies of scale and spread risks across projects.

¹⁴⁰ Acemoglu, D., Zilibotti, F., 1997. Was Prometheus unbounded by chance? Risk, diversification and growth. *Journal of Political Economy* 105, 710-751

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

suboptimal. In terms of figure 7.6, that will be the case is the curve describing the social value of the innovation (such as the dashed curve) lied above the curve describing the market value of the innovation (solid curve).

As a first approach to this question, consider Figure 7.1 again. In that example, the monopoly profit (area b) falls short the social benefit of the invention (area $a+b$). The later accounts for the impact on “consumer surplus”, which, in the case of a horizontal innovation, measures the efficiency-enhancing effect in production, due to the arrival of a new intermediate input. The fact that the monopolist does not fully appropriate all the benefits of its innovations to the society is called the *appropriability effect*¹⁴¹. The appropriability effect implies that a socially beneficial innovation may fail to occur, even if perfectly excludable. In terms of Figure 7.1, this will be the case when the cost of the innovation (annualized) lies between the areas (b) and $(a)+(b)$.

A second reason why the social benefits of innovations exceed the private benefits relates to the fact that new knowledge builds on previous knowledge (the “standing on shoulders effect”). As an example, consider the Toyota “lean production system”. This was achieved by Toyota after carefully studying the Ford production system and improving upon it. Some of Toyota developments were later imitated by American firms¹⁴².

The “standing on shoulders” effect may arise even if the technical details of the previous idea are unknown. Even when the new technology is not well assimilated by followers, the simple propagation of the idea may inspire other researchers to develop alternative ways of achieving similar results. The difference is that in this alternative model of technological diffusion, all the details have to be reinvented. As an example, consider the personal computer industry. The first graphical operating system was first introduced by

¹⁴¹ In the case of a drastic vertical innovation, the increase in consumer surplus also implies a social gain larger than the private gain. In Figure 7.2, we see that ex post profits are given by the area [SMTR] while the increase in social welfare is the area [QC₀MRT].

¹⁴² Mukoyama, T. 2003, Innovation, imitation and growth with cumulative technology, Journal of Monetary Economics 50, 361-380.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Apple in 1984. But Microsoft adapted the idea and came out with its own graphical operating system in 1985, the Windows 1.0. The underlying idea was the same, but the programming language was completely different.

The fact that simple ideas may induce independent efforts to achieve similar results implies that even well maintained trade secrets face the threat of competitors free riding on the idea diffusion. An obvious example is the case of Coca-cola: even though its formula is a trade secret, the concept is not. With no surprise, other firms, such as Pepsi Co and Canada Dry, have entered the market with close substitutes. Another example is the atomic bomb: historians are still debating whether the Russian atomic bomb was based on detailed blueprints stolen from the Americans or instead it was the diffusion of the idea that induced the Soviets to engage in an independent project where the principles of the bomb were reinvented.

The “standing on shoulders” also applies across industries. Technical innovations originated in a particular industry often find important applications, as well as instigate further technological change very far from the original invention’ starting industry. A classical example is the invention of the transistor by the telephone company AT&T. Although this firm was rewarded for its R&D effort through higher margins in the production of telephone devices, many other firms took the opportunity offered by the new invention, namely to develop better radios and better television sets.

The appropriability effect and the standing on shoulders effect imply that innovating firms will not, in general, appropriate all the benefits of their innovations to the society. Thus, in a decentralized economy, they will not innovate as fast as it would be socially desirable.

7.8.2 Subsidies to private R&D

Wherever ex post monopoly rents that the innovator can capture fall short the social welfare created by the invention, private firms will not innovate as fast as it would be socially desirable. In that case, there is scope for the government to support innovation.

A common policy instrument is the subsidy. In terms of figure 7.6, a government subsidy to R&D will cause the horizontal schedule describing the fixed cost to move downwards, rising the incentive to R&D. Government subsidies can be attributed either to

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

specific innovation projects, to particular innovation activities or more generally to research projects in particular industries.

Some authors argue that government subsidies should target differently different industries, on the ground that they are more needed in more competitive environments (because there are no profits) and where potential for technological spillovers are larger. Designing different subsidies to different industries involves however, a large level of discretion. In a world where governments face important information failures, a question of level-playing-field arises: firms or sectors benefiting from a government subsidy may obtain an undesirable competitive advantage against their competitors at the expense of the taxpayers. For this reason, international agreements and some domestic competition laws (such as in the U.S. and in the E.U.) have been limiting narrowly focused subsidies and state-aids to particular firms. By contrast, broad-based subsidisation mechanisms to particular activities, such as R&D tax credits, because they do not depend on government selection of particular projects or industries, are inherently less distortionary and hence more tolerated by the domestic competition laws and international trade agreements.

7.8.3 Government funded R&D

So far, we have analysed mechanisms of government support to R&D that are market based. By establishing and enforcing a system of property rights and by subsidizing innovative activities, government may help firms appropriate more of the social benefits of their innovations, inducing R&D efforts more aligned with the social interest.

Not all research, however, is driven by market concerns. For example, advances in *basic sciences*, such as geography, economics, mathematics and physics cannot be patented. Hence, there are no rents to extract. And yet, because of the large externalities involved, advances in basic science are of great importance for the progress of human kind.

On the other hand, even when particular types of knowledge are suitable for exclusion, it may be socially preferable to make them freely available. Remember that, because knowledge is non-rival, the social cost of having more agents sharing the same idea is zero. Given the cumulative nature of knowledge – i.e, new discoveries build on old discoveries – there is a good case to let relevant pieces of knowledge to become freely available, even when the alternative of patenting is possible.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

As with public goods in general, governments may have a role in supporting directly the creation of knowledge. One possibility is to reward with prizes and research grants the creation of knowledge that becomes public property. For instance, academic and government scientists do not work with the primary objective of profit-maximisation. Their main incentive is to disclose the product of their research in order to receive rewards. This kind of support is known as “patronage”¹⁴³.

Governments may also promote research and development through “procurement”. In this case, a public body contracts out in advance for a specific piece of research to be undertaken. With this mechanism, the government absorbs some or all the risks that the private firm would otherwise have to address. Depending on the interests of the government, the findings of the research undertaken under procurement may become publically available or not. In the case of military research and big space programmes, such as those managed by NASA in the USA, disclosure is not in general allowed.

Government funded R&D accounts for between one-third of total R&D expenditures in US and one half Europe¹⁴⁴.

7.9 Key ideas of chapter 7

- Private agents dedicate valuable resources to the development of new technologies because they expect to be rewarded in case of successful innovation. In the market, the reward of successful R&D depend on the possibility of making the new technology marketable and excludable.
- Monopoly rents made possible by R&D depend on the size of the market, on how long the technological advantage will last, and on the emergence of competing innovations.

¹⁴³ David, P., 1992. Knowledge, property, and the system dynamics of technological change. Proceedings of the World Bank Annual Conference on Development Economics, 215-248.

¹⁴⁴ Keely, L, and Quah, D., 1998. “Technology and Growth”, Centre for Economic Performance Discussion Paper N° 391, London School of Economics, May.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

- When entrepreneurs achieve a successful innovation, they often destroy rents of non-innovating firms. This competitive nature of R&D is labelled Creative Destruction and resembles the Darwin theory of natural selection.
- The fact that the reward to innovation comes in the form of monopoly profits does not necessarily imply that less competition is good for innovation. True, a market with low competition will be more attractive for newcomers, so through this “Schumpeterian effect”, less competition is good for innovation. However, high product market competition at the frontier also creates the incentives for firms in that market to innovate as a form of “escape competition”.
- Investments in R&D involve certain costs to obtain uncertain outcome. The larger the required investment adjusted for the probability of success, the less competitive the industry must be for R&D investment to break even. On such equilibrium is not possible, innovations will not spring up.
- A technology can be made exclusive by market mechanisms, such as secrecy, lead-time, and customer loyalty. When these are not enough, innovators may secure property rights through legal mechanisms, such as patents, copyrights and trademarks. In practice, patents are likely to be an important source of excludability in few industries, such as chemicals and pharmaceuticals. Some authors contend that the dynamic gains achieved with the patent system are not enough to offset the static costs.
- The market mechanism may deliver too little R&D. Even if private incentives exist, financial constraints may prevent a talented researcher to do so. A main reason is that the output of R&D is uncertain and immaterial, and hence it does not provide a feasible collateral. Equity capital is a key form of R&D financing.
- Private researchers do not in general fully appropriate the social benefits of their inventions, even when property rights are fully enforced. This means that the government may have a role in stimulating the research activity through subsidies and research grants.

Appendix 7.1. The crowding out effect

In this appendix, we show how the model with all sectors monopolized looks like, when labour productivity λ differs across sectors. Substituting (7.14) in (7.3), the aggregate demand for labour is as follows:

$$N_Y = \left[\frac{B(1-\beta)^2}{w} \right]^{\frac{1}{\beta}} \sum_{j=1}^m \lambda_j^{\frac{1-\beta}{\beta}} \quad (7.16a)$$

The last term in (7.16a) accounts for cross-sector productivity differences. For mathematical convenience, we define the “average technology”, λ , such that:

$$\sum_j \lambda_j^{\frac{1-\beta}{\beta}} = \sum_j \lambda^{\frac{1-\beta}{\beta}} = m \lambda^{\frac{1-\beta}{\beta}} \quad (7.4a)$$

Replacing (7.4a) in (7.16a), and solving for the wage rate, one obtains (7.16), this time with λ denoting for the “average technology”. Combining (7.16) and (7.14), one obtains another view of the individual labour demands:

$$N_j = \left(\frac{\lambda_j}{\lambda} \right)^{\frac{1-\beta}{\beta}} \frac{N_Y}{m} \quad (7.5a)$$

Equation (7.5a) extends equation (7.5) by allowing productivity improvements in each sector not to be synchronized. According to this expression, employment in sector j will be higher or below average, depending on how sector j ' productivity (λ_j) compares to the economy average, λ . In a sector with productivity equal to the economy' average, $\lambda_j = \lambda$, employment will be equal to the average (7.5). Equation (7.5a) thereby implies a sort of “crowding out effect”, whereby sectors with fast productivity growth expand and absorb workers released from non-innovating sectors, eroding their profits. This “crowding out” effect, mediated by wages, arises as a negative externality of vertical innovations, from innovating firms to laggard firms.

Substituting (7.5a) in (7.2) and then in (7.1) one obtains again the expression of aggregate output as a function of horizontal and vertical innovations, (7.6). Individual profits can be rewritten substituting (7.5a) in (7.15) and using (7.17a):

$$\pi_j = \beta(1-\beta) \frac{Y}{m} \left(\frac{\lambda_j}{\lambda} \right)^{\frac{1-\beta}{\beta}} \quad (7.18a)$$

This equation shows that profits in each sector depend on how the sector productivity (λ_j) relates to the economy' average (λ). Together, equations (7.5a) and (7.18a) reveal that asynchrony in vertical innovations cause labour and profits to reallocate across sectors. Employment and profits will increase in innovating sectors and will decrease in non-innovating sectors.

This extension reveals that cross-sectoral innovation externalities arise both in the form of horizontal and vertical innovations. Profits in each sector depend on the productivity in that sector, but also on the technological developments in other sectors. It is the combined

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

effect of all technological improvements that determines the wage rate and by then, expected profits and the incentives to innovate.

Appendix 7.2. The equilibrium level of R&D

In this appendix, we show how the economy-wide equilibrium level of R&D can be determined in the context of our model. Assume that R&D efforts are aimed to achieve productivity gains along existing product lines (vertical innovations). To abstract from complications related to asynchronized technological change, we focus on the “average sector”, where $\lambda_j = \lambda$, and we assume that all innovations are drastic and proportional¹⁴⁵. As for horizontal innovations, the number of varieties m is assumed to spring automatically in proportion to the size of population¹⁴⁶.

Suppose that the current technological level in the average sector is λ_0 . The probability of the next vintage λ_1 being discovered is assumed proportional to the number of researchers in that sector, $\mu N/m$:

$$q = b\mu[N/m] \quad (7.22a)$$

Since we are assuming that the number of sectors m increases along with the size of population, the number of researchers per variety remains constant. This is an important property of the model: most endogenous growth models are plagued by a scale effect whereby the growth rate of per capita output becomes a function of the size of population. This model gets rid of the scale effect assuming that the number of varieties expands along

¹⁴⁵ More specifically, we assume that $\lambda_2/\lambda_1 = \lambda_1/\lambda_0 = \dots = 1/(1-\beta)$. This is a necessary assumption for the equilibrium μ to be constant over time.

¹⁴⁶ Authors following this direction in growth models include Dinopoulos and Thompson (1998) and Peretto (1998) [Dinopoulos, E. and Thompson, P., 1998. Shumpeterian growth without scale effects. *Journal of economic growth* 3(4), 313-35. Peretto, P. “Technological change and population growth”, 1998. *Journal of Economic Growth*, 3(4), 283-311].

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

with the size of population. This allows the number of researchers per variety to remain unchanged. Since the arrival of new technologies depends on the number of researchers per sector, the average productivity growth is unaffected by the size of population.

Using (7.22a) we can rewriting (7.23) as:

$$\frac{\pi_1}{r + b\mu(N/b)} = \frac{w_0}{b} \quad (7.23a)$$

In (7.23a) the wage rate refers to the period pre-innovation, while profits are generated after innovation¹⁴⁷. Under full monopoly in all sectors, the equilibrium wage rate obeys to (7.17). Using (7.7), this gives:

$$w_0 = \left[\frac{(1-\beta)^2}{1-\mu} \right] y_0 \quad (7.17a)$$

where $y_0 = Y_0/N_0$ denotes for per capita output. This equation relates the wage rate to the proportion of time allocated to R&D at the economy-wide level, μ : because there are decreasing marginal returns to labour (equation 7.6), the more labour in the economy is deviated away from production to R&D, the higher will be the wage rate, and henceforth the higher the opportunity cost of research. With this mechanism, the equilibrium level of R&D will be in general an interior solution.

Monopoly profits are equal to (7.18) in equilibrium. Multiplying and dividing both sides by population, we get:

$$\pi_1 = \beta(1-\beta)y_1 \left(\frac{N}{m} \right) \quad (7.18a)$$

Per capita income evolves according to (7.8). Since we are assuming that the number of varieties increases proportionally to the size of population, the only source of per capita

¹⁴⁷ Also note that, because of the stochastic nature of innovations, the period of time between two successive innovations in this model has a random length.

income growth is λ , that is $y_1 = (\lambda_1/\lambda_0)^{1-\beta} y_0$. Using this to solve together (7.18a) and (7.17a) and substituting in (7.23a), the equilibrium level of R&D becomes:

$$\mu = 1 - \frac{1 + (r/b)(m/N)}{1 + (\lambda_1/\lambda_0)^{1-\beta} \beta / (1-\beta)} \quad (7.24)$$

Equation (7.24) states that the optimal proportion of time devoted to R&D in the economy will be higher, the lower the interest rate, the higher the productivity of R&D, b , and the larger the technological jump, λ_1/λ_0 . It is also apparent that μ is an increasing function of β : the lower the elasticity of the demand curve faced by the intermediate monopolist, the larger the monopoly rents that will be appropriated by successful innovators and hence the larger the incentives to innovate. This accords to the Schumpeter view that market power is good for innovation.

Problems and Exercises

Key concepts

- *Horizontal vs. vertical innovation. Division of labour. Drastic vs. non-drastic innovation. Static vs dynamic efficiency. Limit Pricing. The appropriability Effect. Creative destruction. Business stealing Effect. Neck and neck competition*

Essay questions:

- Comment: “Competition is bad for growth”.
- Comment: “Patents are inefficient. One could banish them and still have incentives for R&D”.

Exercises

- 7.1. (Vertical vs horizontal)** Consider a restaurant, where output (Y) refers to the number of meals produced. The production function is $Y = \sum_j^m x_j^{1/2}$, where x refers to the number of tasks used in the production process. Assume that there are 4 employees (N_Y), each one with productivity equal to $\lambda=9$. (a) Assume first that there was only one job profile ($m=1$), consisting in cooking, collecting the costumers' orders, serving and washing the dishes. (that is, each worker does all the tasks). Find the number of chairs produced by the carpenter. (b) Now consider a technological improvement, consisting in splitting the initial job profile into four specialized tasks ($m=4$), each worker

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

becoming specialized in one task. Explain the impact on production. (c) What would be the impact if the productivity of each worker increased from $\lambda=9$ to $\lambda=25$? (d) Referring to the exercise, explain the difference between vertical innovations and horizontal innovations. Is the example referring to process innovations or to product innovations?

- 7.2. (Drastic, non-drastic)** Consider an economy where the aggregate output is assembled with m intermediate inputs: $Y = B \sum_{j=1}^m x_j^{1-\beta}$. The production function of each intermediate input is given by: $x_j = \lambda N_j$ and the total labour force employed in the

intermediate input sector as a whole is expressed as: $N_y = (1-\mu)N = \sum_{j=1}^m N_j$. μ is the

constant fraction of the labour force devoted to R&D. The wage rate (w) is 50, $\beta=1/3$, $B=100$ and $\lambda=2$. (a) If the final good sector was perfectly competitive, what would be the demand for input j ? (b) If only one producer had the right to produce j , what would be the corresponding price? Represent in a graph. (c) If, in alternative, imitators could produce this variety with a marginal product equal to $\lambda F=1.6$, what would be the equilibrium? Explain, with the help of a graph. (d) Assume now that a firm escaping competition developed a more efficient technique to produce good j ($\lambda=2.5$). Would this innovation be drastic or non-drastic? Explain with the help of a graph.

- 7.3. (Value of the horizontal innovation)** Consider an economy where the demand for each intermediate input is given by $p_j = 20x_j^{-0.5}$. In this economy, technology in each sector is given by $x_j = N_j$, the interest rate is $r=25\%$ and the wage rate is $W=1$. (a) Consider the problem of an entrepreneur trying to discover a new intermediate input, j . (a1) How do you classify this innovation? (a2) What mechanism does this type of innovation influence per capita income? (a3) If the innovator becomes a monopolist in this sector, what will be the optimal amount and price? (a4) and profit? (a5) Represent graphically the welfare gain associated with this innovation, in the case of monopoly. (a6) Identify in the figure the additional welfare gain after the monopoly is eliminated. (b) Consider the entrepreneur's problem again, but before starting the R&D activity. Analyze the incentives for innovation, knowing that the activity involves an initial fixed cost of $F=80$ and a probability of success at the end of the year equal to $b=50\%$, knowing that the patent would last: (b1) eternally? (b2) one year? (b3) two years? (b4) What is the ideal duration of a patent (in years), taking into account the static and dynamic benefits? (b5) Explain.

- 7.4. (NPV of vertical innovation)** Consider an economy where demand for each intermediate product is given by $p_j = 100x_j^{-0.5}$. Initially the market for this product is competitive, being $\lambda_F = 0.5$. It is also known that the interest rate is $r=5\%$ and that initially the salary in this economy is $W=1$. (a) Consider the problem of an entrepreneur who has discovered a new way to produce this good, given by $x=2N$. (a1) How do you classify this type of innovation? If the innovator becomes a monopolist in this sector, what will be (a2) the optimal production; (a3) the price? (a4) profit? (a5) Will this innovation be drastic or not drastic? (a6) Represent graphically the welfare gain resulting from this innovation, under monopoly. (a7) Identify in the figure the additional welfare gain when the monopoly is eliminated. $x = 2N$ (b) Assume now that the entrepreneur was still thinking in inventing this product. Analyze the decision of engaging or not in R&D, taking into account a fixed cost $F=1500$ and the probability of

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

discovery equal to $b=10\%$. In particular, discuss the cases in which: (b1) the innovation becomes immediately available to all competitors; (b2) the probability of arrival of a superior technology each year is 0%; (b3) the probability of arrival of a superior technology each year is 20%. (b4) Which of these situations is more interesting from the firm point of view? Represent in a graph (b5) Which of these situations is more efficient from the dynamic point of view? Discuss, taking into account the different possible effects.

7.5. (Value of horizontal innovation, division of labour) Consider an economy where the final good sector (Y) has the following production function $Y = 80 \sum_j^m x_j^{0.5}$, where m

denotes for the number of intermediate products. In this economy, technology in each sector is given by $x_j = 0.25N_j$, the interest rate is $r=25\%$ and initially $W=1$. (a) Assuming perfect competition in the market for the final product, and defining p_j as the price of intermediate input j . Find out the demand for each input. (b) Consider the problem of an entrepreneur trying to find a new intermediate product, j . (b1) How do you classify this type of innovation? Define it. Assuming that the entrepreneur achieved a monopolist position, find out: (b2) The optimal production: (b3) The optimal price; (b4) The profit (c) (c1) What would be the value of this innovation if the technology became available at zero cost immediately after the innovation? (c2) In that case, what would be the equilibrium price and the quantity demanded? (c3) Represent in a graph, comparing with (b). (c4) From the social point of view, would be more efficient the outcome (c) or (d)? Discuss. (d) Assume that the entrepreneur still didn't find the technology described in (b). Analyse the decision of trying to invent it, taking into account that the probability of inventing was $b=10\%$, with a certain fixed cost equal to $F=20$. In particular, what will be the optimal decision if: d1) The probability of the profits disappearing after period 1 was 15%. d2) The probability of the profits disappearing after period 1 was one. d3) Which probability would turn the entrepreneur indifferent? d4) In the real life, which factors influence that probability? (f) Return to you results in (b): e1) If in this economy there were $N_Y = 1600$ workers and sectors were all alike, how many sectors would exist in the economy? How much would be per capita in income in that case? e2) All else equal, if the number of sectors increased to $m=25$, what would happen to the number of workers in each sector and to per capita income? e3) What is the effect underlying the productivity change? Explain it (1 line) e4) Find out the demand for labour in each sector as a function of the wage rate. Compute the impact of the increase in m to 25 on the wage rate.

7.6. (Vertical innovation, Competition) Consider the market of an intermediate input, which individual production function and market demand are described by $x_i = \lambda_i N_i$ and $p = 1.5x^{-1/3}$, respectively. Also assume that this market is small relative to the rest of the economy, with $W=1$. Initially, this product is produced under perfect competition, with $p = W/\lambda = 0.75$. (a) Assume that an entrepreneur achieved a technology to produce this good with $\lambda=1.6$. Is this innovation vertical or horizontal? The innovation just described is drastic or non-drastic? Explain the optimal strategy of the entrepreneur and the corresponding profit. Represent in a graph. (b) For this strategy to materialize, which condition shall be verified? Identify real world mechanisms that help this condition to be verified. (c) Assume now that more entrepreneurs were able to achieve the same technology $\lambda=1.6$, and divided equally the

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

implied profits. Discuss the effect of the increasing competition in this market on the incentives to innovate, distinguishing entrepreneurs from the fringe and entrepreneurs already using the leading technology.

- 7.7. (Vertical innovation, neck and neck)** Consider the market of an intermediate input, which individual production function and market demand are described by $x_i = \lambda_i N_i$ and $p = 20x^{-1/2}$. Also assume that this market is small relative to the rest of the economy, with $W=5$. In this market, R&D costs are paid by each firm to their own research departments, but in the form of an *annual royalty* (F). (a) Assume that this market was explored by an incumbent monopolist holding the technology $x = 2N$. What would be its optimal price and operating profit? What would be the net profit, after paying the annual royalty amounting to $F=5$? (b) Suppose that a second entrepreneur managed to join the frontier and share the profits with the incumbent, at the annual cost $F=5$. (b1) would that innovation be profitable from the private point of view? (b2) and from the social point of view? (b3) which externalities are involved in this case? (b3) would be profitable for a third entrepreneur to join the frontier? (c) Suppose that the original incumbent estimated at $F=15$ the (flow) cost of achieving $\lambda = 2.5$. Would this investment be worthwhile? Explain the underlying effect.
- 7.8. (N&N, leakages)** Consider the market of an intermediate input, which total demand is given by $p = 4x^{-1/2}$. Assume that this market operates initially under perfect competition, and that the production function of each individual firm is given by $x_i = \lambda_i N_i$, where $\lambda_F = 1$ stands for the technological level at the fringe. Further assume that this market is small relative to the rest of the economy, with $W=1$. Inventions are exclusive to each firm, but the reward of the research department by the firm consists on annual royalty (F), that is only paid in case the technology generates profits. (a) In the initial equilibrium: (i) price is equal to 1; (ii) the total demand for labor is 16 (iii) total demand is $x=16$; (iv) all the above. (b) Assume that an entrepreneur discovered a way of producing the product with the technology $x = 4N$, and kept it secret. This innovation will be: (i) a drastic product innovation; (ii) a non-drastic product innovation; (iii) a drastic process innovation; (iv) a non-drastic process innovation. (c) In the equilibrium after innovation: (i) the entrepreneur's operational profits will be 16; (ii) production will be $x=16$; (iii) employment will be 64; (iv) none of the above. (d) Assume that the use of the new technology involves an annual royalty equal to $F=24$. This innovation will come at: (i) private gain, social loss; (ii) social loss, private loss; (iii) private gain, social gain; (iv) private loss, social gain. (e) Departing from (g) assume that a second entrepreneur was able to discover technology $\lambda = 4$, running an annual R&D cost equal to $F=3$. Assuming that profits would be equally shared by the two entrepreneurs, this innovation will be: (i) socially efficient; (ii) socially inefficient; (iii) unprofitable from the private point of view; (iv) we can't say. (f) Sticking with the assumption that the R&D cost needed for each entrepreneur to join the frontier was $F=3$ and that profits at the front were equally shared, what would be the maximum number of competitors at the front so that innovating to join it was still interesting? (i) 3; (ii) 5; (iii) 6; (iv) can't say. (g) Now, consider the incentives for an entrepreneur at the frontier spending $F=25$ in the patent for the next vintage, $\lambda = 8$. What would be the critical number of competitors at the frontier so that it became attractive for this entrepreneur to innovate and escape? (i) 1; (ii) 2; (iii) 3; (iv) 4. (h) Returning to g), assume that after technology $\lambda = 4$ was discovered ($t=0$), knowledge started leaking at zero cost to the competitive fringe, so that technology at the fringe

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

λ_F expanded at the pace of $\lambda_{F,t} = \lambda_{F,t-1} + 0.5(\lambda - \lambda_{F,t-1})$. In that case, what would be the right time for the leader to spend $F=25$ in the new patent and adopt the next vintage $\lambda = 8$? (i) $t=3$; (ii) $t=4$; (iii) $t=5$; (iv) $t=5$.

- 7.9. (Value of Vertical, N&N)** Consider an economy where the demand for a particular sector is given by $p = 16x^{-0.5}$. It is also known that the interest rate is $r=25\%$ and that initially the salary in this economy is $W=1$. (a) Assume that initially the market for this product was competitive, being $\lambda_F = 0.5$. a1) Determine the price and quantity produced in this market. (b) Consider the problem of an entrepreneur who has discovered a new way to produce this product given by $x = 0.8N$. (b1) How do you classify this innovation? Define it (1 line) Assuming that the innovator became monopolistic in this sector: (b2) Determine the optimal amount of production: (b3) The optimal price of the monopolist (b4) Will this innovation be drastic or not drastic? (b5) Show in graph. (b6) Find out the operating profit (b7) Was this innovation socially desirable? Identify the possible gains and losses for each group (consumers, producers, entrepreneur). (c) (c1) What would be the value of this innovation if the probability of profits disappearing in the following year was $q=1$? c2) And if the probability of profits disappearing each year is $q=0.25$? Comment, comparing with c1 (1 line). (c3) In real life, what factors influence the probability q ? (d) Now consider the problem of a second entrepreneur who wanted to discover exactly the same technology (joining the frontier), to share the market with the incumbent for only one year. d1) Would such innovation result in social welfare gain? What are the two externalities at stake here? d2) Admit that imitation involved a fixed cost of $F=17.5$ and a probability of success of $b=100\%$. In this case, would it be interesting for the entrepreneur to try the imitation? d3) What if there was already another entrepreneur under equal circumstances in the race? (d4) What do you conclude about the intensity of competition and incentives for innovation? What effect is at stake here? This is the general case?
- 7.10. (Equilibrium R&D partial equilibrium)** Consider an economy where the demand for each intermediate input is given by $p_j = 2x_j^{-0.5}$. In this economy, the interest rate is $r=7.5\%$ and the wage rate is $W=1$. (a) Consider the problem of an entrepreneur trying to discover a new intermediate input, j . (a1) How do you classify this innovation? (a2) Through which mechanism does such an innovation impact on per capita output? (b) Assume that the production function for this new product is expected to be $x_j = N_j$. In case the innovator succeeds and becomes monopolist, how much will be its output and profits? (a4) Represent in a graph. (c) Analyze the decision of engaging or not in R&D, taking into account a fixed cost $F=1$, the probability of discovery $b=0.1$, and the probability of being displaced by a drastic innovation, q equal to: (c1) $q=0$; (c2) $q=0.005$; (c3) $q=0.25$? What would be the corresponding NVP? (d) Finally, assume that $q = b\mu = 0.1\mu$. (d1) explain this assumption; (d2) If fixed costs were the same for all firms, what would be the equilibrium level of R&D in this case? Explain, with the help of a graph.
- 7.11. (Equilibrium R&D)** Consider a product, which production is carried by an incumbent that is monopolist in the product market and price taker in the labour market. The demand curve for this product is given by $p = 2x^{-1/2}$ and the production function is equal to $x = \lambda N_Y$, where $N_Y = (1 - \mu)N$ is the proportion of working time that workers in this sector devote to production. The workers' remaining time is devoted to private

R&D, in an attempt to achieve a vertical innovation and displace the incumbent. When 1 unit of labour is devoted to R&D, the probability of achieving a vertical innovation consisting multiplying the previous λ by four is $b=1\%$. The total working time in this sector is constant and given by $N=5$. (a) Consider first the problem of the incumbent monopolist, who achieved a productivity level (λ) equal to 4. Taking into account that the wage rate (w) is equal to 1, find out the selling price and the production level that maximize the incumbent's profits. Compute these profits and represent the monopolist's optimal solution in a graph. [0.5, 16] (b) Taking now the wage rate and the productivity parameter as unknowns, find out the general expression of the operational profits in this sector. (c) Taking now the wage rate and the productivity parameter as unknowns, find out the general expression of the demand for labour in this industry. Show that the wage rate is a negative function of $(1-\mu)$. What will be the demand for blue collars when $\lambda=4$ and $W=1$? (d) Consider now the problem of a research worker that is trying to discover technology $\lambda=16$. If he succeeded and became monopolist (displacing the incumbent), how much would be his profits? (hint: don't forget the impact of innovation on wages, and consider $(1-\mu)$ as unknown). €Taking into account the probability of achieving a vertical innovation ($b=1\%$) and assuming that the discount rate is equal to $r=7\%$. what proportion of his time should he devote to R&D?

8 The extent of the market

“...I do not mean the Big Push is really the right story of how development takes place (...). What I do mean is that the unconventional themes put forth by the high development theorists – their emphasis on strategic complementarity in investment decisions and on the problem of coordination failure – did in fact identify important possibilities that are neglected in competitive models”. [Paul Krugman]

Learning Goals:

- Understand why economies of scale are a source of multiple equilibrium and of divergence.
- Understand the various arguments underlying the big push idea.
- Understand the implication of economies of scale for international trade and factor mobility.
- Acknowledge the key role of transport costs in shaping the economic geography.
- Understand the fundamental role of geography for economic development.

8.1 Introduction

An argument that has been put forward by many economists is that poor countries are unable to adopt modern technologies due to the small size of their domestic markets. This reasoning has a long tradition in economic thinking. Backing from Adam Smith (1776), economists have been arguing that industrialization involves the use of modern technologies, which are potentially more productive than traditional technologies, but that entail some form of economies of scale. Because the adoption of these technologies requires a minimum level of sales, countries with small domestic markets or with limited access to international trade may fail to industrialize. This reasoning inspired development economists in the 1950s and the 1960s, who defended that governments in poor countries should promote massive investment plans, so as to boost demand and achieve self-sustained industrialization. This idea, coined as “big push” by the Austrian economist Rosenstein-Rodan, inspired many development economists at that time and still inspires today.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

In this chapter, we examine the role of economies of scale and of complementarities as a cause of coordination failures and poverty traps. The chapter does not address technological change in the sense of explaining why new technologies are invented. It however reviews different theories that have been put forward to explain why some countries are able to industrialize while others are not and the potential role of the government in overcoming the existing circularities. Because in this literature the initial conditions play a key role in determining the equilibrium, we take the opportunity to review the argument that geography has played a key historical role in determining the initial conditions.

Section 12.2 reviews the original big push argument. In this model, the number of product varieties is fixed, so the critical question is whether the market is large enough for firms to break even. Section 12.3 introduces a different paradigm, where free entry implies that all firms will exactly break even. This model introduces the relationship between the size of the market and the division of labour. In Section 12.4 addresses the possibility of reverse causation from the division of labour to the extent of the market. Section 12.5 discusses the implication of factor mobility in models with economies of scale, to briefly explain the circularities involved in the so-called New Economic Geography. Section 12.6 reviews the Geography Hypothesis and the institutions vs geography debate. Section 12.7 concludes.

8.2 The Big Push model

Consider an emerging economy, closed to international trade, where a fixed number of consumer products, m , is produced and consumed. This economy is engaged in traditional methods of production, using labour input only, and faces the problem of adopting modern technologies, that were developed somewhere else and are freely available for any entrepreneur in this economy.

In each sector, the adoption of the modern technology involves a fixed cost, that shall be equally spend in all sectors in the economy. The existence of a fixed cost implies that the adoption of the modern technology in each sector will only be profitable if the market is large enough. The circularity arises in that the size of the market depends on the number of sectors that adopt the new technology.

8.2.1 The demand side

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

As for the demand side, let's assume that all households are all equal. Hence, one can simply describe the problem as one of a single consumer maximizing a utility function subject to a budget constraint. The utility function is as follows:

$$U = \sum_j^m \ln x_j \quad (8.1)$$

The budget constraint is given by:

$$Y = \sum_j^m p_j x_j \quad (8.2)$$

Where Y denotes for real (aggregate) income. Because the utility function is additively separable and all products have the same weight, the share of expenditure that will be devoted to each product will be equal. Formally, it is easy to obtain the consumer demand for each product:

$$x_j = \frac{Y/m}{p_j} \quad (8.3)$$

8.2.2 Traditional production

In this economy, the only primary input to production is labour, which is available in a fixed amount, N . Each variety can be produced with one of two technologies, which are equal across sectors: a “traditional” technology with constant returns, and a “modern” technology with increasing returns.

Initially, production takes place using the traditional technology, which is postulated as follows:

$$x_j = N_j, \quad (8.4)$$

where N_j is the amount of labour used in the production of the intermediate input j .

8.2.3 Equilibrium under cottage production

Because in cottage production returns are constant returns to scale, a natural assumption to make is that firms operate under perfect competition. Each small producer

takes the output price (p_j) and the wage rate ($W=1$) as given, and maximizes its profits given (8.4), implying:

$$p_j = W = 1 \quad (8.5)$$

Hence, profits are zero and the only source of income is wage income.

Since the price of all goods is equal to 1, from (8.3) we see that there will be an equal demand for each variety:

$$x_j = \frac{Y}{m}, \quad \forall j = 1, \dots, m, \quad (8.6)$$

Thus, the employment level in each sector will be:

$$N_j = \frac{N}{m} \quad j = 1, \dots, m. \quad (8.7)$$

Since under cottage production there are no profits, total income will be equal to labour income:

$$Y = mWN_j = N \quad (8.8)$$

and per capita income will be equal to $y = Y/N = 1$.

8.2.4 Modern production

In the modern sector, labour productivity is higher than in cottage production:

$$x_j = \lambda N_j, \quad \text{with } \lambda > 1, \quad \forall j = 1, \dots, m. \quad (8.9)$$

You may think modern production as requiring the existence of a factory, which flow cost is assumed equal to F . This fixed cost consists in services that have to be purchased on equal amount from all the sectors in the economy (that is, each modern sector j will buy F/m units of production from all sectors, including its own).

Since production with the modern technology involves a fixed cost, a natural monopoly arises.

An entrepreneur that *escapes competition* by investing in a plant would like to charge an infinite price on its good. The reason is that households spend a fixed amount on each good, irrespectively of the quantity (eq. 3) while labour costs increase linearly with <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

production (eq. 6). Since the entrepreneurs faces the potential competition of cottage producers, however, the best he can do is to set the *limit price* (8.5) and undercut the cottage producers. In other words, adopting the modern technology involves a non-drastic product innovation.

The convenient implication of innovations being non-drastic is that revenues will evolve proportionally to output. The profit function of an entrepreneur engaging in modern production will be:

$$\pi_j = \left(1 - \frac{1}{\lambda}\right)x_j - F = \nu x_j - F \quad (8.10)$$

Where $\nu = 1 - \frac{1}{\lambda}$ stands for the mark-up over wage costs.

8.2.5 Industrialization and the extent of the market

Because of economies of scale, there is a critical level of production above which it pays for the economy to engage in modern technologies.

To see this, first note that when all sectors are engaged in cottage production, total output is given by (8.8). When, in contrast, all sectors are engaged in modern production, total production (value added) will be equal to:

$$Y = mx_j - mF = m(\lambda N_j) - mF = \lambda N - mF \quad (8.11)$$

Due to the fixed cost, under modern production per capita income will be an increasing function of the population size:

$$y = \frac{Y}{N} = \lambda - \frac{mF}{N} \quad (8.12)$$

Hence, it will only pay for the economy as a whole to engage in modern production in case (8.11) exceeds (8.8), that is, if $\lambda N - mF > N$, or

$$N^c > \frac{mF}{\lambda - 1} \quad (8.13)$$

Thus, whether this (closed) economy is better served with modern production or with cottage production depends on the size of its population, N : if the economy was run by a central planner concerned with the amount of consumption goods available for its

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

constituents, she would command all sectors to industrialize if the size of the population exceeded the breakeven level (13), and all sectors to remain traditional otherwise.

8.2.6 *Why the decentralized economy may fail*

We just saw that the modern technology will economize labour (and hence it will be preferable) relative to the cottage technology if condition (13) holds. A different question is whether market forces alone will be *sufficient* to induce industrialization when that would be desirable.

The argument goes back from Rosenstein-Rodan (see box). The argument is that while lack of industrialization may be a consequence of an insufficient market size, an insufficient market size might itself be a consequence of low industrialization.

To understand this, one must de-link the concept of market size from that of population size: if a country market size was defined by the size of its population only, then economies with larger populations, such as China and India, should be more industrialized than, for instance, Germany and France, and this is not the case. In fact, it is not the size of population that matters, but the size of aggregate demand, which in turn is determined by income: a given population size will translate into more or less aggregate demand, depending on the size of population (employment) combined with productivity. An economy with large population and low productivity may not generate income enough to feed the fixed costs of industrialization.

Box 8.1. Rosenstein-Rodan and the shoes factory

The term Big-Push was coined by Rosenstein-Rodan. In his 1943 article, *Problems of industrialization of Eastern and South-Eastern Europe*¹⁴⁸, the author imagined a country in

¹⁴⁸ Roseinstein-Rodan, P., 1943. Problems of industrialization in eastern and south-eastern Europe, *Economic Journal*, 53, 202-211.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

which 20,000 workers are taken from agriculture and put into a modern shoe factory, earning wages higher than their previous income. Rosenstein-Rodan argued that such investment would not enlarge the overall market size, because workers of the shoe factory are not expected to spend all their income in shoes only. Hence, the investment in the shoe factory would be most probably unprofitable in isolation.

It could, however, become profitable if accompanied by simultaneous investments in many other industries: according to the author, a coordinated investment effort, spread over a large number of industries (the “Big Push”), could solve the problem of insufficient demand, because each industry would act as each other’s buyers. In that case, each entrepreneur would find it profitable to incur the fixed costs of industrialization, even if no sector could break-even when industrializing alone.

Roseinstein-Rodan was largely inspired by the Marshall Plan in the Post-WWII Europe and used this idea to call for a large-scale external assistance to Eastern Europe. The argument was echoed by other development economists at that time and became very popular in development economics since then. The theory had however to wait until 1989, to be properly formalized in an economic model, by Murphy, Shleifer and Vishny¹⁴⁹.

8.2.7 The profit function again

In the following analysis, let’s assume that equation (13) holds, so that industrialization would be socially desirable. Our question is whether industrialization will be naturally achieved under *laissez faire*.

To analyse the individual incentives to engage in modern production, let’s consider again the profit function, (8.10). This profit function depends on sales, which shall be equal to households’ demand plus the demand by the other modern sectors in the economy:

¹⁴⁹ Murphy, K., Shleifer, A, Vishny, R., 1989. Industrialization and the big push, Journal o Political Economy 97, 1003-1026 .

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

$$x_j = \frac{Y}{m} + \frac{\tilde{m}}{m} F \quad (8.14)$$

Where $0 \leq \tilde{m} \leq m$ stands for the number of sectors that adopted the modern technology. Total income, on the other hand, will consist in wage payments plus the profits originated in the modern sectors, that is:

$$Y = N + \tilde{m} \pi_j \quad (8.15)$$

Replacing this in the profit function, we get:

$$\pi_j = v \left[N_j + (\tilde{m}/m) \pi_j + (\tilde{m}/m) F \right] - F$$

Solving for π_j , we obtain:

$$\pi_j(\tilde{m}) = \frac{1}{1 - (\tilde{m}/m)v} (vN_j - [1 - v(\tilde{m}/m)]F) \quad (8.16)$$

This function depend positively on the degree of industrialization, $\tilde{m}$: the larger the number of firms that adopt the modern technology, the higher the demand for each particular sector, and hence the higher will be individual incentive to adopt the modern technology.

8.2.8 What do we mean by equilibrium?

For an allocation to be an equilibrium, all agents must be happy with their choices. In other words, an economy starting out with that allocation will remain with that allocation.

Since in this model all entrepreneurs face the same technology and demand conditions, if any finds it profitable to invest in a plant, all will find it profitable. Hence, here are only two possible equilibria: one in which all industries remain traditional (and aggregate demand is low); and one where all industries adopt the modern technology (and aggregate demand is high).

8.2.9 Cottage for sure

Consider first the case where the economy starts out industrialized, that is, with $\tilde{m} = m$. In that case, profits in each industry will be:

$$\pi_j(m) = \frac{1}{1-v} (vN_j - [1-v]F) \quad (8.17)$$

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

If (8.17) is negative, then having a plant will not be profitable for any entrepreneur, even when all others are industrialized. Hence, industrialization will not be profitable at all.

Solving (8.17) for $\pi_j(m) < 0$, we get: $N < mF/(\lambda - 1)$. In other words, if the critical condition (8.13) is not met, the economy will naturally gravitate towards the cottage equilibrium, which is after all the more efficient one in that case. Condition (8.13) is a necessary condition for industrialization.

8.2.10 Industrialization for sure

In the following analysis, let's assume that the necessary equation (8.13) holds, so that industrialization would be socially desirable. Our question is whether industrialization will be naturally achieved under laissez faire.

Consider the case in which all x-producers start out with the traditional technology, that is, $\tilde{m} = 0$. In that case, as we already know, output in each industry will be equal to: $x_j = N/m$, and the demand to each sector will be exactly $x_j = Y/m = N/m$.

In that case, would it pay for any individual entrepreneur to industrialize?_

From (8.16), we see that when $\tilde{m} = 0$, profits will be:

$$\pi_j(0) = vN_j - F \quad (8.18)$$

In case $\pi_j(0) > 0$, this means that it will be profitable for any entrepreneur to adopt the modern technology, even if all others sectors remained traditional. Since all industries are equal, in this case all firms will engage in the new technology. Of course, because profits are a positive function of $\tilde{m}$, profits for all will increase with industrialization, implying that there will be a unique equilibrium with all industries adopting the modern technology.

Solving (19) for $\pi_j(0) > 0$, we see that the sufficient condition for industrialization will be: $vN > mF$, that is, if

$$N > \frac{\lambda}{\lambda - 1} mF = \lambda N^C. \quad (19)$$

The analysis above suggests that the existence of benefits in industrializing (as implied by 13) is a necessary but not sufficient condition for the economy to get industrialized under the laissez faire.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

8.2.11 Multiple equilibrium

The case with multiple equilibrium arises when the necessary condition (8.13) is met, but the sufficient condition (19) is not:

$$N^C < N < \lambda N^C. \quad (8.20)$$

In that case, industrialization will be socially desirable, but market forces will not provide incentives enough for individual entrepreneurs to industrialize. Thus:

- If the economy started out industrialized, the fact that condition (8.13) is met implies that $\pi_j(m) > 0$, meaning that it will be profitable to all entrepreneurs to remain industrialized.
- If however the economy started out under traditional production, then the fact that condition (8.19) does not hold implies that it will not be profitable from any individual entrepreneurs to industrialize.

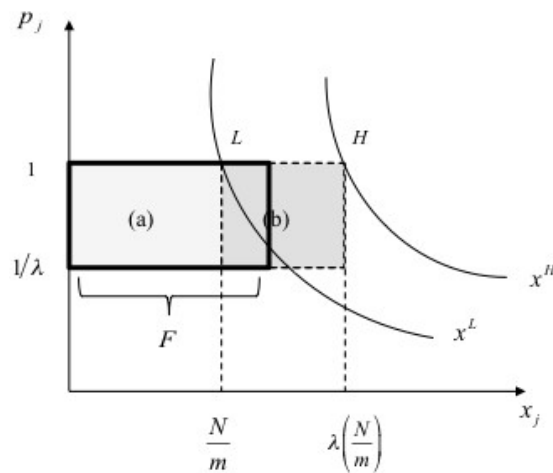
Thus, when (20) holds, the economy will remain under cottage production if it started out under cottage production and will be industrialized if it started out industrialized. Since the equilibrium without industrialization is inferior to the equilibrium with industrialization, it can be interpreted as an underdevelopment trap.

The key aspect in this model is that industrialization by each sector generates an additional demand for all other sectors, in the form of purchases of intermediate inputs. Technically, the coordination failure arises because firms engaged in modern production capture only a fraction of the total benefit of their investment.

8.2.12 Graphical illustration

Figure 1 illustrates the case with multiple equilibrium. The figure displays the market for a given product j , assuming two alternative scenarios regarding the level of demand.

Figure 8.1: Multiple equilibrium in the Big Push Model



Consider first the case in which all sectors start out traditional. Such allocation will be an equilibrium if no entrepreneur find it profitable to set up a factory when all other sectors remain traditional. In the figure, this possibility is illustrated by point L: because all other industries remain traditional, the demand for industry j is low and the operational profits running a plant (shaded area a) are insufficient to cover the fixed cost (F). Since no entrepreneur is breaking even, they will all decide to remain cottage and the economy fails to industrialize.

Now consider the case where all sectors are already modern. This choice will be an equilibrium if all entrepreneurs find it profitable to remain with the modern technology. In the figure, this possibility is illustrated by point H: because all other industries are modern, the demand for industry j is high, so the operating profits running a plant (the shaded areas a+b) are more than the fixed cost (F).

In sum, for both L and H to be equilibria, one needs the fixed costs F to lie between the areas a and a+b in Figure 1. Formally, the following condition must hold: $(1 - 1/\lambda)(N/m) < F < (1 - 1/\lambda)\lambda(N/m)$. Solving for N , you get exactly condition (20).

If condition (8.19) is not satisfied, there will be a unique equilibrium, with industrialization or with cottage production, depending on the case.

8.2.13 The coordination failure

The Big Push model is about a coordination failure: when the level of economic activity is too low, no individual firm finds it profitable to invest, because other complementary investments are not made. If however a sufficiently large number of firms

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

invested at the same time, each investment by each one firm would expand the market of all others firms, allowing them to break-even and making industrialization self-sustained.

The problem is that in a decentralized economy no such coordination occurs: because each single investment is too small to influence the size of the market, no individual firm will find it profitable to invest unilaterally and the economy remains in the trap. Eventually, the government could solve the coordination problem by convincing a large number of players to invest at the same time

8.2.14 Expectations

The model just described suggests an important role for expectations in solving the coordination failure. Suppose you belong to a society stuck in equilibrium L. If, at a certain moment, everybody expected everyone else to invest, it could become individually worthwhile to invest because the new investment would be matched with the higher market size resulting from everyone else's investment. So expectations can move the economy out of the trap. However, each entrepreneur will not invest if he were to believe that others would not invest. In both cases, expectations are self-fulfilled.

This discussion suggests that the government does not need to take a direct intervention to get the economy out of the trap: what he needs is to convince firms to invest simultaneously.

8.2.15 Infrastructures

Another source of economies of scale are infrastructures, such as ports, railroads, power supply, and training facilities. These infrastructures involve in general large fixed costs. Hence, a minimum demand from potential users is required for these infrastructures to be profitable. Assuming that each potential user shares the fixed cost of the infrastructure, the larger the number of potential users, the higher the probability of a given location being served with the infrastructure. This in turn, may induce the establishment of more users in that location, in a virtuous cycle.

Although in principle industrialization and infrastructures go along, coordination failures may prevent essential infrastructures from springing up. Indeed, as long as there is the possibility of a "bad equilibrium" with no industrialization (such as point L), a risk <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

adverse infrastructure builder may prefer not to build if he is not sure on whether the economy will actually industrialize. Thus, the infrastructure may not be built - and hence industrialization will not take place - even if the conditions existed for industrialization.

The failure of an efficient infrastructure to be provided suggests a scope for government intervention: however, subsidizing the infrastructure may not be sufficient: if the economy does not industrialize, then there will be no users and the infra-structure will become a classic “white elephant”. Hence, a coordinated move towards industrialization may require *both* the subsidy to the infra-structure and a coordinated investment in modern factories¹⁵⁰.

Again, the argument is circular and presumes the existence of multiple equilibria: an economy under cottage production (bad equilibrium) would be better off moving to factory production (good equilibrium) but some kind of coordination failure prevents entrepreneurs from making such a move, and the economy gets trapped in the inferior equilibrium.

Box 8.1: Strategic Complementarities

A strategic complementarity occurs when, if one single agent takes some action, it becomes more profitable for another agent to take a related action. Strategic complementarities imply that individuals have incentives to do what the others are doing.

Complementarities belong to the general class of externalities. However, they are not the same. Positive and negative externalities refer to the case in which individual actions impact positively or negatively on the welfare of others. This does not necessarily induce others to take a similar action or a complementary activity. For instance, when one individual issues pollution, this does not necessarily induce the others to issue pollution. This is a simple externality. In alternative, if someone builds a railroad ending in a beautiful beach, this may

¹⁵⁰ Murphy et al, (1989). The novelty of the case with an infrastructure is that it does not require the economy to be closed.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

induce an entrepreneur to build a hotel. Building a road and building the hotel are complementary actions.

Box 8.2 Paul Krugman and the “The evolution of ignorance”

The Nobel Laureate Paul Krugman argued that economists tend to disregard what they don't know how to model, and that this may create “blind spots”.

To illustrate the argument, he remembered how European maps of the African continent evolved from the 15th to the 19th centuries:

“You might have supposed that the process would have been more or less linear: as European knowledge of the continent advanced, the maps would have shown both increasing accuracy and increasing levels of detail. But that's not what happened. In the 15th century, maps of Africa were, of course, quite inaccurate about distances, coastlines, and so on. They did, however, contain quite a lot of information about the interior, based essentially on second- or third-hand travelers' reports. Thus the maps showed Timbuktu, the River Niger, and so forth. Admittedly, they also contained quite a lot of untrue information, like regions inhabited by men with their mouths in their stomachs. (...). Over time, the art of mapmaking and the quality of information used to make maps got steadily better. The coastline of Africa was first explored, then plotted with growing accuracy (...). On the other hand, the interior emptied out. The weird mythical creatures were gone, but so were the real cities and rivers. In a way, Europeans had become more ignorant about Africa than they had been before. It should be obvious what happened: the improvement in the art of mapmaking raised the standard for what was considered valid data. (...). Only features of the landscape that had been visited by reliable informants equipped with sextants and compasses now qualified. And so (...) there was an extended period in which improved technique actually led to some loss in knowledge”. (...)

Krugman argued that something similar happened in economics. The ideas of complementarity, circular causation and poverty traps were patent in the work of the founders of Development Economics in the 1950s and 1960s, like Myrdal, Hirschman, Rosenstein-Rodan, and Nurkse. In the decades after, however, the economic science became progressively more orthodox, relying more and more on formal models. And a problem arose in that nobody knew how to incorporate *internal* economies of scale in a general equilibrium framework: if bigger firms face lower costs, then there should be a tendency for firms to get <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

bigger and bigger until capturing all the market and this is not what we see in reality. This unconformity with real life lead mainstream economics to abandon increasing returns in formal models, dedicating more attention to the case with perfect competition, because this was the model economists knew how to build.

It was only when Avinash Dixit and Joseph Stiglitz came along with their seminal article showing how to incorporate the Chamberlain' monopolistic competition model in a general equilibrium framework, in 1977, that economies of scale returned to the top of the research agenda. The main innovation of the Dixit-Stiglitz model is that it introduces a mechanism offsetting the tendency towards industry concentration: the taste for variety. The Dixit-Stiglitz model revolutionized the economic theory in a number of branches, including industrial organization, international trade, economic geography, economic growth and business cycles.

8.3 The extent of the market and the division of labour

By now, we have been assuming that the number of intermediate inputs (m) in the economy is fixed. The implication of this assumption is that monopoly profits are not dissipated by free entry. In the remaining of this chapter, let's assume instead that positive profits in the intermediate input sector create the incentive for new firms to enter in the market, bringing new varieties. Thus, the "division of labour" will be driven by the size of the market.

8.3.1 Expanding varieties

Consider a closed economy, where the only primary input is labour, which total supply is fixed and equal to N . Output (Y) is produced using an input (X) which, in turn, is assembled with m intermediate inputs:

$$Y = X^{\frac{1}{1-\beta}} \quad (8.21)$$

$$X = \sum_i^m x_i^{1-\beta} \quad \text{with } \beta < 1. \quad (8.22)$$

This technology exhibits constant returns to scale, because increasing the use of all intermediate inputs in the same proportion leads to a proportional increase in Y ¹⁵¹. Each intermediate input is produced using labour, only. There is a fixed cost (F units of labour) and a marginal cost of production. The corresponding production function (also equal across sectors j) is:

$$N_j = F + \frac{x_j}{\lambda}, \quad \text{with } \lambda > 1. \quad (8.23)$$

As before, it is assumed that a competitive fringe using technology (8.4) forces the limit price (8.5). Thus, profits in each sector are given by (8.10). Free entry implies that profits are zero each moment in time. Imposing this on (8.10) and solving for x , we obtain the level of output in each industry:

$$x_j = \frac{F}{1-1/\lambda} \quad (8.24)$$

Using (8.24) in (8.23), we verify that the total labour use in each variety will be exactly:

$$N_j = \frac{F}{1-1/\lambda} \quad (8.24)$$

These equations state that production in each sector will be exactly the break even. Contrasting to the model with a fixed number of varieties, in this model the expansion of the size of the market (as captured by the labour force) does not translate into more production in

¹⁵¹ Another important property is that the marginal product of each variety increases with the number of varieties. Also note that the direct partial elasticity of substitution between every pair of varieties is equal to $1/\beta$. The restriction $\beta < 1$ implies that no intermediate input is essential to production.

each sector and lower unit costs. Instead, production in each sector remains constant (point Q), while the number of varieties increases. Thus, in this model:

$$x_j = N_j = N/m \quad (8.25)$$

The (endogenous) number of varieties can be found using (8.25) in (8.24), to obtain:

$$m = \frac{N}{F} \left(1 - \frac{1}{\lambda} \right) \quad (8.26)$$

This equation establishes an important link between the division of labour and the size of the market, as captured by the size of population: a larger population rises insipiently the incumbents' profits, inducing free entry. Stating in the other way around, the number of intermediate inputs (the division of labour) is determined (*limited*) by the extent of the market¹⁵².

In the final good sector, Y-producers operate under perfect competition, so they take both the output price p_Y and the price of each intermediate input p_j as given. Profit maximization in the final good sector then implies the following demand for each intermediate input j :

$$x_j = \left(\frac{p_Y}{p_j} \right)^{1/\beta} Y, \quad j=1, \dots, m. \quad (8.27)$$

Since all prices are equal to 1, $x_j = x, \forall j$. Using this in (8.21) and (8.22), the, production for final goods becomes:

$$Y = m^{\frac{1}{1-\beta}} x. \quad (8.28)$$

¹⁵² You may remember from the discussion in chapter 7, that a link between horizontal innovations and the population size is how Schumpeterian growth models remove the scale effect. In fact, you may interpret this model as complementing the model described in Chapter 7. That is, while the model in Chapter 7 basically addresses vertical innovations to the R&D effort, this model offers a complementary to understand what drives horizontal innovations.

This equations imply that a larger output in each variety x leads to a larger aggregate income and that this feeds-back to the demand for each variety. Thus, there is much to gain if intermediate inputs are produced with the modern technology¹⁵³.

From (8.27) and (8.28), one obtains an expression for the price of the final good:

$$p_Y = m^{\frac{-\beta}{1-\beta}}. \quad (8.29)$$

This equation states that an increased availability of intermediate inputs leads to a lower output price, even though the price of each input j remains constant (equal to one). This effect captures the *pecuniary externality* steaming from the division of labour: when a new variety arises, the productivity of all existing varieties increases, so the cost of producing one unit of output decreases. The implication is that a larger population, by inducing an expansion in the number of varieties and by then, a higher productive efficiency, will enjoy lower output prices and therefore higher real wages (remember that the nominal wage is the numeraire in this model).

Finally, (8.28) and (8.25) imply that per capita income in this economy is a positive function of the number of varieties:

$$y = \frac{Y}{N} = m^{\frac{\beta}{1-\beta}} \quad (8.30)$$

This equation shows the link between the division of labour and per capita income. So, the larger an economy is, the better¹⁵⁴. Note that this is just another incarnation of the *weak scale effect* described before: a growing population in this model will produce exogenous growth.

¹⁵³ This illustrates well the following statement by Young (1928), p. 533, 534: “the size of the market is determined and defined by the volume of production. (...) an increase in the supply of one commodity is an increase in the demand for other commodities”.

¹⁵⁴ Note that because all profits are zero, income consists only on wages. So per capita output is equal to real wages.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Box 8.4. Pecuniary externalities

A critical feature of the model with increasing returns is that it accounts for the presence of externalities: the investment by each one firm expands the market for other firms. These externalities are however transmitted through the price mechanism. For this reason, they are called *pecuniary* externalities, and are distinct from the *pure* (technological or Marshallian) externalities that we have used in chapters 6, 10 and 11.

The distinctive feature of pecuniary externalities is that they do not alter the technological relationship between inputs and output (the A) at the firm level. Still, they also impact on the average costs of each firm, creating the incentives for firms to cluster together.

An often referred example of a pecuniary externality is the access to a large pool of specialized inputs. In the real world, some production processes are highly sophisticated and require specialized inputs or specific support services. These are not available everywhere. If an individual firm does not provide enough market to attract these specialized inputs, the only solution is to locate where the required inputs are already available. Having a large pool of specialized inputs nearby is a pecuniary externality for a single firm, because the firm will be able to hire these inputs *at lower costs*. Hence, the externality is mediated by the market mechanism. Conversely, there are incentives for these specialized inputs to move to (or to develop in) areas where more potential employers are located. This brings “cumulative causation”: specific inputs go (or develop) where firms are and firms go where specific inputs are.

Some authors have argued that what we normally assume to be technological externalities may be in fact a pecuniary externality. For instance, it may be that most “knowledge spillovers” taking place across firms in a given location occur through inter-firm labour mobility (that is, the leaking knowledge is embodied in workers that switch between

jobs), rather than through demonstration effects or occasional face-to-face contacts. Thus, what may appear to be a pure knowledge externality may in fact be mediated by the labour market¹⁵⁵.

8.3.2 The new trade theory

Long ago, Adam Smith argued that a main advantage of international trade is that, by enlarging the size of the market, allows firms to take opportunity of the division of labour, achieving higher productive efficiency. This theory was first formalized by Paul Krugman, in its 1979 seminal paper.

To illustrate the argument in terms of the model in this section, suppose that, instead of one economy, there are two economies, say East and West. These economies have equal technologies to produce the same final good, but may differ in terms of population (N^E and N^W , respectively). The question is: what happens if the two economies were initially isolated from each other and then became able to trade freely intermediate goods?

To analyse this question, let's assume that labour is immobile between economies. In autarky, the number of varieties in each region is determined by the corresponding labour force. From equation (8.26), the number of varieties in the East and the West will be, respectively, $m^E = N^E(\lambda - 1)/\lambda F$ and $m^W = N^W(\lambda - 1)/\lambda F$. Without trade, the larger economy will produce a wider range of varieties than the smaller region and therefore will enjoy a higher level of per capita output (equation 8.30).

With trade (and assuming away transport costs), the same set of intermediate inputs will be available in both regions at the same price. This means that a final good firm operating in an open economy will be able to use a higher number of varieties (and hence will achieve higher productive efficiency) than in a close economy.

¹⁵⁵ Breshi and Lissoni (2001): “the rationale for co-localization may have less to do with knowledge spillovers mediated by physical proximity, than with the need to access a pool of skilled workers and to establish transaction-intensive relationships with suppliers and customers”.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

From (8.30), per capita income in the free trade area will be equal to:

$$y = \frac{Y}{N} = (m^W + m^E)^{\frac{\beta}{1-\beta}}. \quad (8.31)$$

This is more than each country can achieve in autarky.

In this model, the gains from trade do not arise from differences in technology or in endowments giving rise to comparative advantages. Trade and gains from trade may occur even if the two economies are exactly equal. The reason is that they are realized through the enlargement of the market and the division of labour. Moreover, because the smaller economy has a lower output per capita in autarky, it is the one that has more to gain with trade openness.

This model with trade also sheds some light on the apparent unconformity of the weak scale effect implied by (8.28) with the real world facts: the *scale effect* suggests that countries with large populations, like India and China, should enjoy higher levels of per capita income than countries with smaller populations, like UK and France. However, according to equation (8.30), it is not a country's population that matters, but rather the *size of the market*. Because UK and France are well integrated in the world economy, they may well enjoy larger markets and higher levels of specialization than countries like China and India, less integrated in the world economy and with segmented internal markets¹⁵⁶.

Some authors have argued that the Industrial Revolution took place when a series of reductions in trade costs (between some British regions, first, and then between Britain and other countries) allowed the size of previously autarkic regions to be combined, enlarging the market and allowing these regions to expand through the division of labour¹⁵⁷.

¹⁵⁶ Matsuyama (1992, p. 323), "...a large country does not necessarily mean a large economy. It may simply consist of a large number of regional economies".

¹⁵⁷ See, for instance, Ventura (2005).

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

8.4 The division of labour and the extent of the market

In industrial countries, firms use roundabout production methods, making use of many different specialized intermediate inputs, including machinery and producer services. Many developing countries, however, have not explored the benefits of the division of labour. In poor countries, producers tend to rely on traditional technologies, that are more intensive in “raw” labour, with little use of specialized inputs. A question that naturally arises is why don’t developing countries produce more intermediate inputs and adopt more indirect methods of production?

This section will explore two reasons. First, in a closed economy, an insufficient demand for the final good may prevent intermediate inputs from springing up, which in turn will deliver low productivity in the final product and low demand, in a vicious cycle. Second, in an open economy, comparative advantages may dictate a specialization pattern that is less favourable to the development of intermediate inputs and the benefits of the division of labour.

8.4.1 Vertical complementarities and circular causation

An explanation for why poor countries explore less the division of labour than rich countries follows directly from the theory of circular causation formulated by Allyn Young (Box 12.5): In poor countries, the low demand for final goods implies that the economy will produce only a small number of intermediate inputs (the division of labour is limited by the extent of the market). The lack of local support industries leads to the adoption of relatively simple (labour intensive) methods of production in downstream industries. Low productivity in the final goods sector, in turn, imply a low demand for intermediate inputs (the extent of the market is limited by the division of labour). Thus, an economy that inherits a narrow range of intermediate inputs may find itself trapped into a lower stage of economic development, with limited incentives for new varieties to spring.

To see this formally, consider a model similar to that presented in Section 12.3, with two novelties. First, assume that both raw labour and intermediate inputs can be used in final good production. That is, the production of the final good is given by:

$$Z = F(X, N_z), \quad (8.32)$$

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

where Z refers to the final good, X is defined as in (8.22) and N_Z is the amount of raw labour used in the production of the final good (note that in this version of the model, the labour market equilibrium condition is given by $N = N_Z + mN_j$). Second, assume that the elasticity of substitution between raw labour and X is greater than one.

In this case, it can be proved that there is a critical number of varieties (m) below which it pays more for a firm to rely on raw labour than on intermediate inputs¹⁵⁸. The reason is that, with a low number of varieties, production efficiency will be low and so will be real wages. The price of X , in turn, will be high (remember equation 8.29). Hence, firms will optimally use raw labour intensively. Thus, if the economy inherits fewer varieties than the critical level, it will tend to use production techniques more intensive in raw labour.

If, however the economy achieved a critical mass in support industries, a virtuous cycle would take place: as the number of varieties increased, the relative price of X would decline and real wages would increase, inducing producers to substitute raw labour for intermediate inputs, adopting more indirect methods of production. This movement increases the size of the market for new varieties, and the economy achieves a higher division of labour and higher living standards.

The model explains why in developing countries the lack of local support industries induces the use of relatively simple production methods in downstream industries and why such situation is self-sustained. Like the Big-Push model, this model exhibits multiple equilibrium: a good equilibrium with a wide range of intermediate inputs available and high productivity of labour and wages, and a bad equilibrium with low wages and production methods more intensive in raw labour. In this model, however, pecuniary externalities arise through factor substitution, not by income effects, as in the Big-Push model.

Circular causation in this model also differs from that of the Big Push model in that complementarities operate between upstream and downstream industries. This is called

¹⁵⁸ Ciccone and Matsuyama (1996).

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

“vertical complementarity”. The model of Section 12.2, is of “horizontal complementarities”, because introducing modern methods in one industry enhances the profitability of investment activities of other industries operating at the same level.

Box 8.5. The division of labour, from Adam Smith to Allyn Young

In his masterpiece, Adam Smith contented that the move from traditional agriculture to manufactures entails an efficiency gain, that arises through the splitting of productive operations into smaller and more specialized operations. Adam Smith coined this process as the “division of labour”.

The division of labour improves productivity by different reasons. On one hand, specialization avoids the time it takes a worker to switch from one task to another and allows workers to practice and perfect a particular skill. On the other hand, the division of labour stimulates innovation, as workers engaged in specialized routine operations come to see “better ways of accomplishing the same results”¹⁵⁹.

In his reasoning, Smith was mostly concerned with the division of labour across different tasks within a firm. Allyn Young, in its 1928 seminal article, extended the idea. Young contended that an important mechanism through which the division of labour materialises is through the introduction of new intermediate inputs. That is, splitting production processes into a succession of simpler tasks typically involves the use of machinery, which provision leads to the specialization of industries. For instance, producing a chair can be done more efficiently if a blacksmith provides intermediate inputs like hammers and nails. These inputs in turn can be made more efficiently with machinery designed specifically to produce them and so on.

¹⁵⁹ Smith (1776), Book 1, Ch 1: “Men are much more likely to discover easier and readier methods of attaining any object when the whole attention of their minds is directed towards that single object than when it is dissipated among a great variety of things”.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Setting-up a factory to produce the intermediate input may not be profitable however, in the presence of fixed costs¹⁶⁰. Hence, the proposition: the larger the extent of the market (in upstream industries), the greater the scope for the division of labour (the springing up of downstream industries). This is the type of effect we have analysed so far.

Another novelty in the Young contribution is that he added a two-way causality: “*the division of labour depends upon the extent of the market, but the extent of the market also depends on the division of labour*” (p. 539).

The argument runs like this: when the market is too narrow, it doesn’t pay for a producer of intermediate inputs (e.g. nails) to invest, because he will not sell enough to recover the fixed cost. Therefore, an insufficient demand will prevent a network of supporting industries to spring up. But a narrow range of specialized inputs will also prevent the final goods sector (e.g., chairs) from expanding through the division of labour. That is, the size of the market for upstream industries limits the development of downstream industries, and then the absence of downstream industries limits the extent of the market for upstream industries, in a vicious cycle. This is the argument analysed in section 12.4.

Box 8.6. Backward and forward linkages

The Big Push, as initially advocated by Rosenstein-Rodan, basically consisted on a broadly based investment programme. Albert Hirschman, at the time a professor in Yale, refined the idea, arguing that government intervention to rescue an economy out of a poverty trap should take into account the complex structure of vertical relations in an economy¹⁶¹.

Hirschman distinguished *backward linkages* (which occur when expanding the production of one good rises the demand for an upstream industry enabling it to breakeven) and *forward linkages* (when expanding the production in one sector reduce the costs of potential downstream users of its products pushing them over the threshold). The impact in

¹⁶⁰ Young (1928), p. 530: “it would be wasteful to make a hammer to drive a single nail”.

¹⁶¹ Hirschman (1958).

existing industries induced by a newly established industry would then be given by the sum of its backward linkage effects with the forward linkage effects (the total linkage effect). Note that the story of backward linkages and forward linkages is not different from the circular causation mechanism of Allyn Young: input producers do not invest where there is no downstream demand for them and assembly does not take place where there is no upstream supply. So, there is scope for a Big Push.

Hirshman contended that, in an heterogeneous world, instead of engaging in broad based investment programs, governments should take into account the complex relationship of backward linkages and forward linkages across industries. Thus, the government should first select a few number of key sectors in the economy (that is, those with strong linkages to other sectors) and then tackle other sectors to correct the imbalances generated by the previous investments, and so on. According to the author, a policy of “engineered scarcities and imbalances”, creating shortages and tensions to which the market is expected to respond should provide a sound basis for promoting growth.

8.4.2 Traps and trade

A reason why a country may be found itself locked in a low productivity trap is because of an unfavourable specialization pattern. To see this in terms of the model with pecuniary externalities, let's turn again to the open economy case.

Section 8.3 examined the implications of trade openness in a world with a unique final good and free trade among intermediate inputs. This section considers the opposite case, in which intermediate inputs are non-tradable and trade occurs between two different final goods.

In the real world, many intermediate inputs that are critical for the development of downstream tradable goods have low international mobility¹⁶². For instance, many producer services like banking, auditing, accounting, advertising, engineering, legal supports, wholesale services, transport and communication services, equipment repair and maintenance are mostly non-tradable or have to be supplied near to the final producer. The likelihood of these services to develop in a given economy depends however on the existence of downstream industries, which in turn, may spring or not, depending on the country specialization pattern. This raises the possibility of circular causation between the specialization pattern and the development of upstream industries.

To examine this argument, assume that, instead of one, there are two final goods, Y and Z produced according to (8.2) and (8.30) respectively. The difference between Y and Z is that the production of Y does not use raw labour, while to produce Z some raw labour is required.

Of course, if a central planner could decide the country specialization pattern, he would prefer the country to export Y and to import Z. The reason is that Y uses intermediate goods intensively. Hence, a specialization in Y implies a larger demand for intermediate inputs, leading to a wider availability of intermediate inputs in the domestic economy and hence more production efficiency through the division of labour and higher wages than when the economy is specialized in Z.

Changing the specialization pattern is not, however, a simple stroke of the pen policy: in this model, the specialization pattern and comparative advantages are mutually reinforced. Remember that, as the number of varieties m increases, the price of X declines (equation 8.20). Since Y uses X more intensively than Z, when m increases the cost of producing Y decreases relatively more than the cost of producing Z. Hence, a country that starts out specialized in Y, will enjoy a relatively lower cost of producing Y. A country that starts out

¹⁶² Porter (1992) argued the domestic presence of suppliers is an important determinant of the comparative advantage of nations. The following argument draws on Rodriguez-Clare (1996), and Rodrik (1996).

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

producing Z, in turn, will support a smaller number of varieties and is not likely to develop comparative advantage in Y. Such country will be locked in a low-level equilibrium trap.

This does not mean that trade openness is bad for the country with comparative advantage in Z: in static terms, the economy may be better off in the low-level equilibrium trap with trade than without trade at all. There is however a dynamic argument for restricting international trade: by forcing a country to produce both goods, a temporary import restriction could induce an expansion in the range of available non-tradable intermediate inputs: eventually, if a critical mass of these intermediate inputs was achieved that way, the economy could escape the trap and become specialized in Y, after openness. This reasoning is no more than another incarnation of the *infant-industry argument*^{163 164}.

8.4.3 Multinationals and linkages

Most governments in the world spend large amounts of resources trying to attract FDI by multinational firms. One of the reasons to do so is the belief that multinationals may create important linkages in the host economy. These linkages are easy to describe in terms of the model discussed in this section. In that model, there is no trade of intermediate inputs and there are two tradable final goods, which differ in terms of their intermediate inputs requirements. Depending on initial conditions, a country may be stuck in a low-level equilibrium trap, exporting the good that is less intensive in intermediate inputs and hence

¹⁶³ A model exploring this avenue is Rodrik (1996). Remember that a similar argument was already made in the context of the learning by doing model, with technological externalities.

¹⁶⁴ Trindade (2005) extends the analysis allowing for trade in intermediate inputs too. This reinforces the potential vertical complementarities. The author examined a case where trade openness allows the poor country (South) to become more competitive and specialize in intermediate inputs production. If this allows the South to produce a critical mass of intermediate inputs, then final good assemblers will find it profitable to move to the South, too.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

taking low opportunity of the division of labour. In such a context, multinationals may have a role in tilting the economy from one equilibrium to the other¹⁶⁵.

Box 8.7. The Big push argument revived

During the second half of the twentieth century, the Big Push argument lost a lot of its initial popularity. The view that governments may have a role in shaping the production structure of economies by electing and promoting particular industries was much discredited.

A basic argument is that governments lack the necessary knowledge about the economy to design an appropriate balance between investments in different sectors. On the other hand, planners and bureaucrats may lack the incentives to implement the policy without making things even worse. A reason is that, once a government starts providing support to particular industries, the incentives structure changes: it pays for entrepreneurs to spend resources trying to influence the political decisions.

In the 1990s, the disappointment with state-led industrialization and the collapse of centrally planned economies created the sense that huge government interventions are more likely to lead to waste of resources than to sustained growth. Thus, in light of the Washington Consensus, governments should disengage from policies that target particular sectors, and provide instead broad-based support to all activities in a *sector neutral* way. This includes the provision of public goods, sound money, openness to trade, the rule of law and protection of property rights.

¹⁶⁵ This question was analysed by Rodriguez-Clare (1996a). The key assumption in their model is that multinationals have the possibility of establishing plants in the poor country (thus benefiting from the lower wages there) while using intermediate inputs produced at home (thus overcoming the insufficient supply of upstream industries in the host country). For instance, accountancy and engineering services may be supplied by the multinational's headquarters, while production takes place in a foreign country. Communications costs between the headquarter and the factory imply however some efficiency loss. Hence, there will be some incentive for intermediate inputs to start springing in the host country. If these positive linkages pushed the number of varieties in the host country above a threshold, this could trigger a specialization in the more complex final product, shifting the economy from the bad equilibrium to the good equilibrium.¹⁶⁵ In related seminal paper, Markusen and Venables (1999) add the effect of multinationals on domestic competition, rising the possibility of FDI having a negative effect.

In the last decade, however, the Big-Push argument was brought back to the development agenda. Most notably, the 2005 Millenium Development Project Report (United Nations, 2005) says: "The key to escape the poverty trap is to raise the economy's capital stock to the point where the downward spiral ends and self-sustaining economic growth takes place. This requires a big-push of basic investments between now and 2015 in public administration, human capital (nutrition, health, education), and key infrastructure (roads, electricity, ports, water and sanitation, accessible land for affordable housing, environmental management)"(p.18).

Supporters of the Big Push contend that this phenomenon fits well in many historical episodes. This includes the industrialization of continental Europe in the nineteenth century (e.g, France, Germany), and the recent Southeast Asian' growth miracles (South Korea, Taiwan), that relied heavily on government intervention to catch up. Leading advocates of this view include Dani Rodrik and Jeffrey Sachs¹⁶⁶.

The Big Push idea remains, however, very controversial. Many economists remain suspicious about the ability of governments to implement successful industrial policies. William Easterly, for instance, argued recently that: "Implementing the plan requires the design of proper incentives which may not be an easy task in corrupt bureaucracies". (...) "the recent stagnation of the poorest countries appears to have more to do with awful government than with a poverty trap"¹⁶⁷.

Box 8.8. Doomed to choose

Advocates of the Washington Consensus claim that governments should abstain from selecting particular industries and should instead focus on broad-based sector-neutral policies, such as the rule of law, macroeconomic stability and provision general infrastructure. To this view, Ricardo Hausmann and Dani Rodrik responded that industrial

¹⁶⁶ Rodrik (2005). Sachs et al (2004), Sachs (2005).

¹⁶⁷ Easterly (2006).

policy is unavoidable: in the real world, it is impossible for governments to be sector neutral¹⁶⁸.

The reasoning is as follows. The adoption of many new technologies depends on the provision of critical complementary inputs by the government. This includes specific pieces of regulation and specific infrastructures that serve a narrow range of activities only. For instance, accreditations, safety rules, pollution restrictions, regulation to clarify roles and responsibilities, can be very industry specific. Since some of these inputs are hard to design, providing them involves significant costs for the government. And yet, absence of these critical ingredients will deter investors, either because private returns without these complementary investments are too small or simply because the risks of an adverse regulation in the future are too high.

A problem arises in that government resources are limited. Even though governments have ministries of agriculture, industry and energy, and specialized agencies to regulate food safety, financial markets, professional accreditations, and so on, they have not the resources nor the technical capacity to fix all the standards, to generate all the regulatory pieces and to address all the infrastructures needed to accommodate any potential new activity. Hence, Hausmann and Rodrik contend that public policy cannot be sector neutral: by deciding to provide pieces of regulation that are specific to some activities and useless to others, governments will end up benefiting differentially different economic activities. Governments are “doomed to choose”, they conclude.

8.5 Centripetal and centrifugal forces

8.5.1 International factor mobility

¹⁶⁸ Hausmann and Rodrick (2006).

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

It has long been recognized that factor mobility and trade may act as substitutes for one another. The same is true in the presence of economies of scale. The difference is that, with factor mobility, the economies of scale are realized through a process of agglomeration.

To see this, let's return to the trade model introduced in Section 12.3. Assume however that transport costs are prohibitive, so there is no international trade. Labour, however, is free to migrate across economies.

If the two economies were of equal size, then, with all the rest equal, real wages would be equal in the two regions. Hence there would be no incentive for labour to migrate.

Now assume that one of the regions experiments a small increase in the number of its inhabitants (for instance, $N^W > N^E$). In that case, the larger region will achieve a larger variety of locally produced intermediate inputs. This, in turn, translates into higher productivity and higher real wages, by lowering the price of the final good. Thus, there will be incentives for workers to migrate from the smaller region to the larger region. With free labour movements, a process of cumulative causation takes place: as labour moves from the smaller region to the larger region, there is an expansion in the number of varieties in the larger region and a contraction in the number of varieties in the smaller region and a further widening of the real wage gap. The larger region will become richer and richer in a virtuous cycle and the smaller region will shrink and get poorer, in a vicious cycle. In the limit, the whole population will end up in the largest region and the world per capita income will be equal to (8.29).

Note that the agglomeration process may be triggered solely by an initial difference in the market size. When the two regions have the same productivity levels, it does not matter which region ends up with the whole population. But if the two regions differed in terms of technology (say region W has higher fixed and variable costs), it could be that all population moved to that economy, just because it started out with a larger population size. In that case, the migration process would have delivered the undesirable outcome.

Also note that the welfare effects with labour mobility are in sharp contrast to the case with free trade and no factor mobility: with free trade, workers in both regions gain with trade openness and those in the smaller region gain more. With factor mobility and no trade, the workers of the bigger region gain and the workers that remain in the smaller region loose.

8.5.2 The core-periphery model

By now we have assumed extreme assumptions regarding trade openness and factor mobility. A more complicated story arises when one looks at intermediate cases, where some international trade and some labour mobility are allowed.

A particularly interesting case occurs when only some labour is free to migrate across regions and international trade is subject to transport costs¹⁶⁹. Suppose, for instance, that some workers are tied to location-specific activities, such as agriculture or mining. In that case, location decisions for firms have to trade-off the benefits of staying in the larger region (agglomeration) against the cost of being too far from peripheral costumers. This problem was first investigated by the Nobel Laureate Paul Krugman (1991) and resulted in a new literature that became known as the New Economic Geography¹⁷⁰. This section briefly describes the original idea, which became known as the “core-periphery model”.

Assume that there are two regions with identical preferences and two kinds of goods, agricultural and manufactured. Agricultural production is homogeneous and produced under CRS and perfect competition. Manufactures exist in a large number of varieties and are produced under increasing returns and monopolistic competition. There are also two types of workers: “blue collars”, who are free to move to the region offering higher wages, and “farmers”, who are tied to specific locations. Both types of workers demand agriculture goods and manufacture goods and have equal tastes. Agriculture goods are traded costlessly, whereas manufacture goods are subject to transport costs. The geographical distribution of

¹⁶⁹ When both trade and labour mobility are allowed without frictions, the proposition that workers in both regions gain and those in the smaller region gain more is recovered. In that case, however, it is not possible to determine location: any geographical distribution of production is an equilibrium. For an integrated model with international trade, factor mobility and economic growth, see Ventura (2005), pp 1427-1442.

¹⁷⁰ The starting point of this theory is the final section of Krugman's (1979) famous article, where he argues that patterns of labour migration can be analysed within the same framework. One year after, Krugman (1980) extended the model so as to account for the role of transport costs. In that paper, he showed that, with positive transport costs, there is an incentive for production to cluster close to the largest markets, because this allows economies of scale to be realized with minimum transport costs. But it was his seminal 1991 article that launched the New Economic Geography.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

farmers is taken as fixed, with half farmers in each location. The problem is to find out how blue collars, and hence manufactures production, will be allocated between the two regions.

From what we have learned so far, it should be more or less intuitive how the core-periphery model works. In this model, there are two “centripetal” forces and one “centrifugal” force, which are realized through migration of blue collars. The centripetal forces are the desire of firms to locate close to the larger market and the desire of blue collars to have access to a larger number of consumer varieties at low transport costs. The centrifugal force is the incentive of firms to move out to serve the peripheral demand. Depending on the parameters of the model, the spatial equilibrium may be more or less concentrated. Intuitively, the stronger the agglomeration effects, the more concentrated will be the economic activities. In this model, transport costs act as a dispersion force, counteracting the agglomeration effect of economies of scale.

Because the model is non-linear, the interplay between centripetal and centrifugal forces is very complicated and may lead to different equilibria, depending on the parameter values and on initial conditions. Krugman (1991) calibrated the model with specific functional forms and made some simulations, obtaining the following conclusions. First, when transport costs are very high, there is little trade of manufactures among regions. In that case, the agglomeration advantages are outweighed by the need to serve the peripheral markets and the economy converges to an equilibrium where manufacturing is equally split between the two regions. Second, when transport costs are very low, there are no significant costs in serving the peripheral demand, so the agglomeration forces dominate. In that case, if one of the regions starts out with a larger scale it will be a more attractive location for blue collars to work. Thus, a process of agglomeration will take place until only agriculture workers are left in the periphery¹⁷¹. Third, at intermediate transport costs, different equilibria may emerge, depending on the initial conditions.

¹⁷¹ Of course, if the two regions have initially the same size, blue collars wages will be equal and there will be no incentive to migrate. This will be, however, an unstable equilibrium.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

An implication of this model is that a decline in transport costs may lead to divergence: if, starting from a situation where the economic activity is equally split across regions, transport costs fall below a critical level, then it will pay for any single worker to move from one region to the other, so that the latter becomes the larger region. Then, a cumulative process takes place leading to a core-periphery spatial structure.

8.5.3 The Krugman-Venables theory

The core periphery model was extended in different directions. An important case arises when labour is immobile across regions¹⁷². This setup looks more appropriated to discuss centripetal and centrifugal forces at the global level in our days: today most labour mobility takes place within countries, with labour mobility between countries playing only a minor role.

A novelty in the Krugman and Venables' specification is that manufactured varieties can be either used as intermediate inputs to produce other varieties or consumed as final goods. So the model accounts for backward and forward linkages within the manufactures sector. Since in that model labour cannot migrate internationally, agglomeration effects leading to a rise in the demand for labour in the centre translate into higher wages there. This, in turn, acts as a centrifugal force: as wages increase in the centre, there is an incentive for manufactures production to move towards the periphery. The functioning of the model is rather intuitive: when transportation costs are prohibitive, each region will be basically self-sufficient and the economic activity will be equally split across the two regions. As transportation costs fall, there is a tendency for manufactures to cluster in the centre, taking opportunity of the better access to markets and manufactured inputs (linkages), satisfying the peripheral demand through exports. Thus, a process of agglomeration takes place, with one region becoming industrialized and the other region relegated to primary production. As transport costs continue to fall, however, the agglomeration advantages vanish (varieties can

¹⁷² Krugman and Venables (1995), Venables,(1996).

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

easily be imported from abroad). In that case, some manufactures production will eventually move to the periphery (drawing workers from agriculture) so as to benefit from the lower wages there. Thus, when transport costs fall below a critical level, a symmetric equilibrium emerges again, with half of manufactures located in each region.

The model with immobile labour offers an interpretation for the Great Divergence, as well as for its reversal starting in the last decades of the twentieth century: along the last centuries, there has been a steady decline in transport costs. According to the model, as transport costs declined, there should be initially a tendency for manufactures production to cluster in the region that was initially more developed (the “North”). In result, a significant wage (income) gap emerged between North and South. Later, as transport costs continued to fall, the agglomeration advantages lost importance relative to the centrifugal effect of labour costs. According to this interpretation, the world should now be engaged in a process of convergence, with increasing international trade in manufactures and reallocation of manufactures production from countries with high labour costs to countries with low labour costs.

8.6 Discussion

In the neoclassical paradigm, the expansion of one activity comes only at the expense of the others. So in light of that model, the price mechanism assures that the economy is self-adjusting. With increasing returns, in contrast, the expansion of one activity does not necessarily come at the expense of the other: there is room for complementarities. When investments are complementary, the profitability of two different investments depends on each other and there is no market mechanism to assure that both investments will take place. Hence, critical investments fail to occur because other complementary investments are missing and the latter fail too because the former do not happen. This is a coordination failure.

In this chapter, two versions of the model with increasing returns were described. The first assumes that the number of varieties is fixed. In that case, the expansion of output by one firm may lead to an increase in profits by other firms. The second version of the model is that of monopolistic competition. In that model, there is free entry, so profits are driven down to zero in the long run. In both cases, we found that a larger market size impacts positively on per capita income. This, in turn, leads to cumulative causation: either through a decline in average costs of producing each variety or through an expansion in the number of varieties, the more an economy produces, the more attractive it will be for new investment.

Coordination failures and cumulative causation may have a role in explaining why some countries are rich while others remained poor. Lack of industrialization and low division of labour may be self-sustained. Along this reasoning, many authors contend that governments should have an active policy towards industrialization, either through a coordinated effort with domestic entrepreneurs or through temporary import protection. Other authors contend, however, that governments have not the technical ability nor the political strength to implement successful industrial policies.

Economies of scale are a “centripetal force”. In the real world, however, one does not observe the concentration of economic activity in one unique geographical site. So, in order to understand the dispersion of economic activity across the territory, one needs to account as well for the role of “centrifugal forces”. The so-called “New economic geography” has focused on transport costs. According to this theory, transport costs create a tension between

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

the agglomeration advantages originated by economies of scale and the advantage of staying close to peripheral markets.

The chapter illustrates the difficulty in obtaining general propositions in the presence of economies of scale. From model to model, the conclusions differ dramatically with small changes in parameter values or on initial conditions. This is a general feature of economies of scale: sometimes, small changes in the parameters produce small effects, sometimes they trigger a process of cumulative causation that changes the nature of the equilibrium. With increasing returns, it is much more difficult to obtain general propositions than in the case with perfect competition and constant returns.

8.7 Key ideas of chapter 8

- The Big Push argument points to the possibility of coordination failures arising from horizontal complementarities: a single investment may not be profitable even if it would be profitable when coordinated with other investments. In the Big Push model, the size of the market does not necessarily go along with the size of population: the choice of technology matters.
- With free entry, the number of varieties becomes an increasing function of the size of population (horizontal innovations). The implied division of labour gives rise to productivity gains that translate into a positive relationship between per capita income and the population size (weak scale effect). A corollary of this is that openness to international trade, by enlarging the extent of the market, results in productivity gains.
- As the enlargement of the market translates into higher productivity, a higher productivity may feedback on the size of the market. This mutual causation opens a channel through which a poor country that didn't reach a critical stage in the process of division of labour finds itself trapped in a low level equilibrium, with little division of labour and small market size. This is another version of the Big Push argument.
- To the extent that traded goods differ in respect to division of labour potential, the specialization pattern that arises under free trade is not necessarily the one that delivers the highest possible income. In particular, a country specialized in a good that does not favour the division of labour will fail to achieve the implied productivity gains. In this case, it may pay to promote an "infant industry". By the same token, government efforts to attract multinational firms may help trigger the development of upstream industries that in turn will favour the springing up new downstream industries, in a virtuous cycle.
- The Big Push idea inspired many economists to argue that governments should play an active role in industrialization, by coordinating the private investments. Contenders of the Big Push argue, however, that governments lack the knowledge and the appropriate incentives to implement successful industrialization policies.
- When economies of scale are coupled with factor mobility, cumulative causation takes the form of agglomeration economies, whereby mobile factors tend to move

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

from peripheral regions to the centre. Transport costs and immobile inputs may help alleviate this force. The location of economic activities across the space reflects a tension between centripetal forces and centrifugal forces. When labour is immobile across regions, higher wages in the centre act as a centrifugal force.

- A common feature of models with increasing returns is that the equilibrium is very dependent on the initial conditions and on parameter values. Thus, it is very difficult to obtain general propositions.

Problems and Exercises

Key concepts

- *Big push. Pecuniary externalities . Strategic complementarities. Horizontal vs. vertical complementarities. Backward and forward linkages . Coordination failure. The division of labour .*

Essay questions:

- Comment: “the case for big push is much more stringent than that loosely expressed by Rosenstein-Rodan”.
- In light of the big push model, public provision of an essential infrastructure is not a sufficient condition to escape the trap. Explain why.
- Explain: “*The division of labour depends upon the extent of the market, but the extent of the market also depends on the division of labour*”
- Explain how multinationals can help a country escape a low level equilibrium trap.
- Comment: “Industrial policy is not an option: government are doomed to choose”.
- Comment: “According to the core periphery theory”, a fall in transport costs may lead to divergence”.

Exercises

- 8.1. Consider a closed economy where the final demand for each product is equal to $x_j = (Y/m)/p_j$, where Y denotes for income (value added), and $m=100$ refers to the (fixed) number of products. Each product is produced using labour only, according to $x_j = \lambda N_j$. In this economy, the wage rate is $W=1$. There are two possible technologies (equal to all sectors): a traditional one where $\lambda = 1$; and a modern technology, where $\lambda = 2.5$, but a fixed cost $F=1$ must be paid, in the form of equal purchases from all sectors ($1/m$ must be paid to each sector). (a) Assuming that traditional production takes place under perfect competition, find out the equilibrium price in each sector, and the implied profits. (b) Assume that all sectors adopted the new technology. Find out the expressions for income, profits and total demand for each sector as a function of the size of population, N . (c) If population in this economy was equal to $N=40$, would the

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

economy be better served with cottage production or with modern technology? And if population was $N=100$? (d) Let $\tilde{m}$ be the number of sectors that adopt the new technology. Assume that each sector of these sectors operates under monopoly, subject to the limit price determined in (a). Find out the expression relating the profits in sector j as a function of $\tilde{m}$. /e) What would be the equilibrium if population was (e1) $N=300$?; (e2) $N=40$?; (e3) $N=100$? Explain the intuition. (f) Now suppose that the economy started out with traditional production and population was equal to $N=100$. How many sectors should a government convince to invest in order for this economy to escape the poverty trap? Would this effort require the spending of government money?

- 8.2.** Consider a closed economy at time $t=0$, with a constant population equal to $N=1000$, where the final demand for each product is equal to $x_j = (Y/m)/p_j$, where Y denotes for income (value added), and $m=100$ refers to the (fixed) number of products. Each product is produced using labour only, according to $x_j = \lambda N_j$. In this economy, there are two possible technologies (equal to all sectors): a traditional one, where $\lambda = 1$; and a modern technology, where $\lambda = 1.25$, but a fixed cost $F=5$ must be paid, in the form of equal purchases from all sectors ($1/m$ must be paid to each sector). Further assume that labour is the numeraire ($W=1$). (a) Assuming that traditional production takes place under perfect competition: (i) the price of each variety will be equal to 10; (ii) Per capita income will be equal to 10; (iii) national income will be equal to 10 (iv) production of variety j will be equal to 10. (b) Assume that this economy started out with cottage production, only. If an individual entrepreneur adopted the new technology, its profits will be equal to: (i) -3; (ii) 0; (iii) 2; (iv) -5. (c) Assume instead that all sectors adopted the modern technology at the same time. In this case, the profits of each individual entrepreneur will be equal to: (i) 2.5; (ii) 0; (iii) 5; (iv) -2.5. (d) In the following, assume that technology was diffusing freely across borders, and in particular from the industrial world to this small economy, at the pace $\lambda_t = 1.25 + 0.025t$. Further assume that fixed costs of industrialization remain at $F=5$. Also remember that today is $t=0$. € In a laissez fair, what would be the timing of (spontaneous) industrialization in this country? (i) $t=5$; (ii) $t=10$; (iii) $t=20$; (iv) $t=30$. (f) If you were a benevolent planner seeking to eliminate any market failure, when it would be the right timing to induce a coordinated move towards industrialization in this economy?: (i) $t=0$; (ii) $t=5$; (iii) $t=10$; (iv) $t=20$.

- 8.3.** Consider a closed economy, where aggregate output (Y) is obtained using 100 (m) intermediate inputs, according to the following production function: $Y = \left(\sum_{j=1}^{100} x_j^{1/2} \right)^2$.

Each intermediate input can be produced with one of two technologies: (i) a traditional CRS technology that consists in converting one unit of labour input into one unit of intermediate input, $N_j = x_j$; (ii) a modern technology with increasing returns,

$$N_j = F + \frac{x_j}{\lambda}, \text{ where } F=10 \text{ and } \lambda=2. \text{ When technology is traditional, production will be}$$

competitive. When a plant is installed, the entrepreneur becomes monopolist in the market for the corresponding variety. (a) Admitting that the final output sector is perfectly competitive, find out the demand function for each intermediate input. (b) Knowing that intermediate sectors are all alike and that there is free entry in the final

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

output sector, find out the relative price of Y, P_Y/p_j . (c) What is the volume of employment that turns industrialization desirable? (d) Now admit that $N=2500$. Under these conditions, would it be better for the economy as a whole to remain traditional or to get modern? (e) Assume that the blue collar wage (w) is higher than 1. Why should that be? Verify that, in this case, installing a plant is doomed to be a non-drastic innovation. (f) Bearing in mind the answer to (e), compute the quantity demanded for each intermediate input when: All the producers remain traditional sector; All the producers get modern. (g) Find out the profit of an entrepreneur that decided to go modern when: All other producers remain traditional; All other producers are modern. (h) Assume that $w=1.1$. If all other sectors remain traditional, is it worth from an individual entrepreneur in a given sector j to go modern? What if all other sectors were modern? (i) If instead $w=1.6$, would that pay for an entrepreneur to go modern when all other sectors remain cottage? And if all other sectors were modern? (j) And what if $w=1.16$? Repeat the exercise and conclude.

9 Technology adoption

“...it is a matter not of individual inventiveness but of the receptivity of whole societies to innovation”. [Jared Diamond].

Learning Goals:

- Understand the critical role of economic openness in determining a country’ exposure to world-wide technological progress
- Acknowledge how country’ characteristics may influence the pace of adoption of new technologies
- Understand why the adoption of foreign technologies may require deliberate efforts to improve a country’ absorptive capacity
- The extended Solow model with technological catch up

9.1 Introduction

How to improve the state of technology is a policy question that confronts all modern societies. For an industrial country, keeping the lead requires a continuous effort to invent new products and processes. For an emerging economy, however, the issue is not as much of pushing forward the world technological frontier, but mostly to benefit from technological achievements occurred abroad: since the inventing process does not have to be repeated, there is a potential advantage for those who adopt successful technologies without the need to learn from the beginning.

The view that poor countries may improve their living standards by imitating the best practices in rich countries backs from David Hume, but it was popularized by Alexander

Gershenkron, who coined the term “advantage of backwardness”¹⁷³. In reality, however we do not see new technologies flowing automatically from rich countries to poor countries. New technologies have the *potential* to be transferred across agents and country borders, but whether they are implemented or not in each particular environment depends on incentives. These incentives, in turn, differ across the space, depending on the local economic, political, cultural and geographical circumstances. For a laggard country, taking opportunity of the potential generated by technological diffusion is pretty much a question of shaping the country set of capabilities so that it becomes the interest of individuals to invest in new technologies.

This chapter addresses the question of why available technologies do not flow uniformly across the space, and what policymakers can do about it. In Section 9.2, we stress the role of economic openness in determining the exposure of a country to worldwide technological diffusion. In Section 9.3, we discuss how the recipient country’ characteristics may determine its permeability to the world technological diffusion. Section 9.4 turns to the question of heterogeneity in technology, to discuss the costs involved in the selection of the technologies that better match each country’ set of capabilities, and on the eventual need to adapt foreign technologies to fit the country needs. Section 9.5 presents a model of an emerging economy faced with the challenge of adopting technologies developed abroad. In this model, the World technological frontier expands at an exogenous rate, like in the Solow model, and the country’s characteristics and policies determine how close it gets to that frontier. Section 9.6 summarizes the main ideas.

Box 9.1: four breakthrough invention of the human kind

In the neoclassical model, it is assumed that technology spills over instantaneously across firms and countries borders at no cost. In the real life, however, technology does not

¹⁷³ Hume, D., 1758. *Essays and Treaties’ on Several Subjects*. London: A. Millar. Gershenkron, A. , 1952. “Economic backwardness in historical perspective”. In Bert F. Hoselitz (ed.) *The progress of underdeveloped Areas*, Chicago: The University of Chicago Press, 3-29.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

spread instantaneously across the space, people and societies. It instead flows asymmetrically, through specific mechanisms of human interaction. To illustrate this statement, we exemplify with four breakthrough inventions of the Humankind.

The wheel: the wheel was first used for pottery in Mesopotamia by 3.500 B.C, and was adapted three centuries later for transportation on chariots. You may think this invention as almost a public good: once an agent becomes aware of the concept, it will be very easy for him to imitate the invention and use it for own benefit. Nevertheless, it took many centuries for this simple idea to spread around the Middle East, Western Europe, and Asia. In the New World, people had to wait until the XV century before enjoying the benefits of the wheel in transportation. This example suggests that *geographical distance* has a role in determining the pace of technological diffusion.

Making fire: with no question, this technology is easier to hide from imitators than the wheel. Probably, the hominids who first discovered how to make a campfire, around 1.4 million years ago, tried to keep it secret, so as to have an advantage against their competitors. But, either through disclosure or through independent discoveries, the fact is that the ability to make fire became universally known long ago in the human history. This example suggests the *passage of time* has a role in eroding the barriers to technological diffusion.

Writing: this technology was independently developed in Mesopotamia and in Egypt by the third millennial B.C., in China around 2000 B.C., and in Central America around 600 B.C. Writing is a powerful tool that fuels human interactions, but it is a complex technology: it requires considerable individual efforts to be transmitted. Not surprisingly, after writing was invented, it rapidly spread through merchant societies, where the need to register transactions was latent, but it was unable to penetrate in agricultural-based societies, where *economic incentives* to adopt it were absent. Today, even though governments spend large amounts of resources to make this technology universally available, many people do not manage to learn how to read. When knowledge is costly to assimilate, people will only adopt it if they perceive it to be worthwhile.

Democracy: democracy was first implemented in the ancient Greece, in the city-state of Athens, by the year 508 B.C. At that time, democracy was introduced so as to provide peasants engaged in highly productive long-term investments (preparing the fields to cultivate olives) with a political system that minimised the expropriation risk¹⁷⁴. In our days, democracy is still difficult to implement in many environments. Building a democracy depends on collective actions, and on the existence of complementary ingredients, such as the rule of law and free press. Often, the conditions exist to implement democracy, but this is not the interest of those who have the power to decide. This example stresses the fact that lack of *complementary factors* and *vested interests* may delay the pace of technological diffusion.

Taken together, these examples remind us that, although technology has the potential to diffuse across societies, in practice its diffusion is far from automatic. Because of various combinations of geographical distance, secrecy, complementarities, and incentives, technology tends to be differently assimilated across people and societies.

9.2 The key role of openness

9.2.1 A necessary condition

In isolation, it would be impossible for a region to import foreign technologies. Consider, for instance, the Aboriginal Tasmanians before the European discoveries: since they had no contact with other civilizations for more than 10,000 years, they could not acquire any new technology other than what they invented themselves. A necessary condition for a country to learn from abroad is to maintain a minimum contact with the outside world.

¹⁷⁴ Fleck, R., Hansen, A., 2006. The Origins of Democracy: A Model with Application to Ancient Greece The Journal of Law and Economics, 49, 115–146.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

In our days, no country in the world is completely isolated. With the arrival of telecommunications and the internet, people in remote areas are given the opportunity to learn and share ideas with fellows located in the centre. Today, knowledge has the potential to circulate across distant people and societies at a speed without parallel in human history. And yet, technology still differs considerably across countries and country-regions. Openness and access to information are necessary for technology to diffuse, but certainly they are not sufficient.

9.2.2 *The critical role of international trade*

A primary channel of economic interdependence is international trade. There are different mechanisms through which international trade increases an economy's permeability to the world technological diffusion. First, importing equipment from more advanced countries is a direct way of using *embodied* technology, without the need to replicate the research effort. Second, opening the domestic market to the *competition* of foreign firms bringing newer and more sophisticated products compels domestic firms to improve their products and to seek for more efficient ways of producing them. Third, competition in foreign markets provides exporting firms with the discipline of interacting with demanding customers, inducing them to meet high quality standards (*learning by exporting*). Fourth, access to external markets may favour the establishment of new exporting industries, eventually pushed by *foreign direct investment*, that otherwise would not spring in the country. Fifth, a society more exposed to foreign ideas tends to be more demanding in respect to the quality of domestic *policies* and institutions.

Protectionism, in contrast, creates economic rents and therefore the conditions for agents to become organized in interest groups, and to spend resources in pressing the government for more protection, instead of devoting their talents in the search for better technologies. All in all, trade openness plays a key role in shaping the agents' incentives so that it becomes their interest to take opportunity of the world technological progress.

Some authors argued that one reason why the United States of America emerged economically in the 1865-1929 period and surpassed England as the world technological leader is that it became a “free trade club”¹⁷⁵. In this club, members states were not allowed to impose restrictions on imports from (or on technology developed by) other member states. In Europe, in contrast, sovereign states had the legal power to impose barriers to mutual imports, restricting the exposure of countries to each other innovations. It was only after the launch of the European Economic Community, in 1957, that Europe started its move towards a “free trade club”.

Empirically, many studies have supported the claim that openness to international trade speeds up the pace of technological diffusion and helps improves the quality of domestic policies¹⁷⁶. In that literature, some authors have found that it is not only trade openness that matters, but also the *identity* of the trading partner: that is countries trading primarily with technological leaders are likely to benefit more than countries trading primarily with laggard countries.

9.2.3 Foreign Direct Investment

Like international trade, FDI is a vehicle for cross-country technological diffusion. Specific mechanisms through which FDI promotes the transfer of technology include: bringing new machinery and production techniques to the host country; demonstration effects that induce imitation by local firms; increased competition in the domestic market; creation of a demand for high-quality or specific intermediate inputs.

¹⁷⁵ Parente, S. and Prescott, P., 2004. "Barriers to technology adoption and Development", Journal of Political Economy 102(2), 298-321.

¹⁷⁶ Sachs, J. D. and Warner, A. M., 1995. “Economic reform and the process of economic integration”, Brookings Papers of Economic Activity 1, 1-95. Sachs, J. D. and Warner, A. M., 1997. "Fundamental sources of long-run growth". American Economic Review, Papers and Proceedings, May. Bayoumi, T., Coe., D. and Helpman, E., 1999. R&D spillovers and global growth. Journal of International Economics 47, 399-428. Savvides, A., Zachariadis, M., 2005. International technology diffusion and the growth of TFP in the manufacturing sector of developing economies. Review of Development Economics 9 (4). Lichtenberg, F., de La Potterie, B., 1998. International R&D spillovers: a comment. European Economic Review 42, 1483-1491.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

To this, one shall add an important role of foreign investors in promoting face-to-face contacts between workers in the headquarters and in the subsidiaries. These face-to-face contacts are essential to diffuse the so-called *tacit knowledge*, which by nature is not easy to communicate at distance (see Box 6.2).

The empirical evidence has not been, however, very supportive to the idea that FDI by its own generates faster productivity growth. Studies have found an important potential role for FDI, but the extent to which this potential materializes in each country, depends on country characteristics¹⁷⁷. Countries have different characteristics, and these characteristics play a critical role in determining the permeability to the potential diffusion resulting from FDI. In the next section, we turn precisely to the role of country characteristics.

Box 9.2 Trade openness and convergence

The question as to whether trade openness is good or bad for growth has been subject to intensive debate by economists of all times. The general case in models with a widely accepted set of assumptions is that international trade is good for growth. Still, one may find models stressing less common but equally realistic assumptions showing that trade can be detrimental to growth. Models with learning by doing are typically in the second category.

Empirically, most evidence points to the case that trade openness is indeed good for growth. A seminal contribution is from Jeffrey Sachs and Andrew Warner (1995 and 1997). The authors first constructed an “index of trade openness” according to which a country was classified as “open” if it satisfied 5 requirements at the same time¹⁷⁸. Using this index, the

¹⁷⁷ Borenztein, E., De Gregorio, J., Lee, J., 1998. How does foreign direct investment affect economic growth? *Journal of International Economics* 45, 115-35. Xu, B., 2000. Multinational enterprises, technology diffusion and host country productivity growth. *Journal of development economics*, 62 (2), 477-93.

¹⁷⁸ These are: average tariff rates below 40 percent; average quota and licensing coverage of imports of less than 40 percent; a black market exchange rate premium that averaged less than 20 percent during the decade of the 1970s and 1980s; a non-socialist economic system; no extreme controls (taxes, quotas, state monopolies) on exports.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

authors found that, along the period 1970-1989, open economies outperformed closed economies in different dimensions.

Table 9.1 summarizes some of the authors' results. According to the table, 11 out of the 15 "open economies" in the sample expanded above 3.0% per year, while only four of the 74 "closed economies" achieved such a fast rate of economic growth. Controlling for other explanatory variables, the authors found that, on average, open economies grew by 2-2.5 p.p. faster than closed economies. The authors also concluded that open economies tend to exhibit higher investment rates than closed economies (a similar conclusion was not found for investment in human capital).

Table 9.1 Trade openness and economic growth

Developing countries growth and openness, 1970-1989

Growth rate	always open	not always open
Average growth > 3.0	11	4
Average growth < 3.0	4	70

Source: Sachs and Warner (1995), p. 36.

The authors then investigated how this result changes with a country level of economic development. They found that within the group of "developing countries", those that were considered as "open economies" expanded at 4.49% per year, while "closed economies" expanded at 0.69%, only. Among "developed economies", those that are open economies expanded at 2.29%, while closed economies expanded at 0.74%. The authors also concluded that poor countries tend to grow faster than richer countries as long as they are *linked together* by international trade. Closed economies, in contrast, do not display any tendency towards convergence. This suggests that international trade may be an important channel for *international technological diffusion*.

Finally, the authors investigated whether trade openness helps improve the quality of economic policies ¹⁷⁹. The authors found that, among the 73 closed economies, 59 experienced a severe macroeconomic crisis. In contrast, only one open economy experienced a serious economic crisis (Table 9.2). In general, this evidence supports the general claim that openness to trade helps improve the permeability of economies to best practices in other countries, leading to better governance in general.

Table 9.2 Trade openness and quality of policymaking

Developing countries growth and macroeconomic crisis

Growth rate	Open in 1970s	Not open in 1970s
Macroeconomic crisis in 1980s	1	59
No macroeconomic crisis in 1980s	16	14

Source: Sachs and Warner (1995), p. 56.

Note: “Macroeconomic crisis” is defined by one of the following occurrences: a rescheduling of foreign debt; arrears on external payments; an inflation rate in excess of 100 per year.

9.3 Permeability to technological diffusion

Although backwardness carries with it the potential for a country to catch up, the degree to which this potential materializes in each specific environment depends on economic, political and social circumstances. Factors such as the availability of human skills, infrastructures and the quality of domestic institutions, by shaping the economic incentives to implement new technologies, may accelerate or retard technology adoption. This section briefly reviews the role of country characteristics in determining the permeability of a country to worldwide technological change. We also discuss the challenges posed by the fact that new technologies come along with the destruction of existing rents, giving rise to negative reactions by those who have more to lose with the change.

¹⁷⁹ Sachs and Warner (1995): “(...) the international opening of the economy is the *sine qua non* of the overall reform process. Trade liberalization not only establishes direct linkages between the economy and the world system, but also effectively forces the government to take actions on the other parts of the reform program under the pressures of international competition”.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

9.3.1 Complementarities

The productivity of a new equipment does not depend only on its intrinsic efficiency but also on the abundance/adequacy of *complementary inputs* in the hosting economy. The more a new technology matches with a given country's endowments, the higher the likelihood of it to be profitable and therefore adopted in that country.

An obvious complementary input to new technologies is human capital. Poor and unequal countries with low levels of literacy will find it more difficult to adopt sophisticated technologies than countries with high levels of human capital. Some authors have argued that the human capital needed to absorb new technologies is not independent of a country's development stage. For a laggard country, where most growth opportunities involve the imitation and adoption of foreign technologies, the critical ingredient will be a good coverage of primary and secondary education. For a country closer to the technological frontier, however, because much of the growth opportunities involve the adoption of more sophisticated technologies and the development of new ones, it is important to invest in higher education and in high-quality research centres¹⁸⁰.

This means that human capital not only has a direct effect on production as an input in the production function (as captured by the MRW model), but also an indirect effect, in

¹⁸⁰ Theoretical models where technological diffusion is mediated through human capital include Nelson and Phelps (1996), Acemoglu et al (2006), Aghion and Howitt (2005). [Nelson, R., Phelps, E., 1966. Investment in humans, technological diffusion and economic growth. *American economic review* 61, 69-75. Acemoglu, D., Aghion, P., Zilibotti, F., 2006. Distance to frontier, selection and economic growth, *Journal of the European Economic Association*, March 2006, Vol. 4, No. 1, Pages 37-74. Aghion, P., Howitt, P. 2005. Growth with quality improving innovations: an integrated approach. In Aghion, P., and Durlauf, S. (eds), *Handbook of Economic Growth*, North Holland, Amsterdam, Chapter 2, 67-110.]. On the empirical front, the relationship between human capital and technological adoption was investigated by Griliches (1957), Benhabib and Spiegel (1994), Eaton and Kortum (1996), Doms, Dunn and Troske (1997) and Borentzein et al. (1998), Caselli and Coleman (2001), Caselli and Wilson (2004). [Griliches, Z., 1957. Hybrid corn: an exploration in the economics of technological change. *Econometrica*, 25, pp. 501-522. Benhabib, W., Spiegel, M., 1994. "The role of human capital in economic development: evidence from aggregate cross-country data". *Journal of monetary economics* 34, 143-173. Caselli, F., Wilson, D., 2004. Importing technology. *Journal of monetary economics* 51, 1-32].

shaping the permeability of a country to the World technological progress. Conventional growth accounting assessing the contribution of human capital by the elasticity in production only, tends to understate the true impact of human capital on growth.

Complementary inputs other than human capital include physical infrastructure (roads, ports, telecommunication networks, power supply), business services (accountancy, machinery repairs), financial services (banks, insurance, stock markets), government services (property rights protection, regulation), and so on.

The fact that technology requires complementary inputs suggests that the slow adoption of new technologies in developing countries may be an optimal response to differences in endowments, which translate into differences in the efficiency with which new technologies can be used. The implication is that governments have a role in shaping a country's absorptive capability: by promoting the education of people, building infrastructure and promoting a balanced development of the different capabilities, government policies influence decisively the pace at which new technologies are adopted. That said, improving a country's absorptive capability is not an easy task: infrastructures are expensive, human capital can only change slowly and geographical conditions like the climate cannot be changed at all. Hence, the process of turning the country more permeable to the adoption of new technologies is necessarily a slow one.

9.3.2 The vintage capital theory

New technologies replace old technologies that become obsolete. That being the case, one would expect investors to buy the state-of-the-art technology only, and the share of old technologies in the capital stock to gradually decline over time (this view is known and the "vintage capital theory")¹⁸¹.

¹⁸¹ Johansen, L., 1959. "Substitution versus fixed production coefficients in the theory of economic growth: a synthesis". *Econometrica* 27, 157-176. Solow, R., 1960. "Investment in technical progress", in Arrow, K (ed.), *Mathematical methods in the social sciences*, Stanford University Press.]

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

In the real world, however, there are many examples in which investment in frontier technologies only becomes dominant after a period of time during which investment in non-frontier technologies continues to dominate. A historical example occurred in Germany after WWII¹⁸²: with the war, Germany lost a significant part of its merchant fleet. If Germany had rebuilt its merchant fleet with state-of-the-art ships, only, it should have acquired a higher proportion of motor-ships (relative to sail-ships and steamships) than other European countries. However, that was not the case.

The persistent behaviour of investment in old technologies looks a paradox. Why should entrepreneurs insist in buying technologies that are less efficient, instead of investing in state-of-the-art technologies?

9.3.3 *Switching costs*

One reason for the persistence of old, dominated technologies, when newer and more efficient technologies are available is the existence of “switching costs”: costs related to the move from one technology to another.

A first type of switching costs is related to complementary specific investments in physical capital or in infrastructure that are specific to the old technology: when the DVD player replaced the old vinyl record player, consumers had to set aside their collections of vinyl records and start buying new CD records. Naturally, those consumers with bigger collections of vinyl had more to lose with the technological change, and eventually resisted the adoption of the new technology.

A similar reasoning holds for the *value of experience* acquired in dealing with the old technology. Learning how to operate a technology and taking opportunity of its full potential takes time, materializing a “sunk cost” that cannot be recovered once the technology is abandoned. Hence, after this technology-specific investment in human capital is made, the

¹⁸² Comin, D., and Hobijn, B. , 2004. “Cross-country technology adoption: making the theories, facing the facts”. Journal of Monetary Economics 51, 39-83.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

worker has great incentives to stick with the old technology. For a worker, switching to a new technology will imply a decline in the value of his accumulated experience, as well as the need to incur in new learning costs. With no surprise, when technology-specific skills are important, workers tend to resist the adoption of new technologies, irrespectively of how efficient they are (a classical example in Box 9.3).

Switching costs may also result from “network externalities”. A network externality arises when the benefit of using a given technology increases with the number of users of that technology. For instance, consider the telephone: the more people use telephones, the more valuable the telephone is to each user. Thus, when many users are hang-on to an existing technology, it becomes difficult for a newer, superior, technology to emerge.

A special case of network effects arises in the form of “social learning”. Suppose, for instance, that you considered buying a computer with an exotic operating system. Even if this operating system was more efficient than the MS Windows, you would need to balance the costs of obtaining assistance in case no one else was adopting the same software. In contrast, people using MS Windows face lower user costs, because they have a higher chance of interacting with other people using the same software. The more people are using a technology, the higher the likelihood of each new user to interact with a potential teacher and hence the lower the learning cost. When this is so, the conditions exist for a widely used technology to block a new one, even if more efficient.

Box 9.3. Locked to QWERTY

The QWERTY keyword, that you probably find in your laptop, is a rather inefficient one. As you may easily check, the commonest letters are scattered over all rows and concentrated on the left side of the keyboard, so that right-handed people have to use their weaker hand to reach them.

Why was this keyboard designed with such unhelpful and unproductive features? The reason is that, by 1873, when this layout was created, mechanical typewriters jammed easily if two keys were struck in a very quick succession. The QWERTY key layout was therefore a technological response to a problem of jamming. It was purposefully designed with the aim to slow down typists and reduce the frequency of jams.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

What makes this case interesting is that the QWERTY keyboard is still used today, having survived the elimination of the problem of jamming that motivated its creation in the first place. The reason is that, at the time the problem of jamming was fixed, the QWERTY keyboard was already established as a “lingua franca”: users were accustomed to it, and constructors of typing machines and computers kept providing this layout in new equipments, because this was what the market demanded. Attempts to launch new and more efficient keyboards were tried, but without success, because people were already locked-in to the less efficient technology and refused the change¹⁸³.

9.3.4 Leapfrogging

Lock-in effects related to switching costs should, at the first sight, favour economic convergence: countries that are intensive users of the old technology should be the countries that have more to lose by adopting the new technology. Less developed economies in contrast, because they are not heavily committed to any technology, could in principle jump to the frontier by investing in state-of-the-art technologies. The opportunity that laggard countries have to jump ahead of the leaders is dubbed as “leapfrogging”.

In the real life, examples of leapfrogging abound. For instance, many developing countries in the past decade have adopted cellular phones faster than developed countries, because operators were not locked-in to the technology of conventional telephones.

Leapfrogging is not, however, a general case: there is significant evidence suggesting that more advanced economies are not only those that invent new technologies, they are also those who adopt newer technologies first¹⁸⁴.

¹⁸³ David, P. 1985. Clio and the Economics of Qwerty. American Economic Review 75, 332-337.

¹⁸⁴ Examining the diffusion of 25 technologies across 23 industrial countries for the period from 1788 until 2001, Comin and Hobijn (op.cit) found that most technologies are first adopted in advanced economies and then they trickle down to countries that lag economically. The authors also found that leaders in the adoption of a predecessor technology tend to be the leaders in the adoption of the successive technology.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

A possible explanation for this pattern is that part of experience acquired with the old technology is not exactly a sunk cost, being instead *transferable* to the new technology. For instance, it is probably easier for someone with experience in mechanic typewriting to become a computer typist, than someone with no experience at all. The time and effort the mechanic typewriter invested in learning with old technology is not completely lost, because it can be adapted to work with a computer.

This means that the accumulated experience in dealing with an old technology not only has the potential to reduce the user costs in that technology (giving rise to lock-in effects), it may also help reducing the costs of adopting a new and more efficient technology. When the second effect is significant, users of the old technology will have an advantage instead of a disadvantage relative to workers who have no relevant experience in the field¹⁸⁵. Box 9.4 presents an argument in this avenue.

Box 9.4. The Haussman - Klinger forest

The ability of a country to start producing more sophisticated goods depends on a country's set of capabilities. Much of these capabilities, in turn, were acquired, developed and accumulated in response to earlier production needs. For instance, engineers might have been trained to support an existing textile industry. These engineers, in turn, could be adapted to work in a clothing factory. This means the ability of a country to attract new technologies may be conditional on the usefulness of the industry-specific experience and infrastructure that was generated by the country's current specialization pattern.

¹⁸⁵ Jovanovic and Nyarko (1996) build a model of individual decisions, whereby learning by doing provides an agent with information that improves its productivity in the old technology (vertical shifts). In this model, agents may also switch to new technologies (horizontal shifts). The degree of similarity of the new technology to the old one determines how transferable the accumulated knowledge is. The lower the possibility of transferring the accumulated knowledge to use with the new technology, the larger will be the productivity loss faced by workers being asked to move to the new technology. When the technological leap is too large, the expertise loss may be such that a highly skilled agent prefers not to switch, becoming therefore locked in the old technology [Jovanovic, B., Nyarko, Y., 1996, Learning by doing and the choice of technology, *Econometrica* 64, 1299-1310].

To illustrate this idea, Hausmann and Klinger (2007) proposed a metaphor with a forest, where each tree represents a product¹⁸⁶. In that forest, each tree is placed at some distance to the other trees, the distance capturing the degree to which the skills acquired in one productive experience can be used in other productive experience. Because some industries use skills that are common to a large number of industries, some parts of the forest are *denser* than others.

In this metaphor, firms are monkeys that live in the trees, and the process of structural transformation involves the monkeys jumping around from tree to tree. Moving to trees at larger distances involves the need for productive capabilities that have not been previously accumulated. Because some trees generate more income than others, each monkey would like to move to high productivity trees. However, because smaller jumps are less costly than larger jumps, the ability of the “tribe” to engage in superior technologies depends on having a path to nearby trees that are increasingly of higher productivity. If the move towards high productivity trees require larger jumps, the tribe may find itself in a poverty trap, jumping around lower income trees.

With this paradigm, the authors argued that the process of technological change is path-dependent: when a country’ accumulated experience is less valuable, comparative advantages will determine a specialization in industries with low potential to generate new knowledge and spur economic development. If however the country has accumulated experience that is highly transferable to new technologies, it will find it easier to start producing more sophisticated products. This interpretation is consistent with a broad notion of acquired “experience” or “capabilities”, including infrastructure, labour skills, country-specific technical knowledge, specific regulations, and so on.

9.3.5 *Barriers to technological adoption*

¹⁸⁶ Hausmann, R. and B. Klinger, (2007), “The structure of the product space and the evolution of comparative advantage”, CID Working Paper no. 146, April.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Innovations not only have the potential to generate rents, they also have the potential to destroy existing rents. To the extent that technological progress brings about more efficient machinery and production methods, owners of the old machinery will lose. Technological change often comes along with redistributive effects that challenge the balance of political powers.

Not surprisingly, history is full of examples of powerful elites and organized groups seeing their economic, political and social interests threatened by the adoption of new technologies placing obstacles to its diffusion. For instance, in the nineteenth century Austria-Hungary, the elites acted to block industrialization and even the introduction of railways, just because they realized industrialization would reduce their power and privileges¹⁸⁷.

Obstacles put in the path of innovators by established interests are labelled *barriers to technology adoption*¹⁸⁸. An obvious form of barrier is bribery: whenever the use of old technologies generates economic rents, the opportunity exists for established elites to persuade the political power to block innovations by imposing regulatory and legal constraints. More generally, barriers to technology adoption may include violence or threat of violence, worker strikes, etc.

Barriers to technology adoption may also arise in the form of social norms: the implementation of new technologies often requires organizational changes and complementary reforms that challenge beliefs and traditions within a country. Those with a stake in the old technology may find it profitable to support these traditions and beliefs in an attempt to block the adoption of new technologies. Leaders in laggard countries often lack the political power or the political will to confront these traditions.

All in all, the arrival of new technologies may face the resistance of groups who have a stake in the preservation of the status quo. Because of this, many authors content a most

¹⁸⁷ Acemoglu, D. 2003. "Root causes: a historical approach to assessing the role of institutions in economic development". Finance and Development, 27-30, June.

¹⁸⁸ Parente, S. and Prescott, P., 1994, op. cit.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

decisive element influencing the pace of technological diffusion is the quality of political institutions: the less influenceable they are by privileged elites, the easier it will be to find policymakers committed with reforms and able to accept the underlying changes that the new technologies are likely to bring about.

9.4 Matching specific needs

9.4.1 Self-discovery

Because economies differ in terms of endowments, infrastructure, climate and culture, they should optimally adopt different technologies. If information was perfect, investors would always pick up the technology that better matched the target environment. In a world with uncertainty, however, finding out which of the many potential technologies better fits a country's specific circumstances is a process of trial and failure. This process is known as "self-discovery"¹⁸⁹.

The process of "Self-discovery" involves externalities from the innovator to the followers. First, the entrepreneur that adopts a new technology provides valuable information to its competitors: if the entrepreneur succeeds, other entrepreneurs who opted to wait and see will imitate it, eroding the innovator's rents; if it fails, the innovator will bear the costs alone. Because the entrepreneur that first adopts the new technology provides valuable information to other potential entrepreneurs without being compensated for that, an information externality arises. Second, innovators often need to train workers in the use of the new technology. Once the innovator incurs the training costs, the risk exists of competitors to free ride on workers mobility, and of beating the innovator with lower training costs.

These externalities clearly reduce the incentives to innovate. If the expected gain from moving first is not large enough to compensate the innovating firm for its risk taking, the firm

¹⁸⁹ Hausmann R., Rodrick D. 2003 "Economic Development as Self-Discovery" *Journal of Development Economics* 72(2) 603-633.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

will optimally prefer to wait and see, postponing the adoption of the new technology, even if that was socially valuable. This discussion adds to the general case that private returns to innovation tend to fall short the social returns, calling for government intervention.

9.4.2 Mastering foreign technologies

A difficulty in the adoption of foreign technologies by developing countries is that many technologies are invented targeting the conditions of advanced countries. The reason is that researchers must recover the fixed costs involved in R&D, and this will be easier to achieve if the innovation targets a large economy with many wealthy consumers, than if the innovation is directed to a small, idiosyncratic, poor economy. This reasoning explains why many inventors around the world are trying to develop ideas that are useful in the United States, to sell the patent there, rather trying to develop ideas that are specific to their own narrow contexts. This bias, coined as *directed technological change*, implies that many of the available technologies may not be suitable for the needs of poorer countries, with different climate and factor endowments¹⁹⁰. In agriculture, for instance, most innovations relate to cultures of temperate zones – where rich countries are – thus not being suitable to be implemented in developing countries located in the tropics. In manufactures, many innovations in machinery tend to economize labour and are specially designed to improve the productivity of skilled labour, thus not matching the abundance of unskilled labour that is typical in poor countries.

To the extent that innovations do not fit well with the characteristics of developing countries, their adoption “tout-court” would result in productivity gaps that could not be eliminated along time¹⁹¹. Hence, an effective technology transfer may involve some

¹⁹⁰ Acemoglu, D., 2002. Directed Technical Change, *The Review of Economic Studies* 69 (4) 781–809.

¹⁹¹ Atkinson, A., and J. Stiglitz, 1969. A new view of technological change, *Economic Journal*, pp. 573-578. Basu, S., Weil, D., 1998. Appropriate technology and growth, *Quarterly Journal of Economics*, 113(4), pp. 1025-1054.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

investment by the recipient country, in order to master the foreign technology and adapt it to the local environment, preferences and beliefs.

Mastering a foreign technology involves however a fixed cost for the innovator. An entrepreneur will engage in such an effort only if he is able to make profits during a period of time. This may presume a minimum protection from eventual imitators. A problem with many developing countries is that the enforcement of property rights there is so weak that it doesn't pay for innovators to spend resources in adapting foreign technologies, because imitators will free ride on that effort. Instead, talented inventors in developing countries may find it more profitable to design new products targeting the needs of industrial countries, where patented inventions can be sold to reward the R&D costs. A corollary is that the enforcement of intellectual property rights in developing countries is an essential pre-requisite to expand the size of their markets and, by then, to induce investments in technology more directed to the South¹⁹².

9.4.3 Institutions do not travel well

The idea that foreign technologies may not match the conditions of the recipient country does not apply to technology in the engineering sense only. It holds for technology in broad terms, including policies and institutions. Just like the effectiveness of a given machine in a particular location depends on the availability of labour skills, the effectiveness of a given policy or institution may depend on how this new policy or institution interacts with existing policies and institutions, culture, and beliefs. For instance, financial liberalization may help improve the allocation of resources, but it may also be a source of macroeconomic instability if effective supervision is lacking; privatization of utilities can deliver higher efficiency in management, but it can also be welfare reducing if there is no competition

¹⁹² Acemoglu, D., Zilibotti, F., 2001. Productivity differences. Quarterly Journal of Economics 116, 563-606.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

authority protecting the consumers from price abuses; protecting property rights is in general favourable to long term investment, but it will fail to do so if it lacks an effective judiciary.

The implication is that the simple copy of institutions that perform well in a given context does not necessarily deliver the highest possible economic performance in a different context¹⁹³. A suggestive example of how the replacement of old traditions by apparently more efficient policies resulted in welfare loss happened in Bombay Deccan, in the nineteenth century colonial India. The reform consisted in the introduction of civil courts, to improve the effectiveness of contracts in general. These courts however interacted adversely with the credit market for agriculture: before courts were introduced, a traditional practice of risk-sharing existed, whereby lenders subsidized farmers during bad harvests. The newly established civil courts were able to enforce simple debt contracts, but not the complex risk-sharing informal arrangements such as those that proved effective in the past. The reform end up turning farmers more vulnerable to bad harvests¹⁹⁴.

The recognition that the effectiveness of economic reforms is largely conditional on their interaction with existing policies and institutions had lead policymakers and international institutions to search for an appropriate sequence of reforms¹⁹⁵. Along this

¹⁹³ Douglass North (1994, p.8): “(...) transferring the formal political and economic rules of successful western market economies to third world and Eastern European economies is not a sufficient condition for good economic performance”. [North, D., 1994. Economic performance through time”. *The American economic review* 84(3), 359-368].

¹⁹⁴ Kranton, R., Swamy, A., 1999. The hazards of piecemeal reform: British civil courts and the credit market in colonial India. *Journal of Development Economics* 58, 1-24.

¹⁹⁵ Along this reasoning, it has been argued that laggard countries, which are expected to rely more on adoption of foreign technologies than on own innovation efforts, should focus first on institutions to support investments, such as long-term banking finance. In a later stage, when innovation becomes more important, free entry, open competition, trade openness, and flexible labour markets become critical ingredients to provide a selection mechanism to weed out unprofitable projects [Acemoglu, D., Daron Acemoglu & Philippe Aghion & Fabrizio Zilibotti, 2006. "Distance to Frontier, Selection, and Economic Growth," *Journal of the European Economic Association*, vol. 4(1), pages 37-74, 03.].

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

reasoning, it has been argued that the optimal policy often involves adapting the institutional arrangements to fit a country set of characteristics¹⁹⁶. An example of this is in Box 9.6.

Box 9.6. Islamic Finance

The Islamic laws (*Sharia*), which rule the social, political and economic aspects of Islamic societies, encourage hard work, fair dealing, property rights, and the honour of contracts. The law also approves the earning of profits, because profits reward successful entrepreneurship and reflect the creation of additional wealth.

The *Sharia* prohibits however interest payments. The reason is that interest is a predetermined cost that is due irrespective of the business outcome¹⁹⁷. Banishing interest payments, the *Sharia* precludes the use of bonds and the development of banking, at least as designed for industrial economies. This, in turn, leads to insufficient savings, low investment and low growth.

Fortunately, a window was open to adapt the concept of banking so as to make it acceptable by the Islamic rules. This window is labelled *Murabaha*: this consists on a purchase and resale contract, in which the bank purchases goods from the producer with the promise to re-sell them at an agreed-upon date at an agreed-upon inflated price. Although it looks like a debt-instrument, the *Murabaha* is viewed as legitimate by the Islamic laws, because the financier bears risk during the period he owns the goods.

The principles of Islamic Finance were already practiced in Muslim societies throughout the middle ages, but it was after its inception in Egypt, in 1963, in Dubai, in 1975, and the opening of the first Islam bank subsidiary of a Western Bank (the Citibank) in

¹⁹⁶ Rodrik (2006). Rodrik, D., 2006. Goodbye Washington Consensus, Hello Washington Confusion?, *Journal of Economic Literature* 44 (4), 973-987.

¹⁹⁷ Because the Islamic doctrine advocates profit sharing (*qirad*), it foresees instead a kind of “venture capital” called *Mudarabah*: under this scheme, one party provides the capital for a project and the other party provides labour effort. The principle is that providers of funds become partners instead of creditors: if the enterprise succeeds, they share profits; if it fails, they lose the capital and the working time invested. Such a contract reflects the ideal cooperative spirit of Islam: borrowers and lenders share losses as well as rewards.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Barhain in 1996 that Islamic Finance flourished around the world. Today, there are more than three hundred Islamic financial institutions operating in more than 75 countries, including in Europe and the United States. These institutions offer a wide set of instruments targeting the needs of providers and users of funds: *murabaha* (trade with mar-kup financing), *bay' salam* (forward sale), *bay' mu' ajjal* (deferred payment sale), *ijara* (leasing), *mudarabah* (profit-sharing), *musharaka* (partnership). These instruments can then be combined to build a wider range of complex financial instruments.

The emergence of Islamic banking is creating big challenges to policymakers. This includes developing a framework to implement monetary policy and adapting the Western institutions of supervision and regulation. But with no question, Islamic Finance has proved to be a successful form of providing funds to entrepreneurs in compliance to a specific culture. It offers a good example of the principle that sometimes it is better to have a well-adapted institution than trying to imitate the original one without taking into account the specific circumstances¹⁹⁸.

9.5 A simple model of technology adoption

This section presents an adaptation of the Solow growth model to illustrate the problem of a technological follower aiming to absorb technologies developed elsewhere. In this model, country characteristics and adoption efforts determine how far the country gets from the world technological frontier, while its long run growth is linked to the world-wide rate of technological progress¹⁹⁹.

9.5.1 Modelling technology adoption

¹⁹⁸ For a brief description of Islamic finance, see El Qorchi (2005). [El Qorchi, M., Islamic Finance Gears Up, Finance and Development December 2005].

¹⁹⁹ The model adapts from Klenow and Rodriguez-Clare (2005). [Klenow, P. and Rodriguez-Clare, A., 2005. "Externalities and Growth". In Aghion, P., and Durlauf, S. (eds), Handbook of Economic Growth, North Holland, Amsterdam, Chapter 11, 817-866].

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Consider a small emerging economy, which instead of inventing its own technology, adapts and adopts technology invented elsewhere.

Output consists in a homogenous good, Y , produced under a Cobb-Douglas technology:

$$Y = AK^\beta (\lambda N)^{1-\beta} \quad , \quad (9.1)$$

where K includes both Human and Physical capital and λ measures the efficiency of labour.

Output can be spent in consumption (C), Investment ($I=sY$) or in deliberate efforts to adopt new technologies (R). The later includes the spending of resources to discover or to adapt the technologies that better match the country needs. In a broad interpretation, you may also interpret R as including government spending in the provision of public infrastructures and in mass education, that improve the permeability of the country to technological change.

The capital stock evolves as in the basic Solow model with an exogenous saving rate:

$$\dot{K}_t = sY_t - \delta K_t$$

Similarly, it is assumed that a constant fraction of GDP, s_R , is invested in technology adoption:

$$R = s_R Y \quad (9.2)$$

The level of technology is assumed to evolve according to:

$$\dot{\lambda} = \left(b \frac{R}{N} \right) \left(\frac{\bar{\lambda}}{\lambda} \right)^{\frac{1}{\sigma}} \quad , \text{ with } \sigma > 1 \quad \text{and} \quad \lambda \leq \bar{\lambda} . \quad (9.3)$$

In (9.3) b is a positive parameter measuring the “productivity of the adoption effort”, $\bar{\lambda}$ represents the world technological frontier and σ is a parameter measuring the strength of technological diffusion into the country.

The first term in (9.3) captures the impact of the adoption effort²⁰⁰. With such specification, a deliberate effort to adopt (or adapt) foreign technologies is necessary for the country to be permeable to foreign technologies: when no resources at all are spent, there is no assimilation of foreign technologies and the economy does not grow. The impact of the research effort on technological adoption is mediated through the productivity parameter b . In a narrow interpretation, this term may be interpreted as capturing the skills of engineers and scientists. In a broader interpretation, parameter b may be seen as capturing the influence of barriers to technological adoption, such as licensing, legal restrictions and low enforcement of property rights. These barriers increase the costs of adopting foreign technologies, implying that more resources R are necessary to achieve a given improvement in technology.

The second term in (9.3) captures the “benefits of backwardness”: other things equal, the more knowledge remains to be absorbed by the country ($\bar{\lambda}/\lambda$), the higher it will be its rate of technological change. The rationale is that, as the country approaches the frontier, less and more complex ideas will be available for copying, so the cost of achieving a given improvement in technology increases. Countries that are backward relative to the technological frontier will enjoy greater improvements in technology for each unit of output spent in technology adoption. The parameter $\sigma > 1$ imposes diminishing returns on the benefits of backwardness.

Substituting (9.2) in (9.3) and dividing both terms by λ , one obtains:

$$\frac{\dot{\lambda}}{\lambda} = (bs_R \tilde{y}) \left(\frac{\bar{\lambda}}{\lambda} \right)^{\frac{1}{\sigma}} \quad , \quad (9.4)$$

where

²⁰⁰ In (9.3), the country's expenditure in technology adoption, R is divided by population, N , so as to avoid scale effects. Since R depends linearly on Y , the change in technology will depend on per capita income and, by then, on the level of λ . The model still accounts for a “standing on shoulders effect”, but without a scale effect.

$$\tilde{y} = Y/\lambda N = A^{1/(1-\beta)} (K/Y)^{\beta/(1-\beta)}. \quad (9.5)$$

Equation (9.4) reveals that both the technology adoption effort, s_R , and the respective productivity, b , affect the country rate of technological progress, conditional on output per unit of efficiency labour, $\tilde{y}$. This property of the model captures the interactions between the adoption effort, capital availability (human, physical) and efficiency (A). Thus, policies leading to a higher investment in physical or in human capital and policies aiming to improve static efficiency (A) will enhance the adoption activity, translating into a faster technological catch up. The model therefore establishes a causal relationship from aggregate efficiency and capital intensity on the pace of technological diffusion into a country.

The World technological frontier is assumed to expand at the exogenous rate $\bar{\gamma}$:

$$\bar{\lambda} = e^{\bar{\gamma}} \quad (9.6)$$

Arguably, the expansion of the World technological frontier is the result of the cumulative R&D effort of all countries together²⁰¹.

9.5.2 The Steady state

The benefits of backwardness imply that technology absorption is faster when the country' technological gap is larger. Thus, there is a force pulling the country towards the frontier. In the steady state, this force is powerful enough to ensure that its productivity level will grow at the same rate as the world technological frontier. In other words, the country will evolve along a balanced growth path that is parallel to that of the world economy.

The steady state of the model is obtained setting $\dot{\lambda}/\lambda = \bar{\gamma}$ in (9.4). Solving for the technological gap, this implies:

²⁰¹ Klenow and Rodriguez-Clare allow the world technological frontier to be driven by the research efforts of all countries in the model. The student will thank us for skipping that complication.

$$\frac{\bar{\lambda}}{\lambda} = \left(\frac{\bar{\gamma}}{bs_R \tilde{y}} \right)^\sigma \quad (9.7)$$

Equation (9.7) states that the steady state technological gap is positively affected by the world rate of technological progress ($\bar{\gamma}$) and negatively affected by the country's adoption effort (s_R) and the productivity of the adoption effort (b).

The steady state level of per capita output in this model corresponds to the solution of the Solow model, given by:

$$y_t^* = A^{\frac{1}{1-\beta}} \left(\frac{s}{n + \delta + \gamma} \right)^{\frac{\beta}{1-\beta}} \lambda_t. \quad (9.8)$$

Solving together (9.7) and (9.8), using (9.6) and the definition $\tilde{y} = y/\lambda$, one obtains the steady state level of per capita income in this extended version:

$$y_t^* = A^{\frac{1+\sigma}{1-\beta}} \left(\frac{s}{n + \delta + \bar{\gamma}} \right)^{\frac{\beta(1+\sigma)}{1-\beta}} \left(\frac{bs_R}{\bar{\gamma}} \right)^\sigma e^{\bar{\gamma}t}. \quad (9.9)$$

As in the Solow model, country characteristics determine the level of income per capita, but not its steady state growth rate: the long run growth rate is exogenous and given by the world rate of technological progress.

Also like the Solow model, this model does not predict absolute convergence of per capita incomes, but instead conditional convergence: differences parameters translate into cross-country differences in levels. In the long run, countries will evolve along parallel growth paths, an implication that is supported by the general evidence on conditional convergence.

There are however three main differences relative to the Solow model: First, the exogenous rate of technological progress now applies to the world, and the model determines how close the country gets to the world technological frontier. Second, the country's ability to approach the world technological frontier depends on the proportion of income spent in the technology adoption (innovation, adaptation, addressing specific market failures), s_R , and the productivity of these efforts, b (that may be thought as depending on political, social and economic factors). Third, the influence of the "old" parameters, namely aggregate efficiency (A), the propensity to invest in physical (human) capital (s), and the population growth rate

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

(n), is amplified by a factor of σ . This captures the above mentioned interaction effect between capital availability, TFP and adoption effort.

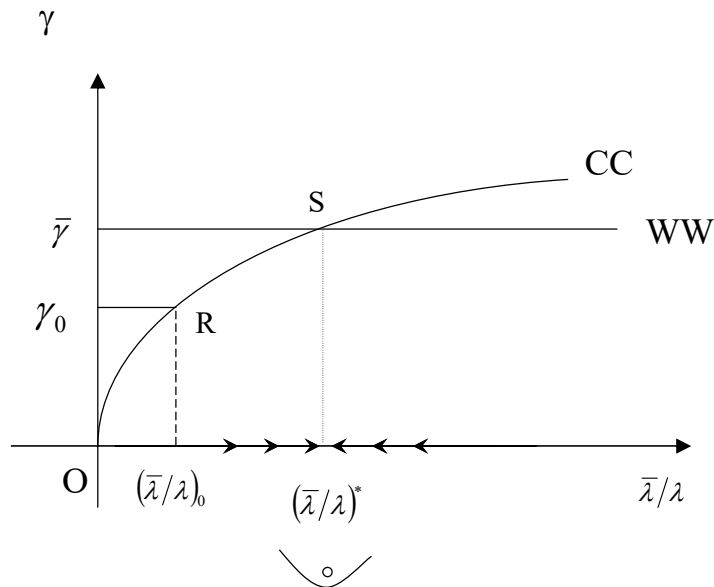
9.5.3 Transition dynamics

As in the basic Solow model, the principle of transition dynamics applies to changes in the exogenous parameters: a favourable change in a parameter will produce a *level effect* and a transitory period during which the country approaches the World technological frontier. During this period, the country will exhibit faster growth than the world average, but this will be temporary. In the long run, despite the differences in the behavioural parameters, the growth rate of per capita income in the country will be equal to that of the World frontier.

To see how the model works in the short run, we depict in Figure 9.1 the country rate of technological adoption, γ , as a function of its technological gap, $\bar{\lambda}/\lambda$ (Curve CC). According to equation (9.4), this curve is upward sloped (a higher technological gap implies a faster rate of technology absorption) through the benefits of backwardness. The figure also depicts the growth rate of the World technological frontier, $\bar{\gamma}$, which is independent of the country technological gap (curve WW).

Now assume that the country is initially in point R, with a technological gap equal to $(\bar{\lambda}/\lambda)_0$. With such a gap, the rate of technological expansion in the economy is smaller than the world rate, that is $\gamma_0 < \bar{\gamma}$. This means that the country will be diverging relative to the world economy. In the figure, the country will move rightwards, from R to S (higher technological gap). As the technological gap widens, the advantage of backwardness shows up more strongly, so the rate of technological adoption (and, thereby, per capita output growth) increases. When point S is reached, the growth rate of the economy is exactly equal to the growth rate of the world technological frontier and the income gap stabilizes. By the same token, if the country starts out on the right-hand side of S, it will converge to S.

Figure 9.1: Transition dynamics and the steady state in the technology adoption model



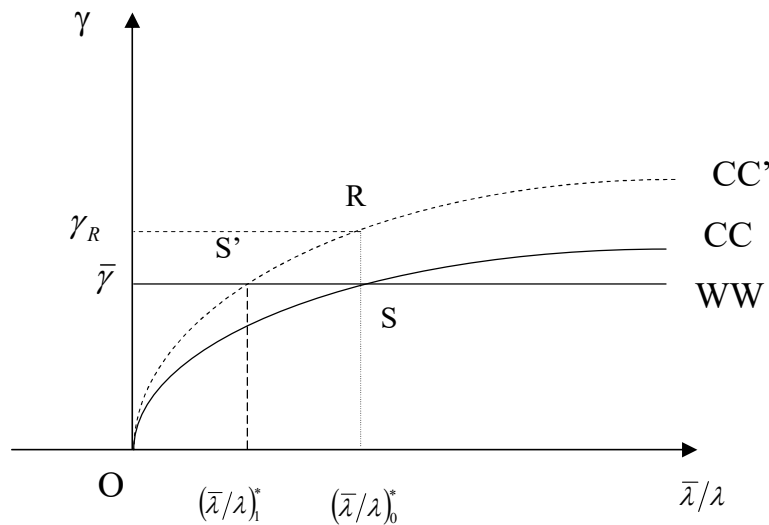
The curve CC describes the benefits of backwardness, for a given adoption effort. The curve WW describes the world rate of technological progress, which is exogenous. In point R, the country's technological gap vis-à-vis the world frontier is too small given its characteristics, implying that technology in this country will expand at a slower pace than at the frontier. In result, the technological gap will increase until the steady state S is reached.

9.5.4 What happens if the technological adoption effort increases?

Figure 9.2 describes the effect of a rise in the country's adoption effort. From (9.4), an increase in s_R causes the CC locus to shift upwards. This means that the steady state moves to a different point (from S to S').

If the country is initially in the original steady state S, at the impact there will be an acceleration in the growth rate of technology adoption (from $\bar{\gamma}$ to γ_R - point R). Since the country technology is now expanding faster than the World frontier, the technological gap starts decreasing, meaning that the country moves leftwards, from R to S'. As the technological gap decreases, the benefit of backwardness decreases, implying a declining growth rate. When the new steady state is reached (S'), the country growth rate equals the world growth rate and the technological gap stabilizes. Thereafter, the economy will evolve in parallel to the rest of the world, but with a lower income gap than initially.

Figure 9.2: The effect of an increase in adoption effort or of a decrease in barriers to technology adoption



Departing from S, an increase in the country permeability to world-wide technological progress will increase its rate of technological change above the world average (point R). As the technological gap decreases, the benefit of backwardness diminishes, until the technological gap stabilizes (S').

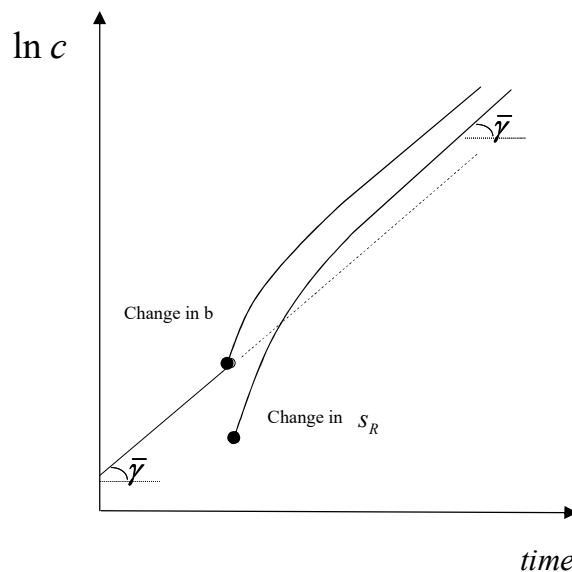
9.5.5 What happens when barriers to technology adoption decline?

Assume now that the economy opened to international trade, enabling the economy to absorb foreign technologies faster than before, for each level of adoption effort. In our model, this is captured by an increase in parameter b . In terms of Figure 9.2, the locus CC shifts up, just as in the earlier section. The adjustment mechanism is similar to the one before: the economy will grow temporarily faster than the rest of the world, but as the productivity gap declines, the growth rate approaches the world rate of technological progress (point S').

Note that an increase in the adoption effort involves the expenditure of resources, while an improvement in the productivity of adoption effort does not. Hence, although s_R and b have similar effects in terms of Figure 9.2, the path of per capita consumption differs in the two cases. Figure 9.3 shows the difference: in the case of an increase in s_R , there is an

initial fall in per capita consumption that may or may not be compensated by the temporary acceleration that follows. In the figure, it is assumed that the acceleration effect dominates, but this is not necessarily true. As in the Solow model there is a golden rule for the optimal spending in technology adoption²⁰².

Figure 9.3: The effect of an increase in adoption effort or productivity



Rising a country' permeability to technological change may involve the spending of resources or an increase in the productivity of the research effort. The difference is that the later implies a trade-off between current consumption and future consumption, and therefore no Pareto improvement.

9.5.6 What happens when the world rate of technological progress increases?

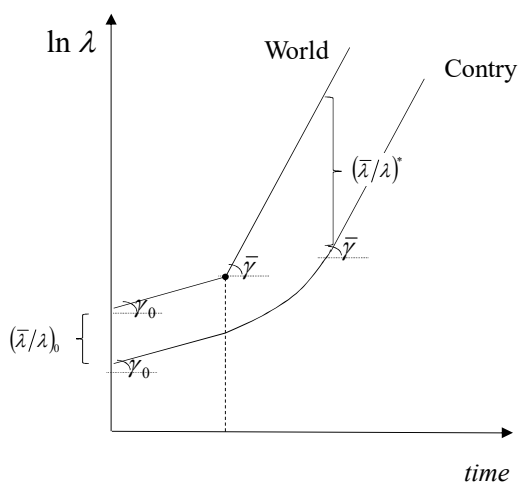
²⁰² To see how the golden rule looks like in this model, you may investigate which values of s_R and s maximize the steady state level of per capita consumption, given by $(1 - s - s_R)y_t^*$. The solution is $s = \beta$ and $s_R = \sigma(1 - \beta)/(1 + \sigma)$.

We now analyse the impact of a shock that is out of control of an emerging country: a change in the world rate of technological progress.

To analyse this, let's refer again to Figure 9.1. Assume that initially the world technological frontier was expanding at rate γ_0 and that our emerging economy was in the corresponding steady state, with a constant technological gap equal to $(\bar{\lambda}/\lambda)_0$. Then suppose that the world rate of technological progress accelerated once-and-for all to $\bar{\gamma}$.

This change implies a shift in the country's steady state, from point R to point S. As explained before, the adjustment to the new steady state is not instantaneous: the faster expansion of the world technology frontier leads to a widening of the country technological gap. This, in turn, impacts positively on the country's rate of technology adoption (due to the benefit of backwardness). When the technological gap is sufficiently large, the country rate of technological absorption becomes equal to the world rate of technological progress $\bar{\gamma}$ and the technological gap stabilizes in its new steady state level, $(\bar{\lambda}/\lambda)^*$.

Figure 9.4: the widening of income gaps



Departing from a steady state, if the world technological frontier starts expanding faster, there will be a period during which the gap of a laggard country relative to the frontier increases. With a larger gap, the benefits of backwardness will increase, fostering the country rate of technological change, until it equals that of the world technological frontier. In the new steady state there will be a larger technological gap.

All in all, the implication of the increase in the rate of technological progress at the frontier is a widening of income disparities. In the long run, the emerging country will grow as fast as the world economy, only because it got sufficiently behind.

Figure 9.4 displays the path of technology in the emerging country and in the world (in logs): before the shock, the two technologies were growing in parallel, with a gap equal to $(\bar{\lambda}/\lambda)_0$; after the shock, the two growth rates depart from each other, resulting in an episode of temporary divergence; in the new steady state, the two technologies evolve again in parallel, but the new technological gap $(\bar{\lambda}/\lambda)^*$ is larger than the initial one.

9.5.7 *The great divergence revisited*

The model just described offers an interpretation for the episode of the Great Divergence: when West Europe and the Western Offshoots entered in modern growth, the world technological frontier started growing faster than before. However, the rest of the world did not enter immediately in modern growth. Because of domestic idiosyncrasies, countries such as India and China entered in modern economic growth two centuries later, only.

According to the discussion above, the acceleration in the rate of technological progress in Western economies should have caused the technological gap of laggard countries to increase. In fact, per capita incomes relative to the leader in India and China fell from 44% and 48%, respectively in 1700 to only 7,9% and 7,2% in 1965. Then, as the income gaps got larger, these countries started benefiting from faster technological diffusion, implying that at some point in time (by the middle of the twentieth century) these countries stopped diverging (see Box 9.7).

9.5.8 *Proximate causes versus ultimate causes once again*

The model stresses the role of exogenous parameters, such as the saving rate (s), the adoption effort (s_R), the productivity of adoption (b) and total factor productivity (A) in determining how close a country gets to the technological frontier.

One should remember, however, that these parameters can hardly be taken as independent from each other. For instance, an increase in the effort of technology adoption,

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

s_R , may induce organizational and political changes in the country, paving the way for the ease of barriers to technological adoption, improving b : that would be a case of transpiration bringing more inspiration.

Similarly, in an economy closed to international trade, where property rights are not enforced and where privileged elites are able to block any attempt to introduce new technologies – hence a country where productivity parameters b and A are very low, one will expect people to save less and to dedicate less resources to the adoption of new technologies (low transpiration because of low inspiration).

In a limiting case where property rights are not enforced at all, no agent in the economy will make any effort to innovate, and the economy will not be growing, despite the technological change at the frontier. This extreme case would capture the world poorest economies, that have not been able to achieve conditional convergence²⁰³.

This discussion brings us again to the discussion of proximate causes versus the ultimate causes of economic growth. A reduced-form model such as the one described above is helpful to understand the mechanics of economic growth and to discuss links between critical behavioural parameters and economic performance. But if one really wants to deepen the question and ask why do some countries save more or invest more in R&D than others, we have to depart from the simple equations outlined above and ask what determines parameters that are taken as exogenous in the model above.

²⁰³ Howitt (2000) proposed a Schumpeterian model of economic growth, where the innovation effort is endogenously chosen by profit maximizing firms. The author shows that only in countries with a *minimum* level of productivity in research and with enough protection of property rights firms will find it profitable to innovate. When these minimum conditions are not met, firms prefer not to innovate, and the economy stagnates. This case intends to capture the situation of the World poorest countries, which growth rates are mainly driven by internal factors. [Howitt, P., 2000. Endogenous growth and cross-country income differences”, American Economic Review 90, 829-846].

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Box 9.7 Miracles and Disasters

Figure 9.5 plots the evolution of per capita incomes in two leader economies, the United Kingdom and the United States, and six followers, Argentina, Portugal, China, India, Botswana and Chad.

The facts in the figure are as follows:

- The two leader countries, United Kingdom and the United States, have evolved mostly in parallel, with differences in levels that you may relate to differences in efficiency in which resources are used. You may think these two countries as driving the technological frontier and sharing equally the benefits of the world technological diffusion.

- China and India diverged relative to the leader countries from the beginning of the sample up to the mid-twentieth century. According to the model above, such divergence could be explained by an acceleration of technological progress in the leader countries.

- By 1970, India is likely to have engaged in a parallel growth path vis-à-vis the leader countries, without being able to catch up. This path is consistent to the idea that, when the laggard economy gets sufficiently behind, the benefits of backwardness prevent further divergence.

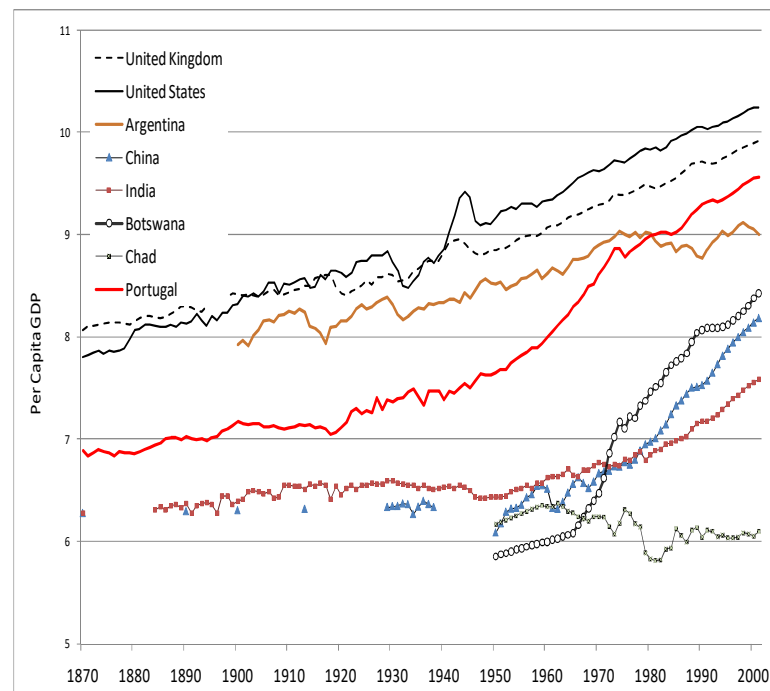
- In the second half of the twentieth century, some countries started approaching the leader countries: Portugal in the early 1950s, China in the early 1960s, Botswana in the mid-1960s. According to the model above, improvements in political, social and economic environments may have helped increase permeability of these countries to the world technological diffusion. These improvements translate into level effects in per capita income and hence to a temporary growth surge. In the long run, each country is expected to stabilize in a parallel growth path vis-à-vis the leader country.

- Along the period, Argentina has diverged relative to the technological frontier. This case is symmetrical to the earlier one: something in the Argentinean polity has evolved in the wrong direction, causing its income gap relative to the frontier to increase.

- Per capita income in Chad has stagnated and even declined in some years: the performance of this country suggests that extremely adverse local conditions, prevented the country from investing in the adoption of new technologies and enjoy the benefits of backwardness. Hence, this economy failed to achieve conditional convergence.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Figure 9.5: Miracles and Disasters



The figure describes the evolution of per capita incomes in different countries, since 1870. In light of the model above, one could interpret the different patterns as describing: technological leaders that have drawn the frontier (US, UK); countries where policy changes determined changes in the respective steady state income gaps (Argentina, China, Portugal, and Botswana); countries that only started growing in parallel when the gap became high enough (India); and a non-converging country (Chad)

9.6 Key ideas of chapter 9

- The view that poor countries may catch up with rich countries by imitating successful technologies without the need to invent them again is labelled the “advantage of backwardness”.
- In general, openness to trade and international factor mobility increase a country exposure to outside innovations, through demonstration effects, enhanced competition, and the transmission of tacit knowledge.
- The absorptive capability of a country depends on local characteristics, such as human capital endowments, infrastructure, institutions, and geography. Governments have a role in shaping a country set of capabilities in order to make it more attractive to technological change.
- Accumulated experience with an old technology may help or retard the adoption of the new technology, depending on how useful the inherited knowledge is to operate with the new technology.
- The arrival of new technologies may destroy existing rents. Vested interests with a stake in the old technology may use their political power to block the adoption of newer technologies.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

- The fact that countries differ in terms of capabilities implies that the appropriate technology differs from country to country. The process of finding out which technology better suits a country set of capabilities is labelled “self-discovery”. The process of self-discovery involves externalities that turn the adoption effort suboptimal in a laissez faire.
- Most innovations are targeted to match the characteristics of industrial countries. The transfer of these technologies to the development context may imply adaptation efforts. This is valid not only for technology in the engineering sense, but also to policies and institutions.
- The model of technological catching up analysed in the chapter distinguishes a world technological frontier and the technological level of an individual country.
- According to the model, international technological diffusion prevents countries from drifting indefinitely apart from each other. How close each country gets to the world frontier depends, however, on its absorptive capacity.
- In light of this model, an acceleration of the world rate of technological progress translates into faster growth in the laggard country too, but this will come up with a lag, implying that during the adjustment process the income gap vis-à-vis the frontier increases. This model offers an interpretation for the Great Divergence.

Problems and Exercises

Key concepts

- *Advantage of backwardness. Tacit knowledge. Vintage capital. Leapfrogging. Barriers to technology adoption. Learning costs . Value of experience . Self-discovery. Directed technological change*

Essay questions:

- Explain how the adoption of a new technology may be retarded by complementarity effects relative to other factors and substitutability effects relative to older vintages.
- Comment: “Institutions do not travel well”.
- Explain why in many poor countries the simple adoption of foreign technologies results in productivity gaps that cannot be eliminated along time.
- “The advantage of backwardness implies that laggard countries are doomed to grow faster than rich countries”.

Exercises

- 9.1. In the "Alpha" economy, the typical company's production function is given by $Y_t = 0.5K_t^{1/2}(\lambda_t N_t)^{1/2}$, where K_t e N_t represent the physical capital and the number of workers in each production unit and λ is a term measuring the quality of the work

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

- factor. In this economy, the savings rate is 25%, population is expanding at 1%, and the depreciation rate is $\delta=0.02$. (a) Assume for a moment that $\lambda_t = e^{0.02t}$. d1) Calculate the per capita product in efficiency units in the at steady state, $\tilde{y}$. d2) Describe the evolution of per capita income in steady state. d3) Graph and discuss the stability of the equilibrium. (b) Now take over that this economy did not produce its own technology, importing ideas from the rest of the world instead. In particular, assume that technology evolves according to $\dot{\lambda}/\lambda = 0.04s_R\tilde{y}(\bar{\lambda}/\lambda)^{0.5}$. f1) Interpret this equation. f2) Calculate the steady state technological gap for the specific case in which $\tilde{y}=1.25$, $s_R = 0.2$ and $\bar{\gamma} = 0.02$. f3) If the initial gap was 9, what would tend to happen? Why? f4) Graph and discuss model stability. (c) Compare the two previous models for the effect of an increase in the savings rate on the trajectory of per capita income. Explain, using graphical analysis.
- 9.2.** Consider the following specification for technological change: $\dot{\lambda}/\lambda = bs_R\tilde{y}(\bar{\lambda}/\lambda)$. Initially, the pace of economic progress in the leader country is $\bar{\gamma} = 0.02$. (a) Interpret the various parameters of the model. (b) Consider a particular economy, B, where $\tilde{y}=4$, $s_R=0.1$ and $b=0.02$. (b1) Compute its technological gap in the steady state. (b2) Represent in a graph. (c) Assume that the country managed to increase parameter b to $b=0.025$. (c1) Which type of reforms can be captured with this change? (c2) Describe in a graph the evolution of per capita income in that country until the new steady state is reached. (d) One implication of the current economic crisis in the industrial countries is that fewer resources are being invested in innovation. Describe, in light of this model, what would be the implications for global growth and for cross country economic convergence.
- 9.3.** Consider a small emerging economy with the following production function: $Y = AK^{0.5}(\lambda N)^{0.5}$, where K includes both human and physical capital and λ measures the efficiency of labour. In this economy the population is constant, the saving rate is equal to $s=0.2$, the depreciation rate is equal to $\delta=0.03$ and $A=0.25$. (a) Assume first that technology in this economy expands at 2% per year. Find the steady state in this economy and discuss the stability of the equilibrium. (b) Examine the implications of an increase in the efficiency parameter from $A=0.25$ to $A=0.5$. Compute the new steady state and explain with the help of a graph. Draw the time paths of income per capita (y) and of the interest rate (r). (c) Assume that technology in this economy evolves according to $\dot{\lambda}/\lambda = bs_s\tilde{y}(\bar{\lambda}/\lambda)^{0.5}$, with $b=0.1$, $s_r = 0.08$ and $\bar{\gamma} = 0.02$. Interpret the expression above. Returning to $A=0.25$, find out the steady-state value of the technological gap and represent it graphically. (d) Assume that A increases to 0.5. Compute the technological gap in the new steady state and explain graphically the adjustment process. Compute the new steady state level of per capita income. Explain. (e) Starting out in a steady state where $A=0.5$, assume that the foreign rate of technological progress decelerates to 0.01. Compute the new steady state level of the gap. Draw the time paths of income per capita (y) and of the efficiency of labour (λ) following this change.

Part III – Government, policies and institutions

10 Government services

“Economic history is overwhelmingly a story of economies that failed to produce a set of economic rules of the game (with enforcement) that induce sustained economic growth”. [Douglas North].

Learning Goals:

- Understand market failures and the role of government.
- Understand the optimal intervention rule in the model with public inputs
- Acknowledge the main sources of government failures and their implication for economic performance

10.1 Introduction

This chapter focuses on the role of government in providing essential goods and services that competitive markets do not generally produce. This includes, a physical dimension (infrastructure, communications systems), and an institutional dimension (for instance the rule of law and regulatory agencies). Without a minimum provision of these goods, the private economy will fail to operate efficiently, giving rise to distortions and bad resource allocation. By providing these services, governments enhance productivity, raising the incentives to produce and invest.

This does not mean that the larger the public provision the better. Government activities are financed with taxes, which crowd out private investment. Hence, a well-balanced intervention shall weight the efficiency enhancing potential of public expenditures against the distortionary effects of taxation. In this judgement, one must take into account the government’ own limitations: because of different types of inefficiencies, there is waste in the process of transforming tax proceeds into public services.

This chapter addresses the trade-offs involved in the public provision of goods and services that are essential to economic activity. Section 10.2 briefly reviews the role of government in the economy. Section 10.3 extends the basic Solow model by adding a government sector that collect taxes and provides a public input. Section 10.4 analyses the trade-offs involved in government intervention. Section 10.5 summarises the main ideas.
<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

10.2 The role of government

10.2.1 Why do we need a government?

In his masterpiece, *Wealth of Nations*, Adam Smith used the metaphor of the “invisible hand” to argue that the public interest would be best served if governments allowed selfish individuals to pursue their own interest. The “profit motive” would lead individuals, competing against each other, to supply the goods other individuals wanted at the lowest possible price. Because only agents producing at the lowest possible cost would survive, in a free-market system resources would not be wasted, and the economy would operate at its maximum level of efficiency²⁰⁴. Smith ideas have influenced nineteenth-century economists, like John Stuart Mill, who advocated the doctrine of the *laissez faire*. According to this doctrine, the government should not interfere with the private sector, regulating or controlling the production. Free competition would serve the best interest of the society. At the other extreme of the economic thinking, the nineteenth-century economist, Karl Marx argued that capitalism leads to grave income inequalities, and advocated a greater role for the state in controlling the means of production.

Economics has progressed a lot since then. With no question, the “invisible hand” argument still has great appeal. In general, there is a much-supported proposition that greater economic freedom is related to better economic performance. But the economic profession also understands government plays an important role as a complement to the market. Although there is now a general belief that markets and private entrepreneurship are at the heart of successful economy, there is also a recognition that an economy without government intervention will hardly work at all. So, fully hedged *laissez faire* is definitely ruled out.

²⁰⁴ Remember that, in terms of the AK model, a higher efficiency parameter (A) leads to faster *growth*. In terms of the neoclassical growth model, this would mean a higher *level* of per capita income in the steady state.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

In general, the free market to produce too much of a range of undesirable outcomes, such as pollution and too little of a range of some essential goods and inputs to production, such as roads and public infrastructure. Economists refer to these problems collectively as “market failures”. Market failures include inadequate provision of public goods, externalities, and imperfect competition, missing markets, information failures and persistent unemployment. When there is a market failure, the market mechanism does not produce the most efficient outcome.

Governments can obviously do things that private agents cannot do. For example, they have the right to force citizens to pay taxes; if they fail to do so, governments can confiscate their property. Governments may also manipulate prices, undertake direct provision, or create institutions or agencies to regulate markets. All in all, governments have the power to interfere and influence profoundly economic outcomes. If the intervention is successful, private incentives will become more aligned with the social interest. This, in turn, will induce a more efficient allocation of resources. As claimed by the Nobel Laureate Douglass North and his co-author Robert Thomas (1973), “getting the prices right” (that is, making individuals capture the social returns to their actions as private returns) is good for growth²⁰⁵.

At the most basic level, governments must provide public order and to set up a coherent system of property rights and enforcement of contracts. The free-market system requires that entrepreneurs who are investing in risky businesses have a high probability of making a profit that rewards their investment and risk. The legal system must protect the right to own property and must protect it from offences and thieves. Having a stake in the future, agents will follow a long-term perspective in their investment decisions. Moreover, enforceability of contracts is a necessary condition for individuals to engage in beneficial exchange. With unsecure property and contract rights, agents will have no incentive to engage in complex long-term operations and to take full opportunity of the benefits of

²⁰⁵ North, D., Thomas, P., 1973. *The rise of the Western World*, Cambridge UK: Cambridge University Press.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

specialization. They will instead tend to adopt shorter-term horizons, investing in inexpensive technologies and relying on bribery and corruption to enforce transactions. Property rights and contract enforcement may be seen as the foundations on which the market economy rests. Without them, there could be no market economy.

10.2.2 Public goods and publicly provided goods

The market mechanism does not adequately provide Public Goods. *Public goods* are a particular category of goods, which differ from private goods in two main aspects: (a) Public goods are non-excludable: if they are provided at all, it will be technically impossible to preclude anyone from consuming it. For instance, it will be virtually impossible to preclude somebody crossing a street from having access to that street's public lighting. Other goods that are non-excludable include clean air, street lighting, radio broadcast, and low inflation. The implication of non-excludability is that the benefits of public goods cannot be confined to those who have paid for it. (b) Public goods are non-rival: one person's consumption does not diminish the amount available to others. For example, if you eat an apple nobody else can eat the same apple. The apple is a rival good. In the case of public goods, extending its provision to an additional user does not diminish the quantity available to other users. For instance, one's benefit with a clean environment does not diminish the enjoyment of others. A clean environment is non-rival. Other goods that are non-rival include cable TV, macroeconomic stability, and the rule of law. A good that is both non-rival and non-excludable is called public good. A classical example of a public good is national defence: once a country is protected from foreign invasion, there is no extra cost in protecting a new citizen. So national defence is a non-rival good. Furthermore, it will be impossible to preclude anyone from that protection.

Private markets do not work at all well when goods are not excludable (characteristic a): whenever it is impossible or difficult to preclude anyone from using a service or consuming a good, it will be difficult to find someone voluntarily paying for its production.

This is because each individual will prefer not to pay and instead to take a *free ride* on any eventual production that does appear. This means that provision of a non-excludable good is not in general profitable²⁰⁶. The implication is that public goods will be under-supplied in a competitive equilibrium, even if they are socially very important. To the extent that the whole economy benefits with public goods, there is scope for government intervention. The government clearly has an advantage over private markets, in that it has the power to coerce citizens to pay taxes. With the tax proceeds, governments can finance the non-excludable goods.

The second characteristic of public goods (non-rivalry) makes exclusion inefficient, even if achievable: if the social marginal cost of one's consumption is zero, why should the consumption of this good be charged at all? Consider, for instance, an un-congested bridge. An un-congested bridge is non-rival in consumption. But it is possible to preclude people from (or charge people for) crossing a bridge. This property makes private provision possible, but not efficient: if the social cost of having an extra individual crossing the bridge is zero (crossing the bridge is non-rivalrous), preventing this will not be, in general, efficient. The government can fix this by publicly providing the bridge. Note however that circumstances will change if, as more and more individuals cross the bridge, the facility may become congested. With congestion, the social marginal cost of an extra individual crossing the bridge (defined in terms of the time lost by the others in attempting to cross it) becomes positive, implying that charging for its use becomes desirable. Many governmental activities, such as highways, water systems, fire services, police and courts are subject to congestion.

A number of government activities are motivated by *information failures*. Information is non-rival. However, not all information is perfectly available to all users. When relevant, the government may intervene to help information diffuse faster. For instance, by forcing

²⁰⁶ Sometimes, spontaneous voluntary associations emerge to collectively assure the provision of public goods. For instance, groups of neighbours in a resident area organize themselves and pay voluntarily for local security. But these associations work better within a small number of members, as they are in general fragile to the free-rider problem.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

firms to label their food products with the true caloric content, by forcing banks to indicate explicitly the effective rate of interest on their loans, by providing weather forecast services and publishing national statistics. By publicly providing this information, governments may improve both the welfare and productivity.

A different category of goods refers to those that are rival in consumption, but which consumption cannot be precluded. Goods in this category are called “common goods”. A classic example is the stock of fish in international waters: the fish is a rival good, but non-excludability in fishing may lead to a coordination failure, called the “tragedy of the commons”: people with access to the “common pool” will try to extract as much as possible without taking into account (because each individual is small) the impact of their actions in the aggregate. This will eventually lead to over-fishing and the depletion of the resource. In this case, excludability would be desirable. Governments can fix this, by coordinating the extraction activity, setting limits to each fisherman, and ensure sustainability.

Goods like education and health services are private goods in technical terms, because the cost of extending the supply to more users is positive (they are rival) and exclusion is relatively easy. Still, due to the positive *externalities* involved (the community as a whole benefits from a higher education level and from a lower incidence of diseases) these services will be under-supplied in a *laissez faire*. In these cases, private provision is feasible (the good is excludable), but government intervention may help boost usage to a level closer to the social optimum.

Some goods are purely private - in the sense that they are rivalrous in consumption and exclusion is feasible - but that tend to be undersupplied in a *laissez faire* because the cost of provision is more than what individuals are willing to pay. An intuitive case arises because of *economies of scale*. Consider, for example, the case of postal services. A mail company seeking for profits will optimally decide not to operate in areas where there are only few users. The government may however consider that the provision of mail services should be equally available to all citizens, irrespectively of their residence. Other cases where private markets may fail to provide desirable goods and services includes unemployment benefits, social security, research and development.

In practice, government spending covers many different areas, including health, education, infrastructures and communication networks, environmental management, water and sanitation, information and communication, scientific research. Because there is no clear-

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

cut distinction between goods that shall only be provided privately and goods that can be provided publicly, the adequate amount of public intervention is a matter of dispute in the economics profession.

10.2.3 *Intervention options*

When markets fail, there is scope for government intervention. The later, in turn, may materialize through different instruments.

One option for governments is to take *direct action*, providing the goods and services themselves. For example, if a government believes there is insufficient supply of education services, it can decide to provide it itself, running public schools. But public provision does not necessarily imply state ownership. For example, governments can contract private schools to offer schooling services and pay directly to the private school. A difficulty in this avenue that procurement contracts shall be properly designed, so as to avoid unnecessary waste and undesirable transfers from taxpayers to contractors. In many countries, infrastructure provision is private, in the sense that the government assigns rights on highways, ports or airports. Still, the location and design of the infrastructure is decided by the government because the market fails to do this properly.

Some market failures may be corrected *at distance*. Through taxes, subsidies, and rewards, the government has the potential to manipulate relative prices so that private incentives become aligned with the public interest. For instance, governments may promote the use of energy efficient cars by taxing more the less efficient cars. In education, the government can subsidize private institutions providing educational services or support directly the students, with education vouchers.

Finally, the government can intervene using regulation and legal sanctions. Governments have the right to create rules that regulate or otherwise restrict private activities, so as to minimise the incidence of undesirable market outcomes. In most countries, government agencies regulate what people can eat and drink, what kind of houses they can live in, how many hours an employee can work at most, how much pollution a factory can produce.

In general, government intervention and regulation impose costs on economic agents, by restricting their choices. Because government intervention is also a source of inefficiency,

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

governments must balance the benefits of tackling market failures with the costs related to intervention they impose in the society.

Box 10.1. Normative versus positive economics

When economists evaluate alternative policies, weighting up their various benefits and costs, they are engaged in *normative* economics. When they describe the economy and construct models to predict effects of policies or how governments will behave, they are engaged in *positive* economics.

Normative economics is concerned with what “should be” or how should governments act. Should they intervene? What are the most effective means? What are the optimal policies? Positive economics is concerned with “what is” or why policymakers do what they do. It incorporates the role of political pressures, institutional constraints, and ideological issues. Normative economics makes use of positive economics.

The market failure approach to the role of government is largely a normative approach. It provides a basis for identifying situations where the government ought to do something. Under positive analyses one should describe the consequences of government actions as well. In this assessment, a critical question is the extent to which the government can *do better* than the market.

10.3 A simple growth model with government spending

This section presents a model to illustrate the efficiency gains arising from public provision of essential inputs to production and the trade-offs involved. The model is formulated assuming diminishing returns to reproducible factors, so the relationship between optimal intervention and efficiency translates into *level* effects. Similar conclusions can however be spelled out in terms of the AK model, with the difference that efficiency will affect growth rates (this alternative formulation is developed in Box 10.4).

10.3.1 Private productivity and public inputs

Consider an economy with a large number of equal firms. Each firm produces a homogeneous consumption good according to the following production function:

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

$$Y_{it} = A_t K_{it}^\beta N_{it}^{1-\beta}, \quad (10.1)$$

where N refers to labour and K refers to private capital (which may include human capital).

Sticking with the assumption of exogenous growth, let's consider again equation (3.1) $A_t = A e^{gt}$. In this equation, we distinguish two components: one related to “efficiency”, the constant A , and the other related to “technological change”, which in this model is assumed to evolve exogenously.

In what follows, public provision is modelled as affecting directly the efficiency parameter, A - say, through lower costs in transacting. Formally, let's assume:

$$A_t = A e^{gt}, \text{ with } A = \left(\frac{G_t}{Y_t} \right)^\eta \text{ and } \eta > 0. \quad (10.2)$$

In (10.2), G refers to the amount of services provided by the government. You may think it as including expenditures related to the enforcement of property rights, the provision of public infrastructures, etc..

Equation (10.2) states that the public input is essential to production: if G is zero, private output will be zero (for instance, without a minimum protection of property rights, production cannot take place). An increase in G raises the marginal products of capital and labour. In this model, G has a non-cumulative nature (that is, public spending must be renewed each period to sustain the private economy).

Because the public good is *non-excludable*, it enters in individual production functions in the form of an externality. That is, firms do not take into account the impact of their own actions on the level of G . The relevant production function for private decisions is (10.1), where A_t is taken as given. Also note that profit maximizing firms will decline any invitation to share the costs of a spontaneous provision of G : since G is *non-excludable*, each firm will prefer to free-ride on any eventual production.

10.3.2 Congestion versus non-congestion

According to (10.2), the output of the average firm i rises with the provision of the public input *relative* to the size of the economy (Y). The implicit assumption is that the public input is rival or subject to congestion: for a given level of public provision, G , the amount of

public input available to each firm declines as output (Y) increases. In other words, when an individual firm expands its production, this acts as a negative externality to other firms.

An alternative specification would be to assume that the public input was such that an increase in the number of users did reduce the amount available to existing beneficiaries. In that case, the public input would be non-rival, and therefore a *pure* public good. To distinguish the two cases, consider the constitutional law and its enforcement. The constitutional law is purely non-rival, in the sense that it will serve a country population, irrespectively of its size. The enforcement of the constitutional law, in turn, is subject to congestion: the larger the population, the more courts and the more police services will be needed to enforce the law, everything else constant.

The distinction between the two cases is more than conceptual. In terms of the model, if the public input was assumed non-rival, the productivity term should depend on G, not on G/Y. In that case, the model would display a scale effect whereby, all else equal, larger economies should be richer or grow faster. Since this does not happen in the real world., we proceed with the case with rivalry.

10.3.3 *The aggregate production function*

To see how aggregate output relates to the availability of public input, just sum the production function (10.1) across firms (remember they are all equal) and substitute (10.2). You'll get:

$$Y_t = G_t^{\frac{\eta}{1+\eta}} K_t^{\frac{\beta}{1+\eta}} L_t^{\frac{1-\beta}{1+\eta}} \quad . \quad (10.3)$$

An interesting feature of this (aggregate) production function is that it still exhibits CRS on both private and government services (note that the sum of the three exponentials is equal to one). Moreover, the public input exhibits decreasing marginal returns: that is, each extra road or mains water pipe has a positive impact on total factor productivity that is lower than of the road or the water pipe before. This rises the question as to whether expanding too much public provision might be inefficient. In particular, if the public input crowds out private capital, beyond a certain point, additional provision will have a negative impact on output. The following model stresses this point.

10.3.4 Factor income shares

In what follows, let's assume that the public provision is financed with a tax proportional to income, τ . Each firm maximizes:

$$\pi_{it} = (1 - \tau)Y_{it} - (r_t + \delta)K_{it} - w_t N_{it} \quad . \quad (10.4)$$

The first order conditions of profit maximization are:

$$\frac{\partial \pi_i}{\partial N_i} = (1 - \tau)(1 - \beta) \frac{Y_{it}}{N_{it}} - w_t = 0, \quad \text{and} \quad (10.5)$$

$$\frac{\partial \pi_i}{\partial K_i} = (1 - \tau)\beta \frac{Y_{it}}{K_{it}} - (r_t + \delta) = 0. \quad (10.6)$$

Since all firms are equal, this leads to the following factor income shares:

$$\frac{w_t N_t}{Y_t} = (1 - \beta)(1 - \tau), \quad \text{and} \quad (10.7)$$

$$\frac{(r_t + \delta)K_t}{Y_t} = \beta(1 - \tau) \quad . \quad (10.8)$$

10.3.5 The market failure

Because government services arise as an externality, private firms do not consider them as an input to production. They *perceive* the contribution of physical capital to output as equal to β (as implied by 10.1), which is higher than the *actual* contribution, $\beta/(1 + \eta)$, (as implied by 10.3). This means “prices are not right”: without intervention (e.g, without the tax on production), factor rewards will *not be aligned* with their effective productivities.

Note that, if no tax was collected, the capital and labour income shares would be, respectively, β and $1 - \beta$. This means that all output would be exhausted on the rewards to these two factors, with nothing left to finance the public input. Since, according to (10.2) the public input is essential to production, without government intervention, this economy would not exist at all.

10.3.6 Getting the prices right

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

As equation (10.7) and (10.8) suggest, the government can use the tax rate so as to get the factor rewards aligned with the public interest. The optimal tax rate, you may guess, is the one that turns the *after-tax* factor income shares, (10.7) and (10.8), equal to the *actual* contributions of capital and labour to output, as stated in (10.3).

Analytically, you may obtain the optimal tax rate τ^G (where the superscript G refers to the Golden Rule), solving the following equation:

$$\frac{\beta}{1+\eta} = \beta(1 - \tau^G)$$

This gives:

$$\tau^G = \frac{\eta}{1+\eta} \quad (10.9)$$

This discussion reveals that not all taxes have adverse effects: in some cases, an appropriate choice of the tax rate constitute an effective tool to get incentives right. The mechanism is simple: we saw that a firm expanding its output imposes a negative externality on other producers, via lower availability of public inputs. Setting a tax that is proportional to output, the government has a perfect mechanism to deal with the congestion problem: a rise in the level of production by an individual firm suffers a penalty equal to the cost it imposes on others. Moreover, this penalty generates government revenues exactly on the amount needed to finance the increase in G that is necessary to compensate the rest of the economy for the erosion of public services per unit of output.

10.4 Intervention trade-offs

10.4.1 Unproductive public expenditures

In order to make the model more interesting, let's assume that a constant fraction ϕ of the tax proceeds are lost in *unproductive* uses. Because of excess bureaucracy, badly designed contracts, corruption, or other inefficiencies, part of the tax proceeds will not translate into the provision of public input with a positive impact on productivity²⁰⁷. Box 10.2 describes different reasons why government actions may result in waste. Box 10.3 shows an attempt to measure the government waste in a sample of OECD countries

In our model, total “unproductive expenditures”, denoted by Φ will be:

$$\Phi_t = \phi\tau Y_t \quad (10.10)$$

In this chapter, we take ϕ as an exogenous and constant parameter. In Chapter 13 we will work out a model where this parameter is endogenous.

The government budget is assumed balanced each moment in time:

$$G_t = \tau(1 - \phi)Y_t \quad \phi \geq 0 \quad (10.11)$$

Equation (10.11) shows that, when ϕ is positive, taxes are higher than the minimum needed to finance a given level of public provision.

Box 10.2. Governments fail, too

The discussion in Section 10.2 suggests that the government *has a role* in altering the working of private markets in desirable ways. A great deal of controversy exists, however, on the extent to which government intervention can *do better* than markets. The reason is that governments have their own failures in achieving their stated objectives. Even assuming that decision makers really want to maximize social welfare, there are good reasons to believe that they may not be able to reach the most efficient outcome.

²⁰⁷ Of course, there “unproductive” expenditures not related to waste. This includes, for instance, supports to culture. The society may however prefer to bear the cost of loss of efficiency in exchange for a more welfare, that this model is not capturing.

In general, one may distinguish four categories of government failures²⁰⁸:

1- *Limited information*: the optimal intervention requires a correct assessment by the government on the nature and the size of the market failure. However, the decision-maker perception may be different from the real world. Due to limited mental capacity by which to process information, governments do not have in general the information required to do what they would like to do. Limited information may preclude the government from correctly distinguishing whether its actions are really needed and to which extent. For example, the government would like to make sure that only disabled people were receiving social assistance. But it is often costly to avoid the free riding of healthy individuals pretending to be disabled. Spending more resources on screening may improve the information available to the government, but at the cost of less resources being available to the social programme.

2- *Limited control over private market response*: the success or failure of programmes in the public sector depends not only on public actions but also on how the private sector responds. For example, by introducing an unemployment benefit, the government does not know the extent to which individuals will adjust, spending more time in unemployment, searching for better jobs. Because the links between policy and outcomes (e.g, multipliers) are not well known, the intervention design and magnitude often fails to be adequate.

3- *Limited control over bureaucracy*: bureaucrats don't face the same kind of pressures on them to cut costs that firms operating in competitive markets have. To the extent that their expenses are not perfectly monitored, they may well become prodigal, spending more than the strict necessary to implement their programmes. Moreover, in many countries, public servants cannot easily be dismissed or rewarded for good performance, so bureaucracies have neither the carrots nor the sticks to provide strong individual incentives. Often, the success of a policy relies on the ability and the honesty of the entrusted officials.

²⁰⁸ See, for instance, Stiglitz, J., Rosengard, J., 2015. *Economics of the Public Sector*. WW Norton and Co,

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

4- *Limitations imposed by political processes*: even if the decision maker perceived the world as it really was, the political process through which decisions are implemented opens the door for deviations from public interest. Representatives often have incentives to act in favour of particular groups or to adopt (populist) policies that the majority of the electorate perceives to be correct, even if they know they aren't. State ownership, subsidized loans, agricultural supports, for example, are often used to serve political goals of governments, at the cost of the social interest.

All in all, while market failures provide a motivation for government intervention, governments should in each case assess the extent to which they can do better than the market. In some cases, such an assessment may lead to the conclusion that the costs of intervention exceed the benefits, so it is better not to intervene after all. Government actions should be directed only to those market failures where there is clear understanding that government intervention can make a significant difference.

Box 10.3 The public services efficiency frontier

The question as to whether governments could make a better use of the resources they use has always attracted the interest of academics and practitioners. An illustration of such an attempt, by Afonso et al, is in Figure 10.1²⁰⁹. Using a sample of 23 OECD economies, the authors first computed, for each country, a measure of “public sector performance”. This measure is defined as an average of seven sub-indicators, measuring the outcomes of intervention in key policy dimensions: the quality of public administration (confidence in the administration of justice, the size of the shadow economy, red tape and corruption); education achievements (secondary school enrolment, scores obtained by students in international tests); health (infant mortality, life expectancy at birth), public infrastructure (quality of communication and transport infrastructures), income distribution (income share of the

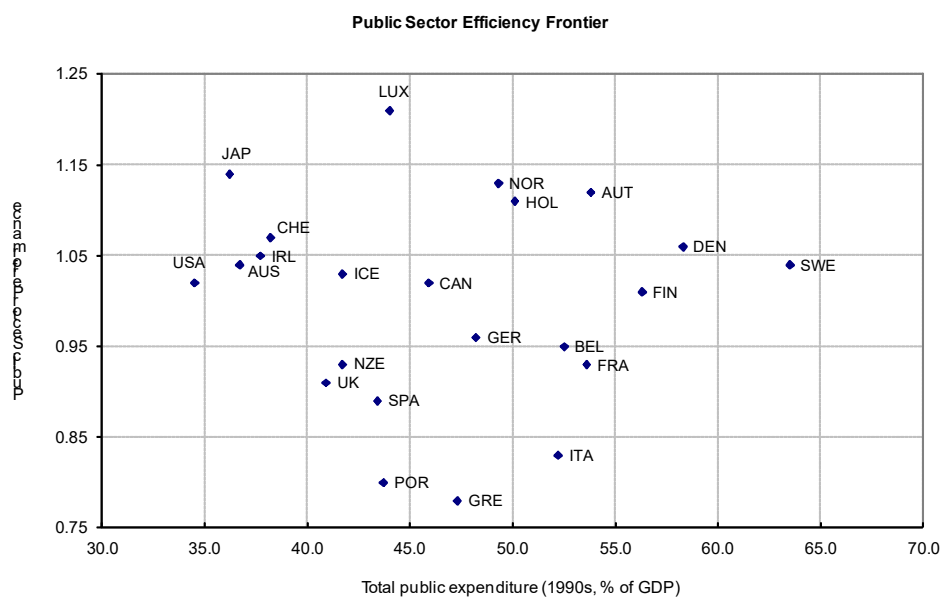
²⁰⁹ Afonso, A., Schuknecht, L., Tanzi, V., 2005. Public Sector Efficiency: an International Comparison. Public Choice 123, 321-347..

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

poorer 40%), macroeconomic stability (volatility of GDP, inflation) and macroeconomic performance (per capita GDP, GDP growth and unemployment rate). The authors then compared the estimated “Public Sector Performance Index” with total public expenditures, which they considered as input in the analysis (the output is public sector performance).

In the figure, the vertical axes measures the “Public Sector Performance Index”. According to these figures, the countries with highest public sector performance are Luxembourg, Japan, Norway and Austria. The countries with the lowest indexes are Greece, Portugal and Italy. The horizontal axes measures the total government expenditure as percentage of GDP. In the figure, we see that countries with larger spending are the Nordic countries (Sweden, Denmark and Finland) while those with smaller spending are the United States, Japan, Australia and Ireland. Note however, the limitations of the exercise: the public sector performance index is computed as a simple average of sub-indexes. If different weights were assigned to the seven dimensions, different efficiency frontiers would be obtained.

Figure 10.1 Public Sector efficiency in 23 OECD Countries (2000)



Source: Afonso et al (2005).

10.4.2 Income and expenditure

In our model, output has four different uses: private consumption (C), private investment (I), government productive expenditures (G) and government waste (Φ):

$$Y_t = C_t + \Phi_t + I_t + G_t. \quad (10.12)$$

Equation (10.12) states that one unit of public input costs the same as one unit of physical capital²¹⁰. However, in the decentralized economy there is no arbitrage condition ensuring the equality between the marginal products of physical capital and of the public inputs. Since G is not excludable, no private agent will find it profitable to produce it. It is up to the government to assure that this condition will hold after intervention.

10.4.3 The flow income chart

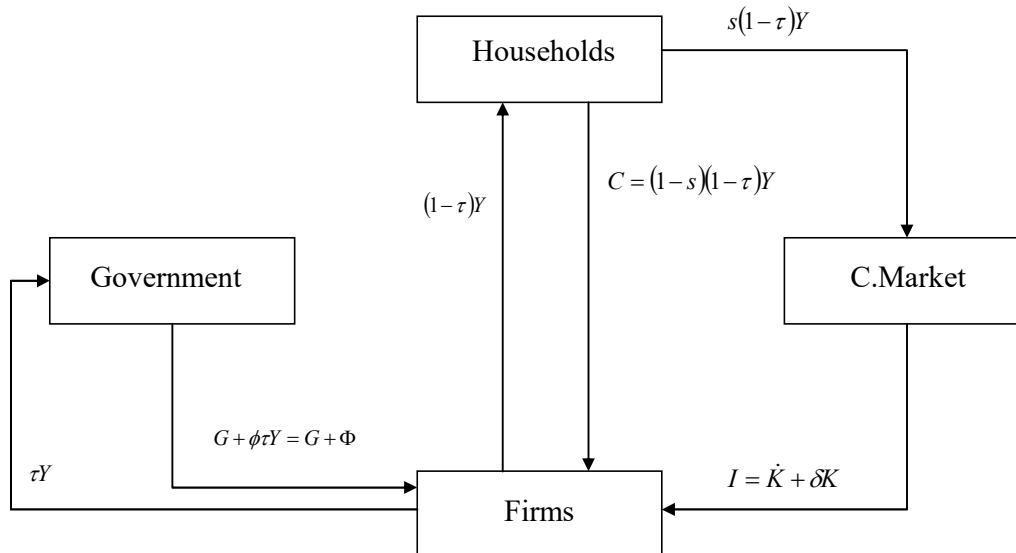
To make a long story short, all flow identities of the model are displayed in Figure 10.2, which describes the flow income chart of this economy²¹¹.

Figure 10.2: The flow income chart

²¹⁰ To keep the model simple, it is assumed that one unit of output can be either consumed or transformed into one unit of capital or to one unit of public input.

²¹¹ In this model, taxes are paid by firms. But it would be equivalent if a *uniform* tax was levied on factor incomes. The only difference would be on *who* was delivering the money to the government. We'll return to this discussion in Chapter 11.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>



10.4.4 The steady state

With the tax rate, households' disposable income declines and so it does the amount of *investment per unit of output*. Comparing to the Solow model, you may guess that the steady state level of per capita income is similar to (3.10), except in that s shall be replaced by $s(1-\tau)$ and A shall be defined as in (10.2). That is:

$$y_t^* = A^{\frac{1}{1-\beta}} \left(\frac{s(1-\tau)}{n+\delta+\gamma} \right)^{\frac{\beta}{1-\beta}} e^{\eta} = \left[\left(\frac{G}{Y} \right)^{\eta} (1-\tau)^{\beta} \right]^{\frac{1}{1-\beta}} \left(\frac{s}{n+\delta+\gamma} \right)^{\frac{\beta}{1-\beta}} e^{\eta}.$$

(10.13)

Comparing to (3.10), you see that government expenditures impact on the “efficiency term” and, by then, on the steady state *level* of per capita income through two different channels:

(i) First, a higher provision of public input raises the productivity of private inputs, raising the steady state level of per capita income;

(ii) Second, a higher tax rate, by reducing the disposable income and henceforth the investment rate, impacts negatively on the steady state level of per capita income.

10.4.5 The golden rule

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Equation (10.13) reveals a trade-off between the benefits and the costs of intervention: on one hand, a larger provision of public input raises the productivity of private capital, inducing a higher level of capital per worker in the economy; on the other hand, taxes impacts negatively on factor incomes and, by then on savings and investment. This section examines the optimal balance between these two effects.

Suppose you are a benevolent planner who wants to maximize the steady state level of per capita consumption in this economy²¹². To do this, first you need to find out the expression for per capita consumption. Using (10.11) in (10.13) and the equation for consumption in Figure 10.2, you get:

$$c_t^* = (1-s)(1-\tau)^{\frac{1}{1-\beta}} [\tau(1-\phi)]^{\frac{\eta}{1-\beta}} \left(\frac{s}{n+\delta+\gamma} \right)^{\frac{\beta}{1-\beta}} e^{\eta t} \quad (10.14)$$

This equation re-states the above-mentioned trade-off, but now in terms of the tax rate, only. It also shows that a rise in government inefficiency, ϕ , by deviating funds to unproductive uses, is equivalent to a decline in the saving rate: it crowds out private investment without any positive impact on productivity. This has an unambiguous negative impact on private consumption per capita.

If you choose the tax rate, τ , so as to maximize (10.14) you'll obtain exactly the “golden rule” tax rate, (10.9)²¹³. Substituting (10.9) in (10.13) and (10.14) you obtain the corresponding golden rule paths of per capita income and per capita consumption.

10.4.6 Efficient bureaucracy

²¹² Note that maximization of per capita consumption in the steady state does not need to correspond to the maximization of social welfare. We abstract, however, from this complication by assuming that the social value of a growth path depends only on the consumption stream associated to it in the steady state.

²¹³ Note that, by abstracting from the consumption-leisure choice, this model does not account for the impact of taxation on the labour supply. This observation illustrates the general claim that the optimal tax structure is highly sensitive to the formulation of the model. Taking this into account, one shall take the model above as a mere illustration of possible intervention trade-offs.

As mentioned before, raising government revenues by the amount needed to provide the optimal level of public inputs - as implied by (10.9) - does not necessarily imply that the government *will* actually provide the optimal level of public inputs: remember from (10.11) that, if ϕ is positive, some fraction of the tax proceeds will be wasted in unproductive uses.

With the tax rate satisfying (10.9), the amount of public input in percentage of total output is:

$$\frac{G}{Y} = (1 - \phi) \frac{\eta}{1 + \eta} \quad (10.15)$$

This means that, if you forgot to use ϕ as a control variable to maximize private consumption (10.14) you did not act as a genuine benevolent planner. If you were really a benevolent planner you should care for your constituents' money, rather than allowing it to be wasted by your bureaucrats in unproductive uses. You should pay very much attention on the way contracts with the private sector are designed, in order to keep the incentives aligned with the public interest.

Thus, if you were really a benevolent planner, you should also set $\phi = 0$, leading to:

$$\left(\frac{G}{Y} \right)^G = \frac{\eta}{1 + \eta} \quad (10.6)$$

To interpret (10.16), note that producing one unit of public services costs the same as one unit of output (equation 10.11). This means that the natural efficiency condition for the size of the government is $\partial Y / \partial G = 1$. According to (10.3), the marginal contribution of G to aggregate output is $\partial Y / \partial G = [\eta / (1 + \eta)] (Y / G)$. Substituting (10.15), this gives $1 - \phi$. Hence, only in case $\phi = 0$ will each resource used in the economy, either in the private sector or in the public sector, worth the equivalent to its opportunity cost and the economy will be operating efficiently.

10.4.7 A Graphical illustration

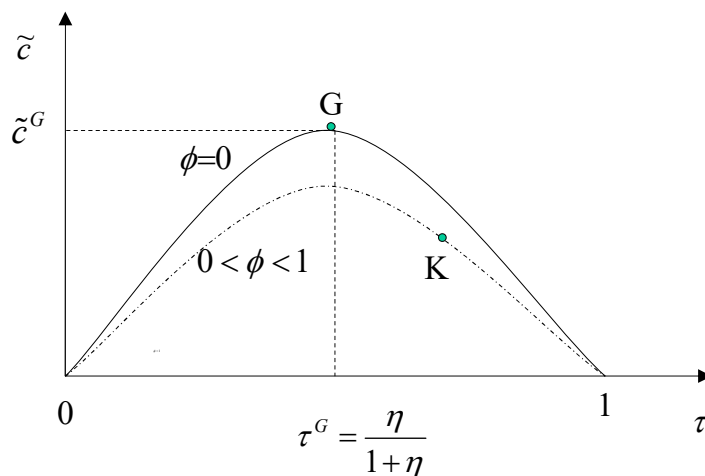
Figure 10.3 plots the steady state level of per capita consumption (per unit of efficiency labour), according to equation (10.14), as a function of the tax rate, for two different levels of ϕ . The upper curve corresponds to a government without failures ($\phi = 0$). The dashed curve corresponds to a case in which $0 < \phi < 1$, with a lower level of per

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

capita consumption for each tax rate. When $\phi=1$, the provision of public input is zero, so per capita income and per capita consumption will also be zero, irrespectively of the tax rate (the curve is flat and coincides with the horizontal axes).

For each given value of ϕ , there is a curve representing the relationship between the government size and the steady state level of per capita consumption (per unit of efficiency labour, L). At lower values of τ , the positive effect (i) described in Equation 10.13 dominates the negative effect (ii), so increasing the size of the government raises per capita consumption. As the size of the government rises, the benefits of expanding further the provision of public services declines, while the negative impact of taxation (ii) rises. At higher levels of taxation, the effect (ii) dominates (i), so a further rise in τ decreases the steady state level of per capita consumption and the curve slopes negatively.

Figure 10.3. The trade-off between public provision and taxation



Box 10.4. The case with an AK production function

This appendix shows how the model changes when the neo-classical production function (10.1) is replaced by an AK production function:

$$Y_{it} = A_t K_{it} \quad (10.1a)$$

As before, it is assumed that productivity depends on the provision of a public input, scaled by the level of activity, due to congestion:

$$A = \left(\frac{G}{Y} \right)^\eta \quad (10.2a)$$

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Considering an exogenous saving rate, s , and the income tax, τ , the growth rate of per capita income will be $\gamma = s(1-\tau)A - (n+\delta)$. Using the government budget constraint (10.11), the equivalent to (10.14) in the AK economy is:

$$\gamma = s(1-\tau)[\tau(1-\phi)]^\eta - (n+\delta) \quad (10.14a)$$

Comparing to the model with diminishing returns in the main text, we see that now government policies impact on *growth rates*, rather than in levels. In this version of the model, maximization of welfare corresponds to maximizing the growth rate of per capita income²¹⁴.

In case savings are endogenous, the Euler equation $\gamma = r - \rho$ applies. From profit maximization, the interest rate is $r = A(1-\tau) - \delta$, and the growth rate of per capita income becomes:

$$\gamma = (1-\tau)[\tau(1-\phi)]^\eta - \delta - \rho \quad (10.14b)$$

In both cases, a figure similar to 10.3 applies, with the difference that in the vertical axes per capita consumption is replaced by its growth rate. As before, there are two opposite effects: (i) a higher provision of public services raises the productivity of private investment, inducing a *faster* capital accumulation; (ii) a higher tax rate reduces the net worth of private investment, inducing a *slower* rate of capital accumulation. The benevolent planner chooses τ and ϕ to maximize (10.14a) or (10.14b), depending on the context. This gives (10-9) and (10.16). Again, the optimal policy corresponds to setting the size of government provision proportional to its impact on aggregate production.

²¹⁴ Barro and Sala-i-Martin, 1995, pp 156.

10.5 Key ideas of chapter 10

- Because of different types of market failures, in general the laissez faire does not deliver in general an efficient resource allocation. Market failures include externalities, public goods, common goods, coordination failures, missing markets, high unemployment, information failures, and imperfect competition.
- A fundamental function that underlies the origin of the state is the establishment of essential institutions, such as the rule of law, protection of property rights, and market regulation.
- This chapter presented an extended version of the productivity term accounting for the role of government in providing services that are non-excludable, but essential to the functioning of a market economy. The implication is that in a laissez faire, there would be no provision and hence no economy at all.
- In the model, the government solves the market failure by coercing people to pay taxes. The optimal intervention involves setting the tax rate so that private prices become aligned with the social interest.
- The model emphasizes the trade-off between the benefits of public provision and the cost of taxation, which in this model lowers private savings. Since government services exhibit diminishing returns, there is an optimal scale of public provision. Beyond that level, intervention becomes counter-productive.
- The first best policy presumes that there are no losses in the process of transforming tax proceeds into public inputs. In practice, different types of “government failures” cause some of the tax proceeds to be wasted in unproductive uses.

Problems and Exercises

Key concepts

- *Market failures. Public goods. Government failures. Intervention trade-offs.*

Exercises

- 10.1. (Levels)** Consider an economy with a large number of equal firms. Each firm i produces a homogeneous consumption good according to $Y_i = AK_i^{0.5}N_i^{0.5}$. Although each firm considers A as an exogenous parameter, in the aggregate the following condition holds : $A = (G/Y)^{0.5}$. Assume that in this economy the population is constant ($n=0$), there is no technological progress ($\gamma=0$), the saving rate is 27% and the capital depreciation rate (δ) is equal to 2%. (a) Compute the aggregate production function in this economy and explain why there is a market failure. What would happen if there was no government? (b) Assume that the provision of public inputs is financed with a production tax τ , but that a fraction θ of the tax revenues is wasted in unproductive uses. Write down the government budget constraint. (c) Find out the expression for the steady state levels of per capita income and per capita consumption in terms of the

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

fundamental parameters. Clue: remember that $c = (1-s)(1-\tau)y$. (d) Explain, with the help of a graph, the dual effect of the tax rate on the steady state level of per capita consumption. (e) How does θ affect the steady-state of private per capita consumption? Explain. (f) Find out the benevolent planner solution. Is this solution intuitive? (g) Examine the implications of a positive waste θ for per capita consumption and G/Y . Discuss.

- 10.2. (Growth)** Consider a closed economy where firms perceive the production function to be of the form $Y=AK$. In this economy the population is constant, the saving rate is equal to $s=0.21$, the depreciation rate is equal to $\delta=0.05$, and there is a tax on production at the rate τ . (a) Consider for the moment that $A=0.4$ and $\tau=0.4$. Using the equality between savings and gross investment, find out the growth rate of per capita income in this economy. (b) Place the main income identities of this economy in a flow income chart. (c) Discuss, with the help of a graph the dynamic properties of this model. Does this model predict convergence among similar economies? (d) Examine the implications of a change in the tax rate from $\tau=0.4$ to $\tau=0.5$, all else equal. (e) Assume now that $A=(G/Y)$, where G is a non-excludable good. Find out the expression for aggregate output. (f) Explain the market failure. (g) Describe the benevolent planner solution.
- 10.3.** Consider a closed economy where firms perceive the production function to be of the form $Y=AK$. In this economy the population is constant, the saving rate is equal to $s=0.24$ and the depreciation rate is equal to $\delta=0.04$. Assume also that the government levies a tax τ_K on household's capital incomes. (a) From the firm's maximization problem, find out the expression for the interest rate in this economy as a function of the tax rate. (b) Find out the expression for the household's disposable income. Place the main income identities of this economy in a flow income chart. (c) Consider for the moment that $A=1/3$ and $\tau_K=1/8$. Using the equality between savings and gross investment, find out the growth rate of per capita income in this economy. (d) Discuss, with the help of a graph the dynamic properties of this model. Does this model predict convergence among similar economies? (e) Examine the implications of a change in the tax rate from $\tau_K=1/8$ to $\tau_K=1/3$. (f) Keeping the tax rate equal to $\tau_K=1/3$, examine now the implications of a change in the efficiency parameter from $A=1/3$ to $A=1/2$. Compare the implications of such a change with a similar change in the context of the Solow model and explain. (g) Assume now that $A=(G/Y)^{0.5}$, where G is a nonexcludable good. Find out the expression for aggregate output. (h) Assuming that the government budget constraint is $G = [\tau_K / (1 + \tau_K)] Y$, compute the growth rate of this economy when $\tau_K=1/8$ and when $\tau_K=1/3$. In light of this new interpretation for A , how would you explain the equilibrium described in e? Compare the 3 solutions with the help of a figure relating the growth rate of the economy with the tax rate (laffer curve). Do any of these correspond to the first best?
- 10.4.** Consider a closed economy where firms perceive the production function to be of the form $Y=AK$. In this economy the population is constant, $\rho=0$, and the depreciation rate is equal to $\delta=0.04$. Assume also that the government levies a tax τ_K on household's capital incomes. (a) From the firm's maximization problem, find out the expression for the interest rate in this economy as a function of the tax rate. (b) Find out

the expression for the households' disposable income. (c) Assume now that $A = (G/Y)^{0.5}$, where G is a public good. Find out the expression for aggregate output. (d) Considering the government budget constraint $G = [\tau/(1+\tau)]Y$, find out the benevolent planner solution. Explain it with the help of a graph.

11 Distortions

“People do what they get paid to do; what they don’t get paid to do, they don’t do”.
[William Easterly].

Learning Goals:

- Identify real world examples of market and government-imposed distortions.
- Understand how taxes and subsidies can be manipulated to “get the prices right”.
- Understand the intervention dilemmas in a second-best framework.

11.1 Introduction

In a well-functioning economy, each resource is valued in the market according to its contribution to social welfare. In the real life, however, different types of impediments drive wedges between market prices and social returns. These cases are referred to as distortions. Whether government-imposed (e.g., taxes) or inherent to certain markets (e.g., externalities, natural monopolies), distortions cause misallocation of resources keeping the economy below its attainable productivity frontier.

This chapter provides a systematic view of the distortions that interfere with capital accumulation and with the allocation of different inputs to production. For mathematical convenience, the underlying framework is the AK model. This means that efficiency losses caused by distortions will translate into lower *growth*. A similar analysis could be spelled out

in terms of the neoclassical model, with level effects. Sticking with the AK model, however, we gain in mathematical simplicity²¹⁵.

In Section 11.2 we discuss the case of distortions that affect consumption-saving decisions. In Section 11.3 we address the case where the distortion affects the relative use of two inputs to production. Section 11.4 extends the analysis in the previous section to the case in which a policy to subsidize one input comes at the cost of a tax in other input. Section 11.5 summarizes.

11.2 Distortion in the consumption-investment decisions

This section gives an example, whereby a distortion affecting the relative price of capital leads to a suboptimal rate of capital accumulation. The simpler manner to model a market distortion is to think it as an unjustified tax that introduces a wedge between private returns and social returns. In a broader interpretation, you can interpret this “tax” as capturing the effect of any policy or institution that reduces artificially the return on investment. This includes bureaucratic barriers to investment, licensing, tariffs on imported equipment, financial market frictions, corruption fees on entrepreneurs, etc.

Assume that the aggregate production function takes the AK form:

$$Y_t = AK_t, \quad A > 0 \quad (11.1)$$

Where K denotes for private (excludable, rival) inputs, only²¹⁶.

It is also assumed that households are able to smooth consumption, according to:

²¹⁵ The chapter draws on Easterly (1993, 2005). Easterly, W. 1993. "How much do distortions affect growth?". *Journal of Monetary Economics* 32, 187-212. Easterly, W. 2005. National policies and economic growth: a reappraisal. In Aghion, P., and Durlauf, S. (eds), *Handbook of Economic Growth*, North Holland, Amsterdam.

²¹⁶ You may interpret K as including both human and physical capital. As long as the tax rate is uniform across capital inputs, there is no gain in modelling the different types of capital separately. Later in this chapter we will address specifically the case with non-uniform taxation.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

$$\gamma = r - \rho \quad (11.2)$$

Because in this model K is a purely private good and consumption-saving decisions are optimal there are no market failures. Thus, there is *no role* for government intervention. The market mechanism delivers the first best outcome and any government attempt to interfere in the price mechanism will be welfare worsening.

Assume now that the government imposes a tax, τ_K , on capital incomes, which proceeds are used for unproductive consumption. Profit maximization by an individual firm i takes the following form:

$$\pi_{it} = AK_{it} - (r_t + \delta)(1 + \tau_K)K_{it} \quad (11.3)$$

The first order condition of profit maximization is:

$$\frac{\partial \pi_{it}}{\partial K_{it}} = A - (r_t + \delta)(1 + \tau_K) = 0$$

With all firms equal, the economy's interest rate will be:

$$r = \frac{A}{(1 + \tau_K)} - \delta \quad (11.4)$$

From (11.4), the households' disposable income (Y_d) will be equal to:

$$Y_d = (r + \delta)K = Y/(1 + \tau_K) \quad (11.5)$$

To see how much the tax rate affects growth, let's first solve the model as if the saving rate was *exogenous*. Because the saving rate applies to Y_d , the equality between savings and gross investment becomes:

$$\frac{sY}{1 + \tau_K} = \dot{K} + \delta K \quad (11.6)$$

Dividing both members of (11.6) by K , rearranging and subtracting n on both sides, one obtains the growth rate of capital per worker (and of per capita income), for each level of the saving rate:

$$\gamma = \frac{\dot{K}}{K} - n = \frac{sA}{1 + \tau_K} - (n + \delta) \quad (11.7)$$

This equation shows that an increase in the tax rate, by reducing the households' disposable income, impacts negatively on capital accumulation and by then, on growth, *for each* level of the saving rate. When the saving rate is endogenous, it will depend on the reward of capital. Because the distortion impacts negatively on capital returns (equation 11.4) consumers will optimally decide to save less.

An interesting implication of (11.7) and (11.8) is that the lower the A , the lower the derivative of growth with respect to the tax rate, τ_K . This means that the distortion is less pervasive when TFP is low. In the words of William Easterly, *bad policies are more likely to be tolerated in low-A countries than in high-A countries*. This, in turn, may help perpetuate the bad policies! This observation has important implications for convergence: if rich countries adopt better policies, then convergence forces like technological diffusion and imitation of best practices might not be enough to induce cross-country economic convergence. If that was the case, the *initial conditions* that shaped the differences in A at the very beginning - natural causes like the quality of arable land and climate should have a key role in explaining the cross-country differences in per capita incomes today.

11.2.1 The case with endogenous savings

Replacing (11.4) in the optimal consumption rule (11.2), one obtains the growth rate of per capita income in the economy with endogenous savings:

$$\gamma = \frac{A}{1 + \tau_K} - \rho - \delta \quad (11.8)$$

Comparing to (11.7), we see that the impact of taxation is much larger in this version of the model. The reason is that, when savings are endogenous, an increase in the tax rate not only decreases the productivity of capital but also the saving rate. To disentangle the two effects, let's equal (11.8) and (11.7) and solve for the (endogenous) saving rate:

$$s = 1 - \frac{\rho - n}{A} (1 + \tau_K) \quad (11.9)$$

This equation shows that an increase in the tax rate leads households to save less out of their disposable income. Thus, equation (11.8) accounts for two effects: on one hand, by reducing the households' disposable income, τ_K reduces the total amount of savings - and by

then investment – for a given saving rate; on the other hand, by reducing the return on investment, τ_K induces a lower saving rate.

It is important to distinguish the circumstances in which, when assessing the growth prospects of a given country, one shall refer to equation (11.7) or instead to equation (11.8). The difference is that (11.8) presumes that the financial system is well developed, so that households are able to transfer income across time through borrowing and lending. This is less likely to be the case wherever financial systems are underdeveloped and households face borrowing constraints. This conjecture reinforces the previous conclusion: bad economic policies (as reflected in lower parameter A) are likely to impact more severely in countries with developed financial systems (with endogenous savings) than in countries with underdeveloped financial systems (where the saving rate is not necessarily responsive to the interest rate).

This discussion adds to the general point that questions like “what happens to per capita income (or to economic growth) when some parameter increases” do not have a unique answer: it depends on specific circumstances.

Box 11.1 Tax on production versus tax on capital

In the model (11.1)-(11-8) we use a tax on capital returns, τ_K . Conclusions do not change if we set a tax on production, τ , instead. In that case, the interest rate will be

$$r = sA(1 - \tau) - \delta, \quad (11.4a)$$

And the growth rate in the case with endogenous savings becomes:

$$\gamma = A(1 - \tau) - \rho - \delta \quad (11.8a)$$

As you may easily check, the tax on production equivalent to a given tax on capital returns is:

$$\tau = 1 - \frac{1}{1 + \tau_K} \quad (11.10)$$

11.2.2 Example: Financial intermediation and economic growth

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

When we described the financial system in the flow income chart, we ignored the frictions that underlie the transfer of funds as between savers and borrowers. In the real world, however, financial transactions involve transaction costs, most of which related to information failures. These include costs related to searching and matching lenders and borrowers (brokerage costs), gathering information on the borrowers to evaluate their risk characteristics (evaluation costs), negotiating and designing contracts (negotiation costs), monitoring the implementation of projects (agency costs) and assuring the enforcement of the contract's obligations (enforcement costs). All these costs imply that the gross return paid by the borrower exceeds the net return received by the lender. The same applies to some types of government intervention in the banking system, such as the imposition of high reserve requirements, or to imperfect competition in the banking system.

In term of our model, the costs related to financial intermediation can be captured by τ_K . That is: for each euro saved, only $1/(1+\tau_K)$ translates into acquisition of new capital; the remaining $\tau_K/(1+\tau_K)$ is retained in the financial system. In that interpretation, a less efficient transformation of savings into investment will translate into a lower τ_K . Wherever financial institutions are fragile, poorly regulated, and contract enforcement is low, one will expects the intermediation spreads to be higher, implying more losses in the process of transformation of savings into investment.

Financial development makes investment more attractive for borrowers and savings more attractive for lenders. By allowing risk to be spread across different assets and reallocated from risk averse agents to risk takers, by pulling together small and short term savings of many households and channelling them to large long term loans to finance big projects, by scrutinizing risks, by collecting and releasing relevant information, the financial system helps deliver more favourable risk-return-maturity combinations for savers while creating incentives for entrepreneurs to engage in long-term and riskier projects, which

otherwise would not be under consideration²¹⁷. Empirically, there is an extensive evidence showing that financial development is a key determinant of economic performance.

Box 11.2: the research of King and Levine

In a series of research papers, Robert King and Ross Levine analysed the relationship between financial development and economic performance. In one of these papers²¹⁸ the authors assessed whether the level of financial development affects real per capita GDP growth, (2) capital accumulation and (3) productivity growth. This study involved 80 countries over the period 1960-1989.

Table 11.1 reproduces some of the author's findings. In the table, dependent variables are in columns and independent variables are in rows. The later group includes the log of initial per capita GDP (capturing conditional convergence), the log of secondary school enrolment, government consumption divided by GDP, inflation, a measure of trade openness and a measure of financial depth, consisting on the amount of a country liquid liabilities (currency plus demand and interest bearing liabilities of financial intermediaries) divided by GDP. As shown in the table, the three regressions suggest that financial development is a good predictor of growth, capital accumulation and TFP.

Table 11.1: Growth and initial financial depth, 1960-1989

²¹⁷ The link between finance and economic growth also runs through the efficiency parameter A : for instance, the function of evaluating and selecting the most profitable projects and of monitoring their implementation tends to enhance the average productivity of capital. Sometimes, governments try to influence banks' decisions. Whenever this translates into the deviation of credit to socially inefficient projects or to loss-making companies, there will be a negative impact on efficiency and thereby on the rate of economic growth.

²¹⁸ King, R., Levine, R., 1993. Finance and growth: Schumpeter might be right. Quarterly Journal of Economics, 108 (3), 717-37

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

	Per capita GDP Growth, 1960-1989	Per Capita Capital Growth, 1960-1989	Per capita Productivity Growth, 1960-1989
Constant	0,035*** [0,001]	0,002 [0,682]	0,034*** [0,001]
Log(real GDP per person in 1960)	-0,016*** [0,001]	-0,004* [0,068]	-0,015** [0,001]
Log(secondary school enrollment in 1960)	0,013*** [0,001]	0,007*** [0,001]	0,011*** [0,001]
Government Consumption/GDP in 1960	0,07* [0,051]	0,049* [0,064]	0,056* [0,076]
Inflation in 1960	0,037 [0,239]	0,02 [0,0238]	0,029 [0,292]
(Imports plus exports)/GDP in 1960	-0,003 [0,604]	-0,001 [0,767]	-0,003 [0,603]
DEPTH (Liquid Liabilities in 1960)	0,028*** [0,001]	0,019*** [0,001]	0,022*** [0,001]
R ²	0,61	0,63	0,58

*significant at the 0,10 level, **significant at the 0,05 level, ***significant at the 0,01 level.

[p-values in brackets]

Observations = 57

Source: King and Levine (1993a)

11.3 Discriminatory taxation

11.3.1 The efficient allocation

The model in Section 11.2 describes the growth effects of policies that impact in the consumption-saving decisions. By pooling together all forms of capital, however, the analysis misses an important category of distortions: those that affect the relative user cost of different inputs. This section and the section that follows address specifically this case.

Assume that there are two types of private capital: human capital (H) and physical capital (K). More generally, you can interpret K and H as simply referring to different types of capital, such as machinery versus vehicles, imported equipment versus domestically produced, etc. The aggregate production function takes the following form:

$$Y = AK^{\beta}H^{1-\beta} \quad (11.11)$$

Also assume that one unit of output can be converted at no cost into either one unit of physical capital or into one unit of human capital: that is, the opportunity costs of investing on unit of output in human capital and in physical capital are the same. It is also assumed that

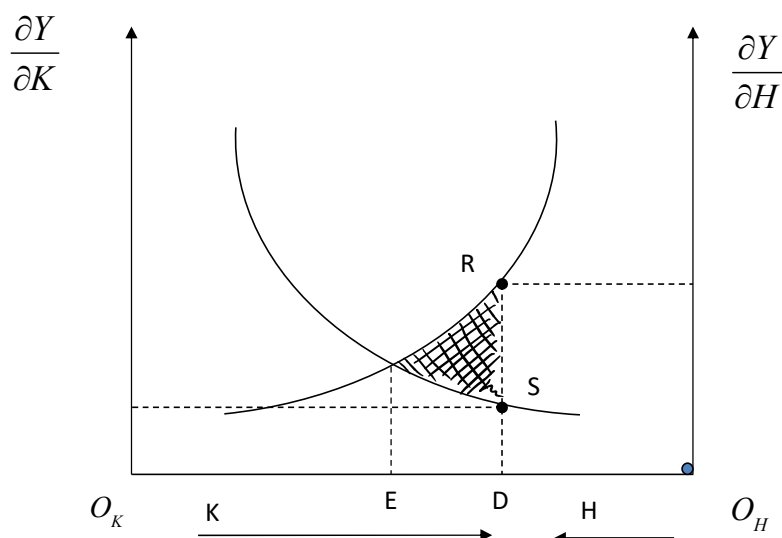
the depreciation rates for the two types of capital are the same. These assumptions are not necessarily realistic, but allow us to abstract from other effects.

In a competitive equilibrium with no market failures, profit maximization ensures that human and physical capital are employed so that their marginal products are equal. That is:

$$\frac{H}{K} = \frac{1-\beta}{\beta} \quad (11.12)$$

Figure 11.1 illustrates this. In the figure, the horizontal axes measures the *total* amount of capital available in the economy at a given moment in time, with the endowment of physical capital being measured from left to right and the endowment of human capital being measured from right to left. The vertical axes measures the marginal products of the two types of capital. The equilibrium described by allocation E corresponds to efficient allocation: if both types of capital cost the same and if their depreciation rates are equal, then profit maximization and absence of arbitrage opportunities will ensure that their marginal products are the same.

Figure 11.1 Marginal products of physical and human capital



In figure 11.1, an allocation like D is inefficient. In allocation D, more physical capital and less human capital are used in comparison to E. Due to diminishing returns, the marginal product of physical capital in D (DS) is lower than the marginal product of human capital (RD). This means that more output could be produced if less physical capital and more human capital were employed in production. The shaded area is a *deadweight loss*.

The implication for economic growth is straightforward. Since in the AK model efficiency has growth effects, allocation E is the one that delivers the maximum possible rate of economic growth²¹⁹. In a context with diminishing returns, a similar conclusion would hold, but in levels.

11.3.2 Distortionary taxes

We now examine how factors affecting the user costs of the two types of capital cause firms to deviate from the socially optimal allocation. We illustrate this assuming that the government taxes differently the returns on physical capital and on human capital. Consider an economy composed by many identical firms, with production functions of the form:

$$Y_i = AK_i^\beta H_i^{1-\beta}. \quad (11.11a)$$

Let τ_K be the tax rate on physical capital incomes and τ_H the tax rate on human capital incomes. These taxes create a wedge between the user cost of capital, paid by firms, and the net worth to households. The individual firm's profits become:

$$\pi_i = Y - (1 + \tau_K)(r + \delta)K_i - (1 + \tau_H)(r + \delta)H_i. \quad (11.13)$$

Profit maximization in that case leads to:

$$\frac{\partial Y_i}{\partial K_i} = \beta \frac{Y_i}{K_i} = (r + \delta)(1 + \tau_K); \quad (11.14)$$

$$\frac{\partial Y_i}{\partial H_i} = (1 - \beta) \frac{Y_i}{H_i} = (r + \delta)(1 + \tau_H). \quad (11.15)$$

²¹⁹ To see this, let $\tilde{k} = K/H$ denote for the ratio of physical to human capital. The question is to find out the value of $\tilde{k}$ that delivers faster growth. To solve this problem, rewrite the production function, as $Y = [\tilde{k}^\beta / (1 + \tilde{k})] \bar{K}$, where $\bar{K} = K + H$ denotes for the total capital in the economy each moment in time. Assuming for a moment an exogenous saving rate, the change in total capital will be given by $\gamma = \dot{\bar{K}} / \bar{K} = sY / \bar{K} - \delta = s[\tilde{k}^\beta / (1 + \tilde{k})] - \delta$. Maximizing this in respect to $\tilde{k}$, you will obtain $\tilde{k} = \beta / (1 - \beta)$.

Due to arbitrage, the *net* rental prices of human and physical capital ought to be equal, implying:

$$\frac{H_i}{K_i} = \frac{H}{K} = \frac{1-\beta}{\beta} \frac{1+\tau_K}{1+\tau_H} \quad (11.16)$$

Equation (11.16) illustrates how discriminatory taxation, affecting the relative use of the two types of capital, impacts on the marginal rate of technical substitution. If both tax rates were zero, the economy would achieve the most efficient allocation, E.

Summing (11.11a) across firms and substituting (11.16), one obtains the aggregate production function as a function of the tax rates:

$$Y_t = A \left(\frac{1+\tau_K}{1+\tau_H} \right)^{1-\beta} \left(\frac{1-\beta}{\beta} \right)^{1-\beta} K_t, \quad (11.17)$$

The novelty in this version of the AK model is that the efficiency component (the average productivity of capital) now depends on two policy parameters, τ_K and τ_H . Solving (11.14) for the interest rate, substituting (11.17) and using the optimal consumption rule (11.2), one obtains the growth rate of per capita income in an economy with endogenous savings:

$$\gamma = \frac{B}{P_I} - \delta - \rho \quad (11.18)$$

where

$$B = A\beta^\beta (1-\beta)^{1-\beta} \quad (11.19)$$

$$P_I = (1+\tau_K)^\beta (1+\tau_H)^{1-\beta}. \quad (11.20)$$

The term P_I is the average user cost of capital.

To interpret equation (11.18), consider first the case in which $\tau_K = 0$ and $\tau_H = 0$. In this case, $P_I = 1$, the marginal products of the two types of capital are equal and the maximum possible growth rate is achieved. Now assume that $\tau_K = \tau_H > 0$, e.g., both taxes are positive and equal. In this case, equation (11.16) becomes equal to (11.12): there is a penalty on investment, but there is no distortion on the *relative use* of the two types of capital. The economy will grow less because of the overall tax on investment (like in 11.8), but there is no efficiency loss in the use of the two types of capital. The economy is still operating in point <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

E. Finally, when taxation is not uniform (say, $\tau_H > 0$ and $\tau_K = 0$), there will be distortions affecting both the relative price of capital P_I , and the relative price of the two types of capital. In terms of Figure 11.1, the economy will be operating in an allocation different from E.

11.3.3 Self-financed subsidies

In light of (11.20), while taxes on capital are harmful to growth, subsidies to capital lead to faster growth. A question might be raised then: should the government give large subsidies to capital accumulation, so as to induce faster growth? To answer this question, one needs to take into account the fact that subsidies must be financed. Since lump-sum taxes are not, in general, available, the positive effect on growth due to a higher subsidy on a type of capital must be balanced against the negative effects of distortionary taxation anywhere else.

To analyse this problem in the context of our model, suppose that the government launches a subsidy to human capital accumulation financed with the proceeds of a tax on physical capital. For a moment, assume that there are no government services ($G=0$). The government budget constraint is:

$$(r + \delta)(\tau_H H + \tau_K K) = 0 \quad (11.21)$$

The main point here is that, when the subsidy to human capital is financed with the proceeds of taxation on physical capital, P_I rises unambiguously. The reason is the policy alters the relative price of the two inputs, according to (11.16), giving rise to substitution effects that shrink the tax base and enlarge the subsidy base. Thus, when the subsidy rate increases, this requires a more than one-for-one increase in the tax rate in order to keep the budget balance unchanged. Thus, the price of capital rises unambiguously, and the rate of economic growth is lower^{220 221}.

²²⁰ Mathematically, replacing (11.21) in (11.20) and using (11.16), we get $P_I = (1 + \tau_H) \left[\beta / (\beta + \tau_H) \right]^\beta$. This price is minimum when $\tau_H = 0$.

Box 11.3: Other Distortions

In the real world, different factors may alter the relative user cost of different inputs to firms, preventing an economy from operating in a point like E. Among others, taxes, tariffs, import quotas, price controls, interest rate controls, dual exchange rates, employment protection rules, nominal rigidities, corruption fees and imperfect competition. The following sections provide some examples.

- Import protection: In many developing countries, import tariffs are used to raise government revenues. In terms of the model outlined above, one may examine the effects of import protection interpreting τ_K and τ_H as tariffs, instead as of taxes. Suppose that H refers to imported equipment while K refers to locally produced equipment. If a tariff is levied on the imported capital, this will lead to a distortion in the relative use of the two types of capital. The same applies to non-tariff import protection mechanisms, such as import quotas, licenses, and technical requirements. All these instruments will cause the relative price of inputs to depart from the competitive equilibrium, causing the economy to move away from E.
- High inflation: Inflation, by affecting nominal variables but not real variables, impacts differently in the user costs of different inputs. In terms of our model, let K denote for physical capital, and H for firms' working capital. In most activities, *working capital* is an essential input to production, consisting in cash, short-term financial instruments, and credit to customers. High inflation, , drives a wedge between the costs of holding working capital (except inventories) relative to, for instance, equipment, which secondary

²²¹ Even if lump-sum taxation was available, a question would arise as to why should the government try to maximize the growth rate of the economy: if there were no distortions and private agents optimally decide their consumption paths, there would be no point in trying to modify the competitive growth rate. Policies that artificially increase the growth rate of the economy are in this context detrimental to welfare and shall be compared to immiserising growth (Easterly, 2005).

market price goes along with inflation. In response, firm will tend to operate with less working capital than in the efficient allocation²²².

- Monopoly: a common source of efficiency-loss is imperfect competition: whenever one side of the market is characterized by a small number of players, these may have the ability to influence prices. In terms of our model, suppose that the supply of human capital was unionized and that the only concern of the union was to maximize its monopsony rents. In that case, the cost of using human capital would rise relative to that of physical capital, inducing a departure from point²²³.
- Segmented labour markets: In many economies, there are barriers to labour mobility that prevent otherwise identical workers get paid equally. Sources of labour market segmentation include differences in the institutional nature of the employer (public vs. private), restrictive labour laws (high severance payments, restrictive conditions for dismissals), dual productive structures (modern vs. traditional, formal vs. informal), barriers to geographical mobility (housing rents, high licensing costs) or the rational for employees to pay wages above the competitive level (efficiency wages). All these factors may prevent real wages to be equalized across workers with similar qualifications. In terms of hour model, these frictions imply different user costs in using different inputs (in this case, two types of human capital), causing again a departure from allocation E.
- *Price controls*: in many countries, governments set maximum price ceilings in some goods, to promote their consumption. Such policy calls for

²²² On the negative relationship between inflation and growth, see Barro (1995, 1997), Sarrel (1996), Gylfason (1998). Bruno and Easterly (1998).

²²³ Comparing to (11.16), we see that a monopolized market for input H is equivalent to a subsidy to physical capital accumulation. This suggests that the government can use taxes and subsidies to remove the distortion.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

budgetary transfers so as to compensate firms from the implied losses. If taxes on other goods are levied to finance the policy, this is equivalent to a tax-cum-subsidy scheme that impacts negatively on efficiency and therefore on growth.

- Ceiling in banks' lending rates: sometimes, central banks impose maximum interest rates on bank loans, giving rise to excess demand in credit markets. In response, parallel markets of finance tend to develop, at very high interest rates. In the official market, the scarce credit is often allocated administratively, favouring "priority sectors" or state-owned enterprises. also induces. All in all, the segmentation of the credit market acts as a tax-cum-subsidy scheme: some types of borrowers are implicitly subsidized at the cost of other borrowers.
- *Dual exchange rates.* In many developing countries, central banks set an official exchange rate, but fail to enforce it in the market, buying or selling foreign exchange. In these cases, the currency is said to be unconvertible. Inconvertible currencies are typically characterized by a dual exchange rate system, with an official exchange rate available to public companies or other priority bodies, and a black market exchange rate for the rest of the economy. Those who are allowed to import goods at the official exchange rate receive an implicit subsidy, while those forced to pay the black market rate pay an implicit tax. In terms of the model, this is equivalent to a tax-cum-subsidy scheme.

11.3.4 Tax evasion

A major problem in government finance is tax evasion. When a significant share of economic activity is informal and doesn't pay taxes, the burden of financing government activities falls on a narrower base, giving rise to unfair competition, wrong incentives, and an implicit transfer from those who pay taxes to those who benefit with the public good without paying taxes. This phenomenon is particularly pervasive in developing countries.

To analyse this, assume that a non-excludable input G is essential to production:

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

$$A = \left(\frac{G}{Y} \right)^\eta \quad (11.22)$$

Substituting (11.22) in (11.11), aggregating up and rearranging, one obtains the economy' production function in terms of the three inputs:

$$Y = G^{\frac{\eta}{1+\eta}} K^{\frac{\beta}{1+\eta}} H^{\frac{1-\beta}{1+\eta}} \quad (11.23)$$

This equation reveals that the *actual* contributions of physical capital and human capital to production are less than what individual firms *perceive* to be, in (11.11a).

The question now is what combination of τ_K and τ_H shall the government use to finance the provision of government services? With public provision, the government budget constraint becomes:

$$(r + \delta)(\tau_H H + \tau_K K) = G \quad (11.24)$$

The rest of the model is equal to that in section 11.3.2: profit maximization at the firm level leads to (11.14) and (11.15) and the growth rate of the economy is just as described by (11.18). The only difference is that now the parameter A (inside B) is related to public provision, according to (11.22). Using (11.14) and (11.15), the government balanced budget constraints simplifies to:

$$\frac{G}{Y} = \frac{\beta \tau_K}{1 + \tau_K} + \frac{(1 - \beta) \tau_H}{1 + \tau_H} \quad (11.25)$$

Substituting this in (11.22), one obtains the expression for A in terms of the two tax rates:

$$A = \left(\frac{\beta \tau_K}{1 + \tau_K} + \frac{(1 - \beta) \tau_H}{1 + \tau_H} \right)^\eta \quad (11.26)$$

The growth rate of this economy is given by (11.18), with the only difference that A shall now be replaced by (11.26).

A benevolent planner in this economy would choose τ_K and τ_H so as to maximize the growth rate of per capita income, as given by (11.18), taking into account that the parameter A (inside B) is determined according to (11.26). As usual, this problem is solved setting the partial derivatives $\partial \gamma / \partial \tau_K$ and $\partial \gamma / \partial \tau_H$ equal to zero. The algebra of the exercise

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

is tedious, but the solution should be, according to our previous findings, intuitive: from (11.14), the share of capital returns (net of taxes) in income is $K(r + \delta)/Y = \beta/(1 + \tau_K)$. Thus, the tax rate that makes this share equal to the contribution of capital to production in (11.23) is $\tau_K = \eta$. Using the same reasoning for human capital, the first best solution will be:

$$\tau_K = \tau_H = \eta \quad (11.27)$$

With no surprise, the first best policy is the one that sets a *uniform tax rate* (this, in turn, is equivalent to a tax on production, as specified in Chapter 10). Substituting the optimal taxation rule (11.27) in (11.25), one obtains the corresponding (optimal) level of public provision:

$$\frac{G}{Y} = \frac{\eta}{1 + \eta} \quad (11.28)$$

Note that this corresponds to the contribution of government services to production, as captured by equation (11.23).

To analyse the implications of tax evasion in this model, let K denote for the *formal* sector (the one that pays taxes) and H denote for the *informal* sector (the one that does not pay taxes). To find the benevolent planner solution in this case, we must impose $\tau_H = 0$ in (11.18) and (11.26) and maximize the growth rate of the economy in respect to τ_K , only. This leads to:

$$\tau_K = \frac{\eta}{\beta} > \eta \quad (11.29)$$

Comparing to (11.27), we see that the tax rate on physical capital is now higher than in the first best: since the government cannot tax human capital, it sets a higher tax rate on physical capital. Also note that, the lower the β , that is, the smaller the size of the formal sector, the higher the tax rate must be.

Substituting (11.29) in (11.25), we get an interesting result:

$$\frac{G}{Y} = \frac{\eta\beta}{\beta + \eta} < \frac{\eta}{1 + \eta} \quad (11.30)$$

This equation states that the optimal provision of government services under tax evasion is lower than the contribution of government services to production, as stated in

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

(11.23). The reason is intuitive: because rising revenues forces the government to impose a distortion in the factor markets, a *benevolent planner will balance the benefits of providing one extra unit of public input against the extra cost resulting from a further move away from E*. Equation (11.30) shows that the optimal balance between these two effects translates into a lower provision of government services than in the case where uniform taxation is available. The public provision deviates from the first best, and the more, the lower the contribution of the formal sector to production (β)²²⁴.

All in all, under tax evasion, the second best solution will imply *those in the formal sector paying very high tax rates and the economy as whole enjoying a very lower level of government services*. No wonder why economies with large informal sectors find it difficult to attract foreign direct investment!

Box 11.5. Second-best decision-making

The case with optimal provision of the public input under tax evasion illustrates a problem of second-best decision-making. The Theory of Second-Best concerns with the optimal choice by the planner when the range of available instruments is somehow restricted. In a second-best framework, *eliminating one distortion does not necessarily leads to higher efficiency*. The reason is that the alleviation of one distortion may impact negatively on other distortions that cannot be removed. Thus, the optimal policy shall involve a balance between the benefits of alleviating one distortion against the costs of aggravating other distortions.

In the example of tax evasion, providing more of the public input implies taxing further the formal sector, exacerbating the distortion in the factor markets. If informality cannot be eliminated (that would be the first best), the second best policy (as summarized in

²²⁴ The Cobb-Douglas production function implicitly postulates a unit elasticity of substitution between H and K . But you could assume instead a CES production function, with a very high elasticity of substitution between the two inputs. In that case, the effects just described would be amplified: the size of the formal sector would shrink even more, the tax rates would be set even higher and the level of public provision would be even lower.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

11.30) involves a balance between the benefits of providing the economy with more of the public input with the cost of further distorting the factor markets.

Box 11.6. Growth diagnostics

The theory of second-best tells us that a policy reform that looks efficiency-enhancing when considered in isolation may end up being counterproductive, due to adverse interaction effects with other distortions. Since in the real world, policymakers do not have the complete knowledge of all prevailing distortions in an economy nor a precise quantification of all possible adverse second-best interactions, the strategy of simply tackling any distortion may reveal unsuccessful. Under this reasoning, Ricardo Hausmann, Dani Rodrik and Andres Velasco proposed a minimalistic approach to economic reform, that they labelled as “growth diagnostics”²²⁵. In a complex world, with many distortions and complex interaction effects between them, finding out the appropriate sequence of reforms is very difficult. The authors argue that, instead of trying to tackle all distortions at the same time (“laundry list approach to economic reform”), policymakers should carefully identify the one or two most important “binding constraints” in a given economy and tackle these constraints.

To help identify the relevant binding constraints, the authors propose a diagnostic analysis, using a conceptual model that can be interpreted in light of the simple AK model. According to that framework, economic activity in a given developing country may be constrained by: (a) Low A (low return on private investment) (b) High τ_K (high cost of finance). When the main problem is a low A (in which case the interest rate is expected to be low and capital is expected to be flowing out), this should be related to low social returns (low human capital, poor infrastructure, bad geography) or to a large gap between social and private returns (distortionary taxation, government failures). When the main problem is instead a high cost of finance, then the country should exhibit high interest rates, and the

²²⁵ “[Getting the Diagnosis Right](https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/)”. Hausmann, R., Rodrik, D., Velasco, A., 2006. Finance and Development 43(1), March 2006.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

distortions should be related to inefficiencies in the financial market, such as low enforcement of loan contracts, restrictions to capital inflows, or imperfect competition on domestic banking. Using this conceptual model, policymakers should identify the one or two most binding constraints to economic growth and adjust the policy to tackle this small number of constraints, only.

11.3.5 Externalities

In the examples above, discriminatory taxation was inefficient because the equilibrium without taxes was efficient. In this section, we argue that, in the presence of market failures, discriminatory taxation may not only be desirable, it may also be the optimal solution.

Assume that there is a positive externality associated to Human Capital:

$$A = C \left(\frac{H}{Y} \right)^\varepsilon \quad (11.31)$$

Equation (11.31) implicitly assumes that the externality is subject to congestion: as the economy gets larger, human capital must increase proportionally to generate the same external effect. Summing the production function (11.11a) across firms and substituting (11.31), one obtains the aggregate production function for this economy:

$$Y_t = BK_t^{1-\alpha} H_t^\alpha \quad (11.32)$$

$$\text{where } B = C^{\frac{1}{1+\varepsilon}} \text{ and } 1-\alpha = \frac{\beta}{1+\varepsilon} \quad (11.33).$$

Comparing (11.11a) and (11.32) we see that there is again a divergence between the *perceived* contributions (betas) and the *actual* contributions (alfas) of physical and human capital to production. Since $1-\alpha < \beta$, this means that the *actual* contribution of physical capital to output is lower than that *perceived* by firms. In turn, the elasticity of human capital, α , is larger than that perceived by firms, $1-\beta$.

From (11.32), we know that the *first best* resource allocation (point E) is:

$$\frac{H}{K} = \frac{\alpha}{1-\alpha} \quad (11.34)$$

From (11.16), we know that a careful choice of taxes and subsidies can be used to push the economy towards (11.34). Setting (11.16) equal to (11.34), the following rule is obtained:

$$\frac{1 + \tau_K}{1 + \tau_H} = 1 + \frac{\varepsilon}{1 - \beta} \quad (11.35)$$

If the two tax rates are set according to (11.35), this means that private firms, following their maximization problem – that is, deciding according to (11.16) – will be induced to choose a relative employment of H and K satisfying the efficiency condition (11.34). For the policy to be self-financed, it must satisfy condition (11.21). Using (11.21), (11.34) and (11.35), one obtains the first best solution $\tau_K^* = \varepsilon$ and $\tau_H^* = -\varepsilon\beta/(1 - \beta + \varepsilon)$. This means that a now unique combination of the *two* instruments satisfies the *two* targets: the correction of the externality and a balanced budget.

11.4 Policies and growth

This chapter reviewed different types of market or government-imposed distortions that impact negatively on economic performance. The chapter also illustrates how government policies can be used to “get the prices right”. It is suggested that a careful intervention using taxes and subsidies has the potential to improve economic efficiency and thereby induce faster growth (or, in alternative, a higher level of per capita income).

The view that policy reforms should address market failures and deadweight triangles in the first place has been at the core of market-oriented reforms prescribed by the main international institutions, since the 1990s, known as the Washington Consensus (see Box 11.7). In practice, the experience has revealed that, despite some heterogeneity in growth outcomes, countries implementing market-oriented reforms have performed better²²⁶.

²²⁶ Recent evidence in Easterly, W., 2019. In search for reforms for growth: new stylized facts on policy and growth outcomes. NBER Working Paper 26318. September.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

With no question, tackling economic efficiency and getting the prices right is a key ingredient to economic success. However, one should not overstate this avenue. Saying that a country is poor because economic policies are wrong may be too short-seeing. At a deeper level, one shall ask why policymakers in developing countries do not implement better economic policies. In the economic debate, it has been recognized that, although good policies are critical for economic performance, implementing the right policies may not be within reach for many developing countries, because they lack the necessary institutions. Institutions set the framework in which policy makers implement their policies. If institutions are poor and allow policymakers to deviate from the public interest, eventually policies will deviate from the public interest.

Empirically, the claim that the quality of institutions determines the quality of economic policies has been supported by extensive evidence in cross-country growth regressions showing that variables measuring the quality of policies (such as trade openness, inflation, etc) tend to lose significance when the quality of institutions is accounted for (an example in Box 5.3). This is not to say that policies are not important: simply, because good institutions tend to generate good policies, in cross-country regressions variables capturing the quality of institutions also capture the quality of economic policies in general.

Box 11.7 The Washington Consensus

The view that policy reforms should address market failures and deadweight triangles in the first place was coined “The Washington Consensus” by John Williamson, in a conference held at the Institute for International Economics, in 1990. In that conference, the author listed 10 policy reforms that synthesized the view: (1) fiscal discipline; (2) reorientation of public expenditures from non-merit subsidies to basic health care, education and infrastructure; (3) tax reform, so as to enlarge the tax base and reduce marginal tax rates; (4) liberalization of interest rates; (5) unified and competitive exchange rates; (6) trade liberalization; (7) openness to inward foreign direct investment; (8) privatisation; (9) deregulation, to ease barriers to entry and exit; (10) secure property rights.

The Washington consensus constituted a departure from the thinking in the 1950s and the 1960s, according to which economic development was a complex process of economic, social, political and historical transformation, requiring a specific prescription for each particular country. Along the 1990s, the Washington Consensus inspired the reforms in many <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

developing countries, namely in Latin America, in Sub-Saharan Africa, and in former socialist republics. By the turn of the century, many developing countries had achieved more open economies, sounder public finance, lower inflation, fewer restrictions on private business, and more efficient financial sectors.

At the beginning, the experience with the implementation of the Washington Consensus revealed different results. In Latin America, for instance, growth rates in the 1990s were on average lower than along 1960-1980, a period, characterized by extensive state intervention and import substitution policies. At the same time, countries like Thailand, Malaysia, China, Vietnam and India were experimenting fast growth, despite their insistence on industrial policies. China and India, in particular, made significant market-based reforms, but maintained high levels of state intervention, import restrictions and targeted industrial policies.

In 2005 the World Bank launched a study, *Economic Growth in the 1990s: Learning from a Decade of Reform*, where it concluded that “some countries managed to sustain rapid growth with just modest reforms, and others could not grow even after implementing a wide range of reforms”. In the document, it is recognized that, while good policies are in general important for growth, their effects may be offset or reinforced by other factors, including the cultural, institutional, social and political environment²²⁷. The report concluded that a correct assessment of each country particular circumstances is essential, in order to define the optimal sequencing of reforms.

More recently, Easterly (2019) presented new evidence pointing to a stronger correlation between policy reforms and growth outcomes than previously estimated by the World Bank. The author contended that earlier judgements on the Washington Consensus happened before the reform process was complete in many countries, and with limited data

²²⁷ The recognition that sound policies work better if embedded in strong institutions had been already materialized in 2001, with a new list of priority reforms that became known as the Monterrey Consensus. This list complemented the original (the so-called *first generation reforms*) with a number of “action points” addressing issues such as governance, corruption and human rights (*second generation reforms*).

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

covering the period post-reform. The author found that those countries struggling with extremely bad policies in the 1980s and early 1990s, namely in Africa and in Latin America became fast reformers and met faster economic growth in the years that followed.

Box 11.8. Easterly: Policies can destroy growth!

Easterly (2005) estimated a number of cross-country growth regressions, using a panel of 5-year averages along 1960-2000. The author considered 6 variables capturing distinct dimensions of policy (details in Table 11.3): the inflation rate (INFL), the government deficit (BB), a measure of real exchange rate overvaluation (LREALOVR), the black market premium on foreign exchange (LBMP), a measure of financial development (M2/GDP) and a measure of trade openness (TRADE). The author then regressed the growth rate of GDP per capita (five years average, 1960-2000) in all these six variables.

Table 11.3 describes Easterly's findings. Column (1) presents the estimated coefficients of a simple regression where the 6 dependent variables are included. All signs are in accordance to the theory and most coefficients are significant²²⁸.

At the first sight, the results in Column (1) suggest that improving the quality of economic policies may have a significant impact on growth. To quantify this, Easterly computed the implied effects on growth resulting from a one standard deviation improvement in each of the policy variables. These effects are displayed in Column (2); the one standard deviation changes in the policy variables are displayed in Column (4). For instance, a reduction in the inflation rate by one standard deviation (that is, by 32 p.p, as showed in Column 4) would augment per capita income growth by 0.6 p.p (Column 2). According to this table, if all variables were improved at the same time, the overall impact on per capita GDP would be as much as 3p.p. This finding suggests that getting the policies right is good for growth.

²²⁸ M2 and TRADE perform less well than the other variables, but the author shows that they become significant once any one of the other five variables is dropped out of the regression equation.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Easterly pointed out, however, that such an exercise has enormous limitations. First, he observed that a one standard-deviation change in any of these six policy variables (Column 4) is outside the experience of most countries: reducing the inflation rate by 32 p.p, improving the budget balance by 5 p.p, improving the M2/GDP ratio in 25 pp, etc, are certainly not easy to achieve for most countries. The reason, he argues, is that these large standard deviations are related to the presence of very *extreme* observations (extreme inflation, extremely high deficits and extremely high black market premiums “on the bad side”, extremely high monetization ratios and extreme degree of openness “in the good side”). These extreme observations influence critically the statistical significance of the policy variables in growth regressions.

To abstract from the presence of these outliers, the author run the regression again, restricting the sample to observations where all six policies variables lie in the range of “moderate” policies, that is, removing the extreme values form the sample (this reduces the sample to roughly one half). Column 3 of Table 11.3 displays the corresponding results. Excluding observations where *any* of the six policy variables are extreme, all policy variables become insignificant!

Table 11.3 – Easterly regressions on polices and growth

Regressions of per capita growth on basic set of 6 policy variables.
Dependent variable : LGDPG (log per capita growth, five year averages, 1960-2000)

	(1)	(2)	(3)	(4)
	Coefficient in growth regression	Change in growth from one standard deviation change in policy (%)	Sample: Moderate Policies, only	Memo: Improvement of one standard deviation in policy variable
INFL	-0.018 (2.61)**	0.6	-0.064 -1.23	-0.325
BB	0.092 (2.81)**	0.5	0.018 0.22	0.054
M2	0.01 1.37	0.3	-0.004 0.27	0.253
LREALOVR	-0.014 (2.97)**	0.5	0.001 0.06	-0.387
LBMP	-0.012 (2.33)*	0.7	-0.038 -0.95	-0.558
TRADE	0.01 1.92	0.5	0.01 1.09	0.454
Constant	0.016 (3.62)**		0.027 (2.52)*	
Observations	422		193	
R-squared	0.18		0.03	

Notes: Robust *t* statistics in parentheses. *Significant at 5%; **Significant at 1%.

(3) Restrictions under moderate policies: INFL between -0.05 and 0.3, BB between -0.12 and 0.02, M2<1.0, LREALOVR between -0.5 and 0.5, trade < 1.20, LBMP between -0.05 and 0.3

Variables used: LGDPG: Log per capita growth rate; INFL: Log (1+ inflation rate); BB: Government budget balance/GDP; M2: M2/GDP; LREALOVR: Log (overvaluation index/100), above zero indicates overvaluation; LBMP: log(1+black market premium on foreign exchange); TRADE: (exports+imports)/GDP.

Source: Easterly (2005)

This finding suggests that the results in column (1) are mainly driven by extreme values. In other words, the results in column (1) may be reflecting mainly the potential for destruction of bad policies. Thus, countries with moderate values of these variables are not expected to obtain large gains by achieving moderate improvements in their policies. Based on this, Easterly concluded that: “Although extremely bad policy can probably destroy any chance of growth, it does not follow that good macroeconomic or trade policy alone can create the conditions for high steady state growth” (pp. 1017).

Key ideas of chapter 11

- A market distortion can be thought as a tax that drives a wedge between private returns and social returns, reducing the level of market activity.
- A first type of distortions analysed in the chapter are those that influence the consumption-savings decision. By affecting the marginal product of capital, these distortions impact on the reward of investment and thereby on saving rates, causing capital accumulation to slow down.
- An example of such type of distortion relates to financial markets. Financial market frictions give rise to transaction costs that arise in the form of an intermediation margin, creating a gap between the marginal product of capital paid by firms and the

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

net return actually received by borrower. The larger this intermediation margin, the lower the return on savings. Financial development, by reducing the costs of intermediation, impacts positively on capital accumulation and economic performance.

- A second type of distortions analysed in this chapter are those that alter the relative price of two inputs, inducing firms to change the proportions in which they are used. Examples of this type of distortion occur with import protection, segmented labour markets, high inflation, monopoly and externalities.
- Some policies act as a subsidy to particular sectors that come as a cost to other sectors, with a negative overall impact on growth. Examples of these “tax-cum-subsidy” schemes include interest rate ceilings, directed banking credits, and dual exchange rates.
- In the presence of tax evasion, there is an incentive for firms to use more of the type of capital that escapes taxation. When this is so, the optimal intervention will involve a lower provision of the public input. This case provides an illustration of second best decision-making.
- The view that reform agendas should focus on market failures and deadweight losses was well reflected in the Washington Consensus. The mixed experience with these market oriented reforms lead economies to call for more attention on the quality of institutions.

Problems and Exercises

Key concepts

- *Distortion. Tax cum subsidy schemes. Second best decision-making. The Washington consensus*

Exercises

- 11.1. (Tax on production)** Consider an economy, where the production function is given by $Y=0,2K$, the population grows at 1% per year and physical capital depreciates at 4%. (a) Assume for the moment that the saving rate is constant and equal to 30%. Describe the main equations of this model and find out the growth rate of per capita income in this economy. Discuss, with the help of a graph the dynamic properties of this model. (b) Consider now that the government imposes a tax on production, which proceedings are used to finance unproductive government consumption. Describe the main income identities of this economy and place them in a flow income chart. Find out the growth rate of per capita income in this economy when: $\tau=0\%$; and $\tau=20\%$. Explain. (c) Now assume that the saving rate is endogenous, so as to satisfy the following inter-temporal consumption rule: $\gamma_t = r_t - 0,15$ (c1) Explain this equation. (c2) Departing from the identity $[\dot{K} + \delta K] = s(1 - \tau)Y$, compute the endogenous saving rate in this economy, as a function of the tax rate. Explain. (c3) Compute again the growth rates of per capita income when $\tau=0\%$; and $\tau=20\%$.
- 11.2. (Tax on Capital income)** Consider an economy where consumers do not face borrowing constraints and where the production function is given by $Y=0,1K$. In this economy, the government imposes a tax on capital incomes, τ_K , which proceedings are used for government consumption. Assume also that $\delta=2.5\%$. The population in this economy is constant. For the moment consider that the saving rate is exogenous and equal to 25%. (a) Describe the main income identities of this economy and place then in a flow income chart. (b) Find out the growth rate of per capita income in this economy when: $\tau_K = 0\%$ and $\tau_K = 25\%$. Explain. (Answer: $\gamma_0 = 0$; $\gamma_t = -0.005$) (c) Consider now that households are able to optimize their intertemporal consumption, implying $\gamma_t = r_t - \rho$, where $\rho = 7.5\%$. Find out the growth rate of per capita income when $\tau_K = 0\%$ and $\tau_K = 25\%$. Explain. (Answer: $\gamma_0 = 0$; $\gamma_t = -0.02$) (d) Solve for the endogenous saving rate.
- 11.3. (Tax cum Subsidy Scheme):** Consider an economy composed by a large number of identical firms, with production functions of the form: $Y_i = 0.25K_i^{1/2}H_i^{1/2}$. In this economy, households are able to optimize their intertemporal consumption, implying $\gamma_t = r_t - \rho$, where $\rho = 5\%$. Population is constant and the depreciation of both physical and human capital is 5%. (a) Find out the growth rate of per capita income in this economy, assuming perfect competition. (b) Now, assume that the government launched a subsidy to the returns of human capital amounting to $\tau_H=-0.25$, to be financed with a tax on capital returns, to τ_K . (b1) From the firm's optimization problem, find out
- <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

the implications for the use of the two types of capital. Explain, with the help of a graph. (b2) Find out the balanced-budget tax on capital incomes. (b3) Quantify the impact of the policy on the growth rate of the economy. (c) Beyond the scope of this exercise, shall we conclude that, in general, the optimal policy involved uniform taxation? Why not?

- 11.4. (Tax evasion)** Consider an economy composed by a large number of identical firms, with production functions of the form: $Y_i = AK_i^{1/3}H_i^{2/3}$. In this economy, the government has the ability to coerce citizens to pay taxes out of their factor incomes. We also know that the saving rate is 12,5%, population is constant and the depreciation of both physical and human capital is 4%. Consider the following optimal consumption rule $\gamma = r - 0,05$. (a) **(distortionary taxation)** Solve the individual firm problem and find out the implied factor income shares. Describe, with the help of a graph, the effects of taxation on the relative use of the two types of capital. (b) Find out the growth rate of per capita income in this economy, depending on the tax rates. (c) Now assume that $A = (G/Y)^{1/2}$, where G are public inputs. Compute the aggregate production function of this economy. Explain the market failure. (d) Obtain an expression for A in terms of the two tax rates. (e) Compute the benevolent planner solution. Graph the equilibrium and explain. (f) Now assume that the government could not tax Human Capital. Find out the optimal tax on physical capital and the corresponding provision of public inputs. Compare with the first best outcome and explain.
- 11.5.** In Unevenland, the aggregate production function can be described as $Y = EK$, where E denotes for aggregate efficiency and K denotes for physical capital. The later depreciates at the rate $\delta=0,05$. Capital markets are perfect and there is no uncertainty, so households can smooth consumption inter-temporally, according to $\gamma = r - 0,1$. (a) **(AK)** Find out the growth rate of per capita income in this economy in terms of E. Describe the dynamic properties of the model. (b) **(Cost of capital)** Elaborating a bit more, suppose that the efficiency term is better described as a ratio of two terms, $E = 0.5A/P_I$, where A is constant and P_I denotes for the relative price of capital goods. Which policies or country circumstances may be captured by P_I ? Assume that initially $A = 1/2$. Compute the growth rates of per capita income in this economy when $P_I = (1/2)^{1/2}$ and when $P_I = (3/2)$. With the help of a graph, compare with the impact of a similar change in the context of the Solow model. (c) **(Public good)** The economy of Unevenland is actually more complex than at the first sight. In particular, each individual firm i faces a production function of the form: $Y_i = AK_i^{1/2}H_i^{1/2}$, where H is human capital and $A = (G/Y)^{1/2}$, where G denotes for a non-excludable good. The government budget constraint is given by $G = (1/2)[\tau_K/(1 + \tau_K) + \tau_H/(1 + \tau_H)]Y$, where τ_K is the tax rate on physical capital incomes and τ_H the tax rate on human capital incomes. P_I becomes $P_I = [(1 + \tau_K)(1 + \tau_H)]^{1/2}$. Compute the aggregate production function and explain why there is a market failure. (d) **(Tax evasion)** The benevolent planner of Unevenland chooses the intervention level G/Y so as to maximize the growth rate of per capita income or - which is the same - the efficiency term, E. For the moment, however, assume that the government can only levy taxes on physical capital τ_K (that is, $\tau_H = 0$). Find out the optimal level of τ_K and the corresponding level of government intervention. (e) **Optimal taxation)** Now assume that the government is

able to impose a uniform tax rate (that is, $\tau_H = \tau_K$). Find out the optimal level of τ_K and the corresponding level of government intervention. Explain the effect on P_I and compare this case to the one in question c. Compare the solutions of d and e and discuss.

- 11.6. (Externality).** Consider an economy composed by a large number of identical firms, with production functions of the form: $Y_i = AK_i^{1/3}H_i^{2/3}$. We also know that the saving rate is 12,5%, population is constant and the depreciation of both physical and human capital is 4%. Assume that there was a positive externality associated to the use of Physical capital, according to: $A = (K/Y)^\epsilon$. Compute the aggregate production function of this economy. Explain the market failure. Suppose that the government want to solve the market failure imposing a tax on human capital, with a balanced budget, so it would use the tax proceeds to finance a subsidy on human capital. What would be the optimal intervention?
- 11.7. (Monopoly)** Consider an economy composed by many identical firms, with production functions in the final good sector of the form: $Y_i = 0.5K_i^{1/3}H_i^{2/3}$, where K and H are physical and human capital. We also know that one unit of output can be transformed in one unit of physical capital or in one unit of human capital and that the depreciation rate for both types of capital is 3%. Consider that each firm i in the final good sector is a price taker in factor markets. Let P_K be the price of physical capital and P_H be the price of human capital. (a) From the profit maximization problem of firms in the final good sector, find out the optimal relation between H and K as a function of factor prices. (b) Now consider the case in which both factor markets are competitive, that is, each firm producing physical and human capital is price taker. Describe the profit maximization problem of a representative firm producing (human or physical) capital, find out the optimal price and the corresponding profits. (c) Now assume that the supply of physical capital is undertaken by a monopolist. Solve its profit maximization problem, finding out its optimal price and the corresponding profits. (d) Compare the relative employment of physical and human capital in the case of perfect competition and in the case of a monopoly. Compare it with the efficient allocation. € Using the optimal consumption rule $\gamma = r - p$, compare the growth rates of this economy in the two cases. (f) How could the government remove this distortion?
- 11.8. (Financial market imperfections)** Consider an economy where aggregate output is produced using two types of capital, according to $Y = AK_1^{1/2}K_2^{1/2}$. The total capital available each moment in time evolves according to $\dot{Y} = [\dot{K} + \delta K](1 + \tau_K)$, where τ_K refers to financial market imperfections. (a) Interpret the equation describing capital accumulation (b) Draw the income flow chart of this economy. (c) Show that, as long as the ratio of the two types of capital is constant, the production function as an AK representation. (d) Find out the ratio K_1/K_2 that maximizes the efficiency level in this economy. Interpret. (e) Suppose that investment decisions for each type of capital were undertaken by two independent firms, which take the other's firm decision as given. Show that, as long as both firms have access to credit at the same interest rate, the maximum efficiency will be achieved. (f) Describe the growth rate of this economy as a function of the exogenous parameters. Interpret.

12 Inequality

12.1 Introduction

So far, we have analysed income disparities across countries focusing on per capita income, and abstracting from income disparities within countries. In this chapter we turn to the issue of economic inequality within countries. Instead of considering only the income of the average citizen in a given nation, we turn to the question of how income is distributed among the citizens of that nation, and the extent to which such distribution influences the income of the average citizen.

The relationship between inequality and economic performance is a complex one. Inequality interacts with many key ingredients of economic growth, including incentives, savings, investment, education, policies and institutions. At the crudest level, income inequality is part of a well-designed incentives system: income inequality is a corollary of the principle that entrepreneurship and innovation must be rewarded, being therefore inexorably linked to economic growth. A problem arises in that inequality does not only reflect individual choices: it also reflects unequal opportunities and bad-luck. While some individuals enjoy favourable initial conditions, others get lost in the way, because of misfortune and exogenous shocks. Unequal opportunities and bad fortune are sources of income disparities that the society in general dislikes. In general, the society is willing to sacrifice some efficiency to achieve a minimum justice in the distribution of income.

Most societies see an important role for the government in income redistribution, by taxing the rich and creating welfare programmes for the poor. Because in general these programmes weaken the market incentives, the typical view is that there exists a “trade-off between efficiency and equity”. More recently, economists have argued that this trade-off is less than certain. The main argument is that the free market fails in delivering an efficient allocation of resources in the first place. Inequality comes along with barriers to entrepreneurship, preventing socially valuable projects from being implemented. Second, inequality interacts with the political process, creating pressures for income redistribution that in general are not favourable to economic growth. Today, most economists agree that inequality is bad for growth.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

This chapter reviews the links through which income or wealth inequality may impact on per capita income. In Section 13.2 we discuss the problem of measuring inequality, and we present some cross-country data on inequality and growth. In Section 13.3 we present a simple model to illustrate how inequality may arise in result of different preferences and unequal opportunities. In that section, we introduce the classical trade-off between equity and efficiency. In Section 13.4 we explain how financial market frictions may prevent the poor from implementing socially valuable projects. Finally, in Section 13.5 we address the argument that high inequality may interact perversely with the quality of institutions, hurting growth. In Section 13.6 we summarize the main ideas.

12.2 Facts on inequality and growth

12.2.1 *Inequality in the personal distribution of income*

Economic inequality is a particular dimension of inequality. In general, inequality refers to fundamental disparities that prevent some individuals in the society to make choices that are available to other individuals in the same society. Economic inequality refers to differential access of people to income. Of course, differential access to income has implications on the type of choices individuals can make, not only regarding the consumption of goods and services that are available in the market, but also regarding opportunities that arise through non-market mechanisms of social selection.

Income inequality results from the process of distributing income. The distribution of income can be categorized as *functional* or as *personal*. The *functional distribution* refers to the way income in an economy is distributed among factors of production, such as capital and labour. The *personal distribution* of income refers to how much income accrues to each family, given their ownership of capital and labour. Even if markets worked perfectly and capital and labour were paid according to their contributions to output, delivering an *efficient* functional distribution of income, differences in individual endowments or in the opportunity to employ these endowments may dictate enormous disparities at the personal level. In earlier chapters, we addressed market failures, such as externalities and public goods, that cause the functional distribution of income to depart from the efficient allocation. In this chapter, it is the personal distribution of income we are interested on.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Income inequality shall not be confounded with poverty. Poverty refers to a condition of deprivation of basic needs, such as safe water, basic food, shelter, health or education. Poverty is therefore an *absolute* concept, that does not depend on how well the other people in the same society are living. Inequality, in contrast, is a *relative* concept. It implies a comparison of standards of living among individuals in the same society. A society can be highly unequal, and still have limited poverty.

The World Bank distinguishes two concepts of poverty: *absolute* (or extreme) poverty and a *relative* poverty. Absolute poverty is measured counting the number of people living with less than \$1.9 a day, at 2011 comparable prices (the *international poverty line*). The World Bank defines *relative poverty* as a condition in which an individual income is less than some proportion of the median income in a society. Relative poverty applies to a particular social context, and is an indicator of inequality. Absolute poverty, in contrast, refers to a unique standard that is consistent over time and across countries.

12.2.2 Measurement problems

Inequality is something that can hardly be measure with a single synthetic indicator. If the society was composed by two individuals (or two groups of identical individuals), a simple measure, such as the share of population earning less than x% of the median income, would be appropriate. But in a society with many and heterogeneous individuals, we need to learn more about the shape of the entire distribution.

Basic indicators of the shape of a distributions include measures of symmetry and measures of dispersion. A simple measure of symmetry is the ratio between the average income and the median income. In a symmetric distribution, like the normal curve, the average and the median are the same. When instead the average income is larger than the median income, this means that a minority of the population has an income that is higher average. In the real world, most income distributions are skewed, with a long right tale reflecting a low fraction of the population with very high incomes.

Dispersion can be assessed by indicators such as the range (difference between higher and lower income) and the standard deviation. However, these measures are influenced by the units of measurement. Because of this, cross-country comparisons rely on relative indexes, such as the coefficient of variation (the standard deviation of incomes divided by the

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

mean) or the share of income earned by the richest $x\%$ divided by share of income earned by the poorest $y\%$ (the so-called Kuznets ratio).

Table 1 shows alternative indicators of inequality, considering four imaginary societies (A, B, C, and D), each one composed by four households (1, 2, 3 and 4). In each of these societies total income is 100. Consider first the society A. Clearly, that allocation is uneven: the income of the 25% richest is four times the income of the 25% poorest. In that society, 25% of the population (household 1) earns an income that is equal or less than 60% of the median income. The richest 25%, on the other hand, earn 40% of the income. Anyone would agree that society A is rather unequal.

Now consider allocation B. Basically, that allocation is similar to A, except that the richest household transferred 10 units of income to Household 1. This is an unambiguous move towards a more equal society. The question is whether this is reflected in the indicators²²⁹. The ratio of incomes between the richest and the poorest declined from 4.0 to 2.33. Also, the coefficient of variation decreased from 0.52 to 0.37. The share of population earning less than 60% of the median income remained unchanged, but the share of population earning less than 40% of the median income decreased from 40% to zero. In general, the different indicators point to less inequality.

More difficult is the comparison between allocations A and C. In allocation C, both households 1 and 4 are better off than in allocation A, at the cost of household 3. The ratio between the richest and the poorest declined to 3, suggesting less inequality. Still, the coefficient of variation increased, and the ratio between the average income and the median income have increased, reflecting the fact that the rich is now richer. Note that the share of income of the bottom 75% of the population is now only 55%. If one wanted to compare distributions A and C, which indicators should we pick up? The fact that the different indicators point to opposite directions is not completely absurd, and reflects the difficulty in

²²⁹ An indicator that improves when there is a transfer from the richest person to the poorest person is said to satisfy the *Pigou-Dalton principle*.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

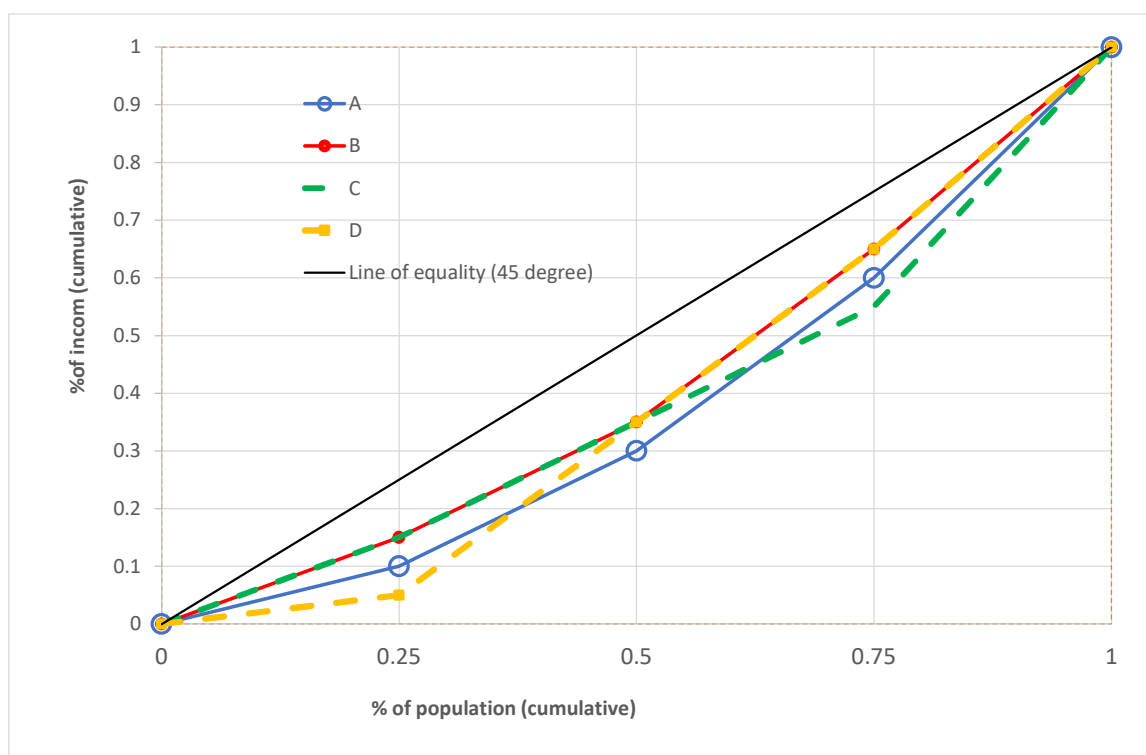
defining inequality itself: shall a society where both the rich and the poor get better at the cost of the middle class be considered more equal or more unequal?

Table 12.1: Measuring inequality in societies with four households

	A	B	C	D
Household 1	10	15	15	5
Household 2	20	20	20	30
Household 3	30	30	20	30
Household 4	40	35	45	35
National Income	100	100	100	100
Average income	25	25	25	25
Median income	25	25	20	30
Share of population with income less or equal to 60% of median	25%	25%	0%	25%
Average income / Median Income	1.00	1.00	1.25	0.83
Income of richer 25% / income of poorest 25%	4.00	2.33	3.00	7.00
Coefficient of variation (Stand.Dev./average)	0.52	0.37	0.54	0.54
Share of bottom 25%	10%	15%	15%	5%
Share of bottom 50%	30%	35%	35%	35%
Share of bottom 75%	60%	65%	55%	65%
GINI	0.25	0.175	0.225	0.225

Finally, consider a move from C to D. In this case, both the richest and poorest lost 10 units of income to the middle class. Again, whether this makes the society more equal is something we cannot say. In terms of indicators, we see that the ratio between the average income and the median income is now lower than 1, revealing the concentration of income in the upper end, and the ratio between the richest and the poorest is higher than ever (7.0). The coefficient of variation however remained unchanged.

Figure 12.1: Lorenz curves corresponding to allocations A-D in Table 1



The synthetic indicator of inequality more frequently used is the Gini coefficient. The Gini index is a measure of concentration, and is calculated based on the *Lorenz curve*. The Lorenz curve relates the cumulative share of income earned by the cumulative fraction of population, with the population arranged in increasing order of income. Figure 1 illustrates the Lorenz curves corresponding to allocations A to D in Table 1. To construct the Lorenz curve, we plot the cumulative share in total income (vertical axes) that is earned by the cumulative fraction of the population arranged in increasing order of income (horizontal axes). Considering, for instance, allocation A (blue line): the 25% poorest (household 1) earn 10% of income, 50% of population earns 30% of income, and 75% of population earns 60% of income. The bowed shape of curve A reflects some deviation from the *line of equality* (in black), that corresponds to the 45° line (25% of income to the bottom 25% of population, 50% to the bottom 50%, etc). Comparing the different curves, we see that allocation B is the closest to the equality line, being therefore the more equal than any other. As for the remaining, they are not easily comparable, because they cross each other. For instance, allocation D reveals more inequality at the bottom 25% than allocation C, but becomes more equal than C when we consider the poorest 75%.

The Gini Index is a single indicator that measures the overall distance between the 45° line and the Lorenz curve. The GINI index is calculated as the area that lies between the Lorenz curve and the line of equality, divided by the area under the line of equality. The Gini index will be equal to zero when the Lorenz curve is equal to the 45° line, and will approach 1 (the limit) as inequality increases. As you can easily check in figure 12.1, the GINI coefficient can be computed decomposing the area below the Lorenz curve as a sum of triangles and rectangles. In the case of allocation A, the area below the Lorenz curve is 0.375. Then, we subtract this amount from the area below the equality line (which is $\frac{1}{2}$), to obtain the area between the two curves, 0.125. Finally, we divide this area by the area below the equality line ($\frac{1}{2}$) to obtain the GINI index of 0.25. In Table 1, we show the Gini indexes calculated for our four societies. According to this indicator, allocation A is the more unequal and allocation B is the less unequal. Allocations C and D have “identical” inequality levels²³⁰.

12.2.3 *Inequality and economic performance in the real world*

We now turn to real world data. Table 1 displays two measures of inequality for a sample of countries: the GNI coefficient, and the share of income accruing to the top 1% richer. In the table, countries are ranked according to the GINI coefficient. In this sample, South Africa is the more unequal society, and Sweden is the less unequal. When we assess inequality by the share in income earned by the top 1%, the ranking changes slightly (Brazil is the more unequal), but the overall pattern is qualitatively the same: South Africa is still amongst the most unequal, and Sweden amongst the more equal.

Table 12.2: Inequality, per capita income and population in 2010 (sample of countries)

²³⁰ A problem with this GINI index is that it weights similarly inequality relative to the bottom segment and to the upper segment. Other formulas for the GINI index avoid this problem, weighting more the observed inequality in the lower segments. There are many methods to calculate the GINI coefficient, but we skip this discussion.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

	GINI	TOP 1%	PER Capita GDP	Population (millions)
South Africa (a)	63.38	18.5	11,832	52.0
Colombia	55.5	20.4	10,637	45.2
Brazil (a)	53.1	29.6	14,968	198
Uruguay	45.32	14.6	15,783	3.4
China	42.06	15.1	9,337	1369
Russia	40.94	20.0	21,754	143.5
United States	40.46	19.8	49,501	309.0
Turkey	38.78	19.9	17,931	72.3
Spain	35.79	8.7	31,611	46.9
Australia	34.94	8.6	44,855	22.2
United Kingdom	34.81	12.6	34,810	63.5
Italy (c)	34.41	9.4	34,728	59.3
India (b)	33.9	20.8	3,795	1218
France	33.78	10.8	35,786	62.9
Canada	33.68	13.6	40,269	34.1
Jordan	33.66	22.0	9,351	7.3
Poland	33.22	12.0	21,006	38.3
Switzerland	32.72	10.6	55,688	7.8
Germany	31.14	13.1	40,627	80.8
Hungary (d)	29.37	9.6	20,478	9.9
Denmark	29.02	6.4	43,416	5.6
Netherlands	28.73	6.4	44,004	16.7
Sweden	26.81	9.0	40,422	9.4

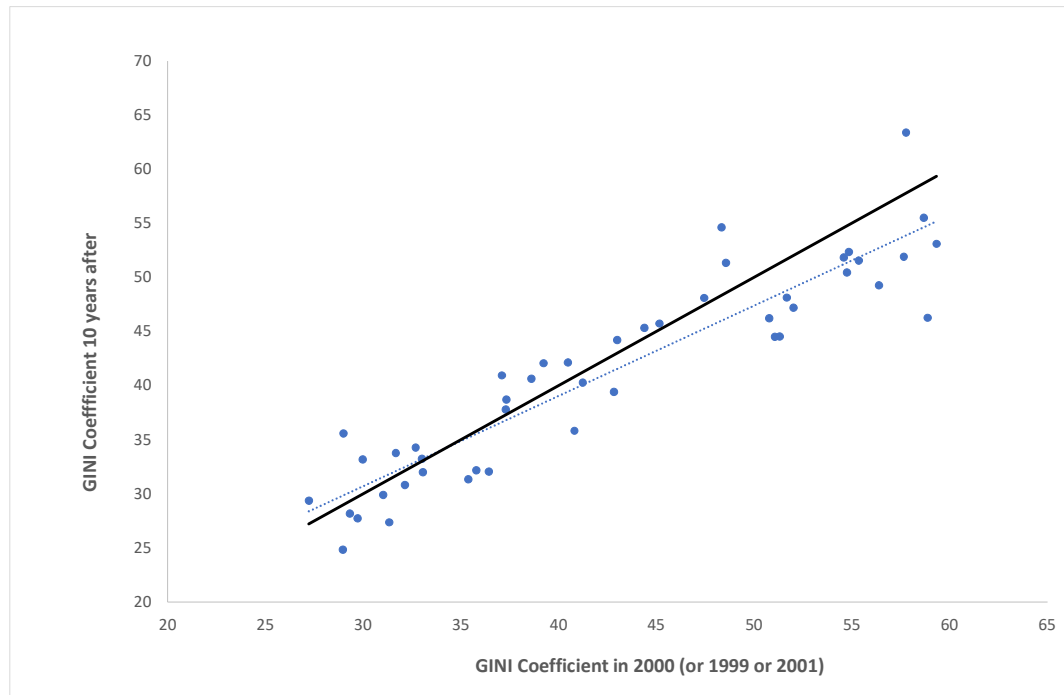
Sources: Max Roser and Esteban Ortiz-Ospina (2013) - "Income Inequality". *Published online at OurWorldInData.org*. Retrieved from: 'https://ourworldindata.org/income-inequality'. GINI indexes are computed by the World Bank, Per capita GDP is from Feenstra et al. 2015, Penn World Tables version 9.1. Population (Gapminder, HYDE 2016 and UN 2019). Notes: Per capita GDP is computed at comparable prices (PPP exchange rates). (a) All data refers to 2011; (b) All data refers to 2009. (c) Data on Top 1% refers to 2009; (d) Data on Top 1% 2008.

In Table 2, most of the data refers to the year of 2010, but in some countries, the information refers to 2008, 2009 or 2011 (see the legend to the figure). The fact that we are using different years in the comparison is not however a big issue. The reason is that inequality tends to change very little over time. To confirm this, in Figure 12.2 we compare, for each country, the GINI around 2008-2011 with the corresponding measure 10 years before. Because of data availability, the analysis is restricted to developing and emerging market economies: if inequality levels did not change at all in a decade, observations should be aligned by the 45° line (straight line). In figure 12.2, we observe small departures from the 45° line, with inequality increasing in some countries and decreasing with others. The dashed line corresponds to the estimated linear relationship. The fitted line points to an overall

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

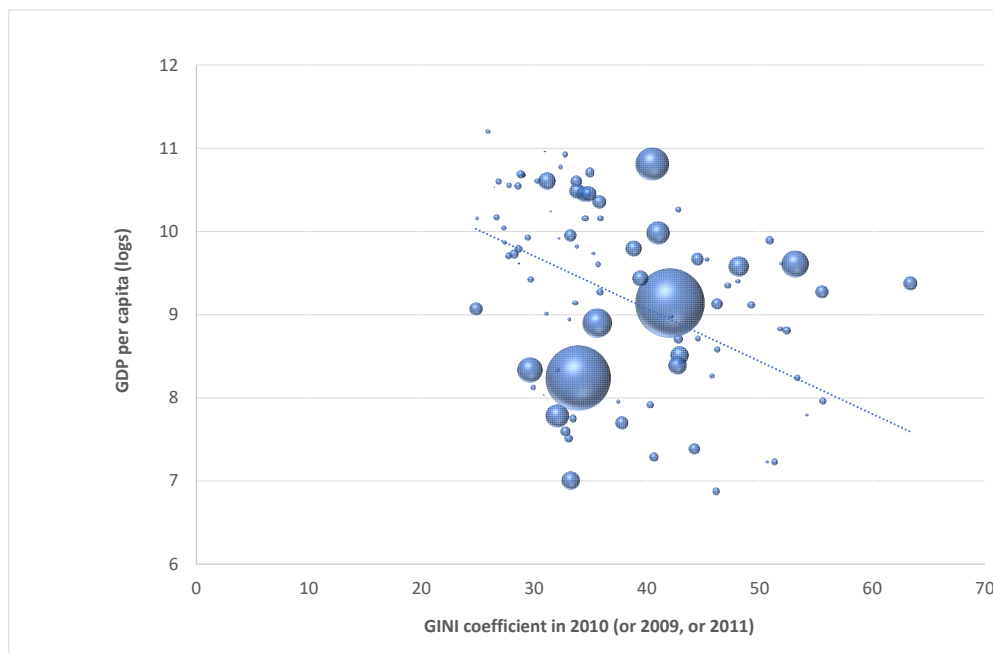
tendency for inequality to decrease in this sample of countries. Still, the correlation between the inequality indexes at 10 years distance is 92%. This evidence illustrates the fact that, once inequality is installed, it is very difficult to eradicate.

Figure 12.2: ten year' change in inequality



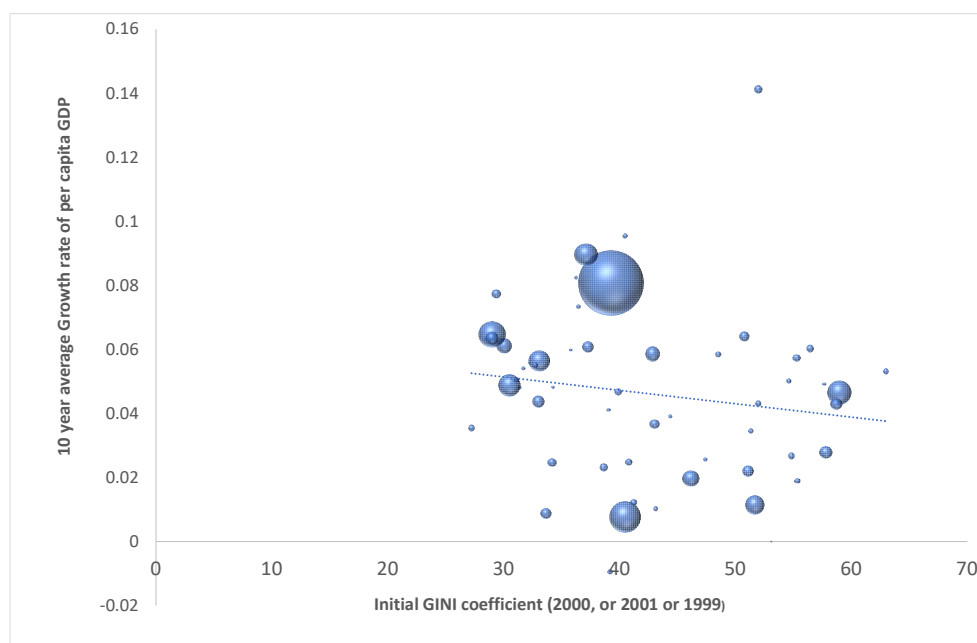
Sources: same as table 2. Sample: 46 countries with per capita income less than \$22,000.

Figure 12.3: Inequality and per capita GDP



Sources: same as table 1. Notes: 92 countries. In the figure, the size of the bubbles is proportional to each country's population.

Figure 12.4: Growth rate of per capita GDP and initial inequality



Sources: same as table 1. Sample 55 countries. . In the figure, the size of the bubbles is proportional to each country's population

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

In table 2, we observe higher levels of inequality in developing countries and in emerging markets, while advanced countries tend to have less inequality. A negative relationship between inequality and per capita income is also apparent in figure 12.3, where we use a sample of 92 countries. The correlation coefficient is -0.38. In figure 12.4, we plot the 10 year' average growth rate of per capita income with the initial inequality. The sample consists in 55 countries, for which data are available. This figure also points to a negative relationship, but correlation is now much lower than in figure 12.3. It is important to note that figures 3 and 4 only show correlations, and correlation does not necessarily implies causality.

In the literature, the early theory regarding the relationship between inequality and economic development is due to Simon Kuznets²³¹. Kuznets theorized that inequality tends to increase at the early stages of economic development, and to decrease at later stages. The pattern in figure 12.3 does not point, however, to an inverted-U shape in the relationship between inequality and per capita GDP: the group of middle income countries includes countries with varying degrees of inequality. In fact, the recent empirical evidence has not being supportive of the idea that economic development by itself influences the level of inequality.²³²

Our main question is actually the reverse one: we want to assess whether inequality impacts on a country' economic performance. A number of authors have investigated the relationship between inequality and income or growth, controlling for the other determinants. In the 1990s, empirical investigations using cross-country growth regressions were able to identify a negative relationship between alternative measures of income inequality (specially of wealth inequality, like the ownership of land) and alternative measures of economic

²³¹ Kuznets, S., 1955. Economic growth and income inequality. American economic review 45, 1-28.

²³² Deininger, K., Squire, L., 1998. New ways of looking at old issues: inequality and growth. Journal of development economics 57, 259-287.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

performance²³³. More recently, greater data availability allowed researchers to include a time dimension to their samples. Some studies using panel data have suggested that the relationship between inequality and growth is non-linear, being conditional on the level of initial income: inequality is good for growth at high levels of income, but bad for growth at low levels of income²³⁴. However, other studies have reported a high sensitivity of the results to the choice of the inequality indicators²³⁵. Measurement problems are still a key issue in the empirical investigation of the relationship between inequality and growth.

12.3 Inequality, efficiency and redistribution

A certain level of income inequality is necessary for an economy to keep the incentives right: if effort and entrepreneurship are rewarded, people seeking for better living standards will work harder and will save more, helping the economy to grow faster. Hence, societies must accept a certain level of income inequality. From this angle, redistributive policies, by rewarding leisure instead of work, weaken the economic incentives.

A problem with this argument is that inequality does not arise merely as the outcome of informed optimal choices by individuals enjoying equal opportunities from the very beginning. Factors that are outside the individuals' control, such as parental wealth, place of birth, and unexpected changes in relative prices may impact quite significantly on the

²³³ Persson, T., Tabellini, G., 1994, Is inequality harmful to growth?, *The American economic review* 84(3), 600-21. Berg, A., Ostry, J., Zettelmeyer, J., 2012. Ostry, J D, A Berg, and C G Tsangarides (2014), "Redistribution, Inequality, and Growth", IMF Staff Discussion Note 14/02. What makes growth sustained? *Journal of Development Economics* 98, 149-166. Easterly, W., 2007. Inequality does cause underdevelopment: insights from a new instrument. *Journal of development economics* 90 (4), 869-87.

²³⁴ Brueckner, M., Lederman, D., 2018. Inequality and Economic Growth: the role of initial income. WB Policy Research Paper 8467. Barro, R., 2000. Barro, R.J. 2000. Inequality and Growth in a Panel of Countries. *Journal of Economic Growth* 5, 5–32.

²³⁵ Panizza, U. (2002). "Income Inequality and Economic Growth: Evidence from American Data." *Journal of Economic Growth* 7: 25-41. Forbes, K.J. (2000). "A Reassessment of the Relationship between Inequality and Growth." *American Economic Review* 90: 869–887. Blotevogel, R., Imamoglu, E., Moriyama, K., Sarr, B., 2020. Measuring income inequality and implications for economic transmission channels. IMF Working Paper 20/164, International Monetary Fund.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

personal distribution of income. Inequality arising because of misfortune and exogenous shocks is in general considered unfair by the society. Thus, the society will be willing to sacrifice some efficiency in order to achieve more justice in the distribution of income. This choice results in the classical *trade-off between efficiency and equity*²³⁶.

12.3.1 Fair inequality

Inequality is inherent to the functioning of a market economy. People work hard, invest and innovate with the aim to obtain rewards. Without these rewards, selfish economic agents would not be willing to engage in costly activities, such as production and investment. Because rewards are, by nature, discriminatory, they are a source of income inequality. To illustrate the case of *fair inequality*, consider a simple model where people choose between consumption and leisure. Assume that each household is endowed with a given amount of time, that can be spent in leisure or in working time:

$$h = N + l \quad (12.1)$$

Without loss of generality, we normalize $h=1$. Households have the following preferences over consumption (C) and leisure (l):

$$U = \alpha \ln C + (1 - \alpha) \ln l \quad (12.2)$$

To make the model as simple as possible, assume that firms are competitive, and that the production function takes the form:

$$x = \lambda N \quad , \quad (12.3)$$

where λ denotes for productivity. Assuming that firms maximize profits, the labour demand will be infinitely elastic at the wage rate,

$$w = \lambda \quad (12.4)$$

²³⁶ This trade-off was exhaustively discussed in the context of the walrazian economy by Arthur Okun (1975). Okun, A M (1975), *Equality and Efficiency: the Big Trade-Off*, Washington: Brookings Institution Press.

and profits will be zero for any level of production.

In this economy, there is a government, which only role is to redistribute income. In particular, the government transfers to some individuals a given amount of the consumption good, T , financing the policy with a tax that is proportional to income, τ .

Each household takes the wage rate, the tax rate and the lump sum transfer as given and maximizes (12.2) subject to the budget constraint:

$$C = T + w(1-t)(h-l) \quad (12.5)$$

This gives the following labour supply

$$N^s = \alpha - \frac{(1-\alpha)T}{w(1-\tau)} \quad (12.6)$$

This labour supply function reveals the impact of taxes and transfers on household choices: in a world without taxes or transfers, $T = \tau = 0$, the labour supply would be inelastic, with the proportion of time devoted to work being equal to the weight of consumption in the utility function, α . When instead $T > 0$, the household is induced to work less: the reason is that the lump-sum transfer gives rise to an income effect, whereby the demand for leisure increases. If, in plus, the individual is taxed on its wage income, this means that the opportunity cost of leisure declines and the demand for leisure increases further, through a substitution effect.

Now assume that in this economy there are two households, differing on preferences, only. In particular, assume that $\alpha_1 > \alpha_2$. From our findings above, we know that in a *laissez-faire* (that is, with $T = \tau = 0$), the first household will be working more than the second: $N_1^s = \alpha_1 > \alpha_2 = N_2^s$. The labour incomes of the two households will be $wN_1^s = \lambda\alpha_1$ and $wN_2^s = \lambda\alpha_2$, and total production in this economy will be

$$x = \lambda(\alpha_1 + \alpha_2) \quad (12.7)$$

Thus, under *laissez-faire*, there will be a disparity in terms of labour incomes, matched by a symmetrical disparity in terms of leisure time. However, there is nothing wrong with such outcome: the perception of inequality only arises if one measures labour incomes based on market transactions. If the individual income was measured, not by labour income, but instead by the total potential income, before consumption of leisure were chosen, there would

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

be no inequality at all. In this example, the competitive equilibrium is efficient, unequal, and fair.

12.3.2 Unfair redistribution

Suppose now that the government wanted income distribution to be more equalitarian, and accordingly decided to tax the income of individual 1 and transfer the proceeds to individual 2. In terms of welfare, individual 2 will be better off and individual 1 will be worse off. As you may see from (12.6), the implication is that individual 1 will work the same, but individual 2 will work even less. Reflecting the efficiency loss, total output (12.7) will decline.

This example illustrates the classic argument against redistributive policies: taxing the rich to transfer money to the poor is rewarding leisure and punishing hard work, undermining the essential incentives to the functioning of a market economy. The redistributive policy in this case is inefficient, and unfair. The society as a whole will lose with redistribution, because there was no unfair income distribution to motivate the income transfer in the first place.

Equation (12.6) illustrates the distortionary effects of taxes and transfers on the choice between working time and leisure. The same argument could be stated in a dynamic perspective: all else equal, more patient individuals invest more and achieve higher incomes in the future. This does not provide a justification for the government to tax those who accumulate capital and subsidize those who consume more. Such transfer would not only be inefficient: it would also be unfair.

A question that arises is why a government should engage in such an undesirable redistributive policy, harming efficiency and economic growth. Arguably, democratically elected governments may engage in redistribution policies because of pressures by voters. When societies are very unequal, the median voter has a below average per capita income and

eventually will push towards redistributive policies²³⁷. The more unlevel the society is, the greater the proportion of people in that society that would benefit from a redistributive policy. If more inequality comes along with more redistribution, greater inequality should be associated with lower per capita income. In light of this argument, there is no direct link between inequality and growth, but an indirect relationship mediated by redistribution policies²³⁸.

12.3.3 *Unfair inequality*

We now turn to the case in which individuals are not given the same opportunities. To focus on this case, assume that all individuals have the same preferences regarding consumption and leisure. In order to make the case as simple as possible, assume that $\alpha = 1$ for all households, implying that all individuals supply exactly one unit of labour. We also assume that financial markets are perfect. Under this assumption, the competitive equilibrium will be efficient.

In this new model, individuals may choose between two working modes: the unskilled mode, whereby one unit of labour is exchanged for a subsistence salary $w=l$; and a modern mode, which delivers a higher income, $w = \lambda > 1$. To adopt the modern mode, the individual must pay in advance an indivisible cost F . We can interpret this cost as the cost of adopting the new technology. This may include a wide range of barriers, such as learning costs, which in turn may depend on the quality of the familiar environment, social connections, and so on. Inequality of opportunities in this model arises in that individuals face different fixed costs, F .

²³⁷ Persson, T., Tabellini, G. 1994. Is equality harmful to growth? *The American Economic Review*, 84 (3), 600-621. Ostry et. al, 2014, provide evidence that more unequal societies tend indeed to redistribute more, though this effect is more robust in OECD nations.

²³⁸ Meltzer, A., Richard, S., 1981. A rational theory of the size of government. *Journal of political economy* 16(3), 914-27. Alesina, A., Rodrik, D., 1994. Distributive policies and economic growth, *Quarterly Journal of economics* 109 (2), 465-90. Persson, T., Tabellini, G., 1994, Is inequality harmful for growth? *The American economic review* 104(5), 979-1009.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

From the individual point of view, it will pay to invest in education as long as $\lambda - F > 1$. Since different individuals have different F , the laissez faire will deliver an unequal society, where some individuals are educated and have a higher income, while others are just unskilled. Such equilibrium will be unequal, but efficient: from the society point of view, it doesn't pay to get educated an individual which fixed cost is too high to break even.

Still, the society may consider this equilibrium unfair: to the extent that inequality is the result of unequal opportunities, the society may be able to accept some loss of efficiency, in exchange for more equity. In this case, there will be a *trade-off between efficiency and equity*, because the competitive equilibrium is not fair in the first place.

12.3.4 The trade-off between efficiency and equity

The distortionary impact of redistributive policy can be easily checked, considering again that the government transfers a fixed amount T to all unskilled individuals, and finances the policy with a proportional tax on the skilled. With taxes and subsidies, the break-even condition that turns investment worthwhile becomes $(1 - \tau)(\lambda - F) > 1 + T$, which solves for $F < \lambda - \frac{1+T}{1-\tau}$. The efficiency loss results from the fact that some individuals who would opt to invest in education in the absence of taxes will now not prefer to remain unskilled. The total efficient loss in the economy will correspond to the projects that fail to be implemented by the mass of agents with F such that²³⁹:

$$\lambda - \frac{1+T}{1-\tau} < F < \lambda - 1 \quad (12.8).$$

Of course, in the real world, governments have more tools than income taxes and lump sum transfers to redistribute income, and some of these tools are less distortive in respect to individual' choices. Taxing goods with negative externalities, for instance, generate

²³⁹ Note that, by abstracting from the choice between consumption and leisure, the model ignores the fact that a transfer will induce individuals to work less, just like in the previous model.

revenue and are efficiency enhancing. On the expenditure side, social benefits can be adjusted to reduce disincentives to work. Government supports to health and education of the poor, public investments in infrastructure and legislation addressing cultural barriers to social mobility may help mitigate differences in F , achieving the twin goal of reducing inequality and promoting growth. Depending on the quality of the overall policy package, the trade-off between efficiency and equity can be more or less severe.

12.3.5 Redistribution and aggregate savings

Another incarnation of the trade-off between efficiency and equity refers to the impact of inequality on savings. The main assumption underlying this reasoning is that rich people have a higher propensity to save than poor people. Hence, a redistributive policy from rich to poor would come along with a decrease in the saving rate. Since in general saving rates are lower than the golden rule (remember the Solow model), such move would be dynamically inefficient.

Along this reasoning, inequality should have a favourable effect on economic growth. By promoting the concentration of wealth amongst the rich, an unequal distribution will be favourable to capital accumulation and by then to economic growth²⁴⁰. Thus, rather than thinking on income redistribution at the early stages of the development process, policymakers should focus on promoting growth in the first place. With a higher per capita income, the poor will be better off, even if their share on income does not change²⁴¹. Distributive policies would then better addressed in a later stage of development.

The idea that rich people save more has a long tradition in the economic theory. In the early literature, the focus was on the functional distribution of income (how much income

²⁴⁰ Kaldor, N., 1960. *Essays on value and distribution*. Glencoe, Ill Free Press.

²⁴¹ Evidence that per capita income growth benefits the poor: Dollar, D., Kraay, A., 2002. Growth is good for the poor. *Journal of Economic Growth* 7(3), 195-225. Bourguignon, F., Morrison, C., 2002. Inequality among world citizens: 1820-1992. *American Economic Review* 92, 727-744.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

accrues to inputs such as capital and labour). Economists such as David Ricardo to Karl Marx theorized that profits are the primary source of capital accumulation. Early models of economic growth, such as Lewis (1954) and Kaldor (1957) assume that only capitalists save, while labour income is basically devoted to consumption²⁴². As you may remember from the Solow model, such assumption imply a saving rate equal to the golden rule.

More recently, economists shifted attention to the *personal distribution* of income (how much income accrues to each family, given their ownership of capital and labour). Arguably, savings can only be built beyond a minimum level (subsistence) of consumption. Although everyone would like to save for the future or to leave a bequest for the heirs, the needs of the present come first, and many households do not afford to save at all. Thus, in general, the poor are expected to save less than the rich.

Empirically, most studies based on households data find evidence that that rich people indeed save proportionally more than the very poor, and that the marginal propensity to bequest is higher for the wealthy²⁴³. However, cross country studies using aggregate data find no evidence that more income inequality comes along with higher saving rates²⁴⁴. An obvious problem in empirical estimates using aggregate data is that savings are the result of individual decisions, and individuals differ in terms of preferences, income, and access to finance. Thus, even if there is a tendency for rich people to save more, aggregate data will be influenced by compositional effects making it difficult to establish a general proposition relating inequality to aggregate savings.

Arguably, societies are not only composed by very rich and very poor, but also by middle class people. The very poor, indeed, may find little usefulness in saving, because the

²⁴² Lewis, A., 1954. Economic development with unlimited supply of labour. The Manchester School 22, 139-191. Kaldor, Nicholas, 1957 A model of economic growth. The Economic Journal 67 (268), 591-624.

²⁴³ Recent evidence from Dynan et al (2004) using data for the United States shows that high-income households save a larger fraction of their permanent income than low-income households. [Dynan, K., Skinner, J., Zeldes, S., 2004. Do rich save more? Journal of Political Economy 112(2), 397-444.

²⁴⁴ Schmidt-Hebbel, K., Serven, L., 2000. Does income inequality raise aggregate saving?, Journal of Development Economics 61: 417-446.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

maximum save they could afford would never be enough to change their living standards. Middle class people, ie, those who have their basic needs already satisfied, may want to engage in high saving, in order to deliver to their children better conditions than what they had when young. Thus, a redistributive policy aiming to get families out of poverty into the middle class with aspirations may actually help promote economic growth. In contrast, in a very unlevel society, where social barriers prevent the majority of the population to get out of poverty, growth prospects will be trapped by the sentiment that savings and work effort by the many will not make any difference.

12.4 Financial market imperfections

So far, we have been discussing the classical trade-off between efficiency and equity. In this section, we explore a different line of reasoning: inequality may be itself a source of inefficiency, by preventing poor individuals from implementing projects that are valuable from the social point of view.

Barriers to entrepreneurship of the poor can be of various order, ranging from cultural or social discrimination, and credit availability. In consequence of these barriers, in societies with high incidence of poverty, many potential investments by talented individuals are lost. This is not only harmful to economic growth, but also to social mobility: those who born poor will remain poor, and those who born rich will be able to invest and remain rich²⁴⁵.

In this section, we illustrate the poverty trap referring to a particular barrier to investment by the poor, which are credit constraints²⁴⁶. A main reason why income inequality can deter entrepreneurship is that many projects involve fixed costs that are indivisible. Agents not affording to pay for these fixed costs may fail to implement the investment. In

²⁴⁵. This vicious cycle in poverty could also be stated in terms of attitude towards fertility: poor families facing Malthusian traps will tend to prioritize the quantity of children rather than investing in their education.

²⁴⁶ Galor, O., Zeira, J., 1993. "Income Distribution and Macroeconomics," *Review of Economic Studies*,. 60(1), pages 35-52.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

principle, agents could turn to the financial markets to borrow. However, credit is not always available. In countries with weak legal systems, loan contracts are more difficult to enforce. When this is so, lenders tend to be more demanding in terms of collateral requests: banks require borrowers to set aside some valuable assets, such as land or a building, that will be taken by the bank in case the borrower defaults on the loan. Under collateral-in-advance constraints, the distribution of wealth determines the distribution of credit in the economy. Poor people lacking the required collateral are banned from the credit market, regardless the quality of their projects. Credit market inefficiencies open a channel through which inequality influences growth.

12.4.1 *A inter-generational model with bequests*

To illustrate the above argument, we consider a model with dynasties. Each dynasty is composed by a sequence of households that transmit wealth across generations, via bequests. Each household has only one member and lives only one period.

To focus on the role of inherited wealth, we abstract from cross-family differences in preferences or in the exogenous parameter F . It will be assumed that all individuals in all generations are alike in terms of capabilities, with $F < \lambda - 1$, implying that for any agent it will pay to make the investment. If for any reason the investment is not made, the households will be paid the subsistence salary $w = I$.

Because wealth is a stock variable, we need a law of motion describing its accumulation. We assume that each household receives a bequest Σ_t from her mother, and delivers a bequest Σ_{t+1} to her daughter corresponding to a fraction s of her total income²⁴⁷. That is:

$$\Sigma_{t+1} = s(\Sigma_t + x_t) \quad (12.9)$$

²⁴⁷ Formally, you may assume an utility function of the form: $U = (1-s)\ln C + s \ln \Sigma_{t+1}$, which is maximized subject to the budget constraint $C_t + \Sigma_{t+1} = y_t$. This gives equation (9).

If the individual does not invest and supplies raw labour, the bequest to the next generation will be:

$$\Sigma_{t+1} = s(\Sigma_t + 1) \quad (12.9a)$$

In case the individual invests, the bequest will be

$$\Sigma_{t+1} = s(\Sigma_t + \lambda - F) \quad (12.9b)$$

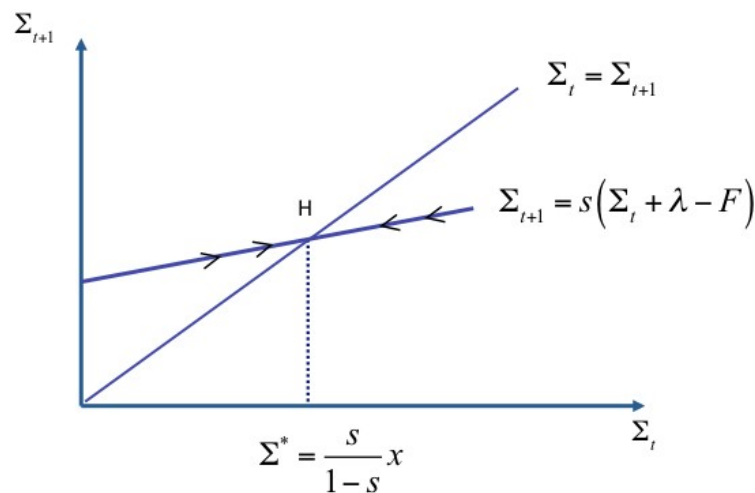
12.4.2 The case with perfect financial markets

For a moment, let's assume that financial markets are perfect. To make the story simple, we set the interest rate equal to zero. Consider the problem of an individual household: since $F < \lambda - 1$, the individual will always find it optimal to invest. In case $\Sigma_t \geq F$, he can pay the fixed cost directly with the bequest, without need for credit. If instead $\Sigma_t < F$, the bequest will not be enough to pay the fixed cost, so she will borrow the amount $F - \Sigma_t$. In any case, her total income will be $F - \Sigma_t + \lambda$, and the bequest for the next generation will be given by (12.9b).

Equation (12.9a) is a dynamic equation whereby the next generation's bequest is a function of the current generation bequest. Since $s < 1$, the state variable converges to a stable equilibrium, where:

$$\Sigma = \frac{s}{1-s}(\lambda - F) \quad (12.10)$$

Figure 12.5 – Dynamic of bequests under perfect financial markets



The equilibrium with perfect financial markets is described in Figure 12.5. As you may easily check, the equilibrium is globally stable. This means that each dynasty will converge to that equilibrium, irrespectively of the initial bequest: any eventual initial inequality in wealth distribution will vanish in the long run, and the society will end up egalitarian.

12.4.3 The case with borrowing constraints

We now turn to the case in which individuals cannot hire a loan exceeding the bequest they have. As before, we assume that $F < \lambda - 1$, so it would pay for any individual to invest, and all projects are socially valuable. Because credit is rationed, however, individuals will only be able to invest in case the inherited bequest is high enough. Hence, her labour income will be

$$x_t = \begin{cases} 1 & \text{if } \Sigma_t < F \\ \lambda - F & \text{if } \Sigma_t \geq F \end{cases} \quad (12.11)$$

Using (12.11) and (12.9), the dynamics of bequests becomes

$$\Sigma_{t+1} = \begin{cases} s(1 + \Sigma_t) & \text{if } \Sigma_t < F \\ s(\lambda - F + \Sigma_t) & \text{if } \Sigma_t \geq F \end{cases} \quad (12.12)$$

Now the “bequest” function exhibits a discontinuity, shifting from “low income” to “high income” when the bequest increases above the fixed cost. In this case, different equilibria may arise, depending on initial wealth conditions. This is illustrated in figures 6-8.

In Figure 12.6, we describe the benign case, where:

$$F < s/(1-s). \quad (12.13)$$

In this case, there is only one steady state, corresponding to point H: even if a dynasty started out very poor – say starting in point P – so that bequests are too small for individuals to invest, bequests will be growing over time and a time will arise when it becomes profitable for individuals in this dynasty to invest: at the time $\Sigma_t \geq F$, the individual will find it profitable to invest and will shift to the upper segment. Then, generations of educated individuals will follow, and the size of bequests will keep increasing over time, until the benign equilibrium (point H) is reached.

The benign outcome H is the only equilibrium in the model, and is stable, because it will be reached whatever the starting point of a dynasty is. If condition (12.13) holds, the society will end up egalitarian. In this case, the only implication of the financial market failure is to delay the convergence of a poor dynasty to the high income equilibrium: in contrast to what happens in Figure 12.5, a poor dynasty has to evolve along the lower segment until the bequest becomes large enough to pay the fixed cost.

Figure 12.6 – Equilibrium with investment

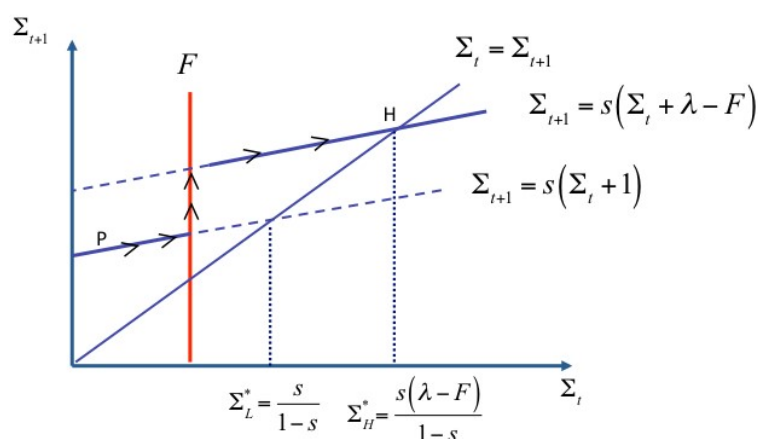
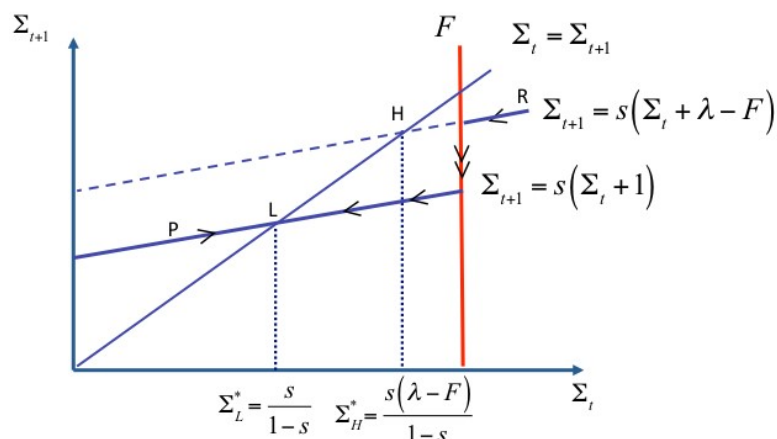


Figure 12.7 describes an alternative case, in which $F > s(\lambda - F)/(1 - s)$, or, which is the same,

$$s < F/\lambda \quad (12.14)$$

In this case, the savings rate is just too small to allow for a benign steady state. Thus, even if a dynasty starts out rich (point R in Figure 12.7), the fact the saving rate is low implies that at some point an individual in that dynasty will not afford the enough to pay for the fixed cost. Hence, the investment will not be made, and the dynasty jumps to the lower segment in (12.12). Thereafter, bequests will keep decreasing over time, until the bad equilibrium “L” is reached.

Figure 12.7 – Equilibrium with no investment



Note that point L is the only equilibrium in Figure 12.7. There is no investment, even if fixed costs are such that all individuals potentially break even. The equilibrium will also be egalitarian, but with poverty.

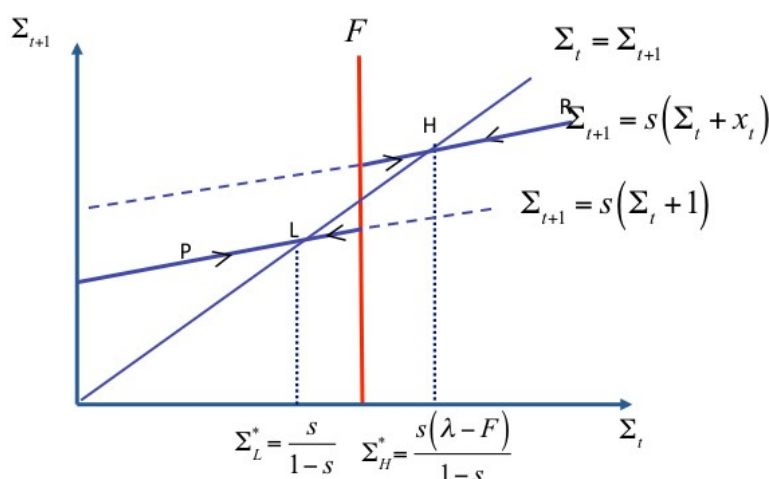
Finally, the case with multiple equilibrium is depicted in Figure 12.8. In this case, none of the conditions (12.13) and (12.14) is verified, so F is in an intermediate level. In this case, a dynasty starting out rich point (R) will approach the high income equilibrium, while the family starting out poor (P) will never inherit the enough to pay for the fixed cost, even if it was efficient to do so. A society that starts out unlevel will remain unlevel. This case fits well with the empirical finding that countries with higher levels of inequality tend to have lower mobility between generations, with parents earnings being the more important determinants of children's earnings²⁴⁸.

Just like the case in Figure 12.7, the equilibrium described by point L in figure 12.8 involves a loss of efficiency, because socially valuable projects ($\lambda - 1 > F$) are not implemented. The difference between cases 7 and 8 is that in the case of Figure 12.8 a temporary transfer to the poor could have permanent effects: if the transfer was large enough for the individual to get educated, then the high income equilibrium will be reached. Note that this will not happen in Figure 12.7, because in that case savings are just too small.

²⁴⁸ Corak, M. 2013. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." *Journal of Economic Perspectives*. Volume 27(3), 79-102.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Figure 12.8 – Multiple equilibrium



12.4.4 Redistribution and efficiency revisited

In the model above, we distinguished two cases: one in which financial markets are perfect, and the second in which an individual cannot borrow in excess of its initial wealth. When financial markets are perfect, the efficient outcome will be immediately reached: the market mechanism will ensure that all socially valuable investments are implemented. Moreover, because all individuals are equally capable, any initial wealth inequalities will vanish over time. Under imperfect financial markets, conclusions are much different. If the borrowing ability of an individual is constrained by initial wealth, valuable projects will not be implemented, and there will be a social loss. A redistributive policy in this case has the scope to improve both equity and efficiency.

In terms of the model, the policy of taxing the rich to transfer to the poor will cause the lower segment in Figure 12.8 to shift upwards, and the rich segment to shift downwards²⁴⁹. The appropriate policy must therefore balance the need to increment the enough the initial wealth of the poor to induce a virtuous cycle, but without forcing the rich into a downward

²⁴⁹ Equation (9a) shall be replaced by $\Sigma_{t+1} = s(\Sigma_t + 1 + T)$ and equation (9b) is replaced by $\Sigma_{t+1} = s(\Sigma_t + (1 - \tau)(\lambda - F))$.

spiral. As argued above, there are redistribution policies other than fiscal: public provision of education and wealth, as well as subsidies to small scale entrepreneurship may reduce F , without distorting the economy so much.

12.4.5 *Human versus physical capital*

Credit constraints affect more investments in human capital than investments in physical capital. The main reason is that human capital is *embodied* in people, while physical capital is not. This differential characteristic determines differences in the way the two types of capital can be utilized and accumulated. While the machinery one' buys can be operated by third parties, human capital can only be used by its owner. Physical capital can be sold in secondary markets, but human capital cannot. The implication is that investments in physical capital are easily collateralized, while investments in human capital are not.

Another difference is that physical capital can be accumulated by a sole individual without bound, while there is a limit in the amount of human capital an individual can accumulate: acquiring human capital is time consuming and requires individual efforts²⁵⁰. Thus, it will not be possible to offset a low investment in human capital by the majority of population with an extremely high investment in human capital by the few rich: for the average level of human capital in an economy to expand, accumulation must be widely spread among individuals in the society. This contrasts to physical capital, that can be expanded in an economy with only few people investing.

Finally, even if it was feasible for a rich person to acquire a gigantic amount of education, that would not be optimal from the individual point of view. The reason is that each individual faces diminishing returns on human capital accumulation: the return of say, 12 years of schooling is expected to be less (in terms of wages) than twice the return of 6

²⁵⁰ Note that this is different from what we assume for human capital in the aggregate: that is, the stock of human capital embodied in the society is allowed to increase without bound, while human capital embodied in individuals with finite lives is bounded by some feasible level.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

years of schooling. The same does not apply to physical capital: an investor can buy a large quantity of physical capital without facing diminishing returns, because physical capital can be rented to other people without losing its effectiveness. Physical capital faces diminishing returns in the aggregate, but at the individual level the rental price of capital is exogenous. From an individual point of view, the optimal investment plan will therefore consist in acquiring human capital only up to the point where the marginal product of human capital equals the interest rate. Before that point, it only makes sense to invest in human capital, and after that point it only make sense to invest in physical capital. The implication is that rich people will invest their marginal wealth in the form of physical capital, only. Poor people, on the contrary, are expected to direct most of their investment to human capital, because at low levels of human capital its marginal return is higher than that of physical capital²⁵¹. Arguably, it would be optimal for the poor to borrow at the exogenous interest rate and invest in human capital up to the point where the private marginal return of human capital is equal to the interest rate. Under incomplete financial markets, however, such move will not be possible, because human capital cannot be collateralized.

What are the implications of all this? In a society where many people are poor, and face borrowing constraints, there will be underinvestment in human capital. In the equilibrium, the marginal product of human capital will be higher than the marginal product of physical capital, the allocation of resources will not be efficient, and per capita income will be lower. Obviously, the size of the distortion depends on the marginal return on human capital. Some authors argued that at initial stages of development the primary engine of growth is physical capital, not human capital. In such an environment, inequality could stimulate growth, by allowing a large concentration of wealth among the rich, and favouring investment in machinery that can be operated by anyone in the society. Then, when the country achieves a minimum level of physical capital, returns to physical capital go down,

²⁵¹ Becker, 1975. Human Capital. NBER. Cambridge. Galor, O., Moav, O. 2004. From Physical to human capital accumulation: inequality and the process of development. *Review of Economic Studies*, 71(4), 1001-26.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

and more sophisticated technologies require human capital. Human capital accumulation becomes the dominating engine of economic growth. The small availability of educated workers will then imply high returns on human capital. In this phase, inequality coupled with financial market imperfections will deprive the poor from investments in human capital that cannot be replaced by the rich, retarding growth. This theory fits the stylized fact that the relationship between inequality and growth is non-linear: inequality is positively correlated with growth at low levels of income, because it favours physical capital accumulation, but retards growth at high levels of income, where human capital is needed²⁵².

This discussion suggests that an appropriate redistributive policy shall include the public provision of education. Education is considered as a public good, because it boosts the economy' permeability to technological change. But it is also a private good, because individuals are able to reap a return on their investments on education. However, people may not have enough money to invest in their education or in their children's'. In a world with credit constraints, public provision of education is a key ingredient for economic development.

Another avenue is to address the financial market imperfections directly. The availability of credit to the poor is a main concern for policymakers in our days. In some countries, governments have promoted the emergence of lending institutions specialized in extending small loans to poor borrowers who lack the collateral to borrow in normal channels. These institutions are said to be specialized in microcredit. "Microloans" are typically conceded to groups of individuals applying for loans together. Depending on the cohesion inside the group, mutual guarantees and peer pressure are expected to help mitigate the problems of individual moral hazard, turning the collateral unnecessary. The first experience in creating a bank specialized in microcredit was at the Grameen Bank, in Bangladesh, established in 1983.

²⁵² Barro (2000), Brueckner and Lederman (2018), op cit.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

12.5 Inequality and political institutions.

In the previous sections, it was argued that inequality may impact indirectly on economic growth via redistribution policies: in an unlevel society, there will be social pressure for fiscal redistribution, and distortionary taxation hurts growth. The argument presumes that the society is run by a well-functioning representative democracy. Politicians seeking to get votes implement policies that they perceive to be desired by the majority of the population.

Non-democratic societies, notwithstanding, also face forces towards income redistribution. Typically, in non-democratic societies, the decisive power is not in the hands of the majority of the population, but instead in the hands of privileged elites. Privileged elites often exert their political influence to extract further income from the poor, placing additional barriers to social mobility and growth, blocking fundamental reforms, and creating extractive institutions. Glaeser et al. (2003) distinguish “Robin Wood” redistribution, that runs from rich to poor and operates through the democratic choice channel, from “King John” redistribution, whereby the rich use bribery and influence to capture the state and influence governmental decisions in their favor²⁵³. This theory adds to the general case that inequality creates pressures for redistribution, either from rich to poor or from poor to rich. In either case, there is a negative influence of inequality on economic growth, that is mediated by government actions²⁵⁴.

Additionally, there are pressures for income distribution running outside the government channel. Poverty, inequality, and perceptions of unfairness impact negatively on trust and on social cohesion, undermining the credibility of a political regime, and creating

²⁵³ Glaeser, E., Scheinkman, J., Schleifer, A., 2003. The injustice of inequality. *Journal of Monetary Economics* 50, 199-222.

²⁵⁴ Banerjee and Duflo (2003) found that changes in inequality in either direction lower growth in the subsequent five-year period. They interpreted this finding as supportive of the notion that redistribution in general hurts growth, at least over short- to medium-run horizon [Banerjee, A., Duflo, E., 2003. Inequality and Growth: What Can the Data Say?. *Journal of Economic Growth* 8, 267–299].

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

the potential for social conflicts. The United Nations (2005) makes the point: “Poverty increases the risk of conflict through multiple paths. (...) Without productive alternatives, young people may turn to violence for material gain, or feel a sense of hopeless, despair and rage. (...). The lack of economically viable options other than the criminal activity creates the seedbed of instability – and increases the potential for violence” (p.6,8). Thus, in unequal societies, pressures for redistribution may arise in the form of violence, destruction of property, social unrest, and political instability. These threats reduce the expected private returns on investment and innovations, and by then economic growth²⁵⁵. Hence, in cases of extreme poverty, an effective redistributive policy may impact positively on aggregate productivity. By levying taxes on the rich to reduce social inequalities, the government will indirectly improve the enforcement of the law and the protection of property rights. With a lower incidence of crime and property offenses, there will be a friendlier economic environment for investment and job creation. In a safer social environment, less private and public resources will be needed to secure property and maintain the public order.

The perverse interaction between inequality and the quality of institutions and of the economic environment is also a source of vicious cycles and virtuous cycles: inequality produces weak institutions, which in turn reinforce the asymmetry in the distribution of income. Thus, just like good institutions are favorable to social mobility and less inequality, an unequal society will tend to produce extractive institutions which will block the social mobility of the poor.

The fact that inequality interacts pervasively with the social environment and the political process provides an argument for Robin Wood redistribution. Even if fiscal redistribution comes at the cost of some loss of efficiency, by reducing the power of the privileged elites, the government may be improving indirectly the enforcement of the law and the protection of property rights, and will tend to implement better policies. With a lower

²⁵⁵ Alesina, A., Perotti, R., 1996. “Income distribution, political instability and investment”. *European Economic Review* 40, 1203-1228. Perotti, R., 1996. Growth, income distribution and democracy: what the data say. *Journal of economic growth* 1(2), 149-87.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

incidence of crime and property offenses, there will be a friendlier economic environment for investment and job creation. In a safer social environment, less private and public resources will be needed to secure property and maintain the public order. In that case the redistributive policy may be seen as included in the public input that leads to higher productivity.

12.6 Key ideas of Chapter 12

- In this chapter, we focused on the relationship between income inequality and economic growth. There are various links through which inequality may affect economic growth.
- The conventional wisdom is that a certain level of inequality is necessary to create the incentives for people to work hard and invest, helping the economy to grow faster.
- Individuals are not all alike. Because of exogenous capabilities, or simply because of luck, individuals will be differently rewarded by the market. In general, the society dislikes inequality caused by exogenous factors and is able to sacrifice some efficiency to achieve more equity.
- Some authors argued that one reason why inequality correlates negatively with economic growth is that citizens in more unequal societies are more likely to press political leaders for redistribution, and redistribution hurts growth. In general, fiscal redistribution weakens the economics incentives and is detrimental to growth. There are however other forms of government intervention, such as public support to education and wealth, that may simultaneously promote equity and growth.
- One reason why a redistributive policy might be costlier is that rich people tend to save more than poor people. That being the case, a redistributive policy from the rich to the poor may deprive the economy from valuable savings to increase the stock of physical capital. Such argument is not expected to be valid at later stages of development when skills, rather than physical capital, become the engine of growth.
- Financial market imperfections may impede lower-income, liquidity constrained households from undertaking projects that are socially valuable. This is especially true for investments in human capital, because human capital cannot be set aside as a collateral for a loan. Borrowing constraints act as a barrier to social mobility, dooming poor people to remain poor.
- In general, income inequality is likely to come along with excess investment in physical capital relative to human capital, giving rise to a static inefficiency. Scarcity of human capital is also a source of dynamic inefficiency, because it reduces the permeability of the economy innovation and technology adoption.
- Inequality is inherently linked to pressures for income redistribution. The traditional view is that redistribution policies aiming to reduce inequality come at the cost of economic inefficiency, thereby having a negative impact on growth. However, in very

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

unequal societies, the state may be captured by privileged elites, and start ruling according to their interests, imposing barriers to competition, and creating economic rents. “Prince John” redistribution is more harmful to growth than “Robin Wood” redistribution, because it operates through dysfunctional institutions.

- High inequality also comes along with pressures for redistribution not mediated by the government. This includes crime and violence, property damage, and political instability, that pay a toll on investment and growth.

12.7 Problems and Exercises

Key concepts

- *Trade-off between efficiency and equity . Efficient redistribution*

Essay questions:

- Discuss: “Inequality is good for growth because rich people save more”.
- Discuss: “Inequality is necessary for the functioning of a market economy”.
- Discuss: “Inequality comes along with a less in the use of physical capital versus human capital”.

Exercises:

- 12.1.** Consider an economy, where households have the following preferences over consumption (C) and leisure (l): $U = \alpha \ln C + (1 - \alpha) \ln l$. Each household is endowed with a given amount of time $h=1$. Labor is homogeneous, firms are competitive, and the production function is $Q = 100N$, where N denotes for labor input. (a) Find out the demand for labour in this economy. (b) Find out the optimal demands for consumption and leisure. (c) Suppose that there were three groups of households, differing according to preferences only. In particular, assume that $\alpha_1 = 0.2$, $\alpha_2 = 0.25$ and $\alpha_3 = 0.5$. In this case, what will be the wage income and the consumption levels of each group? (d) Suppose that total population in this economy amounted to 100 households, of which 50 belonging to group 1, 40 to group 2 and 10 to group 3. In that case, how much would be per capita output? In a democratic system, would this society press for income redistribution? (e) Suppose now that there were 60 households in group 1, 0 in group 2 and 40 in group 3. In that case, how much would be per capita output? Would this society be more equal or less equal than in d)? Would there be pressures for redistribution in this case? (f) Departing from e), assume that the government decided to transfer $T=10$ to each household of type 1, financed by a proportional tax on type 3' wage income. How much will be the required tax rate? Describe the impact of this policy in the welfare of each group, on the respective labour supplies, and on total

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

output of this economy. (g) Would the redistribution policy described in (f) be socially desirable? Why?

- 12.2.** Consider an economy, where households live one period and have the following preferences over consumption (C_t) and bequests for the next generation (Σ_{t+1}): $U = 0.8 \ln C_t + 0.2 \ln \Sigma_{t+1}$. Each household has a certain disposable income, y_t , to be spent in consumption or saved for bequests, that is, $y_t = C_t + \Sigma_{t+1}$. As for production, each household has the opportunity to engage in an unskilled job, earning $w=1$, or to enter in a modern sector, earning $\lambda=2$. Entering in the modern sector involves, however, a fixed cost, F , to be paid in advance. Individuals in this society differ in respect to the level of F . In particular, assume that there are 20 individuals facing $F_1=1.2$, 60 facing $F_2=0.8$ and 20 individuals facing $F_3=0.4$. (a) Interpret the fixed cost F . (b) Find out the optimal demand for bequests, as a function of disposable income. (c) Assume for a moment that capital markets were perfect. Find out the long run income level in each group, as well as the corresponding consumption levels and bequests. (d) Sticking with the assumption of perfect financial markets assume that the government launched a tax $\tau=0.2$, proportional to the value added in the modern sector, $\lambda - F$, to finance a lump-sum transfer to individuals engaged in cottage production. Describe the impact of this redistribution scheme on production, consumption and income of the three groups in the society. Discuss the desirability of this tax in term of efficiency vs equity. (e) Now suppose that there were 100 individuals were of type 3, but capital markets were absent. Further assume that 40 individuals in the society started out with a bequest $\Sigma_t = 0.25$, while the other 60 started out with a bequest of $\Sigma_t = 0.4$. Would this economy approach the efficient outcome? Quantify. Could a temporary redistributive policy be efficiency enhancing in this case? Why? (f) Finally, assume that productivity in the modern sector increased to $\lambda = 2.4$, with the fixed costs being $F_3 = 0.4$. Assuming again that 40 individuals in the society started out with a bequest $\Sigma_t = 0.25$, while the other 60 started out with a bequest of $\Sigma_t = 0.5$, could a temporary redistributive policy be efficiency enhancing in this case? Quantify.
- 12.3.** Consider a dynasty, where bequests are $\Sigma_{t+1} = 0.2(\Sigma_t + 1)$, in case the household does not invest, and $\Sigma_{t+1} = 0.2(\Sigma_t + 2 - F)$ in case the household invests. Further assume that $F = 0.4$. Consider a member of the dynasty starting out with a bequest equal to $\Sigma_t = 0.25$. (a) Which factors can be captured by parameter F ? What will be the outcome of this dynasty, in case: (b) financial markets are perfect; will this allocation be efficient? (b) there are no financial markets; will this allocation be efficient? (c) What do you conclude regarding the effects of poverty on social mobility and on economic performance?

13 Corruption

“How deep is your love?” [Bee Gees]

Learning Goals:

- Understand how departing from the benevolent planner assumption alters the nature of the equilibrium
- Understand why generalized corruption can be much more detrimental to growth than centralized corruption.
- Understand why strategic complementarities may lead to endemic corruption.
- Acknowledge the critical role of institutions as a fundamental factor explaining economic development.

13.1 Introduction

In the previous chapters, we stressed the key role of government policies for economic development. So far, however, we have assumed that policies are designed and implemented by benevolent planners whose interests are aligned with the social interest. This chapter departs from this perspective.

In the real live, public programs are instituted in complicated political processes and implemented through complex bureaucracies. Instead of benevolent planners, political leaders often pursue their own selfish objectives, using their powers to keep themselves in office or to direct resources to their political supporters. In many countries, government officials are primary agents of diversion, seeking to maximize their own benefit through extortion, corruption fees or unduly appropriation of public assets. As a by-product, in societies with high levels of corruption, individuals tend to spend valuable resources in seeking for fast money and special favours, instead of devoting them to production and innovation. When corruption is very high, institutions become dysfunctional, paving the way for corruption to become self-sustained.

In this chapter, we enrich the Solow growth model to examine the implications of corruption and rent seeking on economic performance. This will also provide an opportunity <https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

to discuss the key role of political institutions in aligning the objectives of decision-makers with the public interest. In this discussion, three models of corruption will be considered. In the first model, a *non-benevolent despot* (the kleptocrat) empowered with perfect control over its bureaucracy uses his discretionary power with the aim to maximize his personal theft from the government budget. The only limits he faces are the political, administrative, and legal institutions he cannot change. The second model (*decentralized corruption*) examines the case in which a benevolent leader delegates discretionary power on a large number of non-benevolent public officials which corruption activity cannot be coordinated. In this case, the level of corruption will depend on the ability of the planner to design incentive compatible contracts. Finally, we consider the case of generalized corruption, that affects all levels of the public administration. In this case, there is no benevolent planner seeking to design optimal institutions. The likelihood of detection and punishment decreases dramatically, institutions and policies become highly dysfunctional, and corruption becomes endemic.

The chapter proceeds as follows. Section 13.2 defines corruption and gives some real life examples. Section 13.3 introduces the model with centralized corruption. Section 13.4 addresses the case where corruption is undertaken by a large number of public officers whose activity the benevolent planner cannot control. Section 13.5 briefly reviews how societies in the real-world deal with the problem of corruption and the role of institutions in shaping incentives. Section 13.6 analyses the case where institution become dysfunctional, and corruption becomes endemic. Section 13.7 concludes.

13.2 Corruption

13.2.1 What is corruption?

Corruption may be defined as an “act in which the power of public office is used for personal gain in a manner that contravenes the rule of the game”²⁵⁶. This includes embezzlement, the appropriation of public assets for personal use, the celebration of lucrative contracts to business owed by the public officer’ relatives (actions that the public officer can carry out *alone*), bribery and extortion (actions that necessarily involve *two parties*).

Three conditions are necessary for the existence of corruption: (i) the public official must have the authority to design or administer policies and regulations; (ii) this discretionary power must allow the extraction or creation of economic rents; (iii) the incentives embodied in political, administrative, and legal institutions must be such that the official has incentive to use his discretionary power to extract or create rents²⁵⁷.

The incidence of corruption varies widely across countries. Corruption is more pervasive in the developing world but is a matter of major concern all over the world. Corruption may affect both the lower level of administration and the top levels in the government. Where corruption emerges, it is not because people there are different, but because there are economic or social incentives for it. As stated by Douglas North “if the institutional matrix rewards piracy more than productive activity then learning will take the form of learning to be better pirates”²⁵⁸.

13.2.2 *What is rent seeking?*

Government actions influence the profitability of private agents. To the extent that government officers have the power to set tariffs and subsidies, to buy goods and services, to license industrial activities and to regulate monopolies, they will face pressures from economic agents seeking to obtain favours and special regulations. Among those, permits and

²⁵⁶ Jain, A., 2001. Corruption: a review. *Journal of Economic Surveys* 15(1), 71-121.

²⁵⁷ Aidt, T., 2003. "Economic Analysis of Corruption: a survey". *The Economic Journal* 113, November, 632-653.

²⁵⁸ Douglas North, 1993. The new institutional economics and development”, mimeo Washington University, St. Louis. WUSTL Economic Working Paper Archive (p.6).

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

import quotas face high and inelastic demand by entrepreneurs, so become primary targets for these pressures. Devoting potentially productive resources to persuade politicians and civil servants to take actions that generate income transfers or rents to particular individuals or groups at the cost of the general interest is called “rent seeking”²⁵⁹.

Rent seeking can be either legal or illegal. At the *legal level*, organized lobbies such as trade organizations and unions influence the public decisions, by making pressures and giving financial or electoral support to those parties that better serve their interests. Such influence exists because politicians need votes and financial contributions to their campaigns. At the *illegal level*, agents may influence the decisions of bureaucrats and policymakers by offering them a bribe. Bribery is a form of *pecuniary corruption*: public servants are induced to take actions that deviate from the public interest in exchange for monetary benefit or gifts.

13.2.3 What bribes are for?

There are many things private parties can buy from public officials with bribes or other forms of influence. This includes²⁶⁰:

- Time savings and regulatory avoidance: in many development countries excess bureaucracy and red tape rank very high as an obstacle to doing business. Often firms are given the opportunity to pay bribes to bureaucrats so as “speed up” the bureaucratic process of obtaining the required permits.

- Government revenues: bribes can be used to escape taxes or other payments to the government. A typical case is when a tax inspector accepts bribes in exchange for lower tax collection.

²⁵⁹ More generally, the term rent seeking refers to efforts to obtain wealth transfers without creating any value. A cartel of firms agreeing to raise prices, for instance, is a form of rent seeking that does not involve bribery or pressures on civil servants. In this chapter we are interested on a sub-category, relating to persuading public officers to deviate from the public interest.

²⁶⁰ Gray, C. and Kaufman, D., 1998. “Corruption and Development”, Finance and Development, March.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

- Government benefits: bribes can influence the allocation of benefits to the private sector. This includes monetary benefits (subsidies, pensions) or in kind (food supplies, access to medical care, access to courts, housing, privatizations). For instance, a policeman who is supposed to protect all citizens may be given a bribe to look after a particular interest, only. This comes at a cost of unfair competition, because when the officials can privately sell the protection of property rights to individual firms, they have little interest to provide the public at large with open access to this essential service²⁶¹.

- Government contracts: bribes can be given to influence the choice of private suppliers to the public sector. That is, contracts may be allocated to the firm that pays the largest bribe, instead as to the one that puts the lower bid.

- Influencing legal and regulatory outcomes: bribes can influence how existing laws, rules or regulations are *implemented* with respect to the bribe payers. For instance, bribery can be used to prevent the government from stopping illegal activities, such as pollution and drug dealing, or for a firm to obtain the consent of the competition authorities to charge a monopoly price. At a higher level, bribes can be directed to influence how laws, rules and regulations are *designed*. When bribes are given to parliamentarians to “buy” important pieces of legislation or to government officials to enact favourable regulation, a case of “state capture” is said to take place. This is a deep form of corruption.

13.2.4 *The grease in the wheel's argument*

Many people believe that corruption can be efficiency enhancing: by providing bureaucrats a pecuniary incentive (*speed money*) corruption may help overcome the excess bureaucracy and red tape. As long as bureaucrats give priority to the individuals paying the

²⁶¹ According to Hellman et al. (2003) this mechanism of purchasing *individualized* protection of property rights became very popular in Russia, as a natural response to the general weakness in the rule of law after the collapse of the Soviet Union [Hellman, Joel S., Geraint Jones, and Daniel Kaufmann (2003), *Seize the State, Seize the Day: State Capture and Influence in Transition Economies*, Journal of Comparative Economics, 31, pp. 751-773].

higher bribes and those who offer the higher bribes are those who are carrying the more promising projects, then bribery could be an efficient mechanism to allocate the bureaucrats' scarce time.

The *grease in the wheels* argument fails, however, for various reasons²⁶²: First, this is second-best reasoning: that is, given the rigidity created by the bureaucracy, corruption helps relaxing this rigidity. Clearly, the first best policy would be to address the rigidity itself. Second, since bribery is hidden, it is not equivalent to a competitive bid. Third, the implied "contracts" cannot be enforced by law, so nothing ensures that the higher bids will be actually attended first. Fourth, even if bribery could effectively allocate faster government decisions to those who value it more, the society would be better off if the corresponding revenues were appropriated by the government, rather than by corrupt bureaucrats. Last – but not the least - corruption is often what causes bureaucratic processes to be slowed down, not the other way around: if public officers get rewarded by the existence of red tape and regulations, they will tend to create extra red tape and regulations, just to increase their prey opportunities²⁶³.

Box 13.1 Corruption in the real world

"One of the most extreme real-world examples of theft of productive public infrastructure, according to Abbott (1988, p. 172) involves Luckner Cambronne, a member of the elite that ruled Haiti under the Duvaliers. He apparently had this workman pull up and carefully store the entire rail system lining Port-au-Princes to Verrettes via St. Marc; he then

²⁶² Aidt (2003), op cit.

²⁶³ Kaufman, D. and Wei, S., 2000. Does grease money speed up the wheels of commerce? IMF Working Paper 00/64. Guriev, S., 2004. Red Tape and Corruption, Journal of Development Economics 73, 489-504.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

sold the 150 kilometres of railroad as scrap metal and pocket the money for himself” [Mauro, 2004, p.5]²⁶⁴.

“(…) The bank owned a large stake in one of the country’s most profitable companies. But when the management attempted to sell the stake to the biggest bidder, it was advised by the government to sell the shares to the company’s founder at a quarter of the market price instead. The founder turned out to be a close friend of the country’s president. Where is this bank? It happens to be Crédit Lyonnais in France (…)” [Shleifer and Vishny, 1998, p 1]²⁶⁵.

Box 13.2 The TI Corruption Perceptions Index

The recognition that fighting against corruption and monitoring its progress requires some form of measurement motivated the development of various measures of corruption. A famous indicator is the Corruption Perceptions Index (CPI), produced by Transparency International. The CPI measures corruption perceptions as seen by businessman and country analysts. The index ranges from 0 (highly corrupt) to 10 (highly clean).

Figure 13.1 displays the results of the 2005 survey (159 countries), crossed with per capita income. The positive correlation in the figure reveals a general tendency for the incidence of corruption to be larger in poorer economies. For instance, the Scandinavian bureaucracies rank as the cleanest in the World, while most of the sub-Saharan African bureaucracies rank at the bottom. Whether this reflects a causality from corruption to per capita income or from per capita income to corruption, we don’t know. Arguably, a higher incidence of corruption will give rise to more resources wasted and bad policies, lowering per capita income. Mauro (1995) investigated the correlation between corruption and economic performance, controlling for the possible endogeneity. Using cross-section data for 70

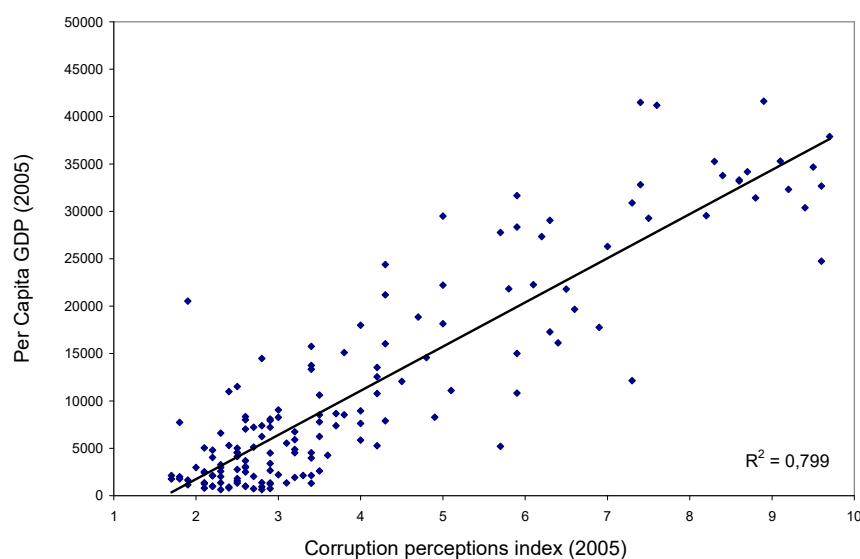
²⁶⁴ Mauro, P., 2004, “The persistence of corruption and slow economic growth”. IMF Staff Papers 51 (1), 1-18

²⁶⁵ Shleifer, K., Vishny, W., 1998. The Grabbing Hand. Harvard University Press.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

countries in the early 1980s, the author found a strong causality from corruption to economic growth. The author also found corruption to be strongly correlated with other indices of bureaucratic and institutional inefficiency, including “political instability” and “inefficiency of the legal system”. Mo (2001) confirmed the impact of corruption on political instability and also found that corruption is associated to lower levels of investment in physical and in human capital, all these channels affecting growth²⁶⁶. A problem however is that fighting corruption and building strong institutions requires resources that may not be easy to rise wherever per capita output is low²⁶⁷.

Figure 13.1 – Corruption perceptions and per capita GDP



Source: Transparency International, <http://www.transparency.org/>.

²⁶⁶ Mauro, P., 1995. Corruption and Growth. *Quarterly Journal of Economics* 110 (3), 681-712. Mo, P. (2001), Corruption and Economic Growth. *Journal of Comparative Economics* 29, 66-79.

²⁶⁷ United Nations (2005, p.16): “Many poor countries without adequate resources for decent salaries - or the checks on political abuse that provide the incentives for performance and the ability to weed out the inept and corrupt – are unable to afford an effective public sector (...)”. [United Nations, 2005. Millenium Development, Project Report, United Nations, New York].

13.3 A model of centralized corruption

This section examines the case of a *non-benevolent leader* (the Kleptocrat), whose only aim is to maximize its personal expenditure. The analysis assumes that the Kleptocrat faces no electoral constraints and is blessed with perfect information and perfect control over the bureaucracy. The underlying model is the Solow model augmented with public inputs already examined in Chapter 10.

13.3.1 Main assumptions

In the private sector, the individual firm production function is given by:

$$Y_i = A_i K_i^\beta N_i^{1-\beta} \quad . \quad (13.1)$$

The productivity term has two components: an exogenous rate of technological progress and an efficiency term related to the ratio of (productive) government expenditures to GDP:

$$A_i = A e^{gt}, \text{ with } A = \left(\frac{G}{Y} \right)^\eta \text{ and } \eta > 0 \quad (13.2)$$

It is assumed that government revenues are raised through a production tax (τ).

In this model, corruption takes the form of *theft from government revenues*, i.e, the planner decides to spend a proportion ϕ of the government revenues “for political reasons”. This translates into a lower provision of productive public inputs:

$$G = \tau(1 - \phi)Y \quad \text{with} \quad \phi \geq 0. \quad (13.3)$$

The total amount of theft is therefore:

$$\Phi = \phi Y. \quad (13.4)$$

In practice, corruption does not need to take the form of theft on the government budget. Corrupt leaders may confiscate assets from the private sector or impose bribes to allocate government resources. From the individual' firm point of view, however, what matters is the total amount paid. Modelling corruption as theft on the government budget is a simple way of capturing many forms of extortion.

13.3.2 The clever Kleptocrat

Assume that, instead of a benevolent planner, you were a kleptocrat whose only objective was to use the government budget for personal expenses. Would you divert *all* tax proceeds to yourself? Clearly, the answer is no: if you were a clever kleptocrat, you would realize that failing to provide essential infrastructure to the economy would impact too badly on aggregate output, and therefore on your tax base.

As an extreme case, just think what would happen if you set $\phi=1$: in that case, there would be no public provision at all ($G=0$). Since government services are essential to production (equation 13.2), there would be no private economy either and you would end up as a bankrupt despot. A clever Kleptocrat must take into account that stealing too much *you may end up eating the egg and the chicken, too*.

Formally, the Kleptocrat problem is to maximize the amount of theft, taking into account how theft impacts on per capita income and by then on government revenues. To see this in terms of our model, just remember the formula for the steady state level of per capita income in the Solow model and adapt it for the existence of a public input. This is equation (10.13), that we reproduce here:

$$y_t^* = \left[\left(\frac{G}{Y} \right)^\eta (1-\tau)^\beta \right]^{\frac{1}{1-\beta}} \left(\frac{s}{n+\delta+\gamma} \right)^{\frac{\beta}{1-\beta}} e^\eta. \quad (13.5)$$

Substituting (13.5) and (13.3) on our variable of interest, (13.4), we get:

$$\Phi = \tau \phi \left[\left((1-\phi)\tau \right)^\eta (1-\tau)^\beta \right]^{\frac{1}{1-\beta}} \left(\frac{s}{n+\delta+\gamma} \right)^{\frac{\beta}{1-\beta}} L_t. \quad (13.6)$$

Now, choosing τ and ϕ to maximize Φ (this is a bit tedious, though not difficult), we obtain (the superscript K stands for the *kleptocrat* solution):

$$\tau^K = \frac{\eta+1-\beta}{\eta+1} \quad ; \quad (13.7)$$

$$\phi^K = \frac{1-\beta}{\eta+1-\beta} \quad . \quad (13.8)$$

Comparing to the benevolent planner' case (10.9), you see that the tax rate is now higher. In terms of Figure 10.3, the equilibrium when the ruler is a fully empowered kleptocrat is represented by point K, with a positive ϕ .

Now, if we substitute (13.7) and (13.8) into (13.3), we realize that there is an agreement between the kleptocrat and the benevolent planner regarding the proportion of output to be spent in public inputs²⁶⁸:

$$\left(\frac{G}{Y}\right)^K = \left(\frac{G}{Y}\right)^G = \frac{\eta}{1+\eta} \quad (13.9)$$

Remember that this fraction corresponds exactly to the contribution of the public input to production, as stated in equation (10.3). The conclusion is that, once you act as a clever kleptocrat, you want resources in your economy to be allocated efficiently. Using Easterly (2001)'s words, you become "solicitous of your victim' prosperity"²⁶⁹

Of course, consumers in this economy will be worse off than in the benevolent planner case. This can easily be demonstrated substituting (13.7) and (13.8) in (10.13) and (10.14).

13.3.3 Dynamic considerations

The above analysis assumes that the kleptocrat maximizes the *steady state* level of theft. However, in the real world, despots do not stay in power forever. Either through democratic elections or through revolutions, non-benevolent leaders may lose their power. Thus, the optimal misappropriation policy from the kleptocrat point of view should also take into account the transition dynamics and length of time in office. Solving such a problem is however beyond the scope of this book.

²⁶⁸ Barro, R., 1990. "Government spending in a simple model of endogenous growth". Journal of Political Economy 98, 103-125.

²⁶⁹ Easterly, W., 2001. "The Elusive Quest for Growth: economist's adventures and Misadventures in the Tropics, Massachusetts Institute of Technology.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Also note that incorporating inter-temporal considerations into the model opens the door for another trade-off: the re-election probability may itself depend on the extent of the theft. That is, if the kleptocrat extorts too much today, he may face a higher probability of losing the chicken of the golden eggs tomorrow (either through democratic elections or through a coup d'état). In this case, the kleptocrat must balance the benefits of more extraction during a shorter period of time with those of less extraction during a longer period of time. Intuitively, this decision shall be dependent on its subjective discount rate: if the kleptocrat is very impatient, he will tend to increase current misappropriation²⁷⁰.

Of course, the probability of dismissal will be more or less sensitive to the extent of the theft, depending on how strong the political regime is: when the kleptocrat leads a strong dictatorship supported by the military cupules, the likelihood of dismissal is lower - and hence theft opportunities are higher - than when the regime is democratic or when generals have not a share in the cake. In a democratic system, the planner may improve the probability of re-election by directing transfers to groups of voters with political influence.

In this judgement it may be wise to find some foreign allies. For instance, suppose you were running an economy endowed with an important mineral resource, such as petroleum. In that case, it would be a good idea to buy extra political stability sharing the cake with a foreign nation with strong military power. You could do so by allowing foreign companies to extract some oil in your territory, in exchange for a military cooperation agreement. If you forgot to do so - and if indeed your mineral resources were significant - then you would most probably face an internal guerrilla, supported by a foreign nation. Being solicitous of your chicken longevity may have a foreign affairs dimension, too.

²⁷⁰ Because government expenditures depend on *contemporaneous* taxation, the model does not allow the kleptocrat to “take all the money and run” ($\phi=1$). But the model could easily be adapted, postulating a one-year delay in the transformation of taxes into government services (e.g, government inputs need one period to enter in the production function). In that case, increasing misappropriation today would impact on the economic performance only tomorrow, giving the Kleptocrat time to take all the money before leaving his post . Such possibility was considered by Ventlou (2002), who analyses the choices of the planner (who maximizes sequential flows of budget misappropriations) and of private agents (who seek to maximize consumption and try to control politicians with voting assessment) [Ventelou, B., 2002. Corruption in a model of growth: political reputation, competition and shocks. Public Choice 110, 23-40].

13.3.4 *Shaping the incentives of the leader*

The theft opportunities of a non-benevolent leader depend on the incentives embodied in institutions that he cannot change. Wherever political, administrative, and legal institutions are not strong enough to persuade politicians from pursuing their own interests, they may take actions that deviate from their constituencies' point of view. Thus, it is the interest of the public to design institutions that reduce the corruption opportunities of non-benevolent leaders.

The most basic instrument to protect citizens from the arrival of selfish political leaders backed by full majority in the parliament is the Constitutional Law. The constitution determines how the different political powers are separate, how high-level public officers are appointed, and what they are allowed to do. Constitutional laws typically can only be changed with a large majority of votes in the parliament, or sometimes with a referendum. At lower levels, different laws can be designed to frame the executive power in different dimensions, and auditing institutions can be created to monitor the compliance of the different political decisions with the law. If the various political institutions are independently appointed and remain uncoordinated, this will favour mutual control.

In complement to formal institutions, there is a role for the *civil society*. A strong civil society backed by a free press that brings watchdogs to the fore helps monitoring the implementation of public policies and in maintaining a continuous pressure on governments to follow policies that best address the people's needs. In some countries, civil society has an explicit consultation role in the decision-making process. Surveillance by civil society will be more effective if there is transparency in decision-making. If government decisions and expenditure programs are publicly known, there will be a further source of social scrutiny over policymakers, inducing them to remain honest. In many countries, government officials are required to make periodic declarations of assets and income sources. This makes more difficult for them to hide illegal revenues.

In general, democracies with strong political institutions, transparent decision making, and a participative civil society tend to be less permeable to corruption than dictatorships where the policy decisions are hidden from public view and the civil society is repressed.

13.4 The model with decentralized corruption

13.4.1 Decentralized corruption and the “tragedy of the commons”

We learned that, under centralized (or organized) corruption, the Kleptocrat looks after the prosperity of its constituencies. In his maximization problem, he takes into account that stealing too much will drive the economy down along the Laffer curve, reducing the tax base. He therefore has incentive to coordinate all the extraction activity, defining the shares each official can have, so that the overall level of corruption does not affect the economy too badly²⁷¹.

A different case occurs when the leader has no control over his bureaucracy. When corruption is undertaken by *a large number* of uncoordinated civil servants, the overall level of corruption will be much higher. The reason is that each corrupt official, being too small to influence the overall outcome, has the incentive to impose as many bribes as he can, without taking into account the shape of the Laffer curve. Like in the “tragedy of the commons”: when the law enforcement becomes too weak, it becomes virtually impossible to preclude any public servant from entering the extraction activity (non-exclusion). However, as more and more people engage in bribery, the amount extracted by each corrupt officer decreases (rivalry). Thus, competition over the common resource would lead to its depletion: in the limit, ϕ would approach 1 and the economy would disappear.

Fortunately, rent seeking is not a costless activity. Bribing, lobbying, matching corrupt officials and corruption opportunities require time and effort. These costs imply that individuals will devote time to rent seeking only to the extent that the reward exceeds the opportunity cost. This mechanism will in general prevent the economy from being totally “exterminated”. The other side of the coin is that rent seeking diverts valuable resources

²⁷¹ Wade (1982) found an interesting example of *organized corruption* in South India: the author observed that each level of the hierarchy in the administration of the irrigation system obtained a fixed percentage of the total bribe [Wade, R., 1982. The system of administrative and political corruption: canal irrigation in South India. *Journal of Development Studies* 18, pp. 287-328].

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

away from production. Private agents, instead of devoting resources to production, will invest part of their talents in seeking for special favours and easy profits. So the economy will be operating below its productivity frontier. The following model addresses these ideas formally.

13.4.2 *Rent seeking as a diversion activity*

To examine the consequences of people devoting part of their effort to rent seeking, let's go back to the Solow model augmented with a public good²⁷². As in the Kleptocrat case, the extraction activity targets the government revenues, τY . The Kleptocrat is however replaced by a *benevolent* leader that cannot control the bureaucracy.

Let ψ be the fraction of time each individual devotes to rent seeking and $1-\psi$ the time devoted to legal work. Output will be determined according to:

$$Y = A_t K^\beta N_Y^{1-\beta}, \quad (13.10)$$

where A is defined as (13.2) and

$$N_Y = (1 - \psi)N \quad (13.11)$$

is the quantity of working-time allocated to production. Since rent seeking takes time, the opportunity cost of rent seeking will be the wage rate.

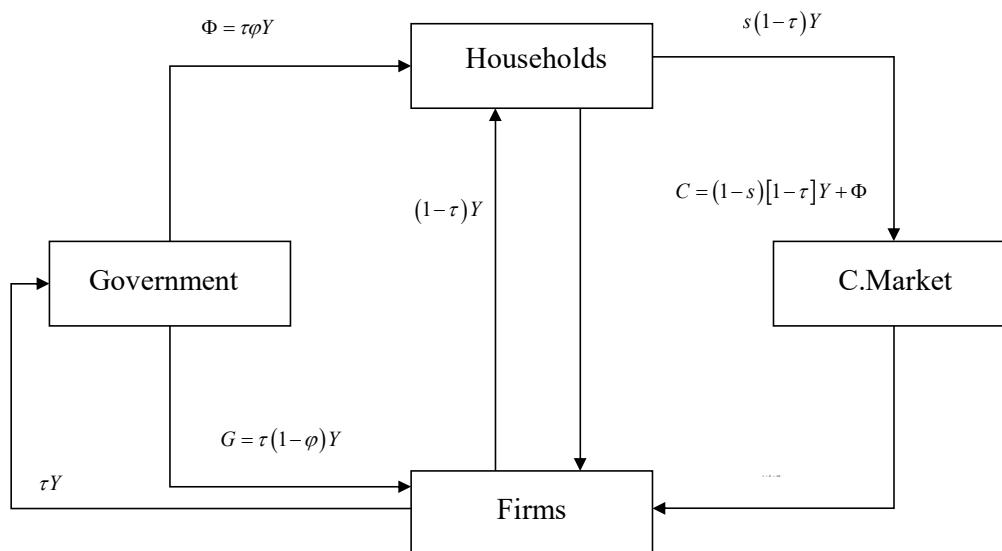
The difference between (13.10) and (13.1) is that the production function is now parametric in the proportion of labour devoted to rent seeking, ψ . In the extreme case in which $\psi=0$, a benevolent planner optimally deciding the tax rate would be able to achieve the first-best outcome, as given by (10.16). Our quest is to find out how much will the production function shift down when the planner has no control over its bureaucracy.

²⁷² The following is an adaptation of the AK model proposed by Mauro (2004), op. cit.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

In this model, we assume that the proportion ϕ of resources deviated from public provision accrues to households, as a reward of rent-seeking. The flow income chart of the economy is displayed in Figure 13.2.

Figure 13.2: The income flow chart with decentralized corruption



13.4.3 The rent-seekers' production function

To find out the equilibrium level of rent seeking, one needs to specify a “production function” relating the time spent in rent seeking to the amount of extraction achieved. To be simple, let's assume that the proportion ϕ of government resources extracted by rent-seekers is a linear function of the proportion of time devoted to rent seeking, ψ :

$$\phi = b\psi \quad (13.12)$$

where b is an exogenous parameter measuring the *effectiveness* of the rent seeking time. In a minute we will discuss how this parameter shall relate to the quality of an economy's institutions.

The total amount of resources deviated from the government budget by rent seekers will be therefore:

$$\Phi = \phi\tau Y = \psi b\tau Y. \quad (13.13)$$

13.4.4 Optimal rent seeking at the individual level

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Households' income in this economy is the sum of the wage bill with total theft:

$$[(1 - \psi)w + \psi bty]N. \quad (13.15)$$

Each household is endowed with a unit amount of time. Thus, the household will allocate its time to legal work or to rent seeking (i.e, he chooses ψ) so as to maximize the term inside brackets in (13.15), taking the wage rate as given. In this model, individuals do not internalise the adverse effect of their appropriative activities on aggregate income and on the wage rate. Abstracting from corner solutions, this leads to the following arbitrage condition:

$$w = b\tau y. \quad (13.16)$$

This condition states that in the optimal allocation of time, spending one extra hour in formal work must pay the same as one extra hour of rent seeking.

13.4.5 The equilibrium level of rent seeking

To determine the impact of individual decisions in the aggregate, we need to determine the wage rate. The wage rate is determined by the labour supply and the labour demand. The labour demand is such that the marginal product of labour is equal to the wage rate. Taking into account the income tax, this is:

$$(1 - \tau) \frac{\partial Y}{\partial N_y} = \frac{(1 - \beta)(1 - \tau)Y}{N_y} = \left[\frac{(1 - \beta)(1 - \tau)}{(1 - \psi)} \right] y = w \quad (13.14)$$

Substituting the wage rate (13.14) in (13.16) and dividing (for convenience) both sides by y , the (macro-level) arbitrage condition becomes:

$$\tau b = \frac{(1 - \beta)(1 - \tau)}{1 - \psi}. \quad (13.17)$$

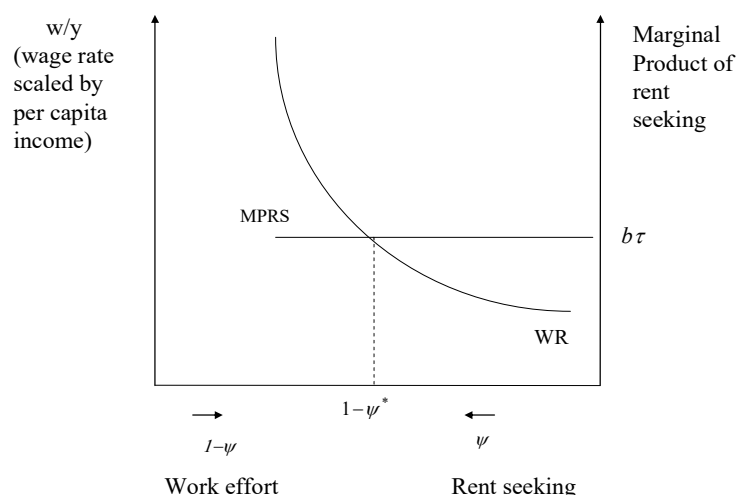
The left-hand side of (13.17) gives the *marginal product of rent seeking* per unit of per capita income. The right-hand side gives the marginal benefit of working time (that is, wages) per unit of per capita income. The equilibrium level of ψ is the one that leaves workers indifferent between the two activities in the margin and solves equation (13.17).

Figure 13.3 illustrates the trade-off between formal work and rent-seeking, as implied by this model. The downward sloping curve WR gives the marginal benefit of working time

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

(the wage rate wages, scaled by per capita income). The curve named MPRS gives the marginal product of rent seeking per unit of per capita income. The equilibrium level of $1-\psi$ is obtained at the intersection of the two curves.

Figure 13.3. The optimal proportion of time devoted to formal work



The equilibrium level of rent seeking obeys to an arbitrage condition whereby the foregone wage rate in the formal sector (that is a decreasing function of the aggregate working time) must equal the marginal product of rent seeking.

In general, the equilibrium level of rent seeking shall be less than one. The reason is that, as more resources are allocated to rent seeking (the economy moves to the left), the reward of formal work rises, because the aggregate production function exhibits diminishing returns. Diminishing returns to labour in the formal sector prevent the scenario of “extermination”.

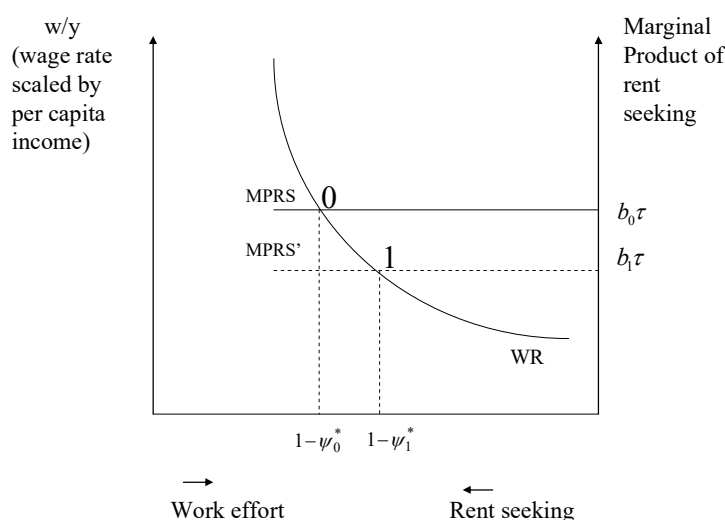
13.4.6 What happens if the effectiveness of rent seeking time declines?

In this model, any policy or institution affecting the effectiveness of a given intensity of rent seeking is captured by a change in the parameter b . For instance, suppose that the government is well succeeded in replacing fifty per cent of their corrupt bureaucrats by honest citizens. In this case, one expects the effectiveness of rent seeking to decline: because rent seekers have a lower chance of being matched to corrupt officials, the time necessary to find a corrupt official increases, on average (or the chance of being caught increases).

The implication of a fall in parameter b is analysed in Figure 13.4. The fall in the marginal product of rent seeking turns formal work more attractive in the margin, so more time will be dedicated to formal work. Then, because of diminishing returns, the equilibrium wage rate falls until it gets equal to the marginal product of rent seeking.

Note that, because individuals devote more time to production, per capita income will rise. In terms of Figure 13.4, the equilibrium moves from 0 to 1. In terms of Figure 3.6, the country switches to a higher parallel growth path, approaching the technological frontier.

Figure 13.4. The effect of a fall in the effectiveness of rent seeking



When the quality of legal institutions deteriorates, delivering a higher marginal product of rent seeking time, the proportion of time deviated away from production increases.

The model thus stresses the relationship between rent seeking and the quality of domestic institutions: the greater the ability of public servants to interpret creatively the law, to apply regulations selectively, to repudiate contracts, to confiscate property or to delay decisions, the more the private agents will be compelled to bribe the officials to guarantee some special treatment. All the rest constant, these countries are expected to observe larger deviation of talents away from production and to enjoy lower levels of per capita income in the steady state.

13.4.7 What happens if the tax rate increases?

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

In this model, we are assuming that the tax rate is chosen by a benevolent leader. To see the implications of setting different levels of taxation, let's see what happens when the leader increases the tax rate from τ_0 to τ_1 . This case is analysed in Figure 13.5.

In light of equation (13.17), there are two distinct effects:

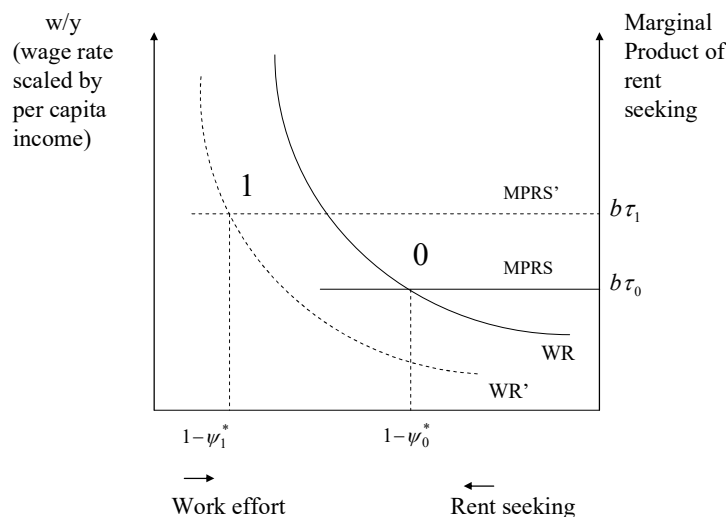
On one hand, the increase in the tax rate increases the marginal return of rent seeking (as given in 13.16). In Figure 13.5, this effect is represented by an upward shift in the marginal product of rent seeking. On the other hand, the tax increase causes a decline in the proportion of income that accrues to households in the form of (net) wages. In Figure 13.5 this is represented by a downward shift of the WR locus. Both effects induce individuals to move further towards informality. In Figure 13.5, the equilibrium moves from 0 to 1.

A different question is whether the increase in the tax rate is still desirable. In this model, taxes are necessary to finance the public input, which is essential to production. Hence, a benevolent planner would choose a positive tax rate for sure. The problem is that a higher tax rate will also induce a higher level of rent seeking. As usual, the optimal policy shall obey to a balance between the benefit of a higher public provision and the negative impact of taxation. In the simple model with public inputs (chapter 10), the only negative impact of taxation was a slower capital accumulation. With corruption, taxes also have the effect of inducing more rent seeking. Intuitively, if a higher tax rate implies a deviation of talents away from production, a benevolent planner will opt to tax less (and hence to provide less public inputs) than in the case without rent seeking. In order to deter corruption, the optimal policy is to set the tax rate and a level of public provision below the first best case (10.9)²⁷³.

Figure 13.5. The effect of a rise in the production tax

²⁷³ Park, H., Philippopoulos, A., Vassilatos, V., 2003. The optimal size of public sector under rent seeking competition from state coffers. Cesifo Working Paper N° 991, Ifo Institute for Economic Research, Munich.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>



When the tax rate increases, the net real wage in the formal sector decreases, and the marginal return of rent seeking increases, causing the equilibrium level of rent seeking to increase.

Box 13.4 Centralized versus decentralized corruption by William Easterly

“Under decentralized corruption there are many bribe takers, and their imposition of bribes is not coordinated among them. Under centralized corruption, a government leader organizes all corruption activity in the economy and determines the shares of each official in the ill-gotten proceeds.” (...). “More generally, a strong dictator will choose a level of corruption that does not harm growth too badly, because he knows his rake-off depends on the size of the economy. A weak state with decentralized corruption doesn’t have this incentive to preserve growth. Each individual bribe taker is too small to affect the overall size of the economy, so he feels little restraint on getting the most out of his victim”. (...) “This tale gives us insight why corruption was more damaging to growth in Zaire than in Indonesia. Zaire is a weak state with many independent official entrepreneurs. Indonesia under Suharto was a strong state that imposed bribes from the top down. Zaire had negative per capita

growth, while Indonesia had exceptional per capita growth (...).” [William Easterly (2001, p. 247-248)]²⁷⁴.

13.4.8 Other costs of corruption

The models outlined above illustrate the pervasive effect of corruption in deviating resources away from production. As any model, however, it cannot address all the negative implications of corruption. Other costs of corruption include:

- Uncertainty: with corruption, the enforcement of the law and regulations becomes more uncertain, turning investment decisions riskier.

- Social unfairness: bribery tends to act as a regressive tax, favouring elites against the general public.

- Distortions: to the extent that corruption fees are not uniformly paid across all agents in an economy, price incentives will be distorted. Corruption affects particularly activities with high fixed costs and long-term profitability, because once the fixed costs become sunk, investors are more vulnerable to additional extortion. It also tends to affect more innovative projects, as these are more likely to require licenses, specific regulation, and so on. Established firms, on the contrary, may use their lobbying efforts to block innovative entrants.

- Low investment in human capital: Since corruption comes along with more inequality and less public provision, poor people will have less opportunity to accumulate education and health. On the other hand, to the extent that corruption discourages investment in new technologies, expected wages for skilled workers will decline, decreasing the education effort and the skill level of the population. More talented workers will tend to emigrate.

²⁷⁴ Easterly, W., 2001. “The Elusive Quest for Growth: economist’s adventures and Misadventures in the Tropics, Massachusetts Institute of Technology.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

- Wrong policies: competition policies, environmental controls, consumer protection rules, and prudential regulations are in principle created to serve the public interest. Through bribes and corruption fees, public servants are induced to alter the way these rules are implemented or even designed in a society, to favour particular interests, at the cost of the general interest.

- Political instability: The use of public offices and public institutions for private advantage undermines the state legitimacy, leading to political instability and violence.

13.5 Coping with decentralized corruption

The problem of a benevolent planner dealing with a large administration is to design institutions that keep the incentives as right as possible. This will include a combination of rules and discretion, checks and balances and a little help from social norms.

13.5.1 Rules versus discretion

An obvious way of limiting corruption opportunities is to restrict the discretionary power of public officers. Examples in the real world include rules for hiring and promoting public personnel, rules for selecting private suppliers, rules to allocated subsidies, and so on. However, this avenue shall not be overlooked. The other side of the coin is that these constraints introduce rigidity: it may become too difficult to fire an incompetent civil servant, to reward a competent one, or to efficiently select a private supplier. Moreover, with excess rules, government officials will have limited capacity to react to new circumstances (i.e., those not accounted for in the rule). To deal with unexpected circumstances, some discretion is desirable.

A well-designed institutional framework involves therefore an appropriate balance between rules and discretionary power. The later must be complemented with other ingredients, such as checks and balances, incentive compatible contracts, and ethical principles. All these ingredients shall be part of a well-designed institutional package.

13.5.2 Sticks and carrots

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

The potential for corruption arises whenever discretionary power is delegated to a bureaucrat. Therefore, it is the interest of the benevolent planner to shape the incentives of the bureaucrat, rewarding good practices and punishing bad practices.

To illustrate the different components of an incentive structure, consider the case of a tax inspector whose job is to investigate whether a firm is liable for taxation²⁷⁵. In case the firm is liable for taxation, the firm shall pay a tax τ . However, the tax collector may report that the firm is not liable in exchange for a corruption fee, ϕ . The tax collector's incentive to misreport will depend on the size of the bribe and also on the penalties he will face in case he gets caught and dismissed, which will happen with probability p . Let g and f be the penalties applying to the firm and to the tax collector, respectively. The firm expected profit in case of misreporting is $\tau - pg - \phi$. A risk neutral firm will be willing to pay the bribe only if:

$$0 < \phi < \tau - pg. \quad (13.18)$$

The actual value of ϕ will be a matter of bargaining between the firm and the tax collector. The interesting question is whether wages in the public sector could be used to deter corruption. To analyse this, let w be the tax collector *wage premium*, defined as the difference between the tax collector wage and the wage he would receive in an alternative job in case of dismissal. The expected return of the tax collector in case of misreporting is $\phi + (1 - p)w - pf$, which he compares to the return of being honest, w . He will accept the bribe if:

$$\phi > p(w + f). \quad (13.19)$$

This last equation shows that the incentive to accept the bribe depends on the institutional setup defined by the following instruments: the wage premium (w), the quality of monitoring (p) and the legal punishments (f , g). The equation shows that wages paid to the tax inspector can be used to deter corruption: by increasing the cost of being dismissed,

²⁷⁵ Aidt, T., 2003. "Economic Analysis of Corruption: a survey". The Economic Journal 113, November, 632-653.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

higher wages turn the inspector less willing to accept bribes. The wage above which the tax inspector becomes honest is:

$$w \geq \frac{\phi}{p} - f \quad (13.20)$$

Together, (13.18) and (13.20) determine the minimum wage the tax inspector should receive, $\bar{w} = \tau/p - (g + f)$. This model points to complementarity between different instruments to deter corruption. The efficiency wage will be lower when the monitoring system is more effective (higher p) and the legal sanctions (f, g) are more severe. In the limit, with an effective monitoring system and sufficiently high penalties, the wage premium could be set equal to zero.

13.5.3 Social norms

A critical ingredient in any institutional setup is the quality of social norms. Social norms are human devised tools that promote the coordination of actions among individuals in a society, reducing the need for law enforcement. The underlying mechanism in the arrival of social norms is that of *strategic complementarities*: by establishing what is socially acceptable, social norms impose a psychological cost on those that take actions that deviate from the common behaviour. This leads individuals to choose actions that conform with what is acceptable to the society.

In the context of the bureaucrat model, it is easy to account for the role of social norms: wherever societies erect social norms against thievery and corruption, the punishment parameters f and g will include a social stigma attached to the event of being caught. On the other hand, in a clean society, the probability of a corrupt official being matched by honest citizens increases, so the probability p of being caught will increase too. Hence, he will feel more constrained to accept bribes. In general, wherever principles in ethics in government are developed, the wage premium necessary to turn public officers honest is lower.

Box 13.5 China and Russia

Dixit (2007, p. 5,6), illustrates the key role of incentives in public administration, taking as examples the cases of China and Russia²⁷⁶: “The research concerning property rights and corruption has yielded some useful conceptual distinctions and implications. The first is the distinction between *de jure* and *de facto* effectiveness of governance. The distinction is most vividly seen by contrasting China and Russia. China, at least until recently, had very little formal legal protection of property rights, especially those of foreign investors. However, in practice it has been able to deliver sufficient security to continue to attract large foreign investments. Russia has a much better legal framework on paper, but reality seems much worse. What explains the difference?” (...) “High officials in Deng Xiaoping’s government understood enough about economics to recognize that growth requires markets and markets require assured property rights. The Communist Party had retained its highly disciplined organization and so was able to prevent self-seeking behaviour by low-level officials.” The top level in Yeltsin’s Russia may have had the same understanding, but presumably lacked the disciplined organization”. (...) “The top level of government, even if itself well-intentioned, needs sufficiently drastic punishments at its disposal to keep the lower and middle-level agents in check” (...) “Stalin had, and used, punishments as drastic as one could imagine, and yet could not get his officials to perform efficiently. What went wrong? Gregory and Harrison (2005) argue that Stalin’s harsh incentives did not work well because his methods for detecting shirking were arbitrary, imprecise, and themselves open to corruption. People found that they ran almost the same risk of being denounced and punished when they worked hard as when they shirked or cheated. Therefore they did not have the incentive to work hard after all. An accurate detection procedure is important for the success of any incentive scheme, including an anti-corruption one”.

²⁷⁶ Dixit, A., 2007. “Governance, Institutions and Development.” P. R. Brahmananda Memorial Lecture, Bank of India, June.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

13.5.4 Optimal corruption

The discussion above suggests that it is *possible* to design an incentive compatible system that fully resolves the agency problems of delegating discretionary power to a potentially corrupt public officer. However, implementing such system in the real world is not that easy: First, public officials often lack the power to dismiss incompetent bureaucrats or to set performance bonuses. Second, setting wages high enough to guarantee the honesty of bureaucrats might become too expensive when monitoring practices are inefficient, requiring high distortionary taxes elsewhere. Third, setting bureaucrat's wages above the private sector wages also leads to a misallocation of talent, as individuals with entrepreneurial skills may feel attracted to become bureaucrats. Finally, increasing monitoring does not necessarily have linear impacts on corruption: anti-corruption agents may turn out themselves corrupt.

These difficulties rise the question as to whether designing a corruption-free bureaucracy would be *desirable*. As in any second-best reasoning, the optimal policy involves a balance between costs and benefits. In general, the fact that achieving a corruption-free bureaucracy would be too expensive implies that even a benevolent planner would optimally decide some corruption to persist. In other words, paying wages to bureaucrats below the critical level may deliver on balance a higher welfare than a system that turns all bureaucrats completely honest.

13.6 When institutions become dysfunctional

The discussion above has pointed to the critical role of institutions in shaping the incentives of public officers in line with the public interest. A problem arises, however, in that the ability of a society to design good institutions is significantly weakened when the level of corruption is already high. This fact brings a vicious cycle: wherever corruption is high because the quality of institutions is low, the quality of institutions tends to remain low because the level of corruption is high. This means that, depending on the starting point, an economy may find itself in a good equilibrium with little diversion and high probability of punishment, or in a bad equilibrium with low probability of punishment and high corruption. This idea was first proposed by the 17th century philosopher, Thomas Hobbes.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

13.6.1 *Why corruption brings more corruption?*

There are many reasons to believe that the social pressure against corruption declines with the level of corruption.

First, when corruption is generalized, matching rent seekers and corrupt bureaucrats is less than a problem. The searching costs for new corruption opportunities are low, increasing the relative reward of rent seeking time. Thus, individuals will have little incentive to go honest. Second, when the number of corrupt civil servants is very large, the probability of any one of them being caught and prosecuted is low: detection does not entail the same social stigma as in a corruption-free society (it is hard to punish someone severely when everyone else is doing the same), auditors may be themselves corrupt, and senior officials have the incentive to make sure that corrupt officials are not punished²⁷⁷. Third, when corruption is widespread, policies envisaging the strengthening of institutions are more likely to be blocked, because those that have the power to create rules may be more interested in making sure that corruption keeps paying off. Finally, corruption undermines people's trust in the political institutions, resulting in a weak civil society, and clearing the way for corruption to flourish.

In terms of the model above, these considerations suggest that some parameters in equations (13.18)-(13.20), rather than exogenous, should depend on the level of corruption. In particular, the parameters measuring the likelihood of detection (p) and the legal punishments (f , g), instead of constant, could be modelled as decreasing functions of the economy-wide incidence of corruption. Thus, just like we concluded that corruption tends to flourish in countries where institutions are poorly designed, checks on discretionary power are missing and punishment is weak, now the reverse causality is also true: countries where

²⁷⁷ Mauro (2004): "Consider, for example, the case of a civil servant in an administration where everybody including his superiors, is very corrupt. It would be difficult for this civil servant to decline offers of bribes in exchange for favours, because his superiors may expect a portion of the bribe for themselves. By contrast, in bureaucracies that are generally honest, a real threat of punishment deters individual civil servants from behaving dishonestly".

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

corruption is generalized tend to have poorly designed institutions, lower checks on discretionary power and lower punishment, making corruption more attractive. Hence, the vicious cycle.

13.6.2 Multiple equilibria

We now adapt the model outlined in Section 13.4, to account for the impact of corruption on the quality of institutions. Remember that the quality of institutions in that model is captured by the parameter measuring the effectiveness of rent seeking time, b . By now, this parameter was assumed exogenous. In the following, let's assume instead that this parameter is an increasing function of ψ : that is, as the society devotes more time to rent seeking, the quality of institutions decreases and hence the higher will be the effectiveness of the rent seeking effort.

This case is depicted in Figure 13.7, where the line describing the marginal product of rent seeking is downward sloping, instead as horizontal. In that case, the model may have two equilibria, one with low corruption (L) and other with high corruption (H): If the economy starts out in L, the corruption level is low, and the probability of corrupt officials being caught is very high. Because the marginal product of rent seeking is low, people dedicate a small fraction of their time to corruption. In this case, the economy will evolve along virtuous cycle with high per capita income. If the economy starts out in H, the probability of someone being charged for corruption is low. This, in turn, induces people to engage in rent-seeking, self-validating the high corruption level and the equilibrium path with low per capita income.

Because there is multiple equilibrium, *history* plays a role where an economy will lie: the economy will be poor and corrupt if it starts out poor and corrupt; the economy will be rich and clean if it starts out rich and clean. Depending on the initial conditions, two-otherwise identical economies may end up with different levels of per capita income.

13.6.3 Escaping the trap

To see how difficult is to escape the bad equilibrium H, suppose that the fraction of time devoted to rent-seeking decreased slightly to a point between H and L. In that case, the marginal product of rent seeking would be temporarily higher than real wages, and the incentives will exist for corruption to increase. Hence, the economy would return to point H.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Because the equilibrium H is stable and is dominated by another possible outcome (equilibrium L), it is a *poverty trap*.

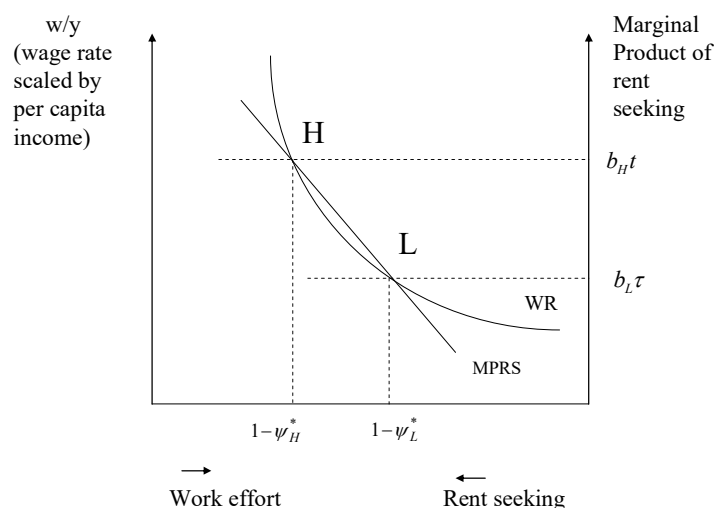
The model provides an explanation of why it is very difficult to eradicate endemic corruption once installed. In light of this model, modest anti-corruption initiatives (like using the police, the courts or even creating ethic offices) are more likely to move the economy somewhere between H and L, without alleviating corruption in a sustained way. For an economy to escape the trap, a huge anti-corruption effort will be needed. The self-sustained aspect of institutions helps understand why many developing countries have failed to improve their living standards despite considerable external support.

The model also implies that an historical accident (like a war, a social unrest, a terms of trade deterioration, a natural catastrophe) may cause the economy to move from the good equilibrium to the bad equilibrium and remain there. Once institutions, informal rules and norms of behaviour have already been adapted to a corrupted system, the high corruption equilibrium will be self-sustained. The economy will be *locked in* the bad equilibrium and will not move back to the original equilibrium, even though the original shock was dissipated. Real life examples of such a path could be the collapse of communism in Russia and, more generally, cases of political instability in Africa²⁷⁸.

Figure 13.7. Multiple equilibrium with rent seeking

²⁷⁸ Murphy, K., Shleifer, A, Vishny, R., 1993. Why is rent seeking so costly to growth? American Economic Review Papers and Proceedings 83 n° 2, May.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>



When the quality of institutions is influenced by the level of rent seeking, there is possibility of multiple equilibrium: an economy that starts out with a high level of corruption will tend to remain corrupt, and an economy that starts out with low corruption will tend to remain with low corruption.

13.7 Discussion

Along this book, we have stressed the key role of human devised government policies for economic development. To a large extent, the analysis has been optimistic regarding the growth opportunities of laggard nations: after all, if achieving economic development was only a matter of adjusting the policy mix, then this should be within reach for any poor country in the world.

This chapter traces a less optimistic view of economic development. The key issue is that policies are embedded in institutions and institutions are not always shaped so as to promote the social welfare. In societies where property rights are not well enforced and where controls over the political elites are lacking, there will be deviation of resources away from public infrastructure, misallocation of talent and bad economic policies.

The discussion in this chapter emphasized the critical role of institutions in shaping the incentives of policymakers, bureaucrats and economic agents in general. A well-balanced institutional design has to include the separation of political powers, transparency in decision-making, a strong civil society, checks and balances, rewards and punishment, and a strike between rules and discretionary power. A well-balanced institutional design shall also take into account that achieving zero corruption would be too expensive. Sound social norms play a critical role in the incentive structure, enhancing the effectiveness of formal institutions and reducing the costs of law enforcement.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

A critical problem is that institutions are not easy to change. Institutions emerge as a result of complex games involving the different groups in a society. When corruption becomes so generalized that institutions, informal rules and norms of behaviour become adapted to it, the economy will find itself *locked in* low-level poverty trap. Poor countries will hardly have the financial means and the political strength to escape this trap, so corruption becomes endemic.

13.8 Key ideas of Chapter 13

- Central planners and civil servants are not necessarily benevolent. Whenever the institutional environment allows for, they may deviate from the social interest to pursue own objectives, at the cost of the quality of the policy environment.
- When corruption affects the high level of the administration only, it is the interest of the corrupt leader to look out at aggregate efficiency and to organize the extraction activity, in order to maximize theft opportunities.
- The theft opportunities of a non-benevolent leader depend on the constraints imposed by the political and legal institutions that he cannot change.
- When corruption affects all levels of the administration, each corrupt official will try to maximize his corruption fees without taking into account the impact on the economy as a whole. This gives rise to a coordination failure that resembles the “tragedy of the commons”. The deviation of resources to corruption comes at a cost of low production in the formal good sector. As long as there are diminishing returns, the higher the corruption activity in the economy, the higher will be the wage rate in formal production, creating the incentives for some fraction of the labour force to remain productive. This prevents the economy from being “exterminated”.
- When corruption is decentralized, a benevolent leader concerned with the social welfare will provide less of the public good than in the case of centralized corruption. The reason is that the required increase in taxation would induce an increase in corruption activity. This case provides another example of second-best decision-making.
- In order to control corruption, benevolent leaders have to strike an ideal balance between the rigidity cost of rules and the opportunities raised with discretion. In the optimal institutional setup, discretionary decisions should be complemented with efficient wages, checks and balances in order to keep the incentives right. The optimal corruption level is not however zero, as that would be socially too costly to achieve.
- Social norms and civil society play a critical role in deterring corruption. However, societies do not naturally gravitate towards good institutions: whenever corruption becomes generalized, strategic complementarities create the incentives for each one individual in the society to remain corrupt. In this case, a society may find itself locked in a low level poverty trap, with high corruption and dysfunctional institutions that nobody has interest to change.

13.9 Problems and Exercises

Key concepts

- *Corruption. Rent seeking. The “Grease in the wheels” argument. Centralized vs. decentralized corruption. Incentive compatible contracts. No-natural gravitation. Strategic complementarities in corruption. Development accounting.*

Essay questions

- Comment: “Corruption is an efficient mechanism to allocate the scarce time of bureaucrats to the most productive uses”
- Comment: “The theft opportunities of a non-benevolent leader depend on the incentives embodied in political, administrative and legal institutions he cannot change”.
- Explain why the provision of key public inputs is lower under decentralized corruption than in the kleptocrat case.
- Comment: “Corruption is bad for growth”.
- Comment: “A well-designed institutional framework involves an appropriate balance between rules, discretionary power, incentive compatible contracts, and well founded ethical principles”.
- Why it pays more to be corrupt when others are corrupt too?
- Explain why sometimes countries hit by an historical accident get locked in high corruption poverty traps.

Exercises

- 13.1** Consider an economy with a large number of equal firms. Each firm i at time t producing a homogeneous consumption good according to: $Y = A_t K_t^{0.5} N_t^{0.5}$, where $A_t = A e^{0.015t}$ and $L_t = e^{\gamma t} N_t$. In this economy the saving rate is 20%, the population is constant and the capital depreciation rate is 3%. a) (**Solow**) Find out the steady state level of per capita income in this economy. Consider $A=0.3$. b) (**Market failure**) Assume now that $A = (G/Y)^{1/3}$, where G are public inputs. (b1) Explain this specification. To which type of public inputs it applies? (b2) Compute the aggregate production function of this economy. (b3) Explain why there is a market failure in this model. (c) (**Waste**) Assume that the provision of public inputs was financed with a production tax τ , but that a fraction ϕ of the tax revenues is wasted in unproductive uses. Write down the government budget constraint. Which factors are captured by parameter ϕ ? Elaborate. (d) (**Steady state**) Find out the expressions for the steady state levels of per capita income and of per capita consumption, in terms of τ and ϕ . (e) (**Benevolent planner**) Assuming that the objective of the planner was to maximize the

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

steady state level of consumption: (e1) find out the optimal values of τ and ϕ . Is this solution intuitive? (e2) Explain, with the help of a graph, the dual effect of the tax rate on the steady state level of per capita consumption. (f) (**Kleptocrat**) Now suppose that you were a kleptocrat, which aim was to maximize your theft on government expenditures. (f1) Which solution would you choose? (f2) would per capita GDP be higher or lower than in (e)? (f3) Is this equilibrium in light with the Kaldor facts? (f4) Compare the benevolent planner and kleptocrat solution in the Solow model graph (g) (**Bureaucracy**) If the kleptocrat could not control its bureaucracy, citizens would be better off or worse off? Why? Which mechanisms can societies implement to limit the corruption opportunities?

- 13.2** In Arcadia the production function faced by individual firms is given by $Y = A_t K_t^{0.5} N_t^{0.5}$. In this economy the depreciation rate is 5%, population and technology are constant, and $s(1-\tau) = 0.2(1-\tau)$, where $\tau = 0.75$ is tax on production. (a) (**Solow**) Consider for a moment that $A=0.5$. Find out the steady state levels of (a1) per capita income (a2) capital per worker; (a3) K/Y . (d4) Represent the equilibrium in a graph and discuss its stability. (a4) To what extent this economy complies with the Kaldor stylized facts? (b) (**Lucas Paradox**) (b1) From the firm's maximization problem, find out the general expression of the interest rate as a function of the tax rate, τ , and of the capital output ratio. (b2) Find out the interest rate in Arcadia. (b3) Suppose that, in another economy, Begonia, the production function was the same, but per capita output was twice that of Arcadia. Could you state that capital should flow from Begonia to Arcadia? In which circumstances? (c) (**Public input**) Assume now that $A = (G/Y)$ where $G = (1-\phi)\tau Y$. (c1) Which factors are captured by parameters ϕ and G ? (c2) Find out the expression of the "true" aggregate production function of this economy. (c3) Explain why there is a market failure in this model. (d) (**Kleptocrat**) Now suppose that you were a kleptocrat, which aim was to maximize your theft on government expenditures: (d1) Which solution would you choose for τ and ϕ ? (d2) What would be your provision of G/Y ? Explain (d3) What would be the level of per capita income? (e) (**Benevolent planner**) Suppose that in Begonia the externality was the same, $A = (G/Y)$, but that country was run by a benevolent planner. (e1) What would be the solutions for τ , ϕ , and G/Y ? (e2) What would be the corresponding level of per capita income? (e3) Would capital flow from Begonia to Arcadia?

- 13.3 (Rent seeking)** Consider an economy with a large number of equal firms. Each firm i at time t producing a homogeneous consumption good according to $Y = AK^{0.5}N_Y^{0.5}$, where $N_Y = (1-\psi)N$ is the total work-time devoted to production and ψ is the fraction of time devoted to rent-seeking. Assume also that the proportion of government resources extracted by rent-seekers (ϕ) is a linear function of the seeking effort (ψ), that is, $\phi = \psi b$, where b is an exogenous parameter that captures the marginal product of rent-seeking time. (a) Considering that the government imposes a tax on income, derive the expression for the wage rate in this economy. (b) Write down the expression for the total amount of resources deviated from the government by rent-seekers. Explain it. (c) Find out the equilibrium level of ψ . With the help of a graph, analyse the implications

of: (c1) A fall in the tax rate. (c2) A decline in productivity of the rent seeking. Which measures could be implemented to achieve this?

- 13.4 (Rent seeking)** Consider an economy where workers are free to decide the time they allocate to formal work and to rent seeking. In this economy, the production function is given by $Y = N_Y^{0.16}$ where $N_Y = (1-\psi)N$ is the time allocated to formal work. The tax on production is equal to $\tau = 1/2$. (a) From profit maximization, find out: (a1) the demand for labour; (a2) the expression relating the wage rate to: per capita income and the fraction of time devoted to rent seeking. (b) Show that individual choices lead to the following condition: $1-\psi = 0.16/b$, where b measures the effectiveness of rent seeking effort. (clue: find out the demand for labour and use the arbitrage condition $w = b\tau y$). (c) Explain this relationship above, illustrating with the following values for b : $b=0,2$ and $b=0.8$. (d) Now assume that the effectiveness of rent seeking is itself a positive function of the level of rent seeking: $b=\psi$. Why should that be? Describe the equilibria of the model, discuss their stability. Explain what this model intends to capture.
- 13.5** Consider an economy where the production function of the individual firm is $Y = AK^{0.5}N_Y^{0.5}$, with $N_Y = (1-\psi)N$ being the amount of labour devoted to formal work and $A = (G/Y)$. In this economy, population is constant ($n=0$), there is no technological progress ($\gamma=0$), and the saving rate is $s=24\%$. Capital depreciates at $\delta=4\%$. (a) Find out the aggregate production function and explain why there is a market failure. (b) Assume that G was financed with a tax on production τ , but a proportion ϕ of the fiscal proceeds was wasted. Which types of failures are captured by this parameter? (c) Find out the benevolent planner optimum assuming that it can decide ϕ and ψ . In particular, quantify: (c1) the optimal tax rate; (c2) G/Y . (c3) private consumption per capita; (c4) Describe the equilibrium in the Solow - (k, y) - diagram. (d) Assume now that the leader could not control its bureaucracy. In particular, assume that the fraction ϕ of the budget was deviated, according to $\phi = b\psi$, with $b=1.5$. (d1) Assuming that the tax rate was $\tau = 0.5$, find out the equilibrium level of rent seeking, as well as the resulting G/Y . (e) With the help of a graph, show how the government could induce zero rent seeking using: (e1) The tax rate; (e2) policy changes impacting on b . Quantify and compare the quality of these two possible equilibria.
- 13.6** Consider an economy with a large number of equal firms. Each firm produces a homogeneous consumption good according to $Y = AK^{0.5}N^{0.5}$. Although each firm considers A as an exogenous parameter, in the aggregate the following condition holds: $A = (G/Y)^{3/2}$. In this economy the population is constant ($n=0$), there is no technological progress ($\gamma=0$), the saving rate is 20% and the capital depreciation rate (δ) is equal to 4%. In this economy, the provision of public inputs is financed with a production tax τ , but a fraction ϕ of the tax revenues is wasted in unproductive government consumption. (a) Describe the profit maximization problem of individual firms and find out the demand for capital and labour, as well as the households' disposable income. (b) **Market failure**: Compute the aggregate production function in this economy and explain why there is a market failure. What would happen if there

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

was no government? (c) Write down the government budget constraint. (d) Which factors are captured by parameter ϕ ? Elaborate. (e) **Steady state**: Find out the expression for the steady state levels of per capita income and per capita consumption, in terms of ϕ and τ . (f) **Benevolent planner**: consider the problem of a benevolent planner, which aim is to maximize the steady state level of private consumption: (f1) compute the optimal values of ϕ and τ . Is this solution intuitive? (f2) Find out the steady state levels of per capita income and of consumption in that case; (f3) Explain, with the help of a graph, the dual effect of the tax rate on the steady state level of per capita consumption. (g) **Kleptocrat**: Now suppose that you were a kleptocrat, which aim was to maximize your theft on government expenditures: (g1) Which values would you choose for τ , ϕ , and G/Y ? (g2) Find out the corresponding steady state levels of per capita income and of consumption. (g3) Explain how this solution compares to that of the benevolent planner. (g4) find out the interest rate in this case. (h) **Decentralized corruption**: Finally, consider the case where the benevolent planner sets the tax rate at 60%, but it cannot control its bureaucracy, so a proportion ϕ of government revenues is appropriated by rent seekers. Further assume that citizens in this economy devote a fraction of their time ψ to rent seeking, and the remaining time to formal work, that is $Y = AK^{0.5}N_Y^{0.5}$. h1) Describe the income flow chart of this economy; h2) Find out the demand for formal labour in this economy, and then re-write it as a function of per capita income; h3) Assume that the private appropriation of government resources is a direct proportion of the time devoted to rent seeking, that is: $\phi = b\psi$, with $b=1$. Find out the equilibrium level of rent seeking in this economy, and explain it with the help of a graph; h4) Find out the level of government provision G/Y in this case, and compare to the previous models; h5) How much will be per capita income and per capita consumption? h6) Which policies could the benevolent planner use to improve the quality of his bureaucracy? Elaborate. (i) **Endemic corruption**: Returning to (h), assume instead that $b = 0.75\psi$. i1) Explain the rationale for such specification. i2) Describe the equilibria and the stability of the model in this case. i3) Interpret.

- 13.7 (Incentives)** Consider a tax collector which job is to investigate whether a firm is liable for taxation or not. If so, the firm has to pay a tax τ . However, the tax collector may report that the firm is not liable for taxation in exchange for a bribe b . Assume also that the government discovers corrupt acts (with probability 0.5) then the firm pays a penalty $g = 1$ and the tax collector pays a penalty $f = 0.5$. (a) If $\tau = 2$, what would be the maximum bribe amount the firm would be willing to offer? (b) Write down the expression for the expected return of the tax collector in case of misreporting. (c) What would be the minimum wage rate that would keep the tax collector honest?

14 Institutions and Geography

“Variations in the economic performance across countries and regions (e.g, earlier industrialization in England than in China) reflect initial differences in geographical factors and historical accidents and their manifestation in variations in institutional, demographic, and cultural factors, trade patterns, colonial status, and public policy”. [Oded Galor]

Learning Goals:

- The relationship between institutions and economic performance
- The various links between geographical variables and economic performance.
- Understand the key role of institutions in explaining episodes of reversal of fortune.

14.1 Introduction

Along the book, we have examined different determinants of economic development. We departed from models describing the mechanics of economic growth, focusing on proximate causes, such as propensities to invest in machinery and in human capital, and the key role of total factor productivity. We then argued that technology can either expand through own R&D efforts, or taking opportunity of learning by doing and knowledge spill overs. We reviewed the idea that, or a laggard country, catching up opportunities arise mostly from the opportunity to adopt and adapt ideas developed in richer countries. We also stressed the pervasive effects of market failures, complementarities, and increasing returns, and the role of government in addressing these failures. At a deeper level, however, the question is why poor countries don't invest more in machinery and in human capital, why they fail to adopt more successful economic policies. Along these questions, we move from the proximate causes to the fundamental causes of economic growth.

In the literature, there are two main views regarding what the fundamental determinants of economic growth are: the view assigning a prominent role for institutions that support property rights and the enforcement of contracts, that are essential for entrepreneurship and innovation; and the view that geography is the key determinant of economic prosperity, through the influence that climate, quality of land, and transport costs

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

exert on individual incentives to produce and invest. Authors supporting the “institutions hypothesis” include Douglas North, William Easterly, Dani Rodrik, and Daron Acemoglu. The argument that geographical factors plays a key role in economic development can be found in the works of Adam Smith, Jeffrey Sachs and Jared Diamond.

At this stage, the reader must be aware that the phenomenon of economic development is too complex to be accounted for any single factor. So, the question shall not be whether the ultimate determinant of economic growth is institutions or geography instead. Still, understanding the relative weight of these two fundamentals has important policy implications: while Geography refers to “forces of nature” that societies cannot change, institutions can be changed by collective actions. From the policymaking point of view, this makes a huge difference.

This chapter reviews the debate regarding the influence of institutions and of geography in a country’ long run economic performance. Section 14.2 describes the main arguments underlying the Institutions Hypothesis. Section 14.3 describes the “geography hypothesis”. Section 14.4 reviews some of the debate regarding the relative importance of Geography versus Institutions. Section 14.5, we summarise de main ideas.

14.2 The Institutions hypothesis

14.2.1 What are institutions?

Human beings compete for scarce resources. For that competition to result in mutual gains, societies need to set up and enforce some supportive framework that coordinates individual actions and expectations. This framework is provided by laws, social norms and conventions, that we generically label as “institutions”. Broadly speaking, institutions are human devised constraints that govern human interactions²⁷⁹. Institutions emerge so that

²⁷⁹ North, D., 1993. The new institutional economics and development. WUSTL Economic Working Paper Archive 9.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

individual actions become more predictable to others, reducing the costs of transacting. Institutions are therefore a natural response of human societies to the need of trade in a world with uncertainty and information asymmetries.

Some institutions are provided by the state in the form of law, regulations, and governmental agencies, such as the judiciary, the competition authority, international agreements, and central banking. Others emerge spontaneously from the civil society or within groups in the society. These include, conventions, norms, and self-imposed codes of conduct. Box 14.1 provides an example of how the insufficient provision of contract enforcement by a government may lead the emergence of parallel institutions performing that function.

Box 14.1 “Don Peppe”

Avinash Dixit, from Princeton University, offers a nice illustration of how institutions that create order to support economic activity and protect property rights may emerge spontaneously in the society, if the legal system administered by the state is inefficient or unable to provide them: “In most economic transactions that can create economic gains for all parties, some or all of them can gain an extra private benefit while hurting the others, by violating the terms of their explicit or implicit agreement. The fear of such exploitation by the other party may deter each from entering into the agreement in the first place. This was brilliantly illustrated by Diego Gambetta in his ethnographic sociological study of the Sicilian Mafia (1993, p. 15). In the course of his interviews, a cattle rancher told him: “When the butcher comes to buy an animal, he knows that I want to cheat him [by supplying a low-quality animal]. But I know that he wants to cheat me [by reneging on payment]. Thus we need Peppe [the third-party mafioso] to make us agree. And we both pay Peppe a commission.” By providing a mechanism of contract enforcement, Peppe makes it possible for the two to enter into a mutually beneficial transaction. And he does this with a profit

motive, exactly as would any businessperson providing any service for which others are willing to pay” (Dixit, 2007, p. 3)²⁸⁰.

14.2.2 How have institutions evolved over time?

Institutions have become increasingly complex along human history. In primitive hunter-gatherer societies the “rules of the game” were mostly encoded in simple traditions and enforced by leaders with extensive discretionary power, often under the supervision of tribe councils. In those times, simple societal structures were enough to resolve most conflicts between community members, though they could not in general support contractual arrangements.

When, around 10.000 years ago, humans moved to agriculture, the emergence of larger and more structured societies turned necessary the creation of new supporting institutions²⁸¹. In these societies, some individuals specialized in agriculture, some in artefact production, some in trade and some other in defending the territory. In order to support investment and to organize exchange and the division of labour in these more complex communities, it became necessary to secure property rights and enforce contracts: if private property was not protected and contracts freely entered into could not be enforced, the incentives to conduct many types of beneficial trade would not exist.

Initially, these services were delivered by coalitions of privileged elites, who had a stake in providing a stable framework within the community and who, at the same time, had the power to assign and enforce their own privileges and property rights. As societies progressed, more open competition and the rise of new elites seeking for the control of the state turned necessary an increasingly complex network of social arrangements. In the

²⁸⁰ Dixit, A., 2007. “Governance, Institutions and Development.” P. R. Brahmananda Memorial Lecture, Bank of India, June.

²⁸¹ North, D., Wallis, J., Weingast, B., 2006. A conceptual framework for interpreting recorded human history. NBER working paper series 12795, December.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

political front, it became necessary to regulate the election of leaders and control their power. On the policy front, it became necessary to regulate markets, to build resilience to economic shocks and to provide socially acceptable income distributions. In the civil society, complementary institutions have emerged, in the form of social networks providing group enforcement, punishment, relevant information and even protection of property rights, wherever government policies revealed less effective or absent.

14.2.3 Economic institutions

Among the broad class of institutions, those that shape the economic incentives, such as the decisions to save and invest, to become educated, or to adopt new technologies can be categorized as “economic institutions”. This category includes both formal law, regulations, and governmental agencies, such as licencing procedures, consumer protection laws, or property registration, or informal (emerged from private enterprise or civil society), such as codes of conduct and traders’ networks, and the enforcement properties of these.

Among the general category of economic institutions, those that protect property rights and ensure the enforcement of contracts are up most important. Without these, there could be no beneficial trade among parties, investment and entrepreneurship. Property rights and contract enforcement are the foundations in which the market economy rests, but are not sufficient to sustain growth. As pointed out by Rodrik et al (2004), economies must build resilience to shocks and achieve socially acceptable mechanisms of risk sharing²⁸². The authors distinguished four types of economic institutions: the institutions that protect property rights and enforce contracts are labelled as “market-clearing institutions”, because without them, markets either do not exist or perform very poorly. The other three types are: “Market-regulating”: those that deal with externalities, economies of scale and imperfect information (consumer protection, competition authorities, regulatory agencies in energy); “Market

²⁸² Rodrik, D., Arvind Subramanian, Francesco Trebbi, 2004. "Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development," *Journal of Economic Growth* 9(2), 131-165.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

stabilizing”: those that reduce macroeconomic volatility and prevent financial crises (central banks, exchange rate regimes, budgetary and fiscal rules); “Market legitimizing”: those that provide social protection and redistribution t (pension systems, unemployment insurance schemes and other social funds). These categories together include many dimensions of government intervention.

By rewarding certain types of behaviour and punishing others, economic institutions shape the incentives to produce and invest in new technologies, thereby influencing decisively the outcome of the competition game over limited resources. Many authors believe that the Rise of the West is much attributable to the development of institutions that allowed economic agents to make their best choices given their talents and skills, exploiting more fully the potential gains from exchange. Of course, institutions may also retard growth. Regulatory agencies may be used to prevent business entry; courts may be commanded to resolve disputes dishonestly instead of ensuring level-playing field; the police may be used to extract bribes from honest citizens; governments may expropriate property from private agents. Not surprisingly, many empirical studies have found a positive relationship between economic growth and measures of institutional quality.

Acemoglu and Johnson (2005) investigated the relative importance of key economic institutions for economic development²⁸³. The authors distinguish institutions that facilitate exchange between private parties, such as laws, courts and regulations (“contracting institutions”) from those that protect citizens from the power of government and elite (“property rights institutions”). Using historical data, the authors found that “property rights institutions” regulating the vertical relations between citizens and the state are more important to explain economic development than “contracting institutions”, that regulate horizontal relations between private bodies. The authors interpreted this, arguing that private agents can reduce the adverse effects of weak contracting institutions, by adjusting the terms in the contracts, by developing alternative mechanisms to enforce their contracts or to

²⁸³ Acemoglu, Daron and Simon Johnson (2005) “Unbundling Institutions.” *Journal of Political Economy*, 113, pp. 949-995.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

cover/reward their risk. When, in alternative, property rights institutions fail to control the state, it is not possible to prevent expropriation because the state is the ultimate arbiter of contracts.

14.2.4 Political institutions

Economic institutions shape the incentives of private agents to produce and invest. Political institutions determine the power of the state to regulate and govern society. They establish how the government is chosen, which part of the government has the power to do what, which mechanisms are available for the civil society to control politicians.

While economic institutions are critical for determining whether a country is poor or prosperous, it is political institutions that determine economic institutions. Economic institutions are determined in the political process, which in turn is governed by political institutions. Political institutions can be pluralist, distribute the powers widely, and support growth-enhancing economic institutions, or instead concentrate the political power and place little constraints on the exercise of this power, allowing for corruption and rent seeking.

Acemoglu et al (2012) used the term “extractive” to characterize political institutions that endow governments with extensive discretionary power with little checks, paving the way for “extractive economic institutions”, that fail to align private incentives with the public and favour the concentration of resources in the hands of privileged elites. By contrast, “inclusive” political institutions will be those characterized by a broad distribution of political powers and high participation of the civil society, favouring the emergence of “inclusive” economic institutions, that allow free entry and ensure level playing field, enabling people to choose the best use of their talents.

Box 14.2 Measuring institutional development

The understanding that institutions are a crucial ingredient for economic performance lead different authors to search for alternative measures of institutional development. A major contribution is Kaufman et al (1999)²⁸⁴. The authors use a broad concept of *governance*, which they define as the “traditions and institutions that determine how authority is exercised in a particular country”.

The authors identified several hundred of cross-country indicators on various dimensions of governance from many different sources and used them compute six compounded indicators of governance, measuring, respectively: “(i) Government Effectiveness: the quality of public service delivery and competence and political independence of the civil service; (ii) Regulatory Quality: the relative absence of government controls on good markets, banking systems and international trade; (iii) Rule of Law: the protection of persons and property against violence and theft, independent and effective judges and contract enforcement; (iv) Control of Corruption: public power is not abused for private gain or corruption; (v) Voice and Accountability: the extent to which citizens can choose their governments and have political rights, civil liberties and independent press; (vi) Political Stability and absence of Violence/Terrorism: the likelihood that the government will not be overthrown by unconstitutional or violent means”.

Table 13.1 shows how different countries have ranked in terms of these six indexes as of 2007. Not surprisingly, countries like Finland, Switzerland and Australia rank at the top, while countries like Chad, Sudan, Somalia and Democratic Republic of Congo and Zimbabwe rank in the bottom, for all dimensions. Kaufman et al (1999b) found a strong causal effect running from improved governance to development indicators.

Table 14.1 – Indicators of Governance

²⁸⁴ Kaufmann, D., Kraay, A., Zoido-Lobaton, P. 1999 "Aggregating Governance Indicators," World Bank Policy Research Working Paper No. 2195 (Washington

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

	(i)	(ii)	(iii)	(iv)	(v)	(vi)
	Government Effectiveness	Regulatory Quality	Rule of Law	Control of Corruption	Voice and Accountability	Political Stability & Absence of Violence/Terrorism
Country	%Rank	%Rank	%Rank	%Rank	%Rank	%Rank
ARGENTINA	52	22	39	43	57	50
AUSTRALIA	97	96	95	95	93	79
BOTSWANA	73	65	70	80	62	78
BRAZIL	53	53	43	52	59	37
CAPE VERDE	66	48	67	74	75	83
CHAD	4	12	6	5	9	6
CHILE	86	91	88	90	77	66
CHINA	61	46	42	31	6	32
DEM. REP. OF CONGO	1	8	1	4	9	2
EGYPT	39	43	52	36	12	22
FINLAND	97	96	97	100	98	99
GERMANY	92	93	94	93	95	81
HAITI	8	20	5	3	26	11
INDIA	57	46	56	47	59	18
INDONESIA	42	44	27	27	43	15
ISRAEL	84	83	73	75	70	13
ITALY	65	74	61	71	87	62
SOUTH KOREA	86	79	75	68	67	62
MAURITIUS	72	68	74	70	73	72
MEXICO	60	64	34	49	49	25
MOROCCO	55	51	51	53	29	27
NEPAL	22	27	31	30	23	3
PAKISTAN	28	29	20	21	19	1
PERU	38	58	27	48	49	20
POLAND	67	72	59	61	72	67
PORTUGAL	80	83	82	84	90	73
RUSSIA	42	35	17	16	20	23
SAUDI ARABIA	51	52	59	58	7	25
SINGAPORE	100	99	95	96	35	90
SOUTH AFRICA	75	66	57	67	69	51
SOMALIA	0	0	0	0	3	0
SUDAN	11	9	4	5	5	2
SWITZERLAND	100	93	100	98	100	99
THAILAND	62	56	53	44	30	17
TURKEY	64	60	53	59	42	21
UNITED KINGDOM	94	98	93	94	94	66
UNITED STATES	91	91	92	91	85	56
ZIMBABWE	3	1	2	4	8	12

Source: World Bank, Governance Matters 2008.

Note: The figures refer to the percentage of countries in the sample with lower performance than the country at hand in each specific dimension.

14.2.5 No natural gravitation

If the quality of institutions is so critical, why don't poor countries just change their institutions so as to achieve better economic performance? A problem is that institutions are not simple strike of the pen choices. Institutions arise in result of *social choices*. These choices influence not only the *level* of a country income but also the *distribution* of income among the members in a society.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Institutions emerge and evolve over time as outcomes of a permanent conflict for income and power among groups in a society. Institutions set the rules of the game and evolve over time reflecting the relative power of the players in that game. Hence, in order to understand how a particular institutional arrangement arises, one shall take into account the motivations and the bargaining powers of those individuals and groups that participate in the game. In many circumstances, those who hold the power to change the rules have more to gain in preserving a status quo that protects their own interests than to change the rules to the good of the population in general: societies do not naturally gravitate towards good institutions.

Azariadis and Stachurski (2005) offer a real-world example:

“Consider, for example, the current situation of Burundi, which has been mired in civil war since its first democratically elected president was assassinated in 1993. The economic consequences have not been efficient. Market-based economic activity has collapsed along with income. Life expectancy has fallen from 54 years in 1992 to 41 in 2000. Household final consumption is down 35% from 1980. Nevertheless, the military elite has much to gain from continuation of the war. The law of the gun benefits those who have guns. Curfew and identity checks provide opportunity for extortion. Military leaders continue to subvert a piece process that would lead to reform of the army.”²⁸⁵

In general, when political institutions do not restrict enough the power of policymakers, policymakers are more likely to pursue their own agendas, shaping economic institutions to create rents and enrich the privileged elites. The existence of economic rents will then help sustain the political power, in a vicious cycle. Historical path dependence is an important element in the evolution of institutions.

Institutions tend to evolve slowly over time, in interaction with the permanent conflict for income and power among the groups in a society. Critical junctures altering the balance of

²⁸⁵ Azariadis, C., Stachurski, J., 2005. Poverty traps. in: Philippe Aghion & Steven Durlauf (ed.), *Handbook of Economic Growth*, edition 1, volume 1, chapter 5.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

powers among groups in a society lead to institutional drift. Institutional changes are easier to accomplish when it becomes the interest of the political elites to do so. This, in turn, may reflect the interest of existing political supporters or the emergence of new groups in the society with power enough to impose the change. When the leaders fail to perceive these movements in civil society, a disruption may occur. As pointed out by Douglas North, revolutions occur when the fundamental conflict between organizations over institutional change cannot be mediated within the existing institutional framework”.

14.3 The Geography hypothesis

14.3.1 Cumulative causation

Along the previous chapters, we reviewed theories pointing towards economic convergence, and theories pointing to divergence. On one optimistic side, capital mobility and technological diffusion create the potential for poorer economies to catch up, importing technologies from rich countries and implementing best practices. However, in the real life, we do not observe capital flowing in a systematic way from rich countries to poor countries. Cross country differences in per capita output are large and persistent (see Box 14.5). This evidence is consistent with the view that the poorest countries in the world are trapped by low productivity levels that cannot be mitigated by simple technological diffusion. Increasing returns, learning by doing, strategic complementarities and coordination failure are among the various barriers that cause economies to remain trapped in low-income equilibrium levels, without any market force pushing towards convergence. On the contrary, bad equilibria tend to be self-reinforced, through adverse population dynamics, low investments in education and wealth, deficient policy making, political instability, and self-sustained predatory institutions.

This reasoning points to a strong role for history in economic development. Countries that start out rich are more likely to remain rich, while countries that start out poor are more likely to remain poor. Thus, if by historical accident, a country is blessed with a small “initial advantage”, this may trigger a process of cumulative causation, through which the initial advantage is magnified over time. A less lucky country, on the contrary, may fail to achieve a minimum competitiveness threshold, risking a continuous drainage of capital and talents to better endowed regions.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

A theory of inequality purely based on cumulative causation looks however rather incomplete. At a deeper level, one may want to understand why some countries started out rich while others started out poor. So, a critical question is to understand where initial conditions came from. Advocates of the Geography hypothesis contend that natural factors played a fundamental role in shaping the “initial conditions”. According to this view, physical factors, such as climate, availability of natural resources, access to navigable waters and endemic diseases played a fundamental role in shaping economic incentives at the very beginning, triggering a process of cumulative causation mediated by adverse fertility decisions, insufficient investment education and wealth, high inequality, wrong policies and poor institutions.

Box 14.3. The tragedy of Moriori

In his famous book *Guns, Germs and Steel*, Jared Diamond contends that geography plays a key role in economic development²⁸⁶. In a brief illustration, the author refers to a particular historical example: the spread of the ancestral Polynesians through the Pacific, 3.200 years ago. In that odyssey, this people encountered and colonized many islands differing dramatically in respect to their area, isolation, elevation, climate, productivity, and geological and biological sources. Within a few millennia, that single ancestral Polynesian society had spawned on those islands a range of diverse societies, from hunter-gatherer tribes to proto empires. One of these groups colonized New Zealand around A.D. 1000, to become the Maori people. A dissident group of Maoris colonized the Chatam Islands, 500 miles east of New Zealand, to become the Moriori.

Although the ancestors of the two groups shared essentially the same culture, language, technology and set of domesticated plants and animals, in the centuries after they evolved in opposite directions. Those that occupied the northern island of New Zealand,

²⁸⁶ Diamond, J., 1998. *Guns, Germs and Steel: a short history of everybody for the last 13,000 years*. Vintage, Surrey, UK.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

found suitable conditions for agriculture. There, they developed new agriculture techniques and invented new tools to grow their crops. The food surpluses allowed population to expand and the society to explore the benefits of specialization, with the emergence of craft specialists, chiefs and soldiers, and political leaders.

Those that occupied the Chatam Islands found a climate that was too cold for the Maori tropical crops to grow. The Moriori had no alternative but to revert to being hunter-gatherers. As hunter-gathered, they did not produce crop surpluses for redistribution and storage, they could not support and feed armies, bureaucracies, and political organization. Hunter-gatherer societies typically organize themselves in small groups with a primitive political structure and do not need to develop sophisticated technologies. Moreover, because the Chatam Islands are relatively small, they could support at most a population of 2000 hunter-gatherers. The result was a small population with simple technologies lacking leadership and organization.

In conclusion, the Moriori and Maori societies developed from the same ancestral society but along very different lines. The Moriori reverted to being hunter-gathers. The North-Island Maori turned to more intensive farming and develop a complex political organization. These two societies lost awareness of each other existence and did not come into contact again for roughly 500 years. Finally, in November 1835, a group of 900 Maori sailed to the Chatham Islands, attacked and exterminated the Moriori people in only one month.

14.3.2 Latitude and per capita income

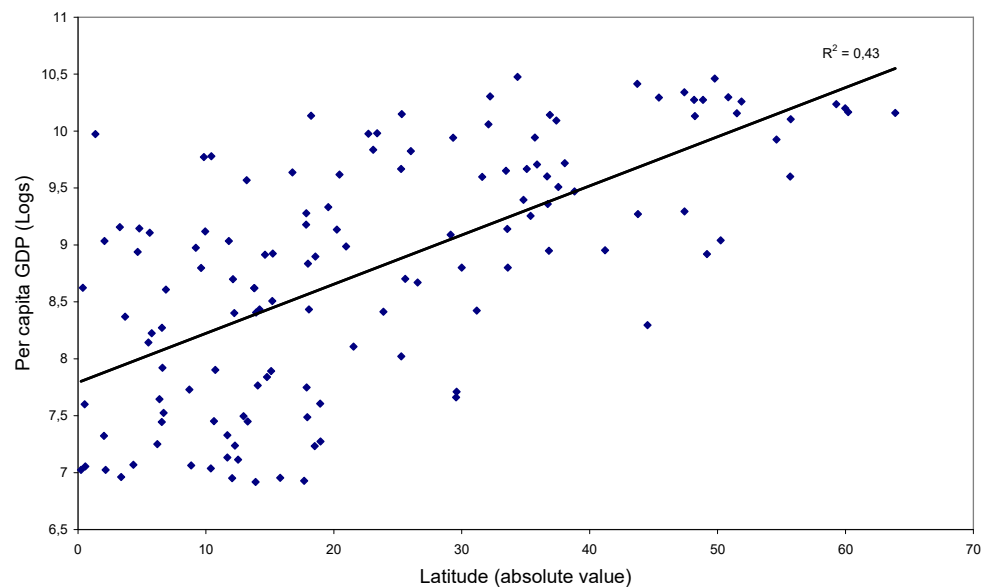
Looking at the World map, we observe that most developed countries are geographically concentrated and tend to be located in tempered areas, while the poorest countries in the World are located in the tropics.

Figure 14.1 provides an illustration. The figure crosses data on per capita GDP for 151 countries as of 1988 with the respective distances to the equator. The figure is suggestive of a systematic relationship, with per capita income being a positive function of latitude. This evidence suggests that climate has a role on economic development: if climate was irrelevant, then per capita GDP should be randomly distributed across latitudes.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Similar evidence is extensive to other geographical variables. In the growth literature, there is extensive empirical evidence pointing to the significant correlation between geographical variables and per capita incomes. Variables that typically correlate with per capita income include the share of the country located in the tropics, the incidence of malaria, the location of a country relative to the sea or to navigable rivers²⁸⁷. This evidence points to an important role of Geography in explaining economic development.

Figure 14.1: Per capita GDP and distance to equator



Data source: Hall, R., and C. Jones, 1999. "Why do some countries produce so much more output per worker than others?", *The Quarterly Journal of Economics* 114 (1), 83-116

The figure shows the cross-country correlation between per capita income and the distance to the equator. The positive slope of the regression line shows that, in general, countries located in temperate areas are richer than countries located in the tropics.

14.3.1 The mark of geography

²⁸⁷ Gallup, J., Sachs, J. and Mellinger, A., 1999. "Geography and Economic Development". *International Regional Science Review* 22 (2), 179-232.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

There are many mechanisms through which natural variables may impact on per capita income.

First, the agro-climatic conditions: a country with low rain, lack of conditions for irrigations and with nutrient-poor soil will obviously face lower productivity in agriculture than a country blessed with fertile land. These unequal conditions determine different abilities to feed large native populations. With no surprise, regions with suitable conditions for agriculture found it easy to feed and attract large populations, engaging in earlier urbanization and taking opportunity of economies of scale, through the division of labour and investments in critical infrastructure.

The argument that availability of suitable conditions for agriculture was the main determinant of the asymmetric development of the different regions in the world until the 15th century was already explored in chapter 1: by the 15th century the region of Eurasia was technological more advanced than the other regions in the World because it benefitted from the availability of a favourable selection of native plants and animal species that could easily be domesticated to produce high yields. This advantage translated into larger populations, increasing the arrival rate of new ideas, storable food surpluses that allowed for productive specialization, development of skills, and the emergence of organized, hierarchic and politically structured societies. The Americas and Australia were isolated from Eurasia and hence could not share these innovations. Although some plants and animals were domesticated by native people in the Americas, the diffusion of these techniques across the continent faced a major barrier: agricultural innovations spread well in the east-west direction, where the latitude is the same, but not in the north-south direction, where the climate changes²⁸⁸. In Eurasia, in contrast, the same innovation – say rice – could be equally productive in China and in the Iberia peninsula.

Another key geographical factor shaping economic incentives is transportation costs. In the last half century, some emerging economies managed to achieve fast economic growth

²⁸⁸ Diamond (1997), op. cit.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

specializing in labour intensive manufacturing exports, made possible by importing equipment from developed countries. Such avenue is not at reach wherever transportation costs are prohibitive. This argument backs from Adam Smith, who pointed out that “the great rivers in Africa are too great a distance from one another to give occasion to any considerable inland navigation”²⁸⁹. Landlocked economies, small islands, economies surrounded by mountains or located far from the major world markets face high transport costs. High transport costs act like tariffs on international trade, reducing the benefits of specialization and world technological diffusion. Closed economies relying on small domestic markets are more likely to be served by local monopolies that can capture the state, locking institutions in a bad equilibrium. In coastal areas, the high responsiveness of output to international trade, governments are more likely to adopt open trade policies that favour institutional change. Landlocked economies, in turn, are more likely to remain closed, and adopt inward oriented policies and barriers to creative destruction.

Where transport costs are not prohibitive, climate may interact adversely with comparative advantages, giving rise to a specialization pattern that is not conducive to economic growth. This argument was already explored in Chapters 1 and 6: specialization in cash crops such as cotton or coffee, where opportunities for technological change and learning by doing are limited, may cause a country to remain locked in an underdevelopment trap, with high inequality, low investment in human capital, adverse population dynamics, and self-feeding extractive institutions.

Another negative impact of geography comes from the prevalence of infectious diseases. Ecology determines differential habitats for the mosquitoes that carry infectious diseases, such as malaria, dengue fever, and yellow fever. These diseases are endemic in tropical areas, being almost absent elsewhere. In general, diseases reduce the quality and availability of the main asset of the poor: working time. In tropical areas, infectious diseases are an important cause of death. By reducing life expectancy, a high incidence of tropical

²⁸⁹ Adam Smith, 1776, the Wealth of Nations, Book I, Chapter III.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

diseases comes along with a smaller expected payback period for investments in human capital. Hence, people will optimally respond investing less in education. Also, in cases of extreme poverty, high mortality rates caused by disease may induce higher fertility rates, delaying the demographic transition. More generally, a high incidence of infectious diseases deters foreign direct investment and constitutes an obstacle to migration and tourism. More likely, local talents will migrate away to regions with higher levels of human capital, self-feeding the process of economic divergence.

Box 14.4: Transport costs

A developing country landlocked or surrounded by forests and mountains will face, everything else constant, higher costs on international trade than a country with a coastal area or with access to navigable rivers. If, as it is often the case, the developing country exports agricultural goods (Y) in exchange for equipment (K), its location will imply a higher relative price of capital as compared to a similar country located in the centre. In that case, high transport costs will act like a tax on physical capital, τ_K : everything else constant, this is expected to reduce the amount of investment that can be obtained out of a given saving rate and also to impact negatively on the saving rate. High transport costs are one of the vehicles through which an adverse geography impacts negatively on economic development.

Other distortions that alter the cost of investing include tariffs in imported equipment, licensing fees, bureaucratic impediments, bribes, political uncertainty, unsecure property rights.

14.3.2 Long lasting effects

Although geographical factors are invariant over time, the relative importance of these factors may change along the development process. This means that some sources of geographical advantage before are not necessarily the same as those of today.

For instance, at the time agriculture was invented, availability of arable land was the critical ingredient. Since at that time transport costs and communications were too costly to support interregional trade, geographical advantage came mainly from location close to highly fertile areas, such as those around the Nile, Tigris and Euphrates. With the progress in transportation, however, the nature of geographical advantage changed: location advantage

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

became related to proximity to coastal areas, such as the Mediterranean and the North Atlantic. Centuries later, the industrial revolution marked a move from geographical-sensitive farming to geographical-sensitive manufacturing. At the outset of the industrial revolution, proximity to key inputs such as coal, or to transport hubs, made a key difference.

In our days, transportation and communication technologies and progresses in medicine are turning location much less important than it was before. Still, even though it is possible in our days to set up a modern society without the need to step through an agriculture stage, today most advanced nations have a wide agricultural basis. By the same token, many great cities today are located near ports, even though the importance of ports is now mitigated by alternative infrastructure, such as highways, railroads and airports.

This evidence is suggestive of long-lasting effects of the initial differential geographical conditions. Even if the geographical conditions that determined the initial cost advantages are no longer important, their effect is perpetuated over time through economies of scale, fertility choices, investments in human capital and the quality of institutions. The long-lasting effects of geography help explain the data in Figure 14.1: Geography still plays an important role in explaining the current location of economic activities, even though the underlying initial advantages of geography are no longer that important.

14.4 The geography vs institutions debate

14.4.1 Natural experiments

Acemoglu and co-authors contended that the historical evidence shows no evidence of the persistence suggested by the geography hypothesis²⁹⁰. Geography is an exogenous and unchanged variable. Thus, if geography was the key determinant of economic development, regions that were initially poor should remain poor, while regions that were initially rich

²⁹⁰ Acemoglu, D., Johnson, S., Robinson, J., 2012. *Why Nations Fail*. Profile Books Ltd: London.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

should remain rich. Fortunately, there are many cases of countries that were poor and became rich and of countries that were rich and entered in divergent paths.

To demonstrate the influence of institutions, the authors focused on particular historical episodes, that can be considered as “natural experiments”: these are episodes where, while other fundamental causes of economic growth were held constant, the variable of interest – in this case the quality of institutions - changes because of some exogenous reason.

One of these “natural experiments” is the split of Korea into South Korea and the People’s Republic of Korea occurred after World War II. The two countries were born out of the same people, share the same culture and had the same per capita income just before separation. The two countries adopted however, different economic systems: while South Korea engaged in a market economy, the Republic of Korea abolished the private property and implemented a centralized command system. Geography, by definition, remained immutable. Three decades after, South Korea was a growth miracle, while the People’s Republic of Korea was amongst the poorest nations in the world. This experiment suggests that institutions overwhelm geography as a driver of economic development.

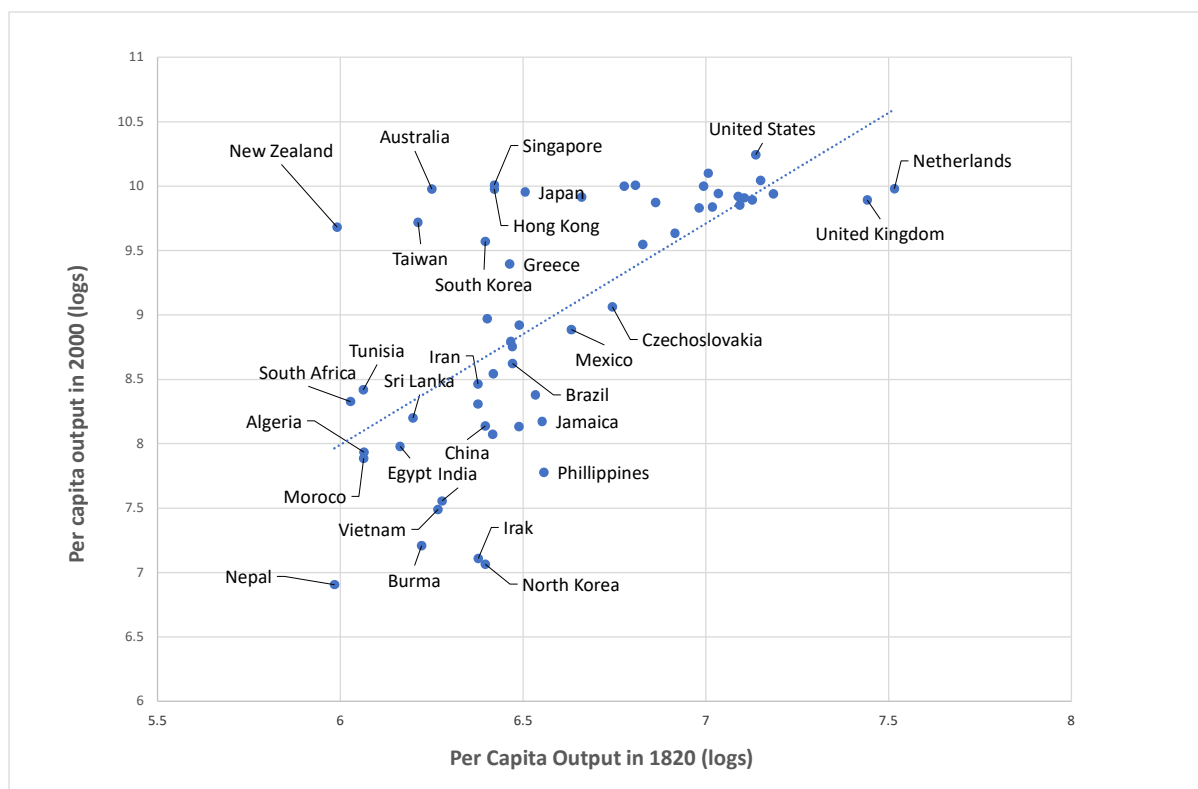
Box 14.5. Path dependence or determinism?

Figure 14.2 plots the levels of GDP per capita in 1820 and in 2000 for 55 countries for which data is available. Most of these countries consist in European, the western of shoots, and Asian countries. Africa is misrepresented, as only data for some North African economies were available as of 1820. The same for most south American economies. Hence, the data is pretty biased, in the sense it omits most countries located in the less successful continents.

Still, the positive correlation between per capita incomes in 1820 and 2000 reveals an important role for path dependence: countries that were already rich in 1820 were more likely to be among the richest by 2000. However, the figure also reveals that destinies varied considerably among the poor: on the positive side, New Zealand, Australia and the South Asian tigers managed to join the club of the rich. By contrast, Nepal, Burma, and North Korea performed worse than what would be predicted by this biased sample.

Figure 14.2: Per capita GDP in 1820 and in 2000 (logs)

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>



Source: Madison historical statistics, University of Groningen.

14.4.2 Reversals of fortune

The colonization of much of the world by Europeans, starting in the 15th century provides another “natural experiment”. European colonization transformed exogenously and abruptly the institutions in the colonized regions, while geography, by definition, remained unchanged. If geography was indeed the main factor determining economic growth, then regions that were already rich at the time of the European colonization should remain rich, while those that were poor should remain poor.

In fact, that was not the case: the European colonisations triggered many episodes of “reversals of fortune”: regions that were relatively developed by 1500, such as the Aztecs and Incas in America and the Mughals in India, were poor by 1995, while some other regions that were initially poor, like New Zealand and North America, became rich.

Daron Acemoglu and the co-authors contended that the “reversals of fortune” can be explained by the institutions that were formed by the European colonizers. In fact, the Europeans followed *different* colonization models and implemented different institutional setups in different colonies. In some regions, Europeans implemented “extractive”

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

institutions”, such as the slave plantations of the Caribbean, Congo and Central America. Extractive institutions were not designed to protect the property rights of the majority of citizens or to constrain the power of elites, but rather to deviate income to the dominant elites. In other regions, like in North America and New Zealand, Europeans founded settler societies, replicating European institutions that protect common citizens and impose constraints on elites.

In the absence of major shocks tilting again the balance of powers, the institutional arrangements founded by the European lasted for long. Colonies gained independence, but the type of institutions remained. In settler colonies that managed to build powerful middle classes and control the elites, post-colonial institutional arrangements became more democratic and more devoted to defend property rights to the majority of the population.

By contrast, in countries where institutions were extractive, the post-colonial elites exploited the pre-existent extractive institutions in their benefit, in some cases making them even more extractive. The case of Burundi is an example of the long-lasting effects of colonial institutions: Burundi is land-locked and has a disease environment that is not favourable to European settling. The Belgium colonialists therefore exploited the country cash crops (coffee, tea, cotton) indirectly, giving power to an ethnic group (the Tutsi) over the other ethnic group (the Hutu). After independence, the Tutsi maintained the power, and implemented a repressive military dictatorship that enforced the continuing deviation of resources to the ruling elite²⁹¹.

The case of Burundi also helps explain *why* the Europeans implemented different institutional models in different colonies. Ironically, the answer to this question lies on geography: European colonialists did not setup institutions for the sake of the society as a whole: they created settler societies wherever it was their interest to do so, and “extractive” institutions wherever it was their interest too.

²⁹¹ Rodrik et al., 2004.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

Thus, in places where the climate and the soil quality made it more effective to produce crops using large plantations and where the disease environment was not favourable to European settlement, the colonialists established plantation systems based on slavery and erected political and legal institutions at the benefit of few landholders against the majority of the population, giving rise to high income inequality and low investment in human capital.

In places where the climate and the soil quality made it more effective to produce using small scale farming, where most of the land was empty and with hospitable climate and germs, Europeans settled in large numbers and developed laws and institutions protecting property rights of the regular citizen. In these colonies, institutions were much more favourable to growth, broad public education and innovation.

Thus, although the reversals of fortune in colonized regions were triggered by disruptions in institutional setups caused by the Europeans, different institutional setups emerged in different colonies in response to the local geographical conditions. There was an “indirect” effect of geography on economic performance, mediated by the quality of institutions²⁹².

14.4.3 The indirect influence of geography

The question whether geography or institutions is the fundamental determinant of economic growth has also been subject to empirical scrutiny using regression analysis. A problem of identification needs however to be accounted for. The fact that the richest countries in the world have institutions that protect property rights and impose constraints on the actions of elites and powerful groups does not prove that institutions are the main cause of economic development: such correlation could reflect omitted factors or reverse causality. By the same token, the fact that the poorest countries in the World are located in the tropics does not prove that geography is the primary influence on economic prosperity. The correlation

²⁹² Sokoloff, K., Engerman, S., (2000). Institutions, factor endowments and paths of development in the new world, *Journal of Economic Perspectives* 14, 217-232.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

between latitude and per capita income observed in Figure 14.1 could be the result of an omitted factor correlated with geography.

Rodrik et al (2004) proposed an empirical strategy to disentangle the direct and indirect influences of the fundamental factors on economic performance²⁹³. The authors considered the influence of geography, institutions and of trade openness. Among these, only geography is entirely exogenous. Geography can affect per capita income directly (via health of population, for instance) or indirectly, either through trade openness (distance from markets) or through the quality of institutions. As for institutions, causality can run both ways: property rights and the rule of law impact positively on per capita income a higher per capita income increases the demand for good institutions. The same for international trade: economic integration influences per capita income through the benefits of specialization and technological diffusion, but it may also be influenced by per capita income, as more educated people may press for openness.

Taking as dependent variable the levels of per capita income in 1995, the authors implemented a cross-section analysis considering three sample sizes (64 countries, 80 countries and 140 countries). Instrumental variables were used to control for the endogeneity of institutions and of trade openness. The authors obtained strong and robust evidence that the quality of institutions overrides everything else in explaining economic performance: institutions (measured by a composite variable capturing the rule of law and protection of property rights) is the only positive and significant determinant of per capita income. The authors found that, controlling for institutions, geography has at most a weak direct effect on per capita income, while trade openness has no effect at all.

On the relationship between the determinants, the authors found a mutual causality between the quality of institutions and trade openness, suggesting that the two variables tend to reinforce each other. Notably, the authors also found a strong influence of geography on

²⁹³ Rodrik, D., Arvind Subramanian, Francesco Trebbi, 2004. "Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development," *Journal of Economic Growth* 9(2), 131-165.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

the quality of institutions. This finding confirms the earlier work by Easterly and Levine (1992) who had implemented a regression analysis on a sample of 72 former colonies, finding that geographical variables (tropical location, exposure to infectious diseases, and an environment favourable to certain cash crops) are an important determinant of the quality of institutions but have other impact on economic development beyond that indirect effect²⁹⁴.

In sum, investigation using econometric analysis seems to confirm the conclusion of the natural experiments based on historical episodes, that the influence of Geography on economic performance is mostly indirect, mediated by the quality of institutions. In light of this reasoning, the correlation between latitude and per capita income observed in Figure 14.1 does not reflect the direct influence of geography on economic performance, but rather the indirect one: to the extent that the historical creation of institutions was correlated with latitude, a statistical relationship between geography and economic performance emerges in the data, that does not necessarily imply causality.

14.5 Key ideas of Chapter 14

- The high persistence of cross-country differences in per capita income suggests a role for history and path dependence on economic development. A question therefore rises as to whether the initial advantages came from.
- In the literature, there are two main views regarding the fundamental causes of economic development. The institutional view contends that what matters is the “rules do the game” in a society, that are human-devised. The geography view contends that “forces of nature” may exert dramatic influence on economic incentives, requiring more action than simple waiting for institutional change.
- Economic institutions are those that shape economic incentives. Among these, those that secure property rights and contract enforcement are essential for investment and entrepreneurship. To sustain economic growth, other economic institutions are needed, namely to regulate markets, stabilize the macroeconomy, and deliver equity.

²⁹⁴ Easterly, W., Levine, R., 2003. Tropics, germs and crops: how endowments influence economic development. *Journal of Monetary Economics* 50(1), 3-39.

<https://mlebredefreitas.wordpress.com/teaching-materials/economic-growth-models-a-primer/>

- Economic institutions are determined by the political process, which in turn is governed by political institutions. Political institutions can be pluralist and inclusive, paving the way for the emergence of inclusive economic institutions, or be repressive and support extractive institutions.
- Institutions evolve reflecting the balance of powers between groups in the society. Societies do not naturally gravitate towards good institutions. When the balance of powers change, institutions can change within the existing framework, or through a disruption.
- There are many reasons to believe that geography played indeed a key role in shaping the initial conditions. Regions with abundant arable land and access to navigable waters emerged as more attractive places to produce and invest, feeding large populations and achieving productivity gains through the division of labour and technological progress, and the demographic transition, in a virtuous cycle. Even if geographical variables are today less important, they may have triggered in the past processes of cumulative causation that help explain income inequalities today.
- To disentangle whether the primary cause of economic progress are institutions or geography, economists have relied on both natural experiments and econometric analysis. The analysis of natural experiments like the political division of the two Koreas and the colonization of much of the world by the Europeans points to a significance incidence of “reversals of fortune” than would be consistent with the idea that the immutable forces of nature are the main driver of economic performance. The fact that many colonized countries experimented of reversals of fortune indicates that geography neither condemns a country to success nor to poverty. Also, the fact that these “reversals of fortune were preceded by abrupt changes in the institutional setup points to a prominent role of institutions as determinant of economic performance.
- The research using econometric analysis corroborates the view that institutions override all other variables as an explanation of economic performance. However, there is a significant indirect role of geography mediated by the quality of institutions. That is, even if geography cannot explain the success a country alone, it has a strong indirect effect, because it influences the quality of institutions that determine economic performance.

Problems and Exercises

Key concepts

- *Economic institutions, Political institutions. Extractive vs inclusive. Non-natural gravitation. Disease burden. Reversals of fortune. The indirect role of geography.*

Essay questions:

- Discuss: “The factor which ultimately better explains economic growth is geography”.

- The colonization of much of the world by Europeans, starting in the fifteen-century delivered different economic performances around the globe. Does this “natural experiment” favour the “geography hypothesis” or instead the “institutions hypothesis?”