Inequality and Growth: a Labor Mobility and Human Capital Analysis

Quentin Dupriez

Thesis submitted for the degree of Ph.D. in Economics

The London School of Economics and Political Science

University of London

January 2002

Abstract

We analyze the multiple connections between inequality and growth from a theoretical perspective. We first consider the debate on poverty, the Kuznets curve and the convergence of developing countries in an open-economy context. The economy is dual as in Kuznets. The two sectors interact in several important ways, however, which makes them complementary. Labor mobility between the two sectors is not perfect, and capital flows from abroad can only be directed to one of the two sectors.

We show that inequality arises as a result of the limited mobility of labor across sectors despite the mobility of capital, and that it decreases monotonically with development. We also show how the "poor" benefit in several ways from an opening of the economy to capital flows, even though they may seem to gain less than the "rich" initially.

While duality may be an appropriate description of many developing countries, it does not fit the structure of developed economies. We then consider an extension of a Schumpeterian model of endogenous growth with a continuum of sectors and where growth occurs through purposeful technological progress. Extending the process of creative destruction to jobs, we show how rigidities in the reallocation of the labor force across sectors generates equilibrium inequality.

We also look at how such rigidities affect steady-state growth, which allows us show that inequality and growth may be related in a non-linear and non-monotonic way. This sheds new light on the growth/inequality nexus and underscores why it may be difficult to find a clear empirical relationship.

Inequality has also tended to increase significantly over the past decades in most industrial countries. Skill-biased technological change is often advanced as the main explanation for such a rise. We note, however, that a large part of the increase in inequality can be attributed to the concentration of income and wealth at the very top of the distribution, which skill-biased technological change cannot explain.

We argue that the increased prevalence of winner-take-all markets may explain this phenomenon and seek to explain why agents may want to acquire human capital to participate in such markets, as opposed to education. We show how incentives may be such that the poor are attracted disproportionately to invest in winner-take-all markets and thus reinforce ex-ante inequality and harm growth.

Acknowledgements

I wish to thank Danny T. Quah for his supervision throughout the past three years, as well as Antonio Ciccone, Giovanni L. Violante and Stephen Redding for accepting to be readers for my viva.

I am particularly grateful to the GPEM of the Universitat Pompeu Fabra and its staff for being so welcoming and making my stay in Barcelona as an EDP student so fruitful and enjoyable. Very special thanks go to fellow LSE and Pompeu Fabra students Giorgia Albertin, Beatriu Canto, Matteo Cervellati, Piergiuseppe Fortunato, Matteo Iacoviello, Matthias Messner, Raoul Minetti, Robert Newland, Thomas Philippon, Ciro Rappaciulo, Philipp Schwalber, Alessandro Secchi, Uwe Sunde, Felipa Uva, Marco Vega, Leopoldo Yanes and Hernando Zuleta for sharing frustrations and enjoyment, as well as for offering comments and encouragement at various stages of my thesis. Many thanks also go to seminar participants in Barcelona, Bonn, London and Louvain-la-Neuve for helpful comments, and the CORE at the Université Catholique de Louvain for kindly hosting me during the summer of 2001.

Last, but certainly not least, I wish to thank the members of the Club Alpí Fabra for making me discover, enjoy and cherish the Pyrénées. Numerous week-end outings brought sheer pleasure as well as proved to be energizing and inspirational. Gràcies a tots, visca el CAF i que per molts anys puguem pujar molts cims, peti qui peti. And to you, U.S., for being my own Gaston Rébuffat and sharing so many fabulous ascents.

Table of Contents

List of Figures

Chapter 1

Introduction

Growth and inequality are in many ways inextricably associated to each other. The abyssal inequality in standards of living between the richest developed countries and the poorest developing ones is a relatively recent phenomenon which has its roots in the "take off" of the former group in the late 18th to early 19th century as a result of the industrial revolution. In a context where world equity and cross-border redistribution remain an utopia, cross-country inequality is thus first and foremost a growth phenomenon.

In order to explain and understand cross-country inequality, one is thus bound to explore the roots and causes of the difference in growth performances between rich and poor countries. While the industrial revolution in the West and political and economic dominance and exploitation are arguably the main reason underlying the initial opening of the gap in standards of living between developed and developing countries, such arguments are no longer sufficient when considering the more recent history.

One is thus forced to search for the factors that are most likely to explain why certain countries persistently grew faster than others, hence widening or closing the gap with the "leaders". In many ways, this constitutes the "Holy Grail" of growth theory, particularly if it allows to recommend and implement policies that are susceptible to sustainably increase growth and hence standards of living.

There has recently been much public controversy regarding the consequences of a "globalized economic system" on world inequality. World trade and finance are presented by some as establishing an exploitative link between rich and poor countries, but also as reinforcing inequality within countries. That is, poor countries are said to be exploited by rich countries, while at the same time, the inequalities between rich and poor within rich and poor countries alike are further reinforced by world trade and finance.

After an initial review of the literature on growth and inequality, chapter 3 of this thesis seeks to analyze the issue of within- and cross-country inequality and globalization. We do not seek to discuss the arguments and contradictions of the anti-globalization advocates. Instead, we seek to address a few focused questions: is globalization, in the restricted sense of an opening up of an economy to capital flows from abroad, good or bad for the poor within an economy? Do the rich within a country benefit from globalization at the expense of the poor in that country? What is the impact of globalization on cross-country inequality? What determines the size and timing of capital flows from abroad? Why do certain countries attract so much more capital than others?

While cross-country inequality appears to account for much of the difference in standards of living across the world, within-country inequality is sizeable in most countries and is an important issue for many reasons. First and foremost are social justice and equity. Such concepts are indeed more easily applicable in a national context and are more likely to lead to specific policy measures (redistributive taxation, public provision of basic goods, etc.) than in an international context. Within-country inequality may be particularly important as well to the extent that it may affect a country's long-run growth performance.

The links between within-country inequality and growth are clearly less obvious than those between cross-country growth performance and inequality. Yet it is important to understand them for policy purposes. The central issue and policy concern surrounding the growth/within-country inequality nexus is whether inequality is an inevitable side-effect of a growth-maximizing economic structure. While the traditional views based on incentives theory tended to point to a trade-off between equity and growth, more recent theories have highlighted channels through which inequality may in fact be harmful to growth. Such theories thus cast some doubt about the growth/inequality trade-off and underline the potential for redistributive policies to be growth enhancing.

After the analysis of chapter 3, which is more focused on developing countries, chapter 4 seeks to understand what determines the level of within-country inequality in developed economies, and to analyze whether a trade-off between inequality and growth necessarily arises.

We build an endogenous growth model where the labor force is allocated across a continuum of sectors, each producing a homogenous final good using a sector-specific intermediate good of a particular technology vintage. Growth is driven by technological progress at the level of the intermediate goods, which are produced by monopolists. This allows us to analyze how the labor force is allocated to the continuum of sectors, how much each sector produces, and how labor is constantly being reallocated across sectors. Introducing frictions in the reallocation process, we relate these to the equilibrium level of inequality in the economy, and highlight a particularly complex relationship between inequality and growth.

The level of inequality within industrial countries remained relatively stable for many decades and significantly lower than in developing countries. Over the recent past, however, inequality has increased again in many OECD countries. Most explanations on the literature reside upon the concept of skill-biased technological change in one way or another.

These explanations, however relevant and elegant, fail to explain the phenomenon in its entirety. A significant part of the rise in inequality has indeed occurred as a result of the concentration of wealth at the very top of the income distribution, and not only among the privileged, but still relatively wide, category of high-skills (educated) people. An additional explanation to the phenomenon of rising inequality is thus called for.

Some have highlighted the increased prevalence of winner-take-all (WTA) markets, which are structured so peculiarly that among all the participants, only the winner (or a few people at the "top") reap significant rewards. In such markets, the losers are little or not rewarded for their efforts (investment), even though the quality and extent of their investment or "performance" might be only marginally different from that of the winners.

Chapter 5 extracts the key characteristics of such WTA markets to model them in the simplest possible way. The aim is to understand why agents are pushed to invest time and resources into such markets, as opposed to education. Contrasting the two investment possibilities, we seek to understand how the choices of the "poor" and the "rich" are determined. We then seek to understand whether WTA markets offer some room for non-distortionary redistributive taxation that may reduce initial inequality and positively affect growth.

Chapter 2

Inequality and Growth, Growth and Inequality

2.1 Introduction

The growth/inequality nexus has been approached in many different ways in the literature. First, following Solow (1956) and Swan (1956) seminal work, a vast amount of research has been dedicated to further investigate and test empirically the issue of *convergence* in standards of living across countries, as predicted by the neoclassical growth framework. This line of work is focused on whether inequality among nations has a "natural" tendency to decrease with time.

A second strand of the literature, following initial work by Kuznets (1955) and Lewis (1954), has investigated the link between a country's economic development (growth) and inequality within its own population. While Kuznets and Lewis' work focused mainly on the evolution of inequality along a country's development path from a rural (poor) economy to an industrial (rich) one, subsequent work in that line has widened to explore a broader set of issues relating growth to within-country inequality. The issues dealt with include whether growth through technological progress impacts income distribution in a consistent way, and how growth affects poverty (distinct from inequality) within a country. Empirical efforts prevail in this area.

A third strand of the literature, particularly important since the renewed theoretical interest generated by research in endogenous growth, seeks to analyze the link from within-county inequality to growth. Of particular interest to this line of research is whether inequality is a necessary "side product" of an economic environment favorable to growth, or whether inequality itself may hamper growth. An important body of theoretical and empirical work has been created in this area.

A fourth strand of the literature, closely related to the second one, has been generated by the observation that inequality has increased significantly in the US, the UK and other industrial countries since the 1970s. An important number of papers have sought to quantify and analyze this phenomenon empirically, and a number of new theoretical investigations into the links between growth and within-country inequality have followed.

This chapter briefly reviews each of these areas of the literature. Given the extent of the literature in each area, we focus on the very essential contributions and those most relevant to the work of this thesis, and refer to other articles for more comprehensive reviews.

2.2 Cross-Country Inequality and Convergence

Inequality, when measured on a world basis, is essentially determined by cross-country differences in standards of living. While there is a significant degree of inequality among the population within the OECD countries and developing countries alike, the most striking differences are between average US citizens, French, Japanese or Germans and Indians, Chinese, Nigerians or Brazilians.

A straightforward application of the neoclassical growth theory generates the conclusion that growth in any given country is proportional to its distance from steady-state: the farther a country is from its own steady-state, the faster its growth rate. If one accepts the neoclassical growth model as a useful simplified representation of the economy and assumes that countries do not differ excessively in the (exogenous) factors that explain long-term growth and other characteristics, one is thus led to conclude that a set of "automatic forces"[1] generate a natural tendency for poor countries to grow faster

[1] As coined by Barro and Sala-i-Martin (1992)

than rich ones, and for the world to converge towards the same level of wealth. That is, in the long run we are all equal.

This idea that poor countries grow faster than rich ones unconditionally is the concept of *absolute convergence*. This concept receives little empirical support, as confirmed by the standard regressions of average per capita growth rate over the past few decades on the logarithm of initial per capita GDP (usually 1960 or 1965). Barro (1991) finds, to the contrary, that the average growth over the period 1960-85 is slightly positively correlated with the 1960 value of real per capita GDP. It seems rather intuitive, indeed, that countries do differ in many characteristics (other than rich/poor), including some that determine growth of the Solow residual, the unexplained growth in total factor productivity.

The concept of *conditional convergence* states more realistically that countries converge in the long run only to the extent that they share the same core characteristics. Under such circumstances, the pace of growth is no longer determined solely as a function of whether a country is poor or rich, but rather depends on each country's distance from its own steady-state.

A vast body of empirical literature has attempted to test the concept of conditional convergence and to estimate the speed of convergence towards steady-state. A good survey is provided by Barro and Sala-i-Martin (1995). We will thus only provide a brief overview of the main results obtained in some of the most relevant papers.

Barro and Sala-i-Martin (1992) test the concept of conditional convergence by looking at the experience of US States[2] and comparing it with findings across countries. While State-specific characteristics that may influence the steady-state do exist, they are of little importance relative to country-specific characteristics, they argue. Carrying out regressions on US States over different intervals covering the period 1880-1988, they find strong evidence of convergence, in the sense that "*poor states tend to grow faster in per capita terms than rich states even if we do not hold constant any variables other than*

[2]Barro and Sala-i-Martin (1995), chapter 11, extend the test to regions in Japan and Europe.

initial per capita income or product."[3] They estimate the β rate of convergence to be around 2 percent, implying a half-life of 27 years.

This provides evidence of convergence when characteristics influencing the steady-state are the same, or at least closely similar. They then proceed to show that convergence is obtained in similar cross-country regressions, but only after controlling for country-specific characteristics that determine the steady-state. Using primary and secondary school enrollment rates in 1960, the average ratio of government consumption expenditure to GDP from 1970 to 1985, proxies for political stability and a measure of market distortion, they find evidence of β convergence of about 2 percent in a sample of 98 countries, while no such convergence is obtained in the absence of variables controlling for country-specific steady-states. They find similar evidence when limiting their sample to 20 OECD countries, with or without control variables.

They take this as strong evidence in support of conditional convergence as derived from the neoclassical model of growth. They do recognize, however, that in order to agree with empirical estimates of the rate of β convergence (about 2 percent), the neoclassical model requires -given reasonable values for the other parameters- a capital share coefficient in the vicinity of 0.8.

Mankiw, Romer and Weil (1992) propose an "augmented Solow model" in which output is a constant returns to scale function of three inputs: physical capital, labor and human capital. Adding human capital to the traditional neoclassical production function, they show that their augmented Solow model yields empirically verified predictions about convergence and speed of convergence. Using the standard Summers-Heston dataset, they conclude that their augmented Solow model "*provides an almost complete explanation of why some countries are rich and other countries are poor*"[4], after controlling for the saving and population growth rates. They also conclude, therefore, to conditional convergence, and find convergence rates in the vicinity of 2 percent as well.

Quah (1996, 1997) questions the 2 percent rate of convergence obtained in most studies by arguing that the regularity may be partly attributed to a statistical uniformity

[3] Barro and Sala-i-Martin (1992), p. 245.
[4] Mankiw, Romer and Weil (1992), p. 408.

linked to the properties of times series with a unit root component. More importantly, he argues that empirical studies of convergence should focus on how distributions of countries across income evolved rather than on specific rates of convergence.

He finds evidence of clustering of countries towards a bi-modal or "twin peaks" distribution, leading to two "convergence clubs": rich countries tend to remain rich or become richer, poor countries tend to remain poor or become poorer, and some movement occurs from initially poor to rich and initially rich to poor (Taiwan, Singapore, ... and Venezuela, respectively). This convergence club empirical result is justified through a model whereby countries endogenously put themselves into groups, as summarized in Quah (1999).

In addition to the issue of convergence, many authors have sought to identify empirically the factors that determine a country's steady-state. Solow's model specifically pinpoints the saving and population growth rates as two main factors, but leaves unexplained the "A" parameter (total factor productivity, TFP) both in terms of level and growth rate. Explaining TFP growth has been the objective of the vast endogenous growth literature, as initiated by Romer (1986, 1987, 1990).

The TFP growth literature is too vast to be surveyed here and is not the focus of this study. Comprehensive surveys are provided in Barro and Sala-i-Martin (1995) and Durlauf and Quah (1998). A particular factor of relevance to chapter 3 of this study needs to be analyzed in some more details, however.

Solow's depiction of neoclassical growth was limited to a closed economy analysis. While providing a useful initial simplification, this has become more and more problematic both empirically and theoretically as countries have significantly opened up to trade and capital flows in recent decades. So much so that trade and capital flows have become key factors in shaping many countries' macroeconomic and growth conditions, particularly across the developing world.

From the international borrowing spurt in the late 1960s and 70s to the debt crisis of the 1980s and from the renewed interest from private investors in the "emerging markets" in the early and mid 1990s to the financial crises of late, capital flows have indeed had a significant influence on developing countries. Lucas (1990) looks at the implications of extending the neoclassical growth model to the open economy and wonders "*why doesn't capital flow from rich to poor countries?*" A simple extension of the model with homogenous final and capital good, homogenous labor inputs, constant returns to scale and free trade and capital mobility generates immediate convergence across countries through massive capital flows from rich to poor countries.

Lucas proposes several alternative explanations for the failure to observe fast convergence and massive capital flows. Introducing heterogeneity in labor inputs by assuming that workers are differentiated by their human capital, he concludes that such heterogeneity is unlikely to be sufficient across countries to explain the small size of capital flows. Building on Lucas (1988), he proposes to explain cross-country differences in the technology parameter "A" by positive externalities on human capital (knowledge spillovers). While he finds empirical estimates of knowledge spillovers large enough to explain equalization of returns to capital across countries despite large differences in capital/labor ratios and hence the failure to observe large capital movements, he remains somewhat sceptical about this explanation in that it rests upon knowledge spillovers being local (not crossing national boundaries). An alternative explanation would rest on capital market imperfections, mostly related to "political risk", i.e. the risk of repayment not being enforceable.

Barro, Mankiw and Sala-i-Martin (1995) make another attempt to reconcile the open-economy neoclassical growth model with empirical observations about capital flows and convergence. They follow Lucas (1990) by introducing human capital in an otherwise standard constant returns production function. By assuming that cumulative external borrowing cannot exceed the stock of physical capital K as a result of capital market imperfections that prevent human capital H to be used as collateral, they show that the steady state of the open economy is not affected with respect to that of the closed economy, but that the speed of convergence is increased. While convergence is faster with capital flows from abroad, market imperfections prevent immediate convergence as predicted from the simple extension of the neoclassical model, and they conclude that "*the main impact of this effect [capital mobility] is likely to be small [on the speed of* *convergence].*"[5]

Beyond reconciling the open-economy neoclassical growth theory with empirical observations on capital flows and speed of convergence, many empirical studies have also sought to measure the impact of openness on growth. Harrison (1996) measures the impact of several measures of openness (to trade, capital flows, etc.) on growth using a panel dataset for developing countries. She finds that her results are sensitive to the time period chosen for analysis, but concludes that "*when openness is statistically significant (...), we find that greater openness is associated with higher growth.*"[6] Other studies, including Barro (1991) and Balassa (1985) tend to show a positive association between openness and growth.

Dollar and Kraay (2001) focus on the effect of openness to trade on growth. They argue that post-1980 "globalizers" have experience higher growth than non-globalizers and OECD countries, hence exhibiting convergence with rich countries. They address issues of measurement errors on trade policy (policy-driven openness to trade), omitted variables and reverse causation and find a strong positive effect of trade on growth. They also argue that trade-induced growth does not systematically favor one type of agent at the expense of the other, i.e. that the poor benefit as much on average from higher growth than the rich.

2.3 Growth and Within-Country Inequality

The literature reviewed in the previous section focused on inequality in average incomes across countries, and whether poor countries can be expected to catch up with rich ones. Kuznets (1955), on the other hand, analyzed the evolution of inequality *within* a country over the course of its economic development.

The first two lines of his paper read as "*the central theme of this paper is the character and causes of long-term changes in the personal distribution of income. Does inequality in the distribution of income increase or decrease in the course of a country's*

[5] Barro, Mankiw and Sala-i-Martin (1995), p. 114.
[6] Harrison (1996), p. 443.

economic growth?"[7] His analysis is thus *not* about a potential trade-off between inequality and (steady-state) growth, but rather about the evolution of income distribution as an economy develops from an agricultural structure to an industrial one.

His results are based on the assumption that "*in earlier periods of industrialization, even when the nonagricultural population was still relatively small in the total, its income distribution was more unequal than that of the agricultural population. (...) The urban income inequalities might be assumed to be far wider than those for the agricultural population which was organized in relatively small individual enterprises.*"[8] He also argues, however, that income in rural areas would increase in relative terms as people migrate from the countryside to cities, and that the distribution of income within the industrial sector would be most unequal at the early stage of industrialization and become more even as the industrial sector settles.

This leads him to conclude that "*one might thus assume a long swing in the inequality characterizing the secular income structure: widening in the early phases of economic growth when the transition from the pre-industrial to the industrial civilization was most rapid; becoming stabilized for a while; and then narrowing in the later phases.*"[9] This is his widely discussed "inverted-U curve" result. Note that Kuznets himself described this inverted-U relationship as one between income inequality and the stage of a country's economic development, not as one between inequality and growth rate directly, as is often studied in the subsequent literature. He goes on interrogating himself whether the pattern of income distribution observed in older developed countries is likely to be repeated in the early phases of industrialization in developing countries.

A vast array of papers attempted subsequently to test empirically Kuznets' inverted-U result.[10] Early empirical studies seemed at first to find a robust inverted-U relationship between income inequality and income levels (development). So much so that Robinson (1976) claims that it "*has acquired the force of economic law*"[11], before presenting a simple two-sector model with different income distributions and changing relative

[7]Kuznets (1955), p. 1.
[8]Kuznets (1955), p. 16.
[9]Kuznets (1955), p. 18.
[10]See Kanbur (2000) for an extensive survey of the literature.
[11]Robinson (1976), p. 437.

population shares working in each sector (secularly declining and rising sectors) that can replicate that "economic law".

Early empirical studies were plagued with poor data quality problems, however, and had to use cross-sectional data to make inferences on a longitudinal relationship for lack of time-series observations. More recent studies as Fields (1994) and Bruno, Ravallion and Squire (1998) based on more comprehensive datasets have questioned the robustness and/or validity of earlier results. Providing and using a new panel dataset of higher quality, Deininger and Squire (1996) conclude that "*our data provide little support for an inverted-U relationship between levels of income and inequality when tested on a country-by-country basis, with no support for the existence of a Kuznets curve in about 90 percent of the countries investigated.*"[12]

Barro (1999), however, disagrees with that conclusion and estimates that the Kuznets curve remains a clear empirical regularity, even though he also finds that the relation cannot explain much of the variation in inequality across countries or over time. A complete consensus on the empirics of the Kuznets curve has thus not been reached yet.

Slightly aside from Kuznets' original argumentation about inequality and development, economists have argued about the effect of growth (and particularly growth-enhancing policies) on the poor. This debate was initiated in particular as a result of developing countries embarking upon structural reforms aimed at restoring macroeconomic balances and growth, and their effect on the poor. Dollar and Kraay (2000) conclude that growth does benefit the poor in general, that "*the effect of growth on income of the poor is no different in poor countries than in rich ones*" and that "*incomes of the poor do not fall more than proportionately during economic crises.*"[13]

On the issue of whether growth is sufficient to reduce (absolute) poverty or whether redistribution is crucial, Quah (2001) concludes that within-country inequalities play a small role in determining global inequality. Adopting a vector stochastic process that jointly determines inequality and growth, his empirical work leads him to conclude

[12] Deininger and Squire (1996), p. 573.
[13] Dollar and Kraay (2000), p. 1.

that "*to understand the secular dynamics of personal incomes against a setting of world inequalities, those forces of first-order importance are macroeconomic ones determining cross-country patterns of growth and convergence. (...) The poor benefit more from increasing aggregate growth by a range of factors, than from reducing inequality through redistribution.*"[14] Such an analysis would thus lead to the conclusion that absolute poverty reduction is most likely to be successful through growth enhancing policies rather than through redistribution, at least in poorer developing economies.

Chapter 3 looks at some of the issues reviewed above. In particular, it seeks to model the evolution of wage inequality as a country develops and tries to provide an answer as to who benefits most from opening the economy to foreign capital; the poor or the rich? It also seeks to understand why developing countries attract a lot less capital from abroad than expected, why some countries attract a lot more capital flows from abroad than others, and why those tend to be relatively richer (in terms of capital/labor ratios) countries, contrary to the predictions of the neoclassical growth theory.

2.4 Effects of Inequality on Growth

The emergence of endogenous growth theory in the 1980s suscitated a revival of interest and research in the underlying causes of the process of economic growth. Along with the drive to explain the exogenous "A" factor in the neoclassical growth theory came the issue of whether inequality is a "necessary evil" to achieve high growth, or whether it may be harmful to growth. The issue is thus no longer whether growth or development generates within-country inequality as in Kuznets or whether cross-country convergence should or should not be expected, but rather resides in the causal link from within-country inequality to growth.

The traditional view was essentially that inequality is a "necessary evil" or side-product of an environment favorable to growth when individuals are endowed with differentiated skills and abilities. Mirrlees (1971) showed how income taxation acts as a disincentive to effort (work) when individuals' abilities are unobservable. Based on

[14] Quah (2001), pp 2-3.

the same approach to incentives, the traditional view essentially held that any type of policy aimed at "manufacturing" less inequality would distort incentives and hence harm growth. Equality (inequality) is thus not necessarily harmful (beneficial) to growth, but "artificially" obtained equality in a setting where agents are endowed with differentiated skills and abilities is bound to hurt efficiency and growth.

Recent empirical and theoretical studies have challenged this incentives-based analysis of the relationship between inequality and growth, however. Persson and Tabellini (1994) regress the average growth rate of per capita GDP (1960-85) over the share of pre-tax income received by the 41st-60th percentile of the population (their measure of inequality) and other control variables. They find that equality is positively related to growth with a highly significant and quantitatively important coefficient.

The theoretical justification of their result is provided in a simple model with skill-differentiated agents who vote on a purely redistributive policy. They proceed to show that in such a context, a political equilibrium represented by the amount of redistribution chosen by the median voter leads to more redistribution if initial inequality is high. They thus identify a channel from more initial equality (lower skill differentiation) to less redistribution, and therefore more investment and faster growth as a result of a better incentive structure. They conclude that "*income inequality is harmful to growth, because it leads to policies that do not protect property rights and do not allow full private appropriation of returns from investment.*"[15]

This view does not fundamentally challenge the traditional incentives view, however, as it essentially claims that initial inequality is harmful to growth to the extent that it leads to distortions in incentives. It leaves unresolved the issue as to why initial inequality may be higher or lower and whether redistribution, by "manufacturing" more equality but also distorting incentives, may have a positive net effect on growth in the long run. Their major contribution is thus empirical in that they show that initial equality (whichever its reasons) is beneficial to subsequent growth.

Alesina and Rodrick (1994) adopt a fundamentally similar approach. They also

[15]Persson and Tabellini (1994), p. 617.

model the impact of how "*an economy's initial configuration of resources shapes the political struggle for income and wealth redistribution, and how that, in turn, affects long-run growth.*"[16] The channels leading from higher inequality to lower growth are thus the same as in Persson and Tabellini, and they conclude that "*there will be a strong demand for redistribution in societies where a large section of the population does not have access to the productive resources of the economy. Such conflict over distribution will generally harm growth.*"[17]

They also leave unanswered the question of how to achieve initial equality in "access to the productive resources". Their empirical results show that inequality in land ownership and income is negatively correlated with subsequent growth, similarly to Persson and Tabellini.

Easterly and Rebello (1993) and Perotti (1996) empirically investigate the net effect of redistributive policies on growth, i.e. the balance between the positive effect of higher equality and the negative effect through incentives. They both find that redistribution (measured by the marginal or average tax rates and/or different types of social spending) seems to have a positive net effect on growth in a cross-section of developed and developing countries.

Deininger and Squire (1996) use their higher quality and wider dataset to reassess some of the results obtained in the earlier literature, including Persson and Tabellini (1994) and Alesina and Rodrick (1994). They find that "*the negative relationship between income inequality and growth evaporates if, for example, we attempt to rerun the regressions by Persson and Tabellini using only the eighteen (out of fifty-five) high-quality observations contained in their sample.*"[18]

In general, their work provides a cautionary note to those who pointed to a clear and robust negative relationship between income inequality and growth or to a positive relationship between redistribution and growth. The main conclusion of their paper is that "*there appears to be little systematic relationship between growth and changes in*

[16] Alesina and Rodrick (1994), p. 465.
[17] Alesina and Rodrick (1994), p. 484.
[18] Deininger and Squire (1996), p. 573.

aggregate inequality. (...) This lack of change suggests that efforts to find systematic links between inequality and aggregate growth may have to be rethought."[19]

Deininger and Squire (1998) take another empirical look at the relationship between inequality and growth. While they do then replicate a negative relationship between initial income inequality and growth with their high-quality dataset, they find that it is no longer significant once regional dummies are included in the regression. They do obtain a robust, negative and significant coefficient for the initial distribution of land, however, which they take as a better indicator of initial wealth inequality than the traditionally used income-based Gini coefficients.

Splitting their sample between democratic and non-democratic countries, they find that initial inequality (in income or land) is not significant in their growth regression for democratic countries. They take this as evidence questioning the validity of political-economy models of inequality and growth of Alesina and Rodrick (1994) and Persson and Tabellini (1994). Benabou (1996) provides additional modeling of the political economy arguments linking initial inequality and growth. He also provides a useful summary and review of the empirical literature.

Forbes (2000) challenges the view that a country's initial level of inequality is negatively related to subsequent growth and seeks address the issue of how a change in inequality will affect growth. Using the Deininger-Squire dataset and panel data estimation techniques, she finds that in the short and medium term, a rise in income inequality is positively correlated with subsequent economic growth. She estimates thus that countries do face a trade-off between lower inequality and higher growth, at least in the short to medium run. She also concludes, however, that "*sufficient data are not currently available to estimate this within-country relationship over periods longer than ten years, and it is possible that the strong positive relationship between inequality and growth could diminish (or even reverse) over significantly longer periods.*"[20]

There is thus no genuine consensus on the empirics of inequality and growth. Significant evidence seems to point towards a negative effect of *initial* asset inequality (more

[19]Deininger and Squire (1996), p. 587.
[20]Forbes (2000), p. 885.

than income inequality) on subsequent growth, however. A lot less certain is the contemporaneous relationship between inequality (or change in inequality, i.e. redistribution) and growth.

The theoretical arguments relating higher initial inequality to lower subsequent growth discussed so far all resided on political economy arguments linking inequality to higher redistributive policies (or, similarly, social conflict) and hence negative effects on incentives and growth. A second, and perhaps more important, theoretical line of thought underlines another channel of interaction. The argumentation rests mostly on credit market imperfections.

Galor and Zeira (1993) first explored the relationship between distribution and growth through human capital investment when credit markets are imperfect. In a simple overlapping generations model where people have to decide either to work as unskilled in both period or to invest an amount h in human capital in period 1 and work as skilled in period 2, they show that the distribution of wealth determines aggregate output when credit markets are imperfect. They proceed to show how whether an agent of a dynasty remains among the skilled workers in the second period of his life depends on a critical amount of bequest he receives, which, in the long run, determines two types of dynasties: rich/poor or skilled/unskilled. They conclude that "*in the face of capital market imperfections the distribution of wealth significantly affects the aggregate economic activity. (...) Hence, growth is affected by the initial distribution of wealth, or more specifically by the percentage of individuals who inherit a large enough wealth to enable them to invest in human capital.*"[21]

While Galor and Zeira's argument resides both on capital market imperfections and investment indivisibilities, other models show that it carries through when investment is a continuous choice but exhibits decreasing marginal returns at the individual level. Benabou (1996) and Aghion and Howitt (1998) show that in the absence of capital market imperfections and with decreasing marginal returns at the individual level, the optimal distribution of aggregate investment is to be equally distributed across the population. Capital market imperfections in the presence of initial inequality prevent such

[21] Galor and Zeira (1993), p. 50.

an outcome to be attained and thus have a negative impact on growth. Aghion and Howitt conclude that "*when credit is unavailable, redistribution to the poorly endowed, that is, to those individuals who exhibit the higher marginal returns to investment, will be growth-enhancing. Correspondingly, more inequality is bad for growth when capital markets are highly imperfect.*"[22]

Note that while investment indivisibilities are unnecessary to their argumentation, it is crucial that marginal returns to investment be decreasing at the *individual* level, which is a strong assumption. Also, investment indivisibilities are sometimes used as an argument as to why, in the absence of perfect capital markets, inequality may be beneficial to growth. Certain investment projects may indeed involve large set-up costs which may only be financed if enough wealth is concentrated among some individuals (if capital markets are not broad enough).

Chapter 4 looks at some of the issues reviewed above. Rather than looking at the effect of initial inequality on subsequent growth, it builds a model that explains the joint determination of growth and income inequality. Growth is determined in a model of endogenous growth in which inequality arises as an equilibrium outcome as a result of rigidities in the reallocation of the labor force across sectors. This allows us to derive a complex but explicit contemporaneous relationship between growth and inequality, through their interaction with labor reallocation rigidities. This complex non-linear contemporaneous relationship sheds new light on the debate and may explain why it is difficult to extract clear empirical relationships between growth and inequality in cross-country regressions.

2.5 Technological Change and Increasing Inequality

The recent rising trend in inequality in certain OECD countries, the US and the UK in particular, is a well documented phenomenon. Murphy and Welch (1992), Gottschalk (1993, 1997), Gottschalk and Smeeding (1997), Atkinson (1996) and Machin (1996) provide comprehensive empirical studies reviewing the phenomenon and the pitfalls in

[22] Aghion and Howitt (1998), p. 286.

comparing data across time and across countries. Despite these pitfalls, the consensus is that inequality (in earnings, income or other measures) has increased over recent decades, even though this is not necessarily true, or true to the same extent, in all countries.

Three main sources of rising inequality are traditionally identified. Autor, Katz and Krueger (1998) and Machin and Van Reenen (1998) highlight the increase in wage differential across educational cohorts and its roots in skill-biased technological change. There have also been rises in within-group inequality, however, as highlighted by the greater wage dispersion for high-school and college graduates, and an increase in age-related wage differentials.

The bulk of theoretical explanations to this phenomenon has focused on trying to explain the observed increase in the skilled labor premium, i.e. the rise in between-group inequality. Two main explanations are advanced: international trade and skill-biased technological change.

One of the most widely known results of the Heckscher-Ohlin-Samuelson model of trade is that countries specialize in the production of goods that require intensive use of the factors in relative abundance. Whether cross-border trade flows actually correspond to such a pattern has nevertheless always been hotly debated in the empirical trade literature. Wood (1994) and Wood and Ridao-Cano (1996) build on that argument and conclude that the globalization of trade flows leads rich countries to specialize in skill-intensive goods and poor countries in low-skill goods. This, they conclude, not only widens the skill premium in rich countries (increasing within-country inequality), but also reduces the skill premium in poor countries, which, they argue, means that *"free trade may not be the developmentally best policy for backward countries, since it retards their accumulation of skills by causing them to specialize in goods of low skill intensity."*[23]

Whether a trade-induced shift in demand away from unskilled labor and towards skilled labor is sufficient to have generated the rise in skill premium in industrial countries is empirically somewhat dubious. Berman, Bound and Griliches (1994) conclude that

[23] Wood and Ridao-Cano (1996), p. 30.

"the shift is due mostly to increased use of skilled workers within the 450 industries in US manufacturing rather than to a reallocation of employment between industries, as would be implied by a shift in product demand due to trade (...)."[24] It could also be, however, that trade would generate reallocation of employment within industries, say from standard textiles products to higher-technology and value-added luxury products. Such a phenomenon would not be captured by Berman & al.'s empirical approach, but could still be the cause of a trade-induced increase in the skill premium.

Acemoglu (1998) seeks to explain the rise of the skill premium through skill-biased technological change. He starts from the assumption that *"new technologies are not complementary to skills by nature, but by design."*[25] He then builds a model in which both the demand for and the supply of skills are endogenous. An initial increase in the supply of skills (as occurred in the US and elsewhere with the increase of college graduates in the 1960s and 70s) depresses the skill premium (substitution effect). The larger supply of skills also increases the size of the market for skill-complementary technologies, which generates a directed technology effect. If this directed technology effect is strong enough relative to the substitution effect, the long-run skill premium may turn out to be larger than initially, even if it is reduced at first. Under such circumstances, he shows that agents can be induced to further increase the aggregate supply of skilled labor in the economy. His conclusion is that *"when there are more skilled workers, the market for technologies that complement skills is larger, hence more of them will be invented, and new technologies will be complementary to skills."*[26]

Krusell, Ohanian, Rios-Rull and Violante (2000) note that skill-biased technological change remains a latent, unobservable variable, which cannot be easily and unequivocally interpreted or measured. They then build a model based on the assumption that the elasticities of substitution between capital and unskilled labor and capital and skilled labor are different, effectively implying substitutability in the first case and complementarity in the second. Skill-biased technological change in their work thus *"reflects the rapid growth of the stock of equipment, combined with the different ways equipment in-*

[24] Berman, Bound and Griliches (1994), p. 367.
[25] Acemoglu (1998), p. 1056.
[26] Acemoglu (1998), p. 1082.

teracts with different types of labor in the production technology."[27] Calibrating and estimating their model, they find that it is able to account for most of the changes in the skill premium over the past 30 years in the U.S. through observable variables (capital equipment, labor inputs).

Aghion and Howitt (1998) build a model of disembodied technological change, where progress takes the form of general purpose technologies (GPT), i.e. innovations that are not sector specific, but that can be productivity increasing for the whole economy. They show how their model can replicate both cycles and a progressive increase in the skill premium as the innovation is adopted.

By assumption, a GPT affects the entire economy, and they claim that its adoption is in general non-linear for reasons related to strategic complementarities between sectors, social learning and other types of externalities. Adoption, they show, is slow initially before picking up quickly. This implies that the demand for skilled labor (able to work with the new GPT) is low at the beginning (below overall supply), implying a non-segmented market between skilled and unskilled labor. As adoption of the GPT picks up, skilled and unskilled labor markets become segmented, which generates a skill premium. In their model, this premium later tends to fall as the entire unskilled labor force is ultimately trained to work with the new GPT (all sectors eventually adopt the new GPT).

Observing that a significant part of the recent increase in inequality has occurred within groups, Aghion, Howitt and Violante (1999) seek to explain this type of inequality through a model of embodied technological change with GPTs. They assume that workers are ex-ante equal but that their ability to adapt to new technologies is subject to history dependent stochastic factors. Three channels that increase the skill premium are identified: faster embodied technological progress raises the premium of adaptable workers; the generality of the technology makes it easier for adaptable workers to transfer their skills to new machines; the generality of the technology implies a lower retooling cost for old machines, which further increases the demand for adaptable workers.

[27] Krusell, Ohanian, Rios-Rull and Violante (2000), p. 1030.

However valid the trade and skill-biased technological change explanations may be, they are only likely to be part of the picture, as underscored by Atkinson (1999). He insists that "*we need to recognize that the distribution of income is subject to a variety of forces, affecting earnings, wealth and incomes. These forces include the policy choices made by governments affecting market incomes and fiscal redistribution. (...) Any single theory, such as that based on a global shift of demand away from unskilled workers, cannot provide a fully adequate explanation.*"[28]

Aside from explanations residing on government interventions that affect market outcomes, it is crucial to note that not only has the rise in within-group inequality been particularly important, but, as Atkinson (1999) himself notes, empirical evidence shows that the dynamics in income distribution responsible for the recent increase in inequality has been concentrated on the upper part of the distribution. In other words, income inequality has not only occurred as a result of the poor becoming poorer (in relative, if not in absolute terms), but also as a consequence of the "middle class" loosing income shares in favor of the very upper class.

This phenomenon of concentration of income is particularly well documented in Feenberg and Poterba (2000) who report on the shares of various measures of income and wages of the upper 0.5 percent of the distribution in the US. They show that this segment of the population mobilized 11.25 percent of national adjusted gross income in 1995, up from 6.1 percent in 1976. Data from the US Census Bureau (2000) confirm this trend showing that the top 5 percent of the distribution represented 21.4 percent of aggregate income in 1998, up from 16.0 percent in 1976.

Virtually no theoretical papers address this issue of rising concentration of income at the very top of the distribution. Frank and Cook (1995) look at how winner-take-all markets have become more prevalent in the US recently. They provide a long list of such types of markets, where many "players" compete for very few but extremely high rewards. They do not provide any modeling of the structure of these markets, however, and the reasons underlying people's participation.

[28] Atkinson (1999), p. 56.

Chapter 5 looks in more detail at the dynamics of income concentration among the very top earners. It then analyses the characteristics and structure of winner-take-all markets and provides a simple modeling tool that aims to explain why agents may be attracted to invest in such types of markets as opposed to education. We then seek to understand whether WTA markets offer some room for redistributive taxation that may reduce inequality without much affecting incentives while positively impact growth.

Bibliography

[1] Acemoglu, Daron (1998). "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." Quarterly Journal of Economics, vol. 113, 4, pp. 1055-1089.

[2] Aghion, Philippe, Eve Caroli and Cecilia Garcia-Penalosa (1999). "Inequality and Economic Growth: the Perspective of the New Growth Theories." Journal of Economic Literature, vol. 38, 4, pp 1615-1660.

[3] Aghion, Philippe and Peter Howitt (1998). Endogenous Growth Theory. Cambridge, MA, MIT Press.

[4] Aghion, Philippe, Peter Howitt and Gianluca Violante (1999). "General Purpose Technology and Within-Group Inequality." Mimeo, University College London.

[5] Alesina, Alberto and Dani Rodrick (1994). "Distributive Politics and Economic Growth." Quarterly Journal of Economics, vol. 109, 2, pp. 465-490.

[6] Atkinson, Anthony B. (1996). "Income Distribution in Europe and the United States." Oxford Review of Economic Policy, vol. 12, 1, pp.15-28.

[7] Atkinson, Anthony B. (1999). "The Distribution of Income in the UK and OECD Countries in the Twentieth Century." Oxford Review of Economic Policy, vol. 15, 4, pp. 56-75.

[8] Autor, David H., Lawrence F. Katz and Alan B. Krueger (1998). "Computing Inequality: Have Computers Changed the Labor Market?" Quarterly Journal of Economics, vol. 113, 4, pp. 1169-1213.

[9] Balassa, Bela (1985). "Exports, Policy Choices and Economic Growth in Developing Countries After the 1973 Oil Shock." Journal of Development Economics, vol. 18, pp.23-35.

[10] Barro, Robert J. (1991). "Economic Growth in a Cross Section of Countries." Quarterly Journal of Economics, vol. 106, 2, pp. 407-443.

[11] Barro, Robert J. (1999). "Inequality, Growth, and Investment." NBER Working Paper 7038.

[12] Barro, Robert J., N. Gregory Mankiw and Xavier Sala-i-Martin (1995). "Capital Mobility in Neoclassical Models of Growth." American Economic Review, vol. 85, 1, pp. 103-115.

[13] Barro, Robert J. and Xavier Sala-i-Martin (1992). "Convergence." Journal of Political Economy, vol. 100, 2, pp. 223-251.

[14] Barro, Robert J. and Xavier Sala-i-Martin (1995). Economic Growth. New York, NY, McGraw-Hill.

[15] Benabou, Roland (1996). "Inequality and Growth." NBER Working Paper 5658.

[16] Berman, E., J. Bound and Z. Griliches (1994). "Changes in the Demand for Skilled Labor Within U.S. Manufacturing: Evidence from the Annual Survey of Manufactures." Quarterly Journal of Economics, 109, pp. 267-297.

[17] Bruno, Michael, M. Ravallion and Lyn Squire (1998). "Equity and Growth in Developing Countries: Old and New Perspectives on Policy Issues." World Bank Policy Research Working Papers, No. 1563.

[18] Deininger, Klaus and Lyn Squire (1996). "A New Data Set Measuring Income Inequality." World Bank Economic Review, vol. 10, 3, pp. 565-591.

[19] Deininger, Klaus and Lyn Squire (1998). "New Ways of Looking at Old Issues: Inequality and Growth." Journal of Development Economics, vol. 57, 2, pp. 259-287.

[20] Dollar, David and Aart Kraay (2000). "Growth *Is* Good for the Poor." World Bank, Development Research Group Working Paper, March.

[21] Dollar, David and Aart Kraay (2001). "Trade, Growth, and Poverty." World Bank, Development Research Group Working Paper, March.

[22] Duczynski, Petr (2000). "Capital Mobility in Neoclassical Models of Growth: Comment." American Economic Review, vol. 90, 3, pp. 687-694.

[23] Durlauf, Steven N. and Danny T. Quah (1998). "The New Empirics of Economic Growth". in Handbook of Macroeconomics, Volume 1A, Taylor John and Michael Woodford, editors, Amsterdam, Elsevier.

[24] Easterly, William and Sergio Rebelo (1993). "Fiscal Policy and Economic Growth: an Empirical Investigation." Journal of Monetary Economics, vol. 32, 3, pp. 417-458.

[25] Feenberg, Daniel R. and James M. Poterba (2000). "Taxing the Rich. The Income and Tax Share of Very High-Income Households, 1960-1995." American Economic Review, vol. 90, 2, pp. 264-270.

[26] Fields, G. (1994). "Data for Measuring Poverty and Inequality Changes in the Developing Countries." Journal of Development Economics, vol. 44, pp. 87-102.

[27] Forbes, Kristin J. (2000). "A Reassessment of the Relationship Between Inequality and Growth." American Economic Review, vol. 90, 4, pp. 869-887.

[28] Frank, Robert H. and Philipp J. Cook (1995). The Winner-Take-All Society. The Free Press.

[29] Galor, Oded and Joseph Zeira (1993). "Income Distribution and Macroeconomics." Review of Economic Studies, vol. 60, 1, pp. 35-52.

[30] Gottschalk, Peter (1993). "Changes in Inequality of Family Income in Seven Industrialized Countries." American Economic Review, vol. 83, 2, pp. 136-142.

[31] Gottschalk, Peter (1997). "Inequality, Income Growth, and Mobility: the Basic Facts." Journal of Economic Perspectives, vol. 11, 2, pp. 21-40.

[32] Gottschalk, Peter and Timothy M. Smeeding (1997). "Cross-National Comparisons of Earnings and Income Inequality." Journal of Economic Literature, vol. 35, 2, pp. 633-687.

[33] Harrison, Ann (1996). "Openness and Growth: a Time-Series, Cross-Country Analysis for Developing Countries." Journal of Development Economics, vol. 48, pp. 419-447.

[34] Kanbur, Ravi (2000). "Income Distribution and Development." in Handbook of Income Distribution, Volume 1, Atkinson Anthony B. and Francois Bourguignon, editors, Amsterdam, Elsevier.

[35] Krusell , Per, Lee E. Ohanian, Jose-Victor Rios-Rull and Giovanni L. Violante (2000). "Capital-Skill Complementarity and Inequality: a Macroeconomic Analysis." Econometrica, vol. 68, 5, pp. 1029-1053.

[36] Kuznets, Simon (1955). "Economic Growth and Income Inequality." American Economic Review, vol. 45, 1, pp. 1-28.

[37] Lewis, W. Arthur (1954). "Economic Development with Unlimited Supplies of Labor." Manchester School of Economics and Social Studies, 22, pp. 139-181.

[38] Lucas, Robert E. Jr. (1988). "On the Mechanics of Economic Development." Journal of Monetary Economics, 22, pp. 3-42.

[39] Lucas, Robert E. Jr. (1990). "Why Doesn't Capital Flow from Rich to Poor Countries?" American Economic Review, Papers and Proceedings, vol. 80, 2, pp. 92-96.

[40] Machin, Stephen (1996). "Wage Inequality in the UK." Oxford Review of Economic Policy, vol. 12, 1, pp. 47-64.

[41] Machin, Stephen and John Van Reenen (1998). "Technology and Changes in Skill Structure: Evidence from Seven OECD Countries." Quarterly Journal of Economics, vol. 113, 4, pp. 1215-1244.

[42] Mankiw, N. Gregory, David Romer and David N. Weil (1992). "A Contribution to the Empirics of Economic Growth." Quarterly Journal of Economics, vol. 107, 2, pp. 407-437.

[43] Mirrlees, James A. (1971). "An Exploration in the Theory of Optimum Income Taxation." Review of Economic Studies, vol. 38, 2, pp. 175-208.

[44] Murphy, Kevin M. and Finis Welch (1992). "The Structure of Wages." Quarterly Journal of Economics, vol. 107, 1, pp.285-326.

[45] Perotti, Roberto (1996). "Growth, Income Distribution, and Democracy: What the Data Say." Journal of Economic Growth, vol. 1, 2, pp. 149-187.

[46] Persson, Torsten and Guido Tabellini (1994). "Is Inequality Harmful for Growth?" American Economic Review, vol. 84, 3, pp. 600-621.

[47] Quah, Danny T. (1996). "Empirics for Economic Growth and Convergence." European Economic Review, vol. 40, 6, pp. 1353-1375.

[48] Quah, Danny T. (1997). "Empirics for Growth and Distribution: Polarization, Stratification, and Convergence Clubs." Journal of Economic Growth, vol. 2, 1, pp. 27-59.

[49] Quah, Danny T. (1999). "Ideas Determining Convergence Clubs." Mimeo, The London School of Economics.

[50] Quah, Danny T. (2001). "Some Simple Arithmetic on How Income Inequality and Economic Growth Matter." Mimeo, The London School of Economics.

[51] Robinson, Sherman (1976). "A Note on the U-Hypothesis Relating Income Inequality and Economic Development." American Economic Review, vol. 66, 3, pp. 437-440.

[52] Romer, Paul M. (1986). "Increasing Returns and Long Run Growth." Journal of Political Economy, vol. 94, 5, pp. 1002-1037.

[53] Romer, Paul M. (1987). "Growth Based on Increasing Returns Due to Specialization." American Economic Review, Papers and Proceedings, vol. 77, 2, pp. 56-72.

[54] Romer, Paul M. (1990). "Endogenous Technological Change." Journal of Political Economy, vol. 98, 5, pp. 71-102.

[55] Solow, Robert M. (1956). "A Contribution to the Theory of Economic Growth." Quarterly Journal of Economics, vol. 70, 1, pp. 65-94.

[56] Summers, Robert and Alan Preston (1988). "A New Set of International Comparisons of Real Product and Price Levels Estimates for 130 Countries, 1950-85." Review of Income and Wealth, vol. 34, pp. 1-26.

[57] Swan, Trevor (1956). "Economic Growth and Capital Accumulation." Economic Record, vol. 32, pp. 334-361.

[58] US Census Bureau (2000). "The Changing Shape of the Nation's Income Distribution, 1947-1998." US Department of Commerce, US Census Bureau.

[59] Wood, Adrian (1994). North-South Trade, Employment and Inequality: Changing Fortunes in a Skill Driven World. Oxford, Clarendon Press.

[60] Wood, Adrian and Cristobal Ridao-Cano (1996). "Skill, Trade and International Inequality." Institute of Development Studies Working Paper 47.

Chapter 3

Wage Inequality and Capital Flows to Developing Countries

3.1 Introduction

Absolute poverty and extreme inequality top the list of the most pressing issues in the developing world. While it seems obvious that an increase in aggregate output in the long run is indispensable to lift people out of absolute poverty, it is often passionately debated whether growth enhancing policies benefit or hurt the poor in the short to medium run.

Such a debate has been particularly active since the implementation of structural adjustment policies in many developing or emerging market economies, mostly but not exclusively under IMF supported programs. A non-exhaustive list of growth-enhancing policies traditionally recommended includes fiscal adjustment, price liberalization (market-oriented economic structure), reforms of the labor markets, privatization, and the opening of the economy to trade, foreign direct investment and capital flows (globalization).

Underlying the idea that (some) growth enhancing policies (may) hurt the poor, one usually finds the view that fighting poverty requires, at least in the short to medium term, redistribution of existing resources more than "expanding the size of the pie", which is relegated to a secondary and longer term means to combat poverty. Growth, in such a perspective, is valuable only to the extent that it also benefits the poor, and growth enhancing policies are viewed critically inasmuch as they have redistributional consequences that may hurt the poor.

Kuznets (1955) argued that an initial rise in inequality is a "necessary step" in a country's process of development from a low-output economy based on agriculture to a high-output one based on industry, but that inequality would also fall subsequently. According to that view, some increase in inequality is thus required for a "poor" economy to turn into a "rich" one. While everybody is likely to be better off in the "rich" stage than in the "poor" one, some people may thus suffer (at least in relative terms) in the intermediate stages.

Initial cross-sectional empirical studies that sought to test Kuznets' inverted U curve hypothesis appeared to validate it to the point of giving it a "stylized fact" status, or even that of "economic law" (Robinson, 1976). More recent empirical work, however, has questioned these results and showed that cross-sectional studies could yield spurious results as a consequence of country-specific determinants of inequality (Fields, 1994, Bruno, Ravallion and Squire, 1998).

Panel data studies that control for fixed effects and with a significant time-series element are difficult to carry out given the lack of data. Recent work, however, shows that Kuznets' postulate should not be classified as a stylized fact, let alone economic law. Deininger and Squire (1996) compile a new dataset of "high quality" observations and find no support for the inverted U curve when tested on a country-by-country basis. Bruno, Ravallion and Squire (1998) obtain an ordinary U relationship and point that *"for most of the range of the data, inequality falls as average income increases."*

While addressing the issue of inequality along a country's development path, Kuznets did not really consider the problem of poverty. Nor did he directly look at the relationship between the growth rate and inequality, i.e. the relationship between the *speed* of development and inequality.

A vast array of theoretical work has sought to explain how inequality may affect growth. The traditional incentives-based theory, inspired by the work of Mirrlees

(1971), argues that growth is maximized when distortionary redistributive policies are minimized. In other words, maximum growth is likely to be associated with higher inequality, inasmuch as agents differ in skills, efforts, etc.

More recent work has showed, however, that initial wealth or income inequality may be harmful to subsequent growth. Persson and Tabellini (1994) and Alesina and Rodrick (1994) argue that high initial inequality is likely to generate political pressure in favor of redistributive policies, hence negatively affecting incentives and growth. Galor and Zeira (1993) and Aghion and Howitt (1998) show that in the presence of capital market imperfections and with decreasing marginal returns to investment at the individual level, initial inequality may yield lower growth without having to invoke distortionary redistributive policies.

The inverse relationship, from growth to inequality, has received less theoretical attention. More work has been done on analyzing the distributional effects of particular growth-enhancing policies than on how growth in general affects inequality and the poor (absolute poor or relative poor). Dollar and Kraay (2000), in a recent empirical study covering 80 countries conclude that income of the poor rises one-for-one with overall growth, hence contradicting the view that growth disproportionately benefits the rich. They also, most importantly, argue that policy-induced growth (opening to international trade, lower inflation, ...) is as beneficial to the poor as it is for the overall economy.[1] This is not to deny, however, the potential distributional effects of certain specific measures as part of an overall package of growth-enhancing policies.

Critically also, Quah (2001) shows that within-country inequalities play a relatively small role in determining world inequality when compared to cross-country inequalities. His main conclusion is that the (absolute or relative) poor in developing countries benefit much more from higher aggregate national output growth than from redistributive policies at the domestic level. His study has the main benefit of refocusing the debate on the fight against poverty in developing countries on the necessity to achieve domestic

[1] They conclude that *"the basic policy package of private property rights, fiscal discipline, macro stability, and openness to trade increases the income of the poor to the same extent that it increases the income of other households in society."* (p. 6).

growth, in priority over distributional issues, which are not negated but put into context in terms of potential effects.

These papers strengthen the view that combatting poverty in the developing world is achievable first and foremost through higher growth and faster convergence toward more advanced countries, more than through redistributive policies. Among the key policies that have recently been advocated to foster growth in developing and developed countries alike, "globalization" and its materialization into freer movements of goods, services, capital and, to a much more limited and restrained extent, labor rank very high. Yet, free trade and capital movements are also the target of the most virulent opposition on the ground of their presumed distributional effects harming the poor. Not only are they assumed to hurt the poor within developing and developed countries alike, but they are also blamed for widening the gap between rich and poor countries, i.e. for generating *divergence* instead of convergence.

This chapter seeks to analyze a few core issues highlighted above through a simple growth model. What happens to inequality as the economy develops? What are the effects of liberalizing capital flows on growth, convergence and inequality? What determines the size and timing of capital flows from abroad?

We model a typical developing economy in a dual way that is more elaborate than in Kuznets in that we allow for interactions and complementarities among the two sectors of the economy. Focusing on the dual aspect of the economy allows us to consider the issue of convergence with free movement of capital across countries and to explain why it is not immediate and why certain countries attract more capital from abroad than others. The distributional effects of opening the economy to capital flows can also be analyzed simply and clearly.

3.2 Domestic Capital, Universal Capital, Social Capital

Kuznets adopted a very simplistic view of dualism in that he assumed that the economy is divided in two sectors, agriculture and industry. In his work, as well as Robinson's (1976) formalization, agriculture and industry are totally separate sectors that do not interact on the production side. Productivity in one sector is independent of output or productivity in the other, and the only channel of interaction is through shifts in labor allocation.

Later views on dualism in developing economies are more sophisticated. Instead of dividing the economy between agriculture and industry, development theory has tended to focus on a division between formal and informal sectors, traditional and modern sectors or, as in trade theory, between tradable and non-tradable sectors. Such a division is a lot more realistic in the sense that it allows for complementarities in production between the two sectors.

Aggregate output in a Kuznets style dual economy would take the form $Y = Y_1(L_1, K_1) + Y_2(L_2, K_2)$, where Y_i, L_i and K_i are output, labor and capital, respectively, in sector i. In the more elaborate version, aggregate output would take the general form $Y = f(L_1, K_1, L_2, K_2)$, which allows a range of interactions between the two sectors and encompasses Kuznets' formulation. We will adopt a specific formulation for Y in the next section and divide the economy in two sectors according to the type of capital that they use, and the way such capital can be accumulated.

One can observe indeed that a wide range of investments cannot be financed by international capital flows, but have to be paid for through domestic saving. Other types of capital infrastructure, on the other hand, are readily financeable by foreign borrowing or through flows of foreign direct investment.

We shall thus assume that two types of capital coexist in the economy: *domestic capital*, and *universal capital*. What will be henceforth defined as domestic capital and universal capital should be understood widely and are in no way limited to physical capital. Also, as will become clear later, the concepts should not be understood rigidly as time or space invariant. To the contrary, the distinction between the two types of capital are likely to be affected by government policies and the institutional framework proper to each economy.

The single most important component of domestic capital is undoubtedly human capital, and the infrastructure allowing its accumulation. Empirically, one can indeed observe that the installation of systems of basic education requires domestic financing and is not paid for by private borrowing or direct investment from abroad. Developing countries that have successfully managed to set up such systems have typically had high domestic saving rates, which generate the necessary domestic resources for such investments. Higher education systems are likely to be classified as domestic capital as well, even though this is a case where the distinction may get blurred. There are instances, for example, of universities setting up schools abroad, which we can consider as a way to finance higher education infrastructure through foreign direct investment. Some of the higher education system could thus be considered as universal capital.

The accumulation or preservation of human capital is also affected by the public health system. The installation and maintenance of basic public health services (vaccination, sanitation, basic care) in developing countries is unlikely to be financeable with external borrowing or foreign direct investment, and yet is an essential element that affects production. Such capital should thus unequivocally be classified as domestic capital. Yet, some types of high technology and sophisticated treatments might be offered through foreign-financed private clinics, once again underscoring the possible blur between domestic and universal capital.

Domestic capital should also be understood widely enough to encompass the concept of social capital. North (1989) argued that much of the growth of productivity that occurred in OECD countries can be attributed to the development of institutions that reduced transaction costs. Similarly, Fukuyama (1995) underlined the importance of trust between social groups or individuals in framing the corporate structure of an economy and in generating the conditions for sustained growth by lowering transaction costs.

Knack and Keefer (1997) attempt to measure social capital, understood as trust, norms of civic cooperation and associations within groups. They proceed to show that their indices of trust and civic cooperation are important in explaining output growth.

Although major problems of definition and measurement remain (and are unlikely ever to be resolved), a relative consensus on the importance of social capital for growth seems to have emerged. Trust, civic cooperation and a low degree of social fragmentation (along ethnic, religious or linguistic lines) are often pinpointed as major elements contributing to high social capital. Strong, stable and transparent political, judicial, legal and administrative institutions are also frequently highlighted as essential elements in establishing the conditions for sustained growth in developing countries.

Clearly, part of a country's social capital pertains to "soft" cultural variables that are hard to define and quantify, and that are not accumulable in the traditional economic sense. Crucially, however, a significant part of the improvement in "trust" and the institutional framework requires real and measurable economic investments. Setting up an efficient and just judicial system could require, for example, increasing the number of judges, improved training, pay rises, computerization, etc. The same argument can be extended to the whole public administration.

Some social capital can thus be accumulated in a traditional economic sense by investment of real resources. The accumulation of the economically measurable portion of social capital is likely to be constrained by domestic saving, so that in most cases we will consider social capital as domestic capital.

Domestic capital should not be understood as pertaining only to human capital or the institutional framework. A significant portion of physical capital may also be considered as domestic capital. One typical example would consist of basic road and communication infrastructure. Foreigners may not be willing to finance such type of capital, either in the form of loans or direct investment, as the returns may not necessarily be financially appropriable, or at too high a cost. It must be recognized also that domestic investors may not be in a position to finance such projects. This would leave room for "useful government spending", as modelled by Barro (1990), or Barro and Sala-i-Martin (1992b).

Certain types of physical capital that one would think of as intrinsically universal might also be restricted to domestic as a result of government regulations specific to each country. Public utilities in most developing countries were, until recently, considered as a sector without possible foreign intervention. Foreign direct investment was banned in many cases in the telecommunication, electricity or water sectors. External borrowing may also not have been possible for domestic companies operating in the sector, either as a result of direct government restrictions, or for other credit constraining reasons. These sectors are typical examples of how government policy may affect the definition of domestic and universal capital.

In recent years, governments in many emerging market economies have started to deregulate the public utilities and infrastructure sectors and have put in place privatization programs together with the opening of these sectors to foreign investment. Examples abound throughout emerging markets and have affected telecommunication companies, the provision or even distribution of electric power, water and sewerage services, and the transportation sector. Toll roads, bridges or power plants are increasingly financed by foreign investors through Build-Operate-Transfer (BOT) or Build-Operate-Own (BOO) schemes, which was unthinkable only a decade or so ago.

Separating the economy into a dual structure along domestic/universal lines seems empirically relevant. It also allows us to analyze formally the issue of convergence in an open economy context, to determine the factors that underpin the size and timing of capital flows, and to consider the impact of such flows on inequality.

Barro, Mankiw and Sala-i-Martin (1995) sought to explain why convergence across countries is not immediate in an open economy context and proceeded along somewhat similar lines as those developed above. They assume that output is a function of physical capital, human capital and raw labor. They proceed to show that if foreign borrowing is constrained by the need for collateral, and that if human capital cannot be used for that purpose, the economy exhibits convergence properties that are more similar to those of a closed economy than those obtained in a simple extension of the Solow neo-classical economy with perfect capital mobility.

The hypothesis that certain types of capital cannot be accumulated through capital flows from abroad is thus similar to Barro, Mankiw and Sala-i-Martin's starting point, but the empirical rationale is somewhat different, and the key difference is that the economy is split in a dual, yet complementary, structure.

3.3 The Model

As explained above, we assume that the economy is dual in the sense that two broadly defined sectors coexist. This coexistence of a "domestic" and "universal" sector can also be thought of along the more usual lines of traditional vs. modern, informal vs. formal or non-tradable vs. tradable. They key assumption is that, in contrast to Kuznets, we postulate that the two sectors are not independent of each other. Specifically, we assume that output of the homogenous final good takes the form

$$Y = A\left[K_d^{\alpha}\left(\left(1-v\right)L\right)^{1-\alpha}\right]^{\gamma}\left[K_u^{\beta}\left(vL\right)^{1-\beta}\right]^{1-\gamma} \tag{3.1}$$

where K_d is domestic capital, K_u is universal capital, and v is a share of the labor force L.

The dynamic equations of the model are:

$$\begin{aligned}
\frac{\dot{L}}{L} &= n && (3.2)\\
v &\leq 1-e^{-\eta K_d} && (3.3)\\
\dot{K}_d &= I_{K_d}-\delta K_d && (3.4)\\
\dot{K}_u &= I_{K_u}-\delta K_u && (3.5)
\end{aligned}$$

where I_{K_x} is the total investment in either type of capital.

The key element of the production function is that the two sectors are complementary in "output", i.e. the combination of labor and capital used in each sector, and not just in the amount of raw capital. An economy with a well developed domestic sector provides a favorable environment for production in the universal sector, and vice versa. Empirically, an economy with a high level of basic education (abundant human capital), good basic transportation infrastructure and an efficient administration (lack of red tape, clear rules and foreseeable decisions, etc.) offers good conditions for a multinational to set up operations. Similarly, a country with a well developed universal sector (good telecommunication and international transportation infrastructures, good financial sector and well developed banks, etc.) offers a favorable environment to the domestic sector.

We also assume, critically, that v is constrained in this economy: the "social planner" chooses v optimally, under the constraint that $v \leq 1 - e^{-\eta K_d}$. In other words, the economy is constrained by the fact that only a fraction of the labor force is qualified, at any given time, to work in the universal sector. The rationale for this assumption is that the universal sector employs more qualified people, who need a minimum amount of training. We assume that v increases with domestic capital: richer and more developed countries have better educational systems that allow them to train the vast majority of the population. As K_d tends to infinity, the constraint is no longer binding as it converges to $v \leq 1$[2]. Any cumulative distribution function could be used instead of the exponential function used here. Each specific functional form would yield to specific speeds and paths of convergence, but would not affect the main results.

We consider a representative agent who maximizes dynastic utility and has a CES utility function:

$$U(C) = \frac{C^{1-\theta} - 1}{1-\theta} \tag{3.6}$$

3.3.1 The Unconstrained Closed Economy

We first solve the model for a closed economy that faces no constraint on v. This is useful to derive the benchmark from which deviations from optimality will be compared. Although we consider a closed economy, we preserve the distinction between K_d and K_u, and capital decisions are not reversible. Without any loss of generality, we ignore population growth and technological progress. The social planner optimizes as follows:

$$Max \int_0^\infty \left[\frac{C^{1-\theta}}{1-\theta}\right] e^{-\rho t} dt \tag{3.7}$$

subject to:

[2]Obviously, we always have $v \in [0,1]$.

$$
\begin{aligned}
Y &= A\left[K_d^{\alpha}\left((1-v)\,L\right)^{1-\alpha}\right]^{\gamma}\left[K_u^{\beta}\left(vL\right)^{1-\beta}\right]^{1-\gamma} && (3.8)\\
\dot{K}_d &= I_{K_d}-\delta K_d && (3.9)\\
\dot{K}_u &= Y-C-I_{K_d}-\delta K_u && (3.10)\\
I_{K_d} &\geq 0 && (3.11)\\
I_{K_u} &\geq 0 && (3.12)
\end{aligned}
$$

Optimization over the choice variables C, v (the share of total labor L allocated to the universal sector), and I_{K_d} yields the following conditions:

$$
\begin{aligned}
C^{-\theta}e^{-\rho t} &= \mu && (3.13)\\
\mu\left\{\begin{array}{c}
A\gamma\left[K_d^{\alpha}\left((1-v)\,L\right)^{1-\alpha}\right]^{\gamma-1}\\
\left[(1-\alpha)\,K_d^{\alpha}\left((1-v)\,L\right)^{-\alpha}(-L)\right]\\
\left[K_u^{\beta}\left(vL\right)^{1-\beta}\right]^{1-\gamma}+A\,(1-\gamma)\\
\left[K_d^{\alpha}\left((1-v)\,L\right)^{1-\alpha}\right]^{\gamma}\\
\left[K_u^{\beta}\left(vL\right)^{1-\beta}\right]^{-\gamma}\left[K_u^{\beta}\,(1-\beta)\left(vL\right)^{-\beta}L\right]
\end{array}\right\} &= 0 && (3.14)\\
\lambda &= \mu && (3.15)\\
-\lambda\delta+\mu\left\{\begin{array}{c}
A\gamma\left[K_d^{\alpha}\left((1-v)\,L\right)^{1-\alpha}\right]^{\gamma-1}\\
\left[\alpha K_d^{\alpha-1}\left((1-v)\,L\right)^{1-\alpha}\right]\left[K_u^{\beta}\left(vL\right)^{1-\beta}\right]^{1-\gamma}
\end{array}\right\} &= -\dot{\lambda} && (3.16)\\
\mu\left\{\begin{array}{c}
A\,(1-\gamma)\left[K_d^{\alpha}\left((1-v)\,L\right)^{1-\alpha}\right]^{\gamma}\\
\left[K_u^{\beta}\left(vL\right)^{1-\beta}\right]^{-\gamma}\left[\beta K_u^{\beta-1}\left(vL\right)^{1-\beta}\right]-\delta
\end{array}\right\} &= -\dot{\mu} && (3.17)
\end{aligned}
$$

Equation (3.14) is the constraint on the marginal productivity of labor. Because we assume for the time being that labor is not differentiated and fully mobile across sectors, the marginal productivities of labor in the two sectors must be equal at all times: $\frac{\partial Y}{\partial(vL)}=\frac{\partial Y}{\partial[(1-v)L]}$. Equation (3.14) implies that v^{opt}, the optimal fraction of total labor employed in the universal sector, is a *constant.*

$$v^{opt} = \frac{(1-\gamma)(1-\beta)}{1-\beta(1-\gamma)-\alpha\gamma} \tag{3.18}$$

One should note, in particular, that v^{opt} does not depend on the amount of domestic or universal capital. It is permanently at its steady-state value, which depends exclusively on the parameters of the model, i.e. the relative size of the domestic sector, γ, and the capital intensities of the domestic and universal sectors, α and β, respectively.

Using equations (3.15), (3.16) and (3.17), we see that the second optimization condition requires that the marginal productivities of domestic and universal capital be equal: $\frac{\partial Y}{\partial K_d} = \frac{\partial Y}{\partial K_u}$. The optimal ratio of domestic to universal capital is thus a constant as well:

$$\frac{K_d}{K_u} = \frac{\alpha\gamma}{\beta(1-\gamma)} \tag{3.19}$$

In order to ensure that this condition holds at all times, we would need to assume reversibility of investment decisions, which we wish to avoid. Equation (3.19) would hold if initial conditions are such that $\frac{K_d(0)}{K_u(0)} = \frac{\alpha\gamma}{\beta(1-\gamma)}$. If that were not the case, the transition dynamics would be such that either $I_{K_d} = 0$ (if $\frac{K_d(0)}{K_u(0)} > \frac{\alpha\gamma}{\beta(1-\gamma)}$), or $I_{K_u} = 0$ (if $\frac{K_d(0)}{K_u(0)} < \frac{\alpha\gamma}{\beta(1-\gamma)}$) in a first phase, before turning positive again. This is not an important consideration here, however, and we can simply assume that initial conditions are such that equation (3.19) holds at $t = 0$ and that there are no subsequent exogenous disturbances (wars, ...) so that it holds at all times.

The influence of the parameters on the steady state values of v^{opt} and $\frac{K_d}{K_u}$ is easily determined:

- $\frac{d(K_d/K_u)}{d\gamma} = \frac{\alpha}{\beta(1-\gamma)} > 0$. The parameter $\gamma \in]0,1[$ is used in this model as an indicator of the relative size of the domestic sector. A larger value of γ, for given capital intensities α and β, implies a larger optimal value of domestic capital relative to universal capital, which one would expect if the domestic sector is large relative to the universal sector.

- $\frac{d(K_d/K_u)}{d\alpha} = \frac{\gamma}{\beta(1-\gamma)} > 0$ and $\frac{d(K_d/K_u)}{d\beta} = -\frac{\alpha\gamma}{\beta^2(1-\gamma)} < 0$. A relatively large capital intensity in any of the two sectors implies a relatively large optimal value of the respective capital stock. In particular, if the universal sector is relatively intensive in capital, a more important fraction of the total capital stock of the economy $\bar{K} = K_d + K_u$ is allocated to that sector at any given time and for a given value of the size indicator γ.

- $\frac{dv^{opt}}{d\gamma} = \frac{-(1-\beta)(1-\alpha)}{(1-\beta+\gamma\beta-\alpha\gamma)^2} < 0$. This confirms that the parameter γ is used as the relative size indicators: a larger value of γ implies, ceteris paribus, that a higher proportion of the labor force is employed in the domestic sector, which one would expect if the domestic sector is large.

- $\frac{dv^{opt}}{d\alpha} = \frac{\gamma}{(1-\beta+\gamma\beta-\alpha\gamma)^2}(1-\beta)(1-\gamma) > 0$ and $\frac{dv^{opt}}{d\beta} = \frac{-\gamma(1-\alpha)(1-\gamma)}{(1-\beta+\gamma\beta-\alpha\gamma)^2} < 0$. A higher capital intensity in any sector implies, ceteris paribus, that a lower proportion of the labor force is employed in that sector.

Equations (3.13) and (3.16) yield the usual dynamic equation for consumption:

$$\frac{\dot{C}}{C} = \frac{1}{\theta}\left\{ \begin{array}{c} A\gamma\left[K_d^{\alpha}\left(\left(1-v^{opt}\right)L\right)^{1-\alpha}\right]^{\gamma-1}\left[\alpha K_d^{\alpha-1}\left(\left(1-v^{opt}\right)L\right)^{1-\alpha}\right] \\ \left[K_u^{\beta}\left(v^{opt}L\right)^{1-\beta}\right]^{1-\gamma} - \delta - \rho \end{array} \right\} \tag{3.20}$$

Given that $\frac{K_d}{K_u}$ is constant, we can deduct also that $\frac{I_{K_d}}{I_{K_u}} = \frac{\alpha\gamma}{\beta(1-\gamma)}$. Constant proportions of total investment $\bar{I} = I_{K_d} + I_{K_u} = Y - C$ are thus devoted to each of the two sectors[3]. This means that the dynamics of the economy can be fully described by analyzing the dynamics of C and K_d (or K_u). Agents optimally choose the level of consumption given the initial condition on the total level of capital $\bar{K}$ present in the economy. For any given level $\bar{K}$, equation (3.19) and $\bar{K} = K_d + K_u$ determine a unique optimal allocation (K_d, K_u). We can thus express the production function in terms of K_d and L only, as v is also a constant (function of the parameters of the model).

[3]The domestic sector receives a share $\frac{\alpha\gamma}{\alpha\gamma+\beta(1-\gamma)}$ of total investment.

$$Y^{closed} = AK_d^{\alpha\gamma+\beta(1-\gamma)} \left[\left(\left(1-v^{opt}\right)L\right)^{1-\alpha}\right]^{\gamma} \left\{\left[\frac{\beta(1-\gamma)}{\alpha\gamma}\right]^{\beta} \left(v^{opt}L\right)^{1-\beta}\right\}^{1-\gamma} \tag{3.21}$$

Since we assumed L to be constant, we have

$$Y^{closed} = A\Omega K_d^{\alpha\gamma+\beta(1-\gamma)} = f(K_d) \tag{3.22}$$

where $\Omega = \left[\left(\left(1-v^{opt}\right)L\right)^{1-\alpha}\right]^{\gamma} \left\{\left[\frac{\beta(1-\gamma)}{\alpha\gamma}\right]^{\beta} \left(v^{opt}L\right)^{1-\beta}\right\}^{1-\gamma}$ is a positive constant. We can easily see that $0 < \alpha\gamma + \beta(1-\gamma) < 1$ because we assume $0 < \alpha < 1$, $0 < \beta < 1$ and $0 < \gamma < 1$. The expression is thus a weighted average of two parameters included between zero and one. This demonstrates that the marginal productivity of K_d is positive but decreasing in K_d[4].

The dynamics of K_d is derived from:

$$\dot{K}_d = I_{K_d} - \delta K_d \tag{3.23}$$

$$I_{K_d} + I_{K_u} = f(K_d) - C \tag{3.24}$$

$$\frac{I_{K_d}}{I_{K_u}} = \frac{\alpha\gamma}{\beta(1-\gamma)} \tag{3.25}$$

From these three equations, we have that,

$$\dot{K}_d = \frac{\alpha\gamma}{\alpha\gamma + \beta(1-\gamma)}\left[f(K_d) - C\right] - \delta K_d \tag{3.26}$$

The dynamic equation for consumption can also be written more succinctly as,

$$\frac{\dot{C}}{C} = \frac{1}{\theta}\left\{\left[\alpha\gamma + \beta(1-\gamma)\right] A\Omega K_d^{\alpha\gamma+\beta(1-\gamma)-1} - \delta - \rho\right\} \tag{3.27}$$

These two equations fully describe the dynamic properties of the closed economy with domestic and universal capital. It does not differ fundamentally in terms of dynamics

[4] $f'(K_d) > 0$ and $f''(K_d) < 0$ because $\alpha\gamma + \beta(1-\gamma) - 1 < 0$

and steady state properties from the traditional Ramsey-Cass-Koopmans economy. In particular, there is a unique steady state C^{SSA}, K_d^{SSA} and the economy has saddle path stability. Because we assumed no population growth and did not consider technological progress, this economy reaches a steady state with zero nominal growth. Introducing steady state growth in per capita variables is obtained straightforwardly by assuming technological progress as in the Ramsey-Cass-Koopmans model.

The steady state values of consumption and domestic capital (and consequently also universal capital) are as follows (where ^{SSA} denotes steady state autarchy equilibrium):

$$C^{SSA} = f(K_d^{SSA}) - \left[\frac{\alpha\gamma + \beta(1-\gamma)}{\alpha\gamma}\right]\delta K_d^{SSA} \tag{3.28}$$

$$K_d^{SSA} = \left[\frac{\alpha\gamma + \beta(1-\gamma)}{\delta + \rho} A\Omega\right]^{\frac{1}{1-\alpha\gamma-\beta(1-\gamma)}} \tag{3.29}$$

where $1 - \alpha\gamma - \beta(1-\gamma) > 0$.

3.3.2 The Constrained Closed Economy

We now consider the case of a closed economy that faces a binding constraint on v, that is an economy for which initial conditions are such that $1 - e^{-\eta K_d(0)} < v^{opt}$. One would obviously expect the economy to cease being constrained on its allocation of labor to the universal sector at some point as v is an increasing function of K_d. In the early stages of development, however, the economy is constrained to set $v = 1 - e^{-\eta K_d} < v^{opt}$.

Optimization over the choice variables C and I_{K_d} yields the following conditions:

$$C^{-\theta}e^{-\rho t} = \mu \tag{3.30}$$

$$\lambda = \mu \tag{3.31}$$

$$-\lambda\delta+\mu\left\{\begin{array}{c} A\gamma\left[K_d^{\alpha}\left(e^{-\eta K_d}L\right)^{1-\alpha}\right]^{\gamma-1} \\ \left[\begin{array}{c}\alpha K_d^{\alpha-1}\left(e^{-\eta K_d}L\right)^{1-\alpha}+(1-\alpha) \\ K_d^{\alpha}\left(e^{-\eta K_d}L\right)^{-\alpha}\left(-\eta e^{-\eta K_d}L\right)\end{array}\right] \\ \left[K_u^{\beta}\left(\left(1-e^{-\eta K_d}\right)L\right)^{1-\beta}\right]^{1-\gamma}+ \\ A(1-\gamma)\left[K_d^{\alpha}\left(e^{-\eta K_d}L\right)^{1-\alpha}\right]^{\gamma} \\ \left[K_u^{\beta}\left(\left(1-e^{-\eta K_d}\right)L\right)^{1-\beta}\right]^{-\gamma} \\ \left[(1-\beta)K_u^{\beta}\left(\left(1-e^{-\eta K_d}\right)L\right)^{-\beta}\left(\eta e^{-\eta K_d}L\right)\right]\end{array}\right\} = -\dot{\lambda} \quad (3.32)$$

$$\mu\left\{\begin{array}{c} A(1-\gamma)\left[K_d^{\alpha}\left(e^{-\eta K_d}L\right)^{1-\alpha}\right]^{\gamma} \\ \left[K_u^{\beta}\left(\left(1-e^{-\eta K_d}\right)L\right)^{1-\beta}\right]^{-\gamma} \\ \left[\beta K_u^{\beta-1}\left(\left(1-e^{-\eta K_d}\right)L\right)^{1-\beta}\right]-\delta\end{array}\right\} = -\dot{\mu} \quad (3.33)$$

Optimization again requires that the marginal returns on each type of capital be equalized. Using equations (3.31), (3.32) and (3.33) we obtain that

$$K_u=\frac{\beta(1-\gamma)}{\alpha\gamma+\eta K_d\left[\frac{e^{-\eta K_d}}{1-e^{-\eta K_d}}-\frac{\gamma(1-\alpha)}{(1-\gamma)(1-\beta)}\right]}K_d \quad (3.34)$$

As in the previous section, this condition could hold at all times with certainty only if we assumed that capital decisions are reversible. We again wish to avoid that assumption, but assume instead that initial conditions are such that equation (3.34) is respected and that there are no subsequent exogenous shocks to disturb it, so that it also holds at all times. Again, assuming initial conditions where equation (3.34) does not hold simply means that investment is temporarily directed only to one type of capital, which is unimportant here.

The term $\alpha\gamma$ in the denominator is the "optimality" term, as it is equal to the denominator term in equation (3.19), which determines the unconstrained optimal level of K_u for any level of K_d. The second term in the denominator is a "distortion from optimality" term. It reflects the impossibility for the constrained economy to be on the first best convergence path. It is easy to verify that $\frac{e^{-\eta K_d}}{1-e^{-\eta K_d}}-\frac{\gamma(1-\alpha)}{(1-\gamma)(1-\beta)}$ is positive for values of K_d below a critical level K_d^O for which the term equals to zero. The function $\frac{e^{-\eta K_d}}{1-e^{-\eta K_d}}$ is indeed equal to infinity for $K_d=0$, and is continuous and everywhere decreasing in K_d. There exists thus a unique level K_d^O for which $\frac{e^{-\eta K_d}}{1-e^{-\eta K_d}} = \frac{\gamma(1-\alpha)}{(1-\gamma)(1-\beta)}$, i.e. a unique level from which the constrained economy rejoins the optimal convergence path. This level is given by,

$$K_d^O = \frac{1}{\eta} \ln \left[\frac{(1-\gamma)(1-\beta) + \gamma(1-\alpha)}{\gamma(1-\alpha)} \right] \tag{3.35}$$

It is obvious also that,

$$v^O = 1 - e^{-\eta K_d^O} = v^{opt} = \frac{(1-\gamma)(1-\beta)}{(1-\gamma)(1-\beta) + \gamma(1-\alpha)} = \frac{(1-\gamma)(1-\beta)}{1-\beta(1-\gamma) - \alpha\gamma} \tag{3.36}$$

The distortion from optimality term is necessarily equal to zero at the level where the optimal v is attained. For any level $v > v^{opt}$, the constraint is no longer binding, and a fraction v^{opt} of the labor force operates in the universal sector.

From this, we can conclude that $\forall K_d < K_d^O : K_u^{constrained} < K_u^{unconstrained}$, and

$$K_u = \begin{cases} \frac{\beta(1-\gamma)}{\alpha\gamma + \eta K_d \left[\frac{e^{-\eta K_d}}{1-e^{-\eta K_d}} - \frac{\gamma(1-\alpha)}{(1-\gamma)(1-\beta)} \right]} K_d & \forall K_d < K_d^O \\ \frac{\beta(1-\gamma)}{\alpha\gamma} K_d & \forall K_d \geq K_d^O \end{cases} \tag{3.37}$$

We can substitute this expression for K_u in the production function again, and obtain that,

$$Y^{closed/constrained} = \begin{cases} h(K_d) & \forall K_d < K_d^O \\ f(K_d) & \forall K_d \geq K_d^O \end{cases} \tag{3.38}$$

where $h(K_d) < f(K_d)$ for $K_d < K_d^O$ and $h(K_d) = f(K_d)$ for $K_d \geq K_d^O$.

We can now show that wage in the universal sector is higher than in the domestic sector $\forall K_d < K_d^O$. The marginal productivities of labor in each sector are given by:

$$\begin{aligned} \frac{\partial Y}{\partial(vL)} &= A(1-\gamma)\left[K_d^\alpha((1-v)L)^{1-\alpha}\right]^\gamma \\ & \left[K_u^\beta(vL)^{1-\beta}\right]^{-\gamma}\left[(1-\beta)K_u^\beta(vL)^{-\beta}\right] \end{aligned} \tag{3.39}$$

$$\frac{\partial Y}{\partial\left[(1-v)L\right]} = A\gamma\left[K_d^{\alpha}\left((1-v)L\right)^{1-\alpha}\right]^{\gamma-1} \left[(1-\alpha)K_d^{\alpha}\left((1-v)L\right)^{-\alpha}\right]\left[K_u^{\beta}(vL)^{1-\beta}\right]^{1-\gamma} \tag{3.40}$$

Comparing $\frac{\partial Y}{\partial(vL)}$ with $\frac{\partial Y}{\partial[(1-v)L]}$ is equivalent, after simplification, to comparing $\frac{(1-\gamma)(1-\beta)}{\gamma(1-\alpha)}$ with $\frac{v}{1-v}$. It is easy to see that $\frac{v}{1-v}$ is everywhere increasing in v (in the restricted domain $0 \leq v \leq 1$) and is equal to zero for $v = 0$. By definition also, $\frac{(1-\gamma)(1-\beta)}{\gamma(1-\alpha)} = \frac{v}{1-v}$ for v^{opt}, which is unique. As a result, we have thus that $\frac{(1-\gamma)(1-\beta)}{\gamma(1-\alpha)} > \frac{v}{1-v}$ for any $K_d < K_d^O$, and $\frac{(1-\gamma)(1-\beta)}{\gamma(1-\alpha)} = \frac{v}{1-v}$ for any $K_d \geq K_d^O$. This implies that the wage differential between the universal and domestic sectors is given by

$$\frac{\partial Y}{\partial(vL)} - \frac{\partial Y}{\partial\left[(1-v)L\right]} = \begin{cases} \frac{Y}{L}\left[\frac{(1-\beta)(1-\gamma)}{v} - \frac{(1-\alpha)\gamma}{1-v}\right] & \forall K_d < K_d^O \\ 0 & \forall K_d \geq K_d^O \end{cases} \tag{3.41}$$

where $\frac{(1-\beta)(1-\gamma)}{v} - \frac{(1-\alpha)\gamma}{1-v} > 0$ when $K_d < K_d^O$, which implies a positive wage differential between the universal and domestic sectors in the constrained economy. Let us denote the wage differential by Δ_w. Because v and K_u are functions of K_d, the wage differential is a function of K_d and the parameters of the model.

$$\frac{\partial Y}{\partial(vL)} - \frac{\partial Y}{\partial\left[(1-v)L\right]} = \Delta_w = \Delta_w(K_d) \tag{3.42}$$

The total wage bill in this economy, for any level of v, is given by $\frac{\partial Y}{\partial(vL)}vL + \frac{\partial Y}{\partial[(1-v)L]}(1-v)L$. The average wage $\bar{w}$ is thus equal to

$$\bar{w} = \left[(1-\gamma)(1-\beta) + \gamma(1-\alpha)\right]\frac{Y}{L} \tag{3.43}$$

The average wage is a constant share of output per worker, where the share is the sector-weighted average of the labor shares $(1-\alpha)$ and $(1-\beta)$. It is easy also to see that the wage rates in the domestic and universal sectors can be expressed as,

$$w_{K_d} = \frac{\gamma(1-\alpha)}{1-v}\frac{Y}{L} \tag{3.44}$$

$$w_{K_u} = \frac{(1-\gamma)(1-\beta)}{v}\frac{Y}{L} \qquad (3.45)$$

In the unconstrained economy, where $v = v^{opt}$, it is straightforward to observe that $w_{K_d} = w_{K_u} = \bar{w}$. In the constrained economy where $v < v^{opt}$, the terms $\frac{\gamma(1-\alpha)}{1-v}$ and $\frac{(1-\gamma)(1-\beta)}{v}$ increase and fall, respectively, as the economy develops (as K_d increases). This implies that the wage differential in the constrained economy is a monotonically decreasing function of K_d. This result is in sharp contrast to Kuznets' hypothesis that inequality increases in the first phases of development. We can see here that allowing for interactions between the two sectors of the economy is sufficient to generate a continuously decreasing level of inequality, which eventually converges to zero.

Equation (3.41) represents the absolute wage differential is this economy. We can also derive a relative differential, a more useful measure of inequality. The wage differential as a percentage of wage in the domestic sector (the low-paid sector) is given by,

$$\frac{\Delta_w(K_d)}{w_{K_d}} = \nabla_w(K_d) = \frac{1-v}{v}\frac{(1-\beta)(1-\gamma)}{(1-\alpha)\gamma} - 1 \qquad (3.46)$$

For any given $K_d < K_d^O$, we have thus that the relative wage differential is influenced as follows by the parameters of the model:

- $\frac{\partial \nabla_w}{\partial \gamma} = -\frac{(1-\alpha)(1-\beta)}{[(1-\alpha)\gamma]^2}\frac{1-v}{v} < 0$. The larger the size of the universal sector (the lower γ), the larger the relative wage differential, for any given value of K_d.
- $\frac{\partial \nabla_w}{\partial \alpha} = \frac{\gamma(1-\beta)(1-\gamma)}{[(1-\alpha)\gamma]^2}\frac{1-v}{v} > 0$ and $\frac{\partial \nabla_w}{\partial \beta} = \frac{-\gamma(1-\gamma)(1-\alpha)}{[(1-\alpha)\gamma]^2}\frac{1-v}{v} < 0$. The larger the capital share in the domestic sector, the larger the relative wage differential, and the larger the capital share in the universal sector, the lower the relative wage differential.

The dynamic evolution of the wage differential is similar even if initial conditions are such that equation (3.34) is not respected. If we assume that initial conditions are such that there is "excess" K_u, all investment will initially be directed towards the accumulation of K_d. As long as we assume that $K_d(0) < K_d^O$, we have a positive wage differential. Because all investment is directed to increasing K_d in a first phase, however, the wage differential falls faster than when investment is directed to the accumulation of both types of capital. As soon as equation (3.34) holds again, the speed of convergence of wages returns to the normal case.

The dynamic properties of this constrained economy can be fully described in the (c, K_d) space by using equation (3.37), which allows us to obtain a production function in terms of K_d only. The constrained economy converges towards the optimal path of the unconstrained economy as $K_d \to K_d^O$. Once this point is reached, the steady-state is reached along a path where wages in the two sectors of the economy are equal, and where a constant fraction of total saving is invested in each type of capital. One should note that the constrained economy eventually reaches the same steady-state as the unconstrained economy.

3.3.3 The Unconstrained Open Economy

We now look at an economy that is open to capital flows, as far as universal capital is concerned. To begin with, we return to the case where the economy faces no constraint on the allocation of labor across sectors, v. We effectively consider the unconstrained closed economy and analyze the consequences of allowing full mobility of universal capital. In other words, we consider the effects of opening the economy to world financial markets. This allows us to first analyze the determinants of capital flows, before looking at their effects on wage inequality in a constrained economy in the following sub-section.

Optimal allocation of labor across the two sectors still requires that the marginal products be equalized. As a result, we obtain the same condition on v as in the closed economy:

$$v^{opt} = \frac{(1-\gamma)(1-\beta)}{1-\beta(1-\gamma)-\alpha\gamma} \tag{3.47}$$

While optimization also required the equalization of marginal returns on the two types of capital in the closed economy, full mobility of universal capital now imposes by arbitrage that the net marginal return on universal capital be equal at all times to the world interest rate:

$$
\begin{aligned}
\frac{\partial Y}{\partial K_u} &= A(1-\gamma)\left[K_d^{\alpha}\left((1-v)L\right)^{1-\alpha}\right]^{\gamma}\left[K_u^{\beta}(vL)^{1-\beta}\right]^{-\gamma} \\
&\quad \left[\beta K_u^{\beta-1}(vL)^{1-\beta}\right] \\
&= r_w+\delta
\end{aligned}
\tag{3.48}
$$

By assuming full mobility of universal capital and a small open economy, the variable K_u effectively ceases to be a state variable. Equation (3.48), the arbitrage condition, does not describe a linear relation between universal and domestic capital: the ratio $\frac{K_u}{K_d}$ is no longer a constant as in the unconstrained closed economy. The level of universal capital in the economy, nevertheless, is a function solely of the level of domestic capital, the parameters and the world interest rate.

$$
\begin{aligned}
K_u &= \left[\frac{A\beta(1-\gamma)}{r_w+\delta}\right]^{\frac{1}{1-\beta(1-\gamma)}}\left(v^{opt}L\right)^{\frac{(1-\beta)(1-\gamma)}{1-\beta(1-\gamma)}} \\
&\quad \left[\left(1-v^{opt}\right)L\right]^{\frac{(1-\alpha)\gamma}{1-\beta(1-\gamma)}} K_d^{\frac{\alpha\gamma}{1-\beta(1-\gamma)}} \\
&= \Psi K_d^{\frac{\alpha\gamma}{1-\beta(1-\gamma)}}
\end{aligned}
\tag{3.49}
$$

We can see that Ψ is positive and $1-\beta(1-\gamma)>0$ for all acceptable values of the parameters. From equation (3.49) we can see that, ceteris paribus:

- *an increase in the world interest rate reduces the level of universal capital in the economy.* If we associate changes in the level of universal capital in the economy to capital flows[5], we have thus that an increase in world interest rate would reduce, ceteris paribus, the level of universal capital from its previous arbitraged level and hence generate an outflow of capital.

- *the level of domestic capital in the economy determines the arbitraged level of universal capital.* Even more importantly, *the rate of increase in domestic capital*

[5]We cannot a priori consider changes in the level of universal capital as equal to capital flows, because universal capital can also be accumulated through domestic saving. Capital flows, on the other hand, do translate one for one into changes in the level of universal capital.

determines the rate of increase in the arbitraged level of universal capita, which we still somewhat loosely identify with capital flows. In other words, an economy that accumulates domestic capital rapidly will attract more capital flows than a similar economy whose domestic capital grows more slowly.

- *an increase in population size generates an increase in the level of universal capital.* Similarly, an economy with a rapidly growing population will experience, ceteris paribus, a higher rate of increase in universal capital (more capital flows). The impact of the technology parameter A is identical.

It is again possible to express the production function in terms of K_d only as v^{opt} is a constant and as we ignore population growth. Substituting $K_u = \Psi K_d^{\frac{\alpha\gamma}{1-\beta(1-\gamma)}}$ in the production function, we obtain,

$$Y^{open} = AK_d^{\frac{\alpha\gamma}{1-\beta(1-\gamma)}} \left[\left(\left(1 - v^{opt}\right) L\right)^{1-\alpha}\right]^{\gamma} \left[\Psi^{\beta} \left(v^{opt} L\right)^{1-\beta}\right]^{1-\gamma} = g\left(K_d\right) \qquad (3.50)$$

It is easy to show again that $0 < \frac{\alpha\gamma}{1-\beta(1-\gamma)} < 1$, so that $g'\left(K_d\right) > 0$ and $g''\left(K_d\right) < 0$.

Using the equations relating universal capital to domestic capital in the closed and in the open economy, we can carry out a comparative analysis of the two economies and analyze the impact of opening an economy to capital flows. We have seen that:

$$K_u^{closed} = \frac{\beta\left(1-\gamma\right)}{\alpha\gamma} K_d \qquad (3.51)$$

$$K_u^{open} = \Psi K_d^{\frac{\alpha\gamma}{1-\beta(1-\gamma)}} \qquad (3.52)$$

where Ψ is a positive constant dependent on L and is likely to be large. Also, $\frac{\beta(1-\gamma)}{\alpha\gamma}$ is expected to be small because one should reasonably assume that α is close to β, and γ (the relative size of the domestic sector) is unlikely to be close to zero. It is thus easy to graph the two equations relating K_u to K_d.

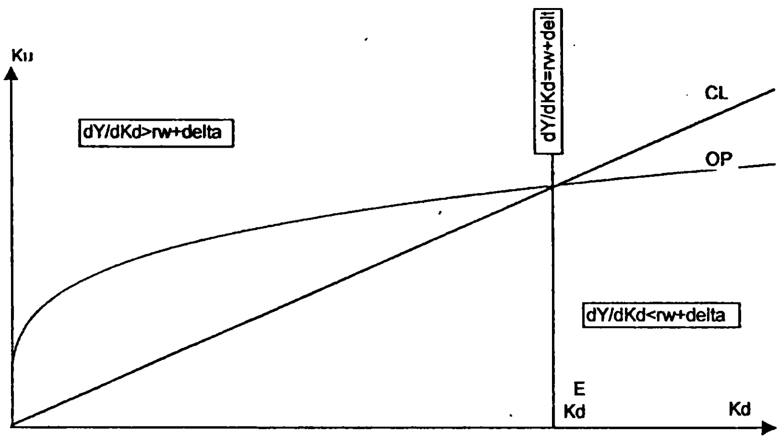


Figure 3.1: $K_u = K_u\left(K_d\right)$ in the open and closed economies

Arbitrage conditions ensure that the closed and the open economies are always on CL and on OP, respectively[6]. At all points of CL, we have that $\frac{\partial Y}{\partial K_d} = \frac{\partial Y}{\partial K_u}$. The OP curve, on the other hand, puts a restriction on the marginal return of universal capital only. At all points of OP, $\frac{\partial Y}{\partial K_u} = r_w + \delta$. The two schedules cross only once because CL is linear while OP is concave and both cross the origin. At that point, the economy respects both arbitrage conditions and $\frac{\partial Y}{\partial K_d} = \frac{\partial Y}{\partial K_u} = r_w + \delta$. This unique level of domestic capital for which both conditions holds can be solved explicitly by using the fact that at that point,

$$\frac{K_d}{K_u} = \frac{\alpha\gamma}{\beta(1-\gamma)} \tag{3.53}$$

$$\frac{\partial Y}{\partial K_d} = r_w + \delta \tag{3.54}$$

$$Y = A\Omega K_d^{\alpha\gamma+\beta(1-\gamma)} \tag{3.55}$$

We have thus that K_d^E, the level at which the net marginal returns on domestic capital and universal capital are both equal to the world interest rate is given by,

[6]The closed economy could be away from CL for some part of the transition process if the initial conditions are such that $\frac{K_d(0)}{K_u(0)} \neq \frac{\alpha\gamma}{\beta(1-\gamma)}$. This is of little interest here, however.

$$K_d^E = \left\{ \frac{A\Omega\left[\alpha\gamma + \beta\left(1-\gamma\right)\right]}{r_w + \delta} \right\}^{\frac{1}{1-\alpha\gamma-\beta(1-\gamma)}} \tag{3.56}$$

Since $1 - \alpha\gamma - \beta(1-\gamma)$ is positive, $\frac{dK_d^E}{dr_w} < 0$. A lower world interest rate implies, ceteris paribus, a higher level of K_d^E, at which CL and OP coincide. This result is quite intuitive, as it simply states that when the world interest rate is low, the level of domestic capital in the economy must be large in order to have the net marginal return to either type of capital equal to the world interest rate. As a consequence of decreasing marginal returns, the net marginal return to K_d (or K_u) in the closed economy is *higher* than the world interest rate for any $K_d < K_d^E$, and *lower* than the world interest rate for any $K_d > K_d^E$.

Consider now the dynamics of a closed economy. Assume that the economy has a steady state at $\left(K_d^{SSA}, K_u^{SSA}\right)$ and that $K_d^{SSA} < K_d^E$. It follows that the autarchy level of interest rate is higher than the world interest rate. Given initial conditions $\left[K_d(0), K_u(0)\right]$ (assumed to be on CL), the economy will move to its steady state along the CL schedule.

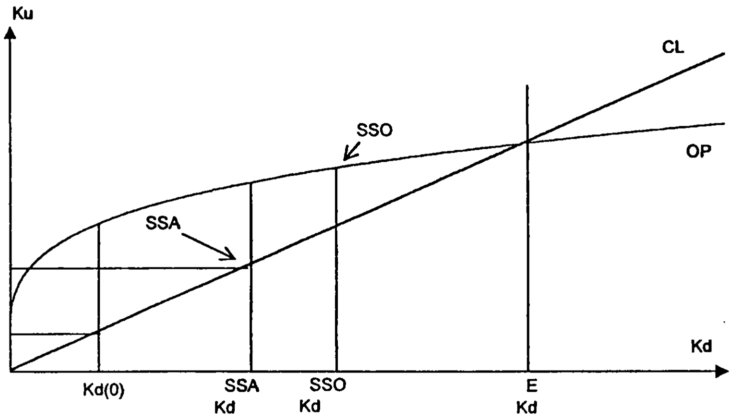


Figure 3.2: Steady-states in open and closed economies

Consider then the impact of opening this economy, starting from the same initial conditions, where $K_d(0) < K_d^E$. At such a point, $\frac{\partial Y}{\partial K_d}_{|K_{d(0)}} = \frac{\partial Y}{\partial K_u}_{|K_{u(0)}} > r_w + \delta$. Opening the economy to foreign capital at such a point results in an immediate inflow of universal capital as a consequence of arbitrage by non residents. The level of domestic capital, on the other hand, remains constrained as a state variable and cannot increase instantly. As a result, the opening of the economy generates an equalization between the net marginal return on universal capital and the world interest rate, and an *increase* in the marginal return on domestic capital. Graphically, the economy instantly moves from a point on the CL schedule to a point on the OP schedule with unchanged domestic capital and higher universal capital.

The economy will now move to its steady state along the OP schedule. One key difference with the closed economy is that the marginal return to domestic capital is higher than the marginal return to universal capital on all points above CL. Because residents have the option to invest in either type of capital, optimization requires the entire pool of domestic saving to be allocated to increases in domestic capital. On all points to the left of K_d^E, the dynamic equation for K_d is thus,

$$\dot{K}_d^{open} = g(K_d) - C - \delta K_d \tag{3.57}$$

while in the closed economy, the dynamic equation for K_d was given by,

$$\dot{K}_d^{closed} = \frac{\alpha\gamma}{\alpha\gamma + \beta(1-\gamma)}\left[f(K_d) - C\right] - \delta K_d \tag{3.58}$$

We know that for any $K_d < K_d^E$, the level of universal capital in the open economy is higher than in the closed economy. Consequently, $g(K_d) > f(K_d)$, and $g'(K_d) > f'(K_d)$. For any $K_d > K_d^E$, on the other hand, the level of universal capital in the closed economy is lower than in the closed economy, and as a result, $g(K_d) < f(K_d)$, and $g'(K_d) < f'(K_d)$. From this, we can conclude that for all points to the left of K_d^E,

$$\dot{K}_d^{open} > \dot{K}_d^{closed} \tag{3.59}$$

$$\left(\frac{\dot{C}}{C}\right)^{open} = \frac{1}{\theta}\left[g'(K_d) - \delta - \rho\right] > \left(\frac{\dot{C}}{C}\right)^{closed} = \frac{1}{\theta}\left[f'(K_d) - \delta - \rho\right] \tag{3.60}$$

The opening of the economy to foreign capital *increases* the speed of convergence for any economy that starts from initial condition $K_d(0) < K_d^E$. The convergence to the steady state is *not* immediate as in the basic open economy Ramsey-Cass-Koopmans model, however. The economy moves progressively towards its steady state, as the entire pool of domestic saving is allocated to increases in domestic capital. Such increases in domestic capital, in turn, generate increases in universal capital, which, in this case, will be entirely financed by inflows from abroad, which means that we can equate changes in universal capital to foreign capital flows. All along the transition path to steady state, we also see that output in the open economy is higher than output in the closed economy (for a given level of K_d).

In the graph above, we assumed that the autarchy steady-state interest rate was above the world interest rate, i.e. that the home economy is more impatient than the rest of the world. In such a case, the unconstrained open economy reaches a new steady-state with $K_d^{SSO} > K_d^{SSA}$. It is not necessarily the case, however, that $K_d^{SSO} = K_d^E$, which means that there could coexist two interest rates in the steady-state economy, with the return on domestic capital being permanently higher than the return on world markets.

The reverse situation, where the home economy is more patient than the rest of the world can also be envisaged. In such a case, opening the economy actually implies an immediate outflow of universal capital and a gradual fall in domestic capital through zero investment and depreciation. The open-economy steady-state would thus imply lower levels of capital.

If we rule out differences in time preferences between home agents and foreign agents, on the other hand, we have that $K_d^{SSO} = K_d^E$. In such a case, opening the economy to foreign capital does not affect steady-state values of capital and output, and domestic and world interest rates are equalized, but convergence to steady-state is faster.

A few key results emerge from this analysis. We have seen that convergence is faster in the open economy than in the closed economy, but that it is not immediate, as predicted by the simple extension of the Solow model. While an immediate inflow of capital follows the opening of the economy, the accumulation of domestic capital is what determines subsequent inflows. That is, domestic investment (i.e. domestic saving) is key in attracting foreign flows: capital attracts capital, "rich" countries may attract more capital flows than "poor" countries. The factors that determine the speed of convergence are thus essentially domestic, even in the open economy setting.

3.3.4 The Constrained Open Economy

In this section, we analyze the effect of allowing perfect international mobility of universal capital when the economy faces a constraint on v, i.e. when $K_d < K_d^O$. We shall also assume that $K_d^O < K_d^E$ always holds, which means that the economy is constrained initially, but that it reaches the optimal convergence path before it reaches steady-state. The corollary to this is that the economy is always fully optimizing in steady-state.

Opening the constrained economy to flows of universal capital implies by arbitrage that the net marginal return to universal capital is always equal to the world interest rate:

$$A\left(1-\gamma\right)\left[K_d^{\alpha}\left(\left(1-v\right)L\right)^{1-\alpha}\right]^{\gamma}\left[K_u^{\beta}\left(vL\right)^{1-\beta}\right]^{-\gamma}\left[\beta K_u^{\beta-1}\left(vL\right)^{1-\beta}\right]=r_w+\delta \tag{3.61}$$

We can express K_u as a function of K_d as in equation (3.49), with the difference that v is no longer a constant:

$$\begin{aligned} K_u &= \left[\frac{A\beta\left(1-\gamma\right)}{r_w+\delta}\right]^{\frac{1}{1-\beta(1-\gamma)}}\left(vL\right)^{\frac{(1-\beta)(1-\gamma)}{1-\beta(1-\gamma)}} \\ &\quad \left[\left(1-v\right)L\right]^{\frac{(1-\alpha)\gamma}{1-\beta(1-\gamma)}}K_d^{\frac{\alpha\gamma}{1-\beta(1-\gamma)}} \\ &= \Psi\left(v\right)K_d^{\frac{\alpha\gamma}{1-\beta(1-\gamma)}} \end{aligned} \tag{3.62}$$

where $\Psi\left(v\right)$ is an increasing function of v between $\left[0,v^{opt}\right]$, which means that $\Psi\left(v\right) < \Psi\left(v^{opt}\right)\ \forall\ v < v^{opt}$.

From this expression, we see that the arbitraged level of universal capital in the economy is constrained directly by the complementarity between the domestic and the universal sectors, but also by the share of the population that is able to work in the universal sector.

Because $\Psi(v) < \Psi(v^{opt}) \ \forall\ v < v^{opt}$, the level of universal capital in the constrained open economy is lower than in the unconstrained open economy $\forall\ K_d < K_d^O$. Once the economy reaches the point where $K_d \geq K_d^O$, the constrained open economy rejoins the convergence path of the unconstrained open economy. The behavior of capital flows is thus equivalent in the two cases, with the difference that the constraint on labor mobility does not allow for as large an initial increase in universal capital. Also, capital inflows are smaller in the constrained open economy as long as the level of domestic capital is not high enough to allow for the optimal share of the labor force to work in the universal sector.

Substituting the expression for K_u in the production function, we obtain that,

$$Y^{open/constrained} = \begin{cases} j(K_d) & \forall K_d < K_d^O \\ g(K_d) & \forall K_d \geq K_d^O \end{cases} \tag{3.63}$$

where $j(K_d) < g(K_d)$ for $K_d < K_d^O$ and $j(K_d) = g(K_d)$ for $K_d \geq K_d^O$. The only difference between $j(K_d)$ and $g(K_d)$ is that the term $\Psi(v)$ is a constant in $g(K_d)$ and is an increasing function in $j(K_d)$. The relation between K_d and K_u in the four cases analyzed is represented graphically as follows.

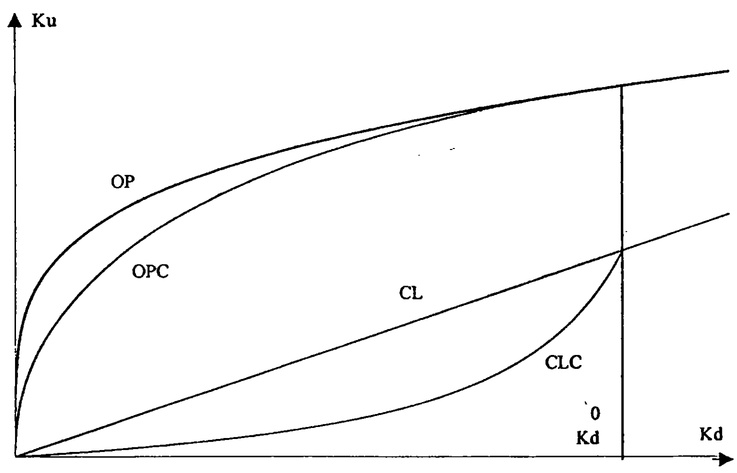


Figure 3.3: Convergence paths in open and closed economies

The labor mobility constraint added in this section means that the level of universal capital in the economy, for any given level of domestic capital, is lower than in the unconstrained open economy. If we assume that $K_d(0) < K_d^O < K_d^E$, opening the constrained closed economy implies by arbitrage an initial jump in the level of universal capital in the economy, like in the unconstrained case. This jump is limited, however, by the availability of labor in the universal sector. Because of our assumptions, all domestic saving is devoted to the accumulation of domestic capital once the economy is opened to capital flows. This allows the constrained open economy to converge towards steady state along the OPC schedule, instead of the CLC schedule. As in the unconstrained case, opening the economy allows a faster process of convergence, but does not affect the steady state if we assume that domestic agents have the same preferences as foreign agents. Also, because we assumed that $K_d^O < K_d^E$, the constrained economy reaches the same convergence path as the unconstrained economy before it reaches steady state.

The main effect of opening the economy on the pattern of convergence is thus essentially the same in the constrained economy as in the unconstrained economy. The constrained economy, however, benefits even more from the opening to capital flows. The process of convergence is indeed doubly affected as being able to devote the entire pool of saving to the accumulation of domestic capital allows not only a faster accumulation of domestic capital itself, but it also allows the economy to reach an optimal allocation of the labor force between sectors at an earlier stage.

The impact of opening the constrained economy on the wage differential is easy to determine. From equation (3.41), we see that the absolute wage differential actually increases at the time the economy is opened to inflows of universal capital. This is obvious from the expression, as output is increased as a consequence of the initial jump in universal capital and v is initially unaffected. This means that in absolute term, the "rich" see their wage increase by a larger amount than the poor, which may be undesirable, either from a social justice perspective, or even from a growth perspective. One should note, however, that the "poor" also see their absolute wage increase, albeit by a smaller amount. It is thus not a case of one group gaining at the expense of the other: opening the economy has favorable effects on both groups.

It is also easy to see from equation (3.46) that there is no initial effect on the relative wage differential. This should be obvious as opening the economy has no immediate effect on v. In fact, the larger absolute wage increase for the "rich" is such that the wage of the poor remains initially at the same percentage of the "rich" wage. This result is rather counter-intuitive as we would expect that a higher level of universal capital, for a given labor force operating in the sector and for a given level of domestic capital, increases the marginal productivity of labor in the universal sector relatively more than in the domestic sector. This is not the case here however, because the two sectors are complementary in "output". This complementarity allows both groups to gain as much in relative terms from the opening of the economy to capital flows.

While opening the economy has no immediate effect on the relative wage distribution in the economy, the "poor" benefit doubly. First, their absolute wage is increased. Second, and probably more importantly, the opening of the economy to inflows of universal capital allows a faster increase in v, which means that the wage differential is reduced more rapidly. In other words, opening the economy allows to reach a more even (exactly even in this simplified case) distribution of income faster.

This simple modeling tool thus shows that relative inequality is not immediately affected by the opening of the economy to capital flows. In fact, the poor gain in absolute terms as well as the rich. Beyond the immediate impact, capital flows allow inequality to fall faster. The poor benefit not only from the economy converging faster to a higher level of output, but also from the wage differential with the rich falling more rapidly.

3.4 Conclusion

A straightforward extension of the Solow-Cass-Koopmans growth model allowed us to obtain several new results on inequality and development and to reach important policy conclusions. In the first place, elaborating on Kuznets' perception of a dual economy to allow for interactions between sectors shows that inequality is likely to decline monotonically with development, instead of exhibiting an inverted U curve behavior. Importantly, we show that inequality can arise in a closed economy if labor is not fully mobile, even though capital can move freely between the two sectors. Capital mobility is thus not sufficient to ensure the equalization of marginal returns to labor.

A second important result that emerges from the model is that convergence is not immediate in the open-economy extension of the Solow-Cass-Koopmans growth model. The speed of convergence is driven mostly by domestic factors, including time preferences and the rate of domestic saving. The size and timing of capital flows from abroad, aside from being affected by the world interest rate, are thus determined by domestic factors.

Given the complementarity between the two sectors of the economy, we have seen that countries with a well developed domestic sector (capital rich countries) are those that are bound to attract large capital flows from abroad. Similarly, countries that accumulate domestic capital at a fast pace, i.e. countries with high domestic investment (high domestic saving), are those that attract large capital flows. In other words, capital attracts capital: one only lends to the rich.

This may explain why, contrary to the simple prediction of the standard open economy growth model, "capital rich" countries in South-East Asia, for example, have consistently attracted more capital flows from abroad than "capital poor" and labor abundant economies elsewhere in Asia, Africa or Latin America. This highlights the importance of domestic investment and domestic saving in a country's process of convergence and questions the unrealistic hopes some put into free movements of capital flows to boost investment and raise the speed of convergence. Capital flows certainly contribute to faster convergence, but their contribution is limited and determined by domestic factors.

The model may thus shed some additional light on the Feldstein-Horioka puzzle. One would in fact expect, particularly if γ (the relative size of the domestic sector) is large, total investment to be significantly correlated with domestic saving, even if there is perfect mobility of universal capital. While capital flows allow the economy to invest more than it saves in aggregate, these flows are in fact determined by the extent to which domestic capital grows, i.e. by the size of domestic saving. Aggregate investment, and even investment in universal capital alone (or capital flows) are thus bound to be significantly correlated with domestic investment.

The main policy conclusion of this chapter is that, just like Dollar and Kraay (2000) stated that "growth is good for the poor", we affirm that "capital flows are good for the poor". Capital flows are good in several ways. To begin with, we saw that opening the economy does not increase relative inequality. While the rich –operating in the universal sector and with higher wages– gain more in absolute terms than the poor –operating in the domestic sector– because capital flows are concentrated on the universal sector, the poor are not worse off in relative terms. Most importantly, they gain in absolute terms from the inflow of capital into the universal sector as a consequence of the complementarity between sectors. In other words, the poor also benefit from foreign investment in electricity generation or water supply systems, not only as a result of better services (not modeled here), but because it makes them more productive.

More importantly still, the model shows that, aside from the immediate gain from opening the economy, the poor stand to gain doubly from capital flows. On the one hand, they benefit because capital flows allow a faster reduction in wage inequality between the two sectors. On the other hand, they benefit because capital flows allow a faster convergence of the economy as a whole toward its (wealthier) steady-state. While they may not appear as the main beneficiaries of opening the economy because they do not operate in the sector receiving the flows of capital from abroad, the poor are thus the ones that stand to gain the most from free movement of capital.

Bibliography

[1] Aghion, Philippe and Peter Howitt (1998). Endogenous Growth Theory. Cambridge, MA, MIT Press.

[2] Alesina, Alberto and Dani Rodrick (1994). "Distributive Politics and Economic Growth." Quarterly Journal of Economics, vol. 109, 2, pp. 465-490.

[3] Barro, Robert J. (1990). "Government Spending in a Simple Model of Endogenous Growth." Journal of Political Economy, 1990, 98(5), pp. S103-S125.

[4] Barro, Robert J. (1998). "Notes on Growth Accounting." NBER Working Paper, No 6654, July 1998.

[5] Barro, Robert J., N. Gregory Mankiw, and Xavier Sala-i-Martin (1995). "Capital Mobility in Neoclassical Models of Growth." The American Economic Review, March 1995, 85(1), pp. 103-115.

[6] Barro, Robert J. and Xavier Sala-i-Martin (1992a). "Convergence." Journal of Political Economy, April 1992, 100(2), pp. 223-251.

[7] Barro, Robert J. and Xavier Sala-i-Martin (1992b). "Public Finance in Models of Economic Growth." Review of Economic Studes, October 1992, 59(4), pp. 645-661.

[8] Barro, Robert J. and Xavier Sala-i-Martin (1995). Economic Growth. New York, McGraw-Hill.

[9] Bénabou, Roland (1996a). "Unequal Societies." C.V. Starr Center for Applied Economics, Economic Research Reports, May 1996, 96-17.

[10] Bénabou, Roland (1996b). "Inequality and Growth." C.V. Starr Center for Applied Economics, Economic Research Reports, July 1996, 96-22.

[11] Bruno, Michael, M. Ravallion and Lyn Squire (1998). "Equity and Growth in Developing Countries: Old and New Perspectives on Policy Issues." World Bank Policy Research Working Papers, No. 1563.

[12] Clarida, Richard H. (1993). "International Capital Mobility, Public Investment and Economic Growth." NBER Working Paper, No 4506, October 1993.

[13] Cohen, Daniel and Jeffrey Sachs (1986). "Growth and External Debt under Risk of Debt Repudiation." European Economic Review, June 1986, 30(3), pp. 526-560.

[14] Deininger, Klaus and Lyn Squire (1998). "New Ways of Looking at Old Issues: Inequality and Growth." Journal of Development Economics, vol. 57, 2, pp. 259-287.

[15] Dollar, David and Aart Kraay (2000). "Growth *Is* Good for the Poor." World Bank, Development Research Group Working Paper, March.

[16] Engel, Charles and Kenneth Kletzer (1989). "Saving and Investment in an Open Economy with Non-traded Goods." International Economic Review, November 1989, 30(4), pp.735-752.

[17] Feldstein, Martin and Charles Horioka (1980). "Domestic Saving and International Capital Flows." The Economic Journal, June 1980, 90, pp. 314-328.

[18] Fields, G. (1994). "Data for Measuring Poverty and Inequality Changes in the Developing Countries." Journal of Development Economics, vol. 44, pp. 87-102.

[19] Fukuyama, Francis (1995). Trust. The Social Virtues and the Creation of Prosperity. The Free Press.

[20] Galor, Oded and Joseph Zeira (1993). "Income Distribution and Macroeconomics." Review of Economic Studies, vol. 60, 1, pp. 35-52.

[21] Harrison, Ann (1996). "Openness and Growth: a Time-Series Cross-Country Analysis for Developing Countries." Journal of Development Economics, 1996, 48, pp. 419-447.

[22] King, Robert G. and Sergio Rebelo (1990). "Public Policy and Economic Growth: Developing Neoclassical Implications." Journal of Political Economy, 1990, 98(5), pp. S126-S150.

[23] Knack, Stephen and Philip Keefer (1997). "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." Quarterly Journal of Economics, November 1997, pp. 1251-1288.

[24] Kuznets, Simon (1955). "Economic Growth and Income Inequality." The American Economic Review, March 1955, XLV (1), pp. 1-28.

[25] Lucas, Robert E. Jr. (1988). "On the Mechanics of Economic Development." Journal of Monetary Economics, 22 (1988), pp. 3-42.

[26] Lucas, Robert E. Jr. (1990). "Why Doesn't Capital Flow from Rich to Poor Countries?" AEA Papers and Proceedings, May 1990, 80(2), pp.92-96.

[27] Mankiw, N. Gregory, David Romer, and David N. Weil (1992). "A Contribution to the Empirics of Economic Growth." The Quarterly Journal of Economics, May 1992, pp. 407-437.

[28] Mirrlees, James A. (1971). "An Exploration in the Theory of Optimum Income Taxation." Review of Economic Studies, vol. 38, 2, pp. 175-208.

[29] North, Douglass C. (1989). "Institutions and Economic Growth: an Historical Introduction." World Development, 1989, 17(9), pp. 1319-1332.

[30] Persson, Torsten and Guido Tabellini (1994). "Is Inequality Harmful for Growth?". The American Economic Review, June 1994, 84(3), pp. 600-621.

[31] Quah, Danny T. (2001). "Some Simple Arithmetic on How Income Inequality and Economic Growth Matter." Mimeo, The London School of Economics.

[32] Rebelo, Sergio (1991). "Long-run Policy Analysis and Long-run Growth." Journal of Political Economy, 1991, 99(3), pp. 500-521.

[33] Robinson, Sherman (1976). "A Note on the U-Hypothesis Relating Income Inequality and Economic Development." American Economic Review, vol. 66, 3, pp. 437-440.

[34] Romer, Paul M. (1986). "Increasing Returns and Long-run Growth." Journal of Political Economy, 1986, 94(5), pp.1002-1037.

[35] Romer, Paul M. (1990). "Endogenous Technological Change." Journal of Political Economy, 1990, 98(5), pp. S71-S102.

[36] Solow, Robert M. (1956). "A Contribution to the Theory of Economic Growth." The Quarterly Journal of Economics, February 1956, LXX, pp. 65-94.

[37] Ventura, Jaume (1997). "Growth and Interdependence." The Quarterly Journal of Economics, February 1997, pp. 57-84.

Chapter 4

Intersectoral Labor Mobility, Equilibrium Inequality and Growth

4.1 Introduction

From Karl Marx to Ross Perot, people have feared that structural economic changes, whether due to capital accumulation, technological progress, comparative advantages or other causes would lead to an absolute loss in employment. Yet, Jacquard's loom, the steam engine, chain work, the automation and computerization of all aspects of -economic- life or NAFTA have not generated massive net job destruction, and US unemployment is at its lowest at a time of rapid technological change. What has indeed happened, however, is that technological changes have caused a large, persistent and continuous reallocation of labor across sectors: new technologies have destroyed many of the early 20^{th} century jobs, but have created many more new ones in a process of creative destruction.

The explanation of economic growth through purposeful R&D and creative destruction can be traced back at least to Schumpeter (1934). More recently, Aghion and Howitt (1992, 1998) and Grossman and Helpman (1991a,b) pioneered the formalization of Schumpeter's ideas in models of endogenous growth. Creative destruction, as presented in these models, can be assimilated to a purposeful Darwinian evolutionary process of survival of the fittest, where profit-driven agents seek the next evolutionary step. R&D allows companies to produce more efficiently and drive out competitors.

This process of creative destruction has been extensively studied as far as monopoly positions are concerned. Typically, a discovery allows a producer to enjoy a monopoly over a certain production process superior to that of all its competitors. This monopolistic position is later destroyed by somebody else's discovery, which generates a yet superior production process.

The Schumpeterian growth literature has so far failed to analyze in detail the process of creative destruction as it applies to jobs, however. It has also little investigated the fact that new technologies require the labor force to adapt before they can be used effectively, and that this "adaptation process" may have an important impact on growth.

A very different strand of the literature has explored, mostly empirically, the effect on growth of the reallocation of labor across sectors. Kuznets (1955) presented some arguments as to how the reallocation of labor from agriculture to industry may affect growth and inequality. Denison (1967) presented a more rigorous empirical analysis of the growth effect of labor reallocation from agriculture to industry in some western countries. Other studies, including Robinson (1971) and Poirson (2000), have further investigated this issue.

This entire strand of the literature, however, has focused on the reallocation of labor from agriculture to industry as a transitory phenomenon. Also, it usually merely postulates that labor productivity is higher in industry than agriculture, and fails to consider general equilibrium effects of labor reallocation, in particular the effect on relative productivities in each sector.

This chapter proposes a refined version of the Schumpeterian model of endogenous growth that allows not only for creative destruction in monopolistic positions, but that also considers the causes and effects of creative destruction in jobs. Introducing a Darwinian process on jobs, we analyze cross-sectoral labor reallocation phenomena and their effects on growth and inequality. This allows us to propose a more satisfying approach to the "labor reallocation effect" literature in that the reallocation process becomes an endogenous equilibrium phenomenon and not just a transitory one, in that we allow for a multiplicity of sectors, and in that the source of labor productivity differentials is clearly explained.

Introducing frictions in the labor reallocation process, our approach also allows us, critically, to analyze the interactions between labor market rigidities, inequality and growth. New light is shed on the issue of inequality and growth. In particular, the model explicitly generates Lorenz curves and Gini coefficients as a function of a few core parameters.

4.2 The Model

4.2.1 Basic Structure

The core structure of the model is standard and builds on Aghion and Howitt (1992, 1998). Let the economy be populated by a continuous and constant mass L of agents (normalized to 1) each endowed with one flow unit of labor, and where work does not provide disutility. For simplicity, we assume that agents have linear intertemporal preferences:

$$U(c) = \int_0^\infty c_s e^{-rs} ds \tag{4.1}$$

which implies that r represents both the rate of time preference and the interest rate.

Output of the homogenous final good Y is produced by a continuum of sectors indexed over $[0,1]$, each of which uses an intermediate good X_i in association with a mass L_i of the total labor force and a public good G_t that is subject to congestion effects. The production function for firm j in sector i at time t takes the form

$$Y_{ijt} = \tilde{A}_{it} X_{ijt}^{\alpha} L_{ijt}^{\beta} \left[\frac{G_t}{f(n)} \right]^{1-\alpha-\beta} \tag{4.2}$$

where $\tilde{A}_{it}$ is the index summarizing technology of sector i at time t, n is the total number of firms producing the final good, and where $\alpha + \beta < 1$. It is possible to show that, under certain conditions, there is a unique, constant and strictly larger than 1 number of firms that maximizes aggregate output in any sector i[1]. The production function for sector i can then be rewritten as

$$Y_{it} = A_{it} X_{it}^{\alpha} L_{it}^{\beta} \tag{4.3}$$

Assuming a continuum of sectors over $[0, 1]$ implies that

$$Y_t = \int_0^1 Y_{it} di \tag{4.4}$$

$$L = 1 = \int_0^1 L_{it} di \tag{4.5}$$

The key element of the production function is that labor is introduced as a complementary input to the intermediate good X_i, and that each sector employs only a partial mass L_i of the total labor force. This divides the economy into independent sectors and captures the idea that workers, being a rival good, operate exclusively in one sector. Any worker is assumed to work with only one intermediate good, which is not fully realistic, but which better captures reality than the postulate that the entire labor force operates with the whole range of intermediate inputs, which is what is assumed in models where the production function takes the form $Y_t = L^{1-\alpha} \int_0^1 A_{it} X_{it}^{\alpha} di$, as in Barro and Sala-i-Martin (1995), or that workers are not used directly in the production of the final good, which is what Aghion and Howitt (1998) assume by using the production function $Y_{it} = A_{it} X_{it}^{\alpha}$.

Having decreasing returns to scale in X and L (i.e. $\alpha + \beta < 1$) is essential to ensure that all sectors are used to produce the final good in equilibrium. This will be demonstrated later as a more complete justification of the assumption is provided.

The rest of the setting is traditional: final good producers operate in a competitive

[1] See appendix 4A

market, while the production of intermediate goods is subject to monopoly. Technological progress takes place at the level of the intermediate goods and is the result of purposeful R&D. The research effort is remunerated by a monopolistic position over the production of intermediate good X_i of a given vintage. Vintages are differentiated by the technology parameter A_{it}, with more recent vintages having a higher A_{it}. We assume that all intermediate goods, regardless of sector and vintage, are produced at a unit cost in terms of final output Y.

Optimization by final good producers requires that

$$\begin{aligned} \frac{\partial Y_{it}}{\partial X_{it}} &= \alpha A_{it} X_{it}^{\alpha-1} L_{it}^{\beta} = P_{it} && (4.6) \\ \frac{\partial Y_{it}}{\partial L_{it}} &= \beta A_{it} X_{it}^{\alpha} L_{it}^{\beta-1} = w_{it} && (4.7) \end{aligned}$$

where P_{it} is the price of intermediate good i at time t and w_{it} is the wage rate in sector i at time t.

Monopolists maximize profit: $\max\limits_{X_{it}} \left[P\left(X_{it}\right) - 1\right] X_{it}$. The equilibrium level of intermediate good is thus given by

$$X_{it} = \left[\alpha^2 A_{it}\right]^{\frac{1}{1-\alpha}} L_{it}^{\frac{\beta}{1-\alpha}} \tag{4.8}$$

Given the demand function of final good producers, intermediate goods are all priced at a constant mark-up over marginal cost (itself equal to one for all sectors and at all times):

$$P_{it} = \frac{1}{\alpha} \tag{4.9}$$

4.2.2 R&D Process and Discoveries

The process of R&D and discovery is modeled in a standard way. There exists a distinct competitive R&D sector for each intermediate good. Discoveries follow a path-independent Poisson process with arrival rate $\lambda \frac{R_{it}}{\eta A_t^{Max \frac{1}{1-\alpha}}}$, where R_{it} is the amount of resources (in units of the final good) devoted to research in sector i at time t, A_t^{Max} is the highest level of technology attained in the economy (in any sector) and $0 < \lambda < 1$ and $0 < \eta < 1$ are parameters. An important assumption is that the probability of discovery in a sector does not depend on the time elapsed since the last discovery, but only on the resources invested in R&D. This implies, together with the earlier assumptions on the sectoral production function, full symmetry across sectors.

Arrival rates across sectors, as explicited above, are fully independent of each other. We wish to allow for the presence of knowledge spillovers, however, meaning that discovery in a sector affects other sectors by contributing to a global pool of knowledge.

The state of the global pool of knowledge is captured by A_t^{Max}, the cutting-edge technology in the economy, i.e. the highest A_i at time t. Knowledge spillover is modeled by assuming that a discovery in any sector i at time t allows the innovator to start producing with technology parameter A_t^{Max}. Absent subsequent discovery, the technology remains constant thereafter, until a new discovery occurs at time $t+s$, in which case the technology parameter jumps discretely from A_t^{Max} to A_{t+s}^{Max}.

For simplicity, we assume that the parameters of the model are such that innovations are drastic. In other words, innovations are such that an incumbent monopolist is always displaced when a discovery occurs in its sector, and the new monopolist applies full monopoly pricing.

The cutting-edge parameter A_t^{Max} itself is assumed to evolve proportionately to the aggregate level of resources devoted to R&D. Let $R_t = \int_0^1 R_{it} di$ represent the aggregate level of research in the economy. We assume that

$$\frac{\dot{A}_t^{Max}}{A_t^{Max}} = \lambda \frac{R_t}{\eta {A_t^{Max}}^{\frac{1}{1-\alpha}}} \ln \gamma, \qquad \gamma > 1 \tag{4.10}$$

Two features of the discovery process need to be underlined. Firstly, $\ln \gamma$ must be thought of as a "size of innovations" parameter. It is a simple extension of the framework where discovery is assumed to make the technology parameter grow by a proportionality factor γ. Say, for example, that there is only one sector, and that discovery means that $A_t = \gamma A_{t-1}$. In such a case, the technology parameter follows a step function with expected growth rate proportional to $\ln \gamma$. In the current setting, the presence of a continuum of sectors in $[0, 1]$ ensures that the actual cutting-edge parameter grows smoothly, even though individual sectors' technology parameters follow step functions.

Secondly, we have assumed that the Poisson arrival rate depends on $\frac{R_{it}}{\eta A_t^{Max\frac{1}{1-\alpha}}}$. An increasing absolute amount of R&D effort is thus needed to keep the sectoral arrival rate (and the growth rate of the cutting-edge technology) constant. This assumption aims to capture the empirical observation that increasing resources are invested into R&D without noticeable accelerating effects on the speed of technological progress. It is also necessary to ensure the existence of a balanced growth path.

Technology parameters are distributed between 0 and A_t^{Max} at any given time t. Innovations imply that the distribution shifts rightward over time. Let us define the relative technology parameter in sector i at time t as,

$$a_{it} = \frac{A_{it}}{A_t^{Max}} \tag{4.11}$$

Relative technology parameters are distributed over $[0, 1]$ and decrease continuously as long as no discovery occurs in a sector. If a discovery occurs, the technology parameter in that sector discretely jumps to A_t^{Max}, taking the relative parameter to 1.

Having showed that the domain of relative parameters is constant, one can show that their distribution is also constant in the long run, the movement of one sector along the distribution always being compensated by the movement of other sectors. This can be proved by letting $F(A, t)$ be the cumulative distribution function (cdf) of absolute technology parameters at any given time t. Noting that $F\left(A_0^{Max}, 0\right) = 1$ and that the density to the left of A_0^{Max} falls at a rate equal to $\frac{dF\left(A_0^{Max}, t\right)}{dt} = -F\left(A_0^{Max}, t\right)\Psi$, where Ψ is the Poisson arrival rate, we obtain a differential equation with given initial condition. Solving the differential equation and using the fact that $\frac{dA_t^{Max}}{dt} = A_t^{Max}\lambda\frac{R_t}{\eta A_t^{Max\frac{1}{1-\alpha}}}\ln\gamma$, we obtain that the cdf of relative technology parameters is constant and given by[2],

[2] A more detailed proof can be found in appendix 4B and in Aghion and Howitt (1998), pp. 115-6.

$$H(a) = a^{\frac{1}{\ln \gamma}} \tag{4.12}$$

This property holds regardless of the initial condition on the distribution of the technology parameters. Although it holds only in the long run, we assume for simplicity that it also holds at time 0. One can see from equation (4.12) that the shape of the distribution of relative parameters depends exclusively on $\ln \gamma$, the size of innovations parameter. This parameter can be calibrated to fit some priors one may have on the structure of the economy. One could sensibly assume that a developing economy, say with larger potential for adaptations of existing foreign technologies, has a large γ, while a more developed economy has less potential for very large innovations. Graphically, the distribution function is then represented as

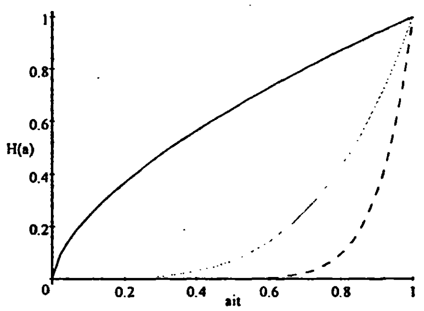


Figure 4.1: $H(a) = a^{\frac{1}{\ln \gamma}}$ with $\gamma = 1.1$ (dots), $\gamma = 1.3$, and $\gamma = 5$ (solid)

Large innovations generate a situation where sectors with low technology coexist with sectors with high technology to a much larger extent than if innovations are smaller. This may indeed fit the description of developing countries where modern high technology sectors coexist with traditional sectors to a larger extent than in industrialized countries. Although we will consider γ as a fixed parameter throughout the model, the framework will allow us to consider what could happen if γ was allowed to change over time. Even though it would be interesting to carry out, endogenizing γ will be left for further research.

4.2.3 Flexible Allocation of Labor Across Sectors

At this stage, we need to make assumptions about the mobility of labor across sectors. As a benchmark, we first analyze an economy where labor is fully and instantaneously mobile across sectors. At a later stage, we introduce rigidities in the allocation of labor, which allows us to derive the main results on equilibrium inequality and growth.

Sectoral Allocation of Labor

If labor is fully and instantaneously mobile across sectors, its marginal product is necessarily equal across sectors and there is a single wage rate in the economy:

$$\frac{\partial Y_{it}}{\partial L_{it}} = w_t \qquad \forall i, t \tag{4.13}$$

The ratio of the labor force allocated to any two sectors is thus given by $\frac{L_{jt}}{L_{it}} = \left[\frac{A_{jt}}{A_{it}}\right]^{\frac{1}{1-\beta}} \left[\frac{X_{jt}}{X_{it}}\right]^{\frac{\alpha}{1-\beta}}$. Using the equilibrium condition for X_{it}, this ratio is equal to

$$\frac{L_{jt}}{L_{it}} = \left[\frac{A_{jt}}{A_{it}}\right]^{\frac{1}{1-\alpha-\beta}} \tag{4.14}$$

where $\frac{1}{1-\alpha-\beta} > 1$ by assumption.

Just as we defined the relative technology parameter a_{it}, we can define the relative labor share as the ratio of the labor force in sector i at time t to the labor force in the sector with the highest technology parameter, L_t^{Max}. Let

$$l_{it} = \frac{L_{it}}{L_t^{Max}} \tag{4.15}$$

L_t^{Max} is unknown but constant at time t. The domain of the relative labor share is also $[0, 1]$. Because equation (4.14) holds for any two sectors, it necessarily holds if sector i is the highest technology sector at time t. This implies that $l_{it} = a_{it}^{\frac{1}{1-\alpha-\beta}}$, i.e. that the relative labor share of any sector is a function of the relative technology of that sector. Because l_{it} is a continuous transformation of the random variable a_{it} with cdf given by equation (4.12), we can derive the cdf of l_{it} as follows:

$$H(l) = l^{\frac{1-\alpha-\beta}{\ln\gamma}} \tag{4.16}$$

While A_t^{Max} grows constantly, L_t^{Max} cannot grow without bounds. In particular, we assumed that the total labor force L is constant: $L = 1 = \int_0^1 L_{it} di$. At any given time t, L_t^{Max} is a constant. It follows then that $\frac{1}{L_t^{Max}} = \frac{1}{L_t^{Max}} \int_0^1 L_{it} di$. This means that $L_t^{Max} = \frac{1}{\int_0^1 l_{it} h(l) dl}$, where $h(l) = \frac{\partial H(l)}{\partial l}$ is the probability distribution function (pdf) of l. Evaluating the integral we find that

$$L_t^{Max} = \frac{1-\alpha-\beta+\ln\gamma}{1-\alpha-\beta} = L^{Max} \tag{4.17}$$

The amount of labor allocated to the sector with the highest technology parameter is thus a constant function of the parameters. In particular, one should note that it does *not* depend on the level of technology A_t^{Max}, and that it is time invariant. Using $l_{it} = \frac{L_{it}}{L_t^{Max}}$ and equation (4.17), the absolute amount of labor allocated to any sector is uniquely determined as a function of its relative technology:

$$L_{it} = a_{it}^{\frac{1}{1-\alpha-\beta}} \left[\frac{1-\alpha-\beta+\ln\gamma}{1-\alpha-\beta} \right] \tag{4.18}$$

Because L_{it} is a continuous transformation of the random variable a_{it}, the cdf of the sectoral allocation of labor is given by

$$H(L) = \left[L_{it} \left(\frac{1-\alpha-\beta}{1-\alpha-\beta+\ln\gamma} \right) \right]^{\frac{1-\alpha-\beta}{\ln\gamma}} \tag{4.19}$$

The distribution is constant over time and driven by the distribution of the relative technology a and the parameters of the model. While the shape of the distribution is constant over time, there is perpetual movement of sectors along the distribution, driven by the movement of sectors along the relative technology distribution as a result of R&D and random discoveries. This obviously means that there is a continuous movement of the labor force across sectors as well.

Two results on the distribution of labor across sectors need to be underlined at this stage. First, the amount of labor allocated to the sector with the highest technology, L^{Max}, is an increasing function of the "degree" of decreasing returns to scale of the production function. Defining $\xi = \alpha + \beta$ as the degree of decreasing returns, we see that $\frac{\partial L^{Max}}{\partial \xi} = \frac{\ln \gamma}{(1-\xi)^2} > 0$. The farther one moves from constant returns to scale (the smaller ξ), the smaller the amount of labor allocated to the highest technology sector. In other words, if there are large decreasing returns to scale, the total mass of labor is allocated in a more widespread manner across sectors, and vice versa. In the extreme case where $\alpha + \beta = 1$, the entire mass of labor is allocated to the one sector with the highest technology. Because this sector changes permanently, this would imply that only one sector (but a different one at each time) is used in equilibrium.

This is obviously not a desirable outcome, which is why we impose that $\alpha + \beta < 1$. The assumption can be justified by the need to have all sectors used in equilibrium. One could avoid that shortcut by assuming that each sector allows the production of a slightly differentiated final good, and that consumers value diversity[3]. This would not add anything to the results, however, and would merely complicate the setting.

Second, beyond determining the amount of labor allocated to the sector with the highest technology, the parameters α, β and γ are crucial in determining the shape of the distribution function of labor across sectors. In other words, these parameters determine whether the labor force is highly concentrated in the top (high technology) sectors, or whether it is more widespread across sectors. Graphically, we have the following representations (Figures 4.2 to 4.5, $\gamma = 1.3$):

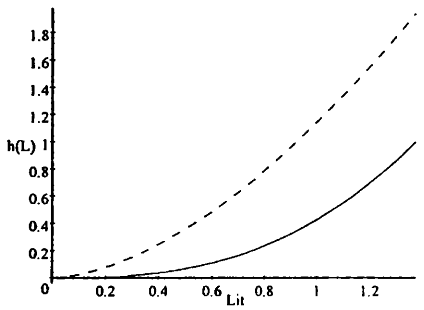


Fig. 4.2: $H(L)$ and $h(L)$ (dots), $\alpha + \beta = 0.3$

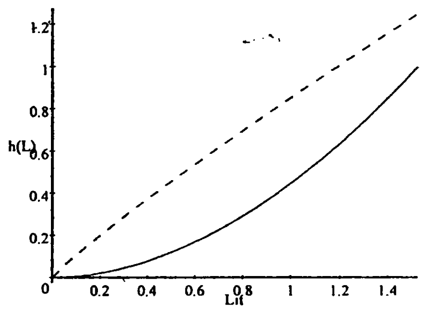
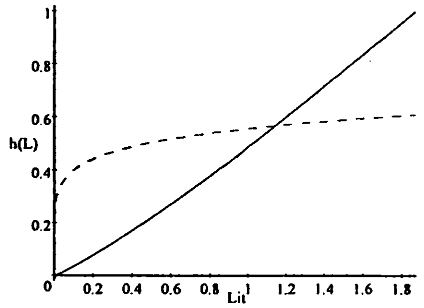


Fig. 4.3: $H(L)$ and $h(L)$ (dots), $\alpha + \beta = 0.5$

[3]Grossman and Helpman (1991a) provide a formalisation of this argument.



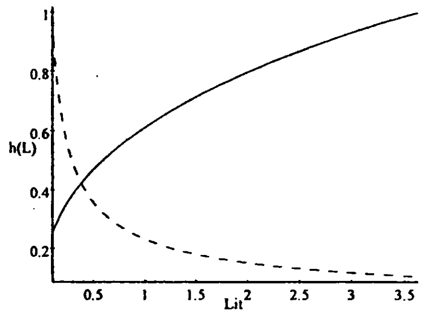


Fig. 4.4: $H(L)$ and $h(L)$ (dots), $\alpha+\beta=0.7$ Fig. 4.5: $H(L)$ and $h(L)$ (dots), $\alpha+\beta=0.9$

The equilibrium wage is unique across sectors and equal to its marginal product, $\beta A_{it}X_{it}^{\alpha}L_{it}^{\beta-1}$. Using equilibrium condition (4.8) for X_{it}, we obtain that $w_t = \beta A_{it}^{\frac{1}{1-\alpha}}\alpha^{\frac{2\alpha}{1-\alpha}}L_{it}^{-\frac{1-\alpha-\beta}{1-\alpha}}$. Because this holds for any sector, it also holds for the highest technology sector. The equilibrium wage is thus given by

$$w_t = \beta\alpha^{\frac{2\alpha}{1-\alpha}}A_t^{Max\,\frac{1}{1-\alpha}}\left[\frac{1-\alpha-\beta}{1-\alpha-\beta+\ln\gamma}\right]^{\frac{1-\alpha-\beta}{1-\alpha}} \tag{4.20}$$

Dynamics

It is clear from the previous section that the dynamics of the model is driven by the R&D process affecting the sectors' relative technological position. From equation (4.18), we find that

$$\frac{\dot{L}_{it}}{L_{it}} = -\frac{1}{1-\alpha-\beta}\lambda\frac{R_t}{\eta A_t^{Max\,\frac{1}{1-\alpha}}}\ln\gamma \tag{4.21}$$

The labor force allocated to a given sector falls at a constant rate (as long as $\frac{R_t}{\eta A_t^{Max\,\frac{1}{1-\alpha}}}$ is constant, which is true on the balanced growth path) proportional to the growth rate of the cutting-edge technology, as long as there is no discovery in the sector. Starting from maximal employment L^{Max}, every sector thus registers a falling labor force as other sectors make discoveries and attract labor. Note that the higher the growth rate of the cutting-edge technology, the faster employment in a sector falls, absent discovery. An innovation generates a discrete jump in employment back to L^{Max}.

Because the equilibrium supply of intermediate good X_{it} depends ultimately on the labor force employed in sector i, it follows the same dynamic pattern as L_{it}. The monopolist thus supplies a quantity X_{it} that falls constantly from an initial maximal level (determined by A_t^{Max} and L^{Max}) at the time of discovery. A subsequent discovery in the sector implies that the incumbent monopolist is displaced by a new one. Using equation (4.8),

$$\frac{\dot{X}_{it}}{X_{it}} = -\frac{\beta}{(1-\alpha)(1-\alpha-\beta)}\lambda\frac{R_t}{\eta A_t^{Max\frac{1}{1-\alpha}}}\ln\gamma \tag{4.22}$$

The dynamics of the monopolist's profit can be calculated easily. Because we assume unit marginal cost of production, $\pi_{it} = P_{it}X_{it} - X_{it}$, where P_{it} and X_{it} are given by equations (4.9) and (4.8). Using equation (4.18), profit is given by

$$\pi_{it} = \left[\frac{1}{\alpha} - 1\right]\alpha^{\frac{2}{1-\alpha}} A_{it}^{\frac{1}{1-\alpha-\beta}} A_t^{Max^{-\frac{\beta}{(1-\alpha)(1-\alpha-\beta)}}} \left[\frac{1-\alpha-\beta+\ln\gamma}{1-\alpha-\beta}\right]^{\frac{\beta}{1-\alpha}} \tag{4.23}$$

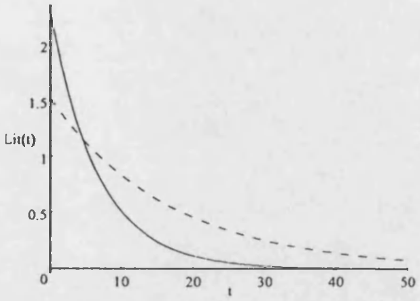
From this, we can see that the level of profit follows the same dynamic pattern as that of X_{it}, absent discovery:

$$\frac{\dot{\pi}_{it}}{\pi_{it}} = \frac{\dot{X}_{it}}{X_{it}} = -\frac{\beta}{(1-\alpha)(1-\alpha-\beta)}\lambda\frac{R_t}{\eta A_t^{Max\frac{1}{1-\alpha}}}\ln\gamma \tag{4.24}$$

The dynamic evolution of the wage rate is derived from equation (4.20) as follows:

$$\frac{\dot{w}_t}{w_t} = \frac{1}{1-\alpha}\lambda\frac{R_t}{\eta A_t^{Max\frac{1}{1-\alpha}}}\ln\gamma \tag{4.25}$$

It is helpful to visualize the dynamic evolution of L_{it} and π_{it} for different values of the parameters, absent discovery (Figures 4.6 and 4.7).



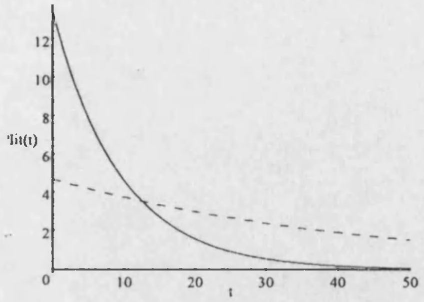


Fig. 4.6: $L_{it}(t)$, $\xi = 0.5$ (dots) and $\xi = 0.8$ Fig. 4.7: $\pi_{it}(t)$, $\xi = 0.5$ (dots) and $\xi = 0.8$

Equilibrium and Labor Reallocation Effect

Solving for the balanced growth path requires us to determine the steady-state level of R&D activity. Free entry into R&D means that by arbitrage its marginal cost is equal to its expected marginal profit. Equation (4.23) gives the level of profit at time t for a monopolist with technological level A_{it}. Any existing monopolist, however, faces the risk of being displaced by some innovator.

At any time $s > t$, the probability that a monopolist "born" on date t still holds its monopoly is equal to $e^{-\lambda \frac{R_t}{\eta A_t^{Max \frac{1}{1-\alpha}}}(s-t)}$, as long as $\frac{R_t}{\eta A_t^{Max \frac{1}{1-\alpha}}}$ is constant over time, which is true on the balanced growth path. According to equation (4.23), the profit at time s for a monopolist whose discovery occurred at time t is given by

$$\pi_{is} = \left[\frac{1}{\alpha} - 1\right] \alpha^{\frac{2}{1-\alpha}} A_t^{Max \frac{1}{1-\alpha-\beta}} A_s^{Max - \frac{\beta}{(1-\alpha)(1-\alpha-\beta)}} \left[\frac{1-\alpha-\beta+\ln\gamma}{1-\alpha-\beta}\right]^{\frac{\beta}{1-\alpha}} \tag{4.26}$$

The expected value of a discovery at time t, V_t, is equal to the discounted stream of profit, taking into account the probability of seeing the monopoly position disappear. We have thus that

$$V_t = \int_t^{\infty} e^{-(r+\lambda\sigma)(s-t)} \Omega A_t^{Max \frac{1}{1-\alpha-\beta}} A_s^{Max - \frac{\beta}{(1-\alpha)(1-\alpha-\beta)}} ds \tag{4.27}$$

where $\sigma = \frac{R_t}{\eta A_t^{Max \frac{1}{1-\alpha}}}$ is constant on the balanced growth path, and where $\Omega =$

$\left[\frac{1}{\alpha}-1\right]\alpha^{\frac{2}{1-\alpha}}\left[\frac{1-\alpha-\beta+\ln\gamma}{1-\alpha-\beta}\right]^{\frac{\beta}{1-\alpha}}$. Given that the arrival rate of discovery is the same across sectors, the expected marginal product of R&D is equal to λV_t. By arbitrage it must then be that

$$\lambda V_t = \eta A_t^{Max\frac{1}{1-\alpha}} \tag{4.28}$$

where the right hand side is equal to the marginal cost of R&D.

Arbitrage and steady-state thus require that

$$\lambda \int_t^\infty e^{-(r+\lambda\sigma)(s-t)}\Omega\left[a_s\right]^{\frac{\beta}{(1-\alpha)(1-\alpha-\beta)}} ds = \eta \tag{4.29}$$

The equilibrium level σ of R&D can be obtained explicitly if one recognizes that $\frac{\dot{a}_s}{a_s} = -\lambda\sigma\ln\gamma$. Solving this differential equation with initial condition $a_t = 1$, we have that $a_s = e^{-(\lambda\sigma\ln\gamma)(s-t)}$. Using this in the arbitrage condition, we obtain

$$\lambda\Omega \int_0^\infty e^{-(r+\lambda\sigma)s} e^{-\frac{\beta\lambda\sigma\ln\gamma}{(1-\alpha)(1-\alpha-\beta)}s} ds = \eta \tag{4.30}$$

which means that $\frac{1}{r+\lambda\sigma+\frac{\beta\lambda\sigma\ln\gamma}{(1-\alpha)(1-\alpha-\beta)}} = \frac{\eta}{\lambda\Omega}$. The steady state level of R&D is thus given by

$$\sigma = \frac{1}{1+\frac{\beta\ln\gamma}{(1-\alpha)(1-\alpha-\beta)}}\left[\frac{\Omega}{\eta}-\frac{r}{\lambda}\right] \tag{4.31}$$

Given this constant level of effective research σ on the balanced growth path, the economy exhibits the following dynamic properties in steady state:

- The cutting-edge technology grows at the constant rate $g = \lambda\sigma\ln\gamma$.
- Output grows at the constant rate $\frac{1}{1-\alpha}g$.
- The labor force allocated to sector i falls at the constant rate $\frac{1}{1-\alpha-\beta}g$ as long as no discovery arises in the sector, while it discretely jumps to L^{Max} at the time of discovery.

- The equilibrium level of intermediate input supplied by the incumbent monopolist in sector i falls at the constant rate $\frac{\beta}{(1-\alpha)(1-\alpha-\beta)}g$ as long as no discovery arises in the sector. The level of profit in the sector falls at the same rate from its initial maximal level at the date of discovery, absent discovery.

- The wage increases at the constant rate $\frac{1}{1-\alpha}g$.

The behavior of the economy is thus quite similar to what is traditionally derived in Schumpeterian models of endogenous growth. A steady-state is obtained where the growth rate of variables is ultimately determined by the amount of equilibrium R&D and hence technological progress. All parameters that affect the incentives to conduct R&D act in the usual way.

What is new in the dynamics of this model is that it implies a continuous reallocation of the labor force across sectors and a gradual decline in the extent to which a sector is used relative to the others in producing final output, absent discovery. At any point of time, some sector innovates and attracts extra labor, which comes from all the sectors that fail to innovate. In addition to creative destruction in monopolistic positions, there is thus also creative destruction in jobs, as a result of the process of technological development. In other words, jobs are currently being lost in, say, the metallurgic industry, but are being replaced by new jobs in the higher technology computer or pharmaceutical industries.

The setting elaborated above allows us to explicitly derive a quantitative measure of the cross-sectoral labor reallocation effect. Equation (4.21) indicates the rate at which L_{it} falls over time, absent discovery. We have also derived the pdf of the sectoral labor allocation, $h(L) = \frac{\partial H(L)}{\partial L}$, with domain $L \in \left[0, \frac{1-\alpha-\beta+\ln\gamma}{1-\alpha-\beta}\right]$. This allows us to calculate the total fall in the labor force employed in the "decreasing sectors" as $\int_0^{L^{Max}} \dot{L}_{it} h(L_{it})\, dL$. Denoting the fall in labor force per unit of time as δ, we have that

$$-\delta = -\frac{g}{\ln\gamma}\left[\frac{1-\alpha-\beta}{1-\alpha-\beta+\ln\gamma}\right]^{\frac{1-\alpha-\beta}{\ln\gamma}} \int_0^{L^{Max}} L_{it}^{\frac{1-\alpha-\beta}{\ln\gamma}}\, dL \qquad (4.32)$$

Hence,

$$\delta = \frac{1}{1 - \alpha - \beta} g \tag{4.33}$$

Given the constant unit mass labor force L, δ represents the rate at which the labor force is reallocated across sectors[4]. Using $g = 0.01$ and $\alpha + \beta = 0.7$ implies that $\delta = 0.033$, which means that 3.3 percent of the labor force is reallocated per unit of time. The labor reallocation effect is increasing (linearly) in g, ceteris paribus, and is increasing (non-linearly) in $(\alpha + \beta)$. A rapidly growing economy thus "needs" a larger amount of labor reallocation per unit of time. If one moves closer to constant returns to scale ($\alpha + \beta$ close to 1), i.e. when there is less widespread use of the sectors in producing the final output and the labor force is more concentrated in the higher technology sectors, the reallocation needs to be larger (Figure 4.8). At the extreme, when $\alpha + \beta = 1$, only one sector produces at a time, and the entire labor force is constantly being reallocated to the highest technology sector.

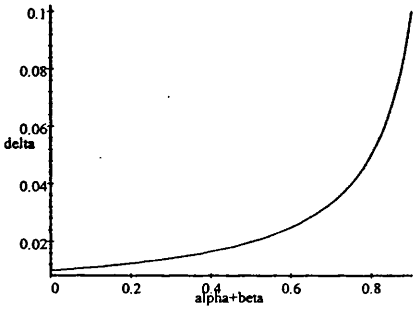


Figure 4.8: $\delta = \frac{1}{1-\alpha-\beta} g$ as a function of $(\alpha + \beta)$

This equilibrium labor reallocation process across sectors naturally leads us to wonder what may happen if perfect and instantaneous mobility of labor is not possible. In other words, what are the effects of labor market rigidities on sectoral labor allocation, R&D, growth and wage? What happens if reallocating a metallurgic worker to the computer industry is not as smoothly feasible as assumed in the previous sections?

[4]It is easy to verify that the increase in the labour force in the sectors experiencing discovery does indeed match the decrease in those not experiencing discovery. One only needs to note that the increase is given by $\int_0^{L^{Max}} \lambda\sigma \left[L^{Max} - L_{it}\right] h(L)\, dL$, since $\lambda\sigma$ is the probability of discovery in any sector. Solving the integral yields $\delta = \frac{1}{1-\alpha-\beta} g$.

4.2.4 Rigidity in Labor Force Reallocation Across Sectors

Rationale for Rigidity

One can think of three main reasons why the actual sectoral labor reallocation may fall short of that necessary to equalize the marginal productivity of workers across sectors:

- following the labor market literature, the process of matching workers with vacancies is time consuming and can only be done through some type of "matching function";
- there are likely to be structural and legal impediments that prevent firings and downsizings, which slow down the reallocation process;
- one should recognize that each intermediate sector may require some specific skills from its labor force. Having workers changing employment across sectors may thus be hampered by such sector specific skills, and is likely to take time for retraining.

Each of these arguments could lead to an alternative way to model rigidity. We do not enter in a detailed modeling of the micro-foundations in order to focus on the effects of introducing rigidity on the behavior of the economy. We assume that, for a combination of the reasons given above, the economy is unable to reallocate the amount of labor δ per unit of time. Instead, rigidities allow only for a fraction ϕ ($0 \leq \phi \leq 1$) of the optimal reallocation δ to take place. The effective, rigidity constrained, labor reallocation is thus given as

$$\delta_r = \phi \frac{g_r}{1-\alpha-\beta} \tag{4.34}$$

where a lower case r indicates a variable within the rigidity setting, as opposed to the flexible setting.

Sectoral Allocation of Labor and Dynamics with Rigidities

We consider the effect of imposing rigidity ϕ on an economy that was previously on its full flexibility balanced growth path. We do not formally analyze the transition path and focus instead on the new rigidity constrained steady-state. The transition process, however, is quite intuitive and can be visualized from the plots of $h(L)$. In order to keep the distribution of $h(L)$ constant over time over the range $[0, L^{Max}]$, exactly the amount δ of labor needs to be reallocated from the whole range of sectors towards the sector experiencing the discovery.

If we introduce rigidities and allow only for $\phi\delta$ of labor to be reallocated across sectors, the movement between sectors is insufficient to preserve L^{Max} constant. In other words, there is not enough mass being redistributed across sectors to fill the "gap" that is opening on the right side of the distribution $h(L)$. This sets a transition process where L^{Max} starts to fall and the distribution function $h(L)$ changes. Obviously, other variables, including particularly the amount of R&D, are also affected.

The transitional change in L^{Max} and $h(L)$ will stop only once the decrease in labor force across sectors exactly matches the "required" increase, as in the previous section. Hence, stability requires that,

$$- \int_0^{L_r^{Max}} \dot{L}_{r,it} h_r(L_r)\, dL_r = \int_0^{L_r^{Max}} \left[L_r^{Max} - L_{r,it}\right] \lambda \sigma_r h_r(L_r)\, dL_r = \phi \frac{g_r}{1-\alpha-\beta} \tag{4.35}$$

where σ_r, g_r are constant on the balanced growth path.

Optimality, as seen above, would require to decrease the labor force in every sector at the constant rate $\frac{\dot{L}_{it}}{L_{it}} = -\frac{g}{1-\alpha-\beta}$. If this amount of reallocation is impossible, the second best is to reallocate as much as feasible. This implies that $\frac{\dot{L}_{r,it}}{L_{r,it}} = -\phi\frac{g_r}{1-\alpha-\beta}$. Hence, the rigidity constrained amount of sectoral labor reallocation is given by

$$- \int_0^{L_r^{Max}} \dot{L}_{r,it} h_r(L_r)\, dL_r = \phi \frac{g_r}{1-\alpha-\beta} \tag{4.36}$$

Determining L_r^{Max} and $h_r(L_r)$ requires "educated guesswork" and recognizing that they need to satisfy the following conditions:

- $H_r\left(L_r^{Max}\right) = 1$;

- L_r^{Max} is constant;

- $-\int_0^{L_r^{Max}} \dot{L}_{r,it} h_r(L_r)\, dL_r = \int_0^{L_r^{Max}} \left[L_r^{Max} - L_{r,it}\right] \lambda \sigma_r h_r(L_r)\, dL_r = \phi \frac{g_r}{1-\alpha-\beta}$;

- if $\phi = 1$, then we have $L_r^{Max} = L^{Max}$ and $H_r(L_r) = H(L)$.

Let us then assume the following forms for $L_r^{Max}$and $H_r(L_r)$:

$$L_r^{Max} = \frac{1-\alpha-\beta+\phi \ln \gamma}{1-\alpha-\beta} \tag{4.37}$$

$$H_r(L_r) = \left[L_{r,it}\left(\frac{1-\alpha-\beta}{1-\alpha-\beta+\phi \ln \gamma}\right)\right]^{\frac{1-\alpha-\beta}{\phi \ln \gamma}} \tag{4.38}$$

It is easy to check that the conditions listed above hold for these equations, where $h_r(L_r) = \frac{\partial H_r(L_r)}{\partial L_r}$.

The labor force allocated to sector i at time t falls, in the absence of discovery, at the constant rate $\frac{\dot{L}_{r,it}}{L_{r,it}} = -\phi \frac{g_r}{1-\alpha-\beta}$ from a maximal amount L_r^{Max} at the instant of discovery as determined by equation (4.37). We can thus determine that $L_{r,it} = \left(\frac{1-\alpha-\beta+\phi \ln \gamma}{1-\alpha-\beta}\right) e^{-\phi \frac{g_r}{1-\alpha-\beta}(t-s)}$, where s is the time when the last discovery occurred in sector i. As in the previous section, however, we continue to assume that $\frac{\dot{A}_t^{Max}}{A_t^{Max}} = \lambda \frac{R_{r,t}}{\eta A_t^{Max \frac{1}{1-\alpha}}} \ln \gamma$, which is constant and equal to g_r in steady-state. It is clear then that $\frac{A_{it}}{A_t^{Max}} = e^{-(t-s)g_r}$. The labor allocation across sectors is thus uniquely determined as

$$L_{r,it} = a_{it}^{\frac{\phi}{1-\alpha-\beta}} \left[\frac{1-\alpha-\beta+\phi \ln \gamma}{1-\alpha-\beta}\right] \tag{4.39}$$

which compares with equation (4.18) in the flexible case. Again, the distribution of relative technology parameters a_{it} means that the distribution function of the sectoral labor allocation is given by equation (4.38).

The core problem of the monopolist is unaffected by the introduction of rigidities in labor force reallocation. Profit maximization still implies that $P_{r,it} = \frac{1}{\alpha}$ and $X_{r,it} = \left[\alpha^2 A_{it}\right]^{\frac{1}{1-\alpha}} L_{r,it}^{\frac{\beta}{1-\alpha}}$. The equilibrium quantity of intermediate input i at time t is thus given by,

$$X_{r,it} = \left[\alpha^2 A_{it}\right]^{\frac{1}{1-\alpha}} \left[\frac{A_{it}}{A_t^{Max}}\right]^{\frac{\phi\beta}{(1-\alpha)(1-\alpha-\beta)}} \left[\frac{1-\alpha-\beta+\phi\ln\gamma}{1-\alpha-\beta}\right]^{\frac{\beta}{1-\alpha}} \tag{4.40}$$

Profit at time t is obtained as before as $\pi_{r,it} = \left(\frac{1}{\alpha}-1\right) X_{r,it}$. It follows that

$$\pi_{r,it} = \left[\frac{1}{\alpha}-1\right] \alpha^{\frac{2}{1-\alpha}} A_{it}^{\frac{1-\alpha-\beta(1-\phi)}{(1-\alpha)(1-\alpha-\beta)}} {A_t^{Max}}^{-\frac{\phi\beta}{(1-\alpha)(1-\alpha-\beta)}} \left[\frac{1-\alpha-\beta+\phi\ln\gamma}{1-\alpha-\beta}\right]^{\frac{\beta}{1-\alpha}} \tag{4.41}$$

The effect of introducing rigidity ϕ on the labor allocation across sectors is thus quite clear. The insufficient amount of labor being freed up from "decreasing" sectors implies that innovating sectors do not receive as much extra labor as they would under full flexibility. This is captured by $L_r^{Max} < L^{Max}$. While a sector innovating at time t has a shortfall of labor at time t and carries it to $t+\varepsilon$, the labor "freed" at time $t+\varepsilon$ will not fill that shortfall, as by then some other sector will have innovated. Within a longer time horizon, however, rigidities imply that the labor force employed in a sector cannot fall as quickly as necessary, which implies that in the longer run, the sector may end up with excess labor. The main implication of introducing rigidities is thus that sectors that have recently innovated suffer a shortfall of labor, while sectors that have not seen a discovery for some time suffer excess employment. Labor is misallocated across sectors, in the sense that marginal productivities cannot be equalized.

The dynamic pattern of labor, equilibrium supply of $X_{r,it}$ and profit, absent discovery, are essentially the same as before and determined by:

$$\begin{aligned} \frac{\dot{L}_{r,it}}{L_{r,it}} &= -\phi\frac{g_r}{1-\alpha-\beta} && (4.42) \\ \frac{\dot{X}_{r,it}}{X_{r,it}} &= \frac{\dot{\pi}_{r,it}}{\pi_{r,it}} = -\phi\frac{\beta}{(1-\alpha)(1-\alpha-\beta)}g_r && (4.43) \end{aligned}$$

Equilibrium Growth and Inequality

The steady-state level of R&D and balanced growth path are obtained as before from arbitrage in the R&D sector. The profit flow of an incumbent monopolist is given by equation (4.41), and the probability of still holding the monopoly position is equal to $e^{-\lambda \frac{R_{r,t}}{\eta A_t^{Max\frac{1}{1-\alpha}}}(s-t)}$, as long as $\frac{R_{r,t}}{\eta A_t^{Max\frac{1}{1-\alpha}}}$ is constant over time, which is true on the balanced growth path. The expected value of a discovery is thus given by

$$V_{r,t} = \int_t^{\infty} e^{-(r+\lambda\sigma_r)(s-t)} \Omega_r A_t^{Max\frac{1-\alpha-\beta(1-\phi)}{(1-\alpha)(1-\alpha-\beta)}} A_s^{Max^{-\frac{\phi\beta}{(1-\alpha)(1-\alpha-\beta)}}} ds \tag{4.44}$$

where $\sigma_r = \frac{R_{r,t}}{\eta A_t^{Max\frac{1}{1-\alpha}}}$ is constant and $\Omega_r = \left[\frac{1}{\alpha} - 1\right] \alpha^{\frac{2}{1-\alpha}} \left[\frac{1-\alpha-\beta+\phi\ln\gamma}{1-\alpha-\beta}\right]^{\frac{\beta}{1-\alpha}}$.

Arbitrage and steady-state then require that

$$\lambda \int_t^{\infty} e^{-(r+\lambda\sigma_r)(s-t)} \Omega_r \left[a_s\right]^{\frac{\phi\beta}{(1-\alpha)(1-\alpha-\beta)}} ds = \eta \tag{4.45}$$

This equation can be solved similarly to equation (4.29) to obtain the steady-state level of research σ_r in the economy with labor force reallocation rigidities:

$$\sigma_r = \frac{1}{1 + \frac{\phi\beta\ln\gamma}{(1-\alpha)(1-\alpha-\beta)}} \left[\frac{\Omega_r}{\eta} - \frac{r}{\lambda}\right] \tag{4.46}$$

The cutting-edge technology thus grows at the constant rate $g_r = \lambda\sigma_r \ln\gamma$, while output grows at the rate $\frac{1}{1-\alpha} g_r$.

A detailed comparative analysis of the flexible and rigidity constrained economies is left for the following section. At this stage, we need to analyze the impact of the sectoral labor misallocation. As noted, rigidities prevent the optimal, marginal productivity equalizing, amount of labor to be allocated to each sector. If labor is paid its marginal product, then, rigidities generate wage inequality.

Since $w_{r,it} = \frac{\partial Y_{it}}{\partial L_{r,it}} = \beta A_{it} X_{r,it}^{\alpha} L_{r,it}^{\beta-1}$, we obtain from equation (4.39) that the wage rate in sector i at time t is given by

$$w_{r,it} = \beta\alpha^{\frac{2\alpha}{1-\alpha}} A_{it}^{\frac{1-\phi}{1-\alpha}} A_t^{Max\frac{\phi}{1-\alpha}} \left[\frac{1-\alpha-\beta}{1-\alpha-\beta+\phi\ln\gamma}\right]^{\frac{1-\alpha-\beta}{1-\alpha}} \tag{4.47}$$

The wage rate is thus *not* equal across sectors, as it was in the economy with full flexibility. In fact, at any given time t, wage is an increasing function of the sectoral technology level A_{it}. One can also observe that $\frac{\dot{w}_{r,it}}{w_{r,it}} = \frac{\phi}{1-\alpha} g_r$, which implies that wages increase at a constant rate across all sectors that do not experience discovery, even though their levels differ.

Wage being equal across sectors in the flexible economy, the average wage $\bar{w}_t$ is equal to the prevailing wage w_t as given by equation (4.20). In the economy with rigidity, on the other hand, the average wage $\bar{w}_{r,t}$ is determined as the labor-weighted average of the respective wages in all sectors, $\int_0^{L_r^{Max}} w_{r,it} h_r(L_r)\, dL_r$. Using equation (4.39), one can express $w_{r,it}$ as a function of A_t^{Max} and $L_{r,it}$ and solve the integral, which gives the average rigidity constrained wage as

$$\bar{w}_{r,t} = \beta\alpha^{\frac{2\alpha}{1-\alpha}} \left[\frac{1-\alpha-\beta}{1-\alpha-\beta+\phi\ln\gamma}\right]^{\frac{1-\alpha-\beta}{1-\alpha}} \frac{1-\alpha}{(1-\alpha)+(1-\phi)\ln\gamma} A_t^{Max\frac{1}{1-\alpha}} \tag{4.48}$$

which is to be compared with equation (4.20) in the flexible economy. Because $\phi < 1$ and $\frac{1-\alpha}{(1-\alpha)+(1-\phi)\ln\gamma} < 1$, we can conclude that $\bar{w}_{r,t} < \bar{w}_t$: the average wage in the rigid economy is lower than in the flexible economy. Also, $\frac{\dot{\bar{w}}_{r,t}}{\bar{w}_{r,t}} = \frac{1}{1-\alpha} g_r$, which means that the growth rate of the average wage in the rigid economy is lower than in the flexible one inasmuch as $g_r < g$, which remains to be determined. Note also that the average wage $\bar{w}_{r,t}$ increases faster than the wage $w_{r,it}$ in sectors not experiencing discovery. Rigidities and inequality indeed generate a catch-up phenomenon whereby the innovating sectors see the wage jump discretely to the top-end of the wage distribution. This phenomenon was absent in the flexible setting, where wage is equal across all sectors.

The wage being different across sectors and changing continuously over time, it is useful to define a relative wage measure whose distribution remains constant over time. Let $\omega_{r,it} = \frac{w_{r,it}}{A_t^{Max\frac{1}{1-\alpha}}}$ be the prevailing wage rate adjusted by a proportional function of the cutting-edge level of technology. We have that

$$\omega_{r,it} = \beta\alpha^{\frac{2\alpha}{1-\alpha}} \left[\frac{1-\alpha-\beta}{1-\alpha-\beta+\phi\ln\gamma}\right]^{\frac{1-\alpha-\beta}{1-\alpha}} [a_{it}]^{\frac{1-\phi}{1-\alpha}} \tag{4.49}$$

Given that a_{it} has a constant distribution function $H(a) = a^{\frac{1}{\ln\gamma}}$, ω_{it} is distributed according to the following time-independent cdf:

$$H_r(\omega_r) = \left[\frac{1}{\beta\alpha^{\frac{2\alpha}{1-\alpha}}\left[\frac{1-\alpha-\beta}{1-\alpha-\beta+\phi\ln\gamma}\right]^{\frac{1-\alpha-\beta}{1-\alpha}}}\omega_{r,it}\right]^{\frac{1}{\ln\gamma}\frac{1-\alpha}{1-\phi}} \tag{4.50}$$

where $\omega_{r,it} \in \left[0, \beta\alpha^{\frac{2\alpha}{1-\alpha}}\left[\frac{1-\alpha-\beta}{1-\alpha-\beta+\phi\ln\gamma}\right]^{\frac{1-\alpha-\beta}{1-\alpha}}\right]$.

The constant distribution of relative wages ω_{it} does not prevent movement of sectors along that distribution. To the contrary, and as before, sectors are constantly moving along the relative wage distribution, but with sectoral movements constantly cancelling each other to keep the distribution constant. The same way backward sectors catch up with high technology sectors through innovations, low relative wage sectors catch up with high relative wage sectors through innovation. There are thus constant changes as to which sectors are high wage or low wage.

4.3 Labor Markets, R&D, Growth and Inequality

As the model elaborated above shows, a relatively simple extension of the Schumpeterian process of creative destruction to jobs generates informative results on a range of issues, which have been analyzed in distinct strands of the literature. We now turn to the main results of the model and relate them to some of the existing literature.

4.3.1 Sectoral Labor Allocation, Rigidity and Growth

Kuznets (1955) first expressed the idea that the reallocation of labor from one sector of the economy to the other may be closely linked to growth and inequality, if a labor productivity gap exists between the two sectors. Denison (1967), and others more recently, argued that the reallocation of labor from agriculture to industry accounts for a significant part of the rapid growth experienced by Western countries in the 1950s-60s, and that it could explain the more recent productivity slowdown as reallocation came to an end. Others, including Squire (1981) and Poirson (2000), have used the same idea to explain the dualism that characterizes labor markets in developing countries or to explain growth.

These papers all suffer from a lack of theoretical foundation, however. In particular, they fail to identify the source of the labor productivity gap between the two sectors and to take into account the effects of labor reallocation on the productivity differential. All also focus on the reallocation of labor from agriculture to industry, which means that their contribution is limited to analyzing growth over a historically short transitionary period.

The model we have developed, on the other hand, shows how labor reallocation across sectors is an equilibrium phenomenon. The very nature of growth through technological change requires permanent labor reallocation. We also show that the relationship between growth and labor reallocation is more complex than depicted by the literature mentioned above, which quickly concludes that fast reallocation generates high growth. As Poirson asserts, *"these countries that reallocate their workers more efficiently over time tend to grow faster, in per worker terms, ceteris paribus."*[5] While this may be true for countries in transition from agriculture-based economies to industrial economies, this may not be the case in general, our model shows.

In order to analyze the impact of labor reallocation rigidities ϕ on growth in our model, we need to compare the equilibrium level of R&D in the flexible and rigid cases, σ and σ_r as given by equations (4.31) and (4.46), respectively. It is easy to verify that $\Omega > \Omega_r$ for $0 < \phi < 1$. The term in brackets on the RHS of equation (4.31) is thus larger than that in the RHS of equation (4.46), which implies that there exists an effect through which increasing flexibility induces a higher level of R&D. Explicitly, we can calculate that $\frac{\partial \Omega_r}{\partial \phi} = \left[\frac{1}{\alpha} - 1\right] \alpha^{\frac{2}{1-\alpha}} \frac{\beta}{1-\alpha} \left[\frac{1-\alpha-\beta+\phi \ln \gamma}{1-\alpha-\beta}\right]^{\frac{\beta}{1-\alpha}-1} \frac{\ln \gamma}{1-\alpha-\beta} > 0$.

There exists a second effect, however, through which *decreasing* flexibility induces a higher level of R&D. This is evident from the denominator in the first term on the RHS of equation (4.46). These two counteracting effects mean that we cannot unequivocally sign $\frac{\partial \sigma_r}{\partial \phi}$. Whether increasing the degree of flexibility of the labor market generates a

[5] Poirson (2000), p. 22.

rise or a fall in the steady-state level of R&D and growth is thus undetermined. In fact, taking the derivative, we can determine that the condition for higher flexibility to generate higher R&D is that

$$\frac{\Omega_r}{\eta}\frac{(1-\alpha-\beta)^2}{1-\alpha-\beta+\phi\ln\gamma} > \beta\frac{r}{\lambda} \tag{4.51}$$

What is needed is that the first effect dominates the second. Intuitively, the first effect, which increases R&D when labor market flexibility rises, occurs from the fact that more flexibility allows the economy to "use" the sector with the highest technology to a larger extent. In other words, flexibility allows the economy to reach a level L_r^{Max} closer to the optimal level L^{Max}. A higher L_r^{Max} obviously implies a higher demand for the monopolist's intermediate input in the period immediately following discovery. The first effect is thus that more flexibility generates a higher demand for a monopolist in the early stage of the "life" of the monopoly, and hence a higher profit.

The second effect is the inverse corollary of the first. We have indeed seen that a lower degree of flexibility slows down the pace at which the labor force can be reallocated from non-innovating to innovating sectors. This means that the labor force, and hence the demand facing the monopolist and its profit, do not fall as rapidly as otherwise.

In brief, rigidity hurts the monopolist in the early stages following discovery of the new vintage as it reduces maximal demand, but it benefits it in the sense that, even though starting from a lower base, the demand falls at a slower pace. Which of the two effects dominates the other depends on the value of the parameters and the degree of rigidity around which the marginal effect of increased flexibility is measured. The impact of rigidity on the equilibrium level of growth can be understood more easily graphically. Using $\alpha = 0.3$ and $\beta = 0.4$, we have (Figures 4.9 to 4.12)

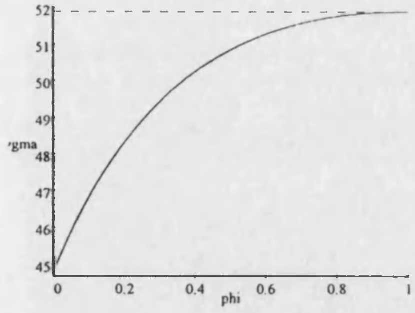


Fig. 4.9: σ (dots) and $\sigma_r(\phi)$, with $\gamma = 1.5$

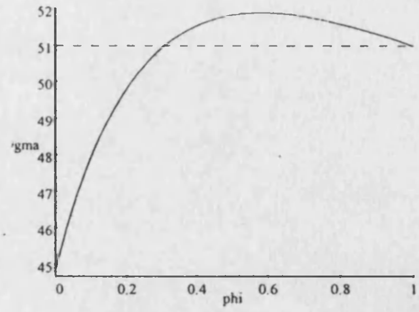


Fig. 4.10: σ (dots) and $\sigma_r(\phi)$ with $\gamma = 1.8$

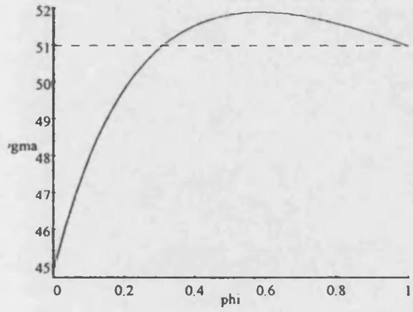


Fig. 4.11: σ (dots) and $\sigma_r(\phi)$ with $\gamma = 2$

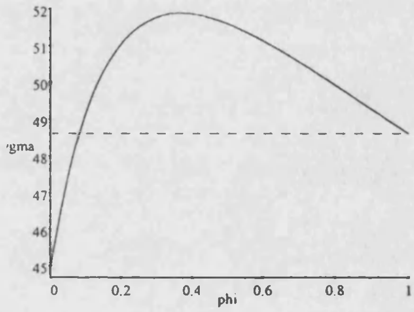


Fig. 4.12: σ (dots) and $\sigma_r(\phi)$ with $\gamma = 3$

We can see that with $\gamma = 1.5$, the maximal level of growth is reached when $\phi = 1$, i.e. when there is full flexibility in labor reallocation. We also see, interestingly, that when $\gamma = 1.8$ or larger, the maximal level of growth is reached within $\phi \in]0, 1[$. Reaching the maximal growth rate thus requires a certain level of rigidity in labor force reallocation across sectors. We can see that the growth-maximizing degree of rigidity increases with γ, the "size of innovation" parameter.

The relation between labor reallocation flexibility and growth is thus more complex than could appear at first analysis, and reaching definite conclusions would require calibrating the model to actual data. The interesting conclusion that can be drawn from the theory is that rigidity, i.e. sub-optimal allocation of labor across sectors, does not necessarily imply lower steady-state growth.

4.3.2 Equilibrium Wage Inequality

While the effect of sectoral labor reallocation rigidities on the equilibrium level of R&D and growth is not unequivocal, the model shows a clear effect of rigidities on wage inequality. We derived that the economy with rigidities in sectoral labor force reallocation generates a continuum of relative wages $\omega_{r,it}$ with time-independent distribution function $H_r(\omega_r)$. This constant distribution of relative wage allows us to derive explicitly the two most traditional measures of inequality as a function of the parameters of the model: the Lorenz curve and Gini coefficient. The average relative wage, which is also equal to the total wage because we assume a unit mass of labor, $\bar{\omega}_r$, is given by

$$\bar{\omega}_r = \beta\alpha^{\frac{2\alpha}{1-\alpha}} \left(\frac{1-\alpha-\beta}{1-\alpha-\beta+\phi\ln\gamma}\right)^{\frac{1-\alpha-\beta}{1-\alpha}} \frac{1-\alpha}{1-\phi+1-\alpha} \tag{4.52}$$

Using equation (4.49), we can calculate the shares of total wage income of any share of the total population. The equation for the Lorenz curve is given by,

$$f(x) = x^{\frac{1-\phi}{1-\alpha}\ln\gamma+1} \tag{4.53}$$

where $x \in [0,1]$ is the share of the total population. One can see that if $\phi = 1$, $f(x) = x$, indicating the absence of inequality. This result should be obvious since we saw in the first part of the model that fully flexible labor force reallocation implies an equalization of wage across sectors.

The Gini coefficient is obtained straightforwardly as

$$G = 1 - \frac{2(1-\alpha)}{(1-\phi)\ln\gamma + 2(1-\alpha)} \tag{4.54}$$

where $G = 0$ if $\phi = 1$, indicating again that full flexibility implies zero inequality.

It is informative to plot the Lorenz curves implied by the model for different values of the parameters (Figures 4.13 to 4.16).

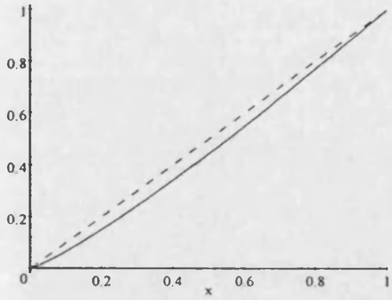


Figure 4.13: $\alpha = 0.3$, $\phi = 0.7$, $\gamma = 1.5$

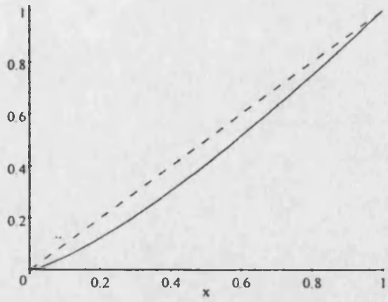


Figure 4.14: $\alpha = 0.3$, $\phi = 0.7$, $\gamma = 2$

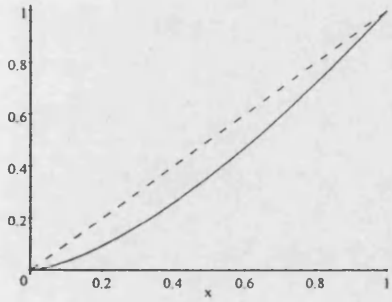


Figure 4.15: $\alpha = 0.3$, $\phi = 0.7$, $\gamma = 3$

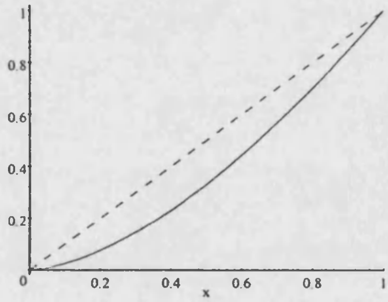


Figure 4.16: $\alpha = 0.3$, $\phi = 0.7$, $\gamma = 4$

We can see from these that a relatively large value of the parameter γ is necessary to generate a large degree of inequality. Similarly, we can see that an increasing degree of labor market rigidity raises the level of equilibrium inequality, ceteris paribus (Figures 4.17 to 4.20).

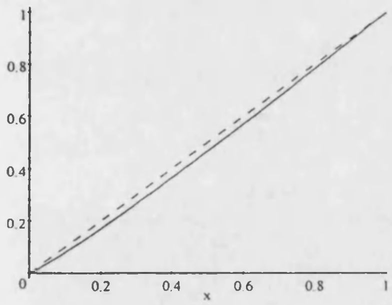


Figure 4.17: $\alpha = 0.3$, $\phi = 0.9$, $\gamma = 2$

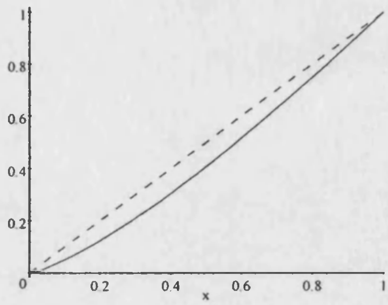
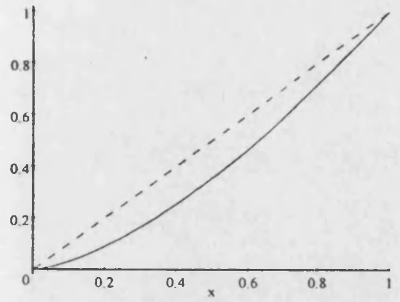


Figure 4.18: $\alpha = 0.3$, $\phi = 0.7$, $\gamma = 2$



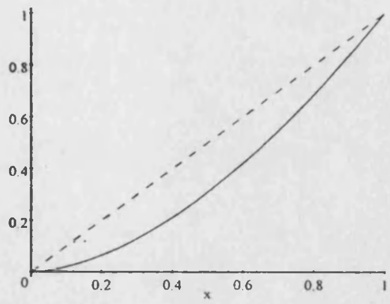


Figure 4.19: $\alpha = 0.3$, $\phi = 0.5$, $\gamma = 2$ Figure 4.20: $\alpha = 0.3$, $\phi = 0.3$, $\gamma = 2$

The following table provides values of the Gini coefficients for different sets of parameters. It provides a clear measure of how the different parameters affect the degree of equilibrium inequality.

Gini coefficients for different sets of parameters

	α=0.3 γ=1.1	α=0.3 γ=1.3	α=0.3 γ=1.5	α=0.3 γ=2.0	α=0.3 γ=3.0	α=0.3 γ=2.0	α=0.5 γ=2.0	α=0.7 γ=2.0
ϕ=0.1	0.06	0.14	0.21	0.31	0.41	0.31	0.38	0.51
ϕ=0.2	0.05	0.13	0.19	0.28	0.39	0.28	0.36	0.48
ϕ=0.3	0.05	0.12	0.17	0.26	0.35	0.26	0.33	0.45
ϕ=0.4	0.04	0.10	0.15	0.23	0.32	0.23	0.29	0.41
ϕ=0.5	0.03	0.09	0.13	0.20	0.28	0.20	0.26	0.37
ϕ=0.6	0.03	0.07	0.10	0.17	0.24	0.17	0.22	0.32
ϕ=0.7	0.02	0.05	0.08	0.13	0.19	0.13	0.17	0.26
ϕ=0.8	0.01	0.04	0.05	0.09	0.14	0.09	0.12	0.19
ϕ=0.9	0.01	0.02	0.03	0.05	0.07	0.05	0.06	0.10
ϕ=1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.1

The model thus clearly shows that rigidities in sectoral labor reallocation generate a certain degree of equilibrium inequality. The degree of equilibrium inequality itself is influenced by three parameters of the model: the degree of rigidity, ϕ, the size of innovation, γ, and the degree of monopoly power, α. Clearly, rigidities in sectoral labor reallocation increase the degree of equilibrium inequality. Also, large innovations generate higher inequality, ceteris paribus, while higher monopoly power (lower α) decreases it.

It is also interesting to reassert a point that was made earlier: for a given level of maximum technology A_t^{Max}, average wage is lower in the rigid economy than in the flexible economy, ceteris paribus. Rigidities thus introduce a *static efficiency loss.* We have seen earlier, however, that the impact of rigidities on steady-state growth is uncertain. It may very well be, in fact, that the rigid economy grows faster than the flexible economy, i.e. that $g_r > g$. The average wage may thus increase faster in equilibrium in the rigid economy than in the flexible one, meaning that rigidities may generate a *dynamic efficiency gain* in terms of wage and growth. There could thus be a trade-off in terms of welfare between wage dispersion and lower average wage as a result of rigidities, and higher future wage growth. In other words, agents may be ready to suffer the possibility of being among the lower wage earners today if that allows them to benefit from potentially higher wages in the future. Such a welfare analysis is beyond the scope of this chapter, and would require strong assumptions about the mobility of workers along the relative wage curve.

4.3.3 Inequality and Growth

The links between growth and rigidities and between rigidities and inequality established above allow us to shed some new light on the debate on inequality and growth initiated by Kuznets (1955) and subsequently developed in many directions in attempts to explain the nexus growth-inequality or inequality-growth. The growth-inequality causality has been the subject of a vast empirical literature. Dollar and Kraay (2000) provide a recent study and review of that literature. The causality from inequality to growth has been the subject of a great deal of recent theoretical studies, highlighting channels through which inequality may harm growth.

The model elaborated above derives a more nuanced and indirect relationship between growth and inequality. We showed that rigidities in sectoral labor reallocation unequivocally increase equilibrium inequality, but that such rigidities may have a positive or negative impact on growth. It is thus impossible to derive a clear-cut relationship between growth and inequality, just as it was impossible to reach definite conclusions on the growth-rigidity nexus.

As is clear from the figures relating equilibrium R&D σ_r and rigidity ϕ, whether reducing inequality (in this case through increasing flexibility) generates a fall or a rise in the equilibrium growth rate depends on the parameters of the model and the initial level of rigidity. The extent to which increasing flexibility reduces inequality as measured by the Gini coefficient is also affected by the parameters α and γ describing the economy.

The growth-inequality nexus thus appears particularly complex, even in the relatively simple setting of the model. In particular, countries characterized by different sets of parameters are likely to exhibit distinct relationships, complicating the derivation of clear results through cross-sectional empirical studies. Allowing for redistributive policies, which we completely abstracted from in the theoretical setting, would make the relationship yet more complex.

4.4 Conclusion

Extending the Schumpeterian concept of creative destruction to jobs allowed us to derive several new and important results on sectoral labor allocation, R&D, growth and inequality. The basic flexible model highlights the extent to which technological progress implies a constant reallocation of labor across sectors.

We then showed how rigidities in the labor reallocation process generate a given level of equilibrium inequality, which is an increasing function of the level of rigidity and other parameters of the model, in particular the size of technological innovations. Although equilibrium inequality is constant in steady-state, there is mobility of individuals along the wage curve.

One important policy implication is that there may be avenues to reduce inequality without need to put in place distortionary policies. In particular, inequality could be reduced through:

- a reduction in the legal measures preventing mobility, including firing costs or bankruptcy procedures;
- general purpose education, which is likely to favor a quicker and smoother adaptation of workers to sector-specific skills;

- on-the-job training programs, which allow workers to gain the necessary sector-specific skills.

While such policies may contribute to reduce inequality, it obviously does not mean that they can replace redistributive policies, but rather that there is likely to be some scope to reduce inequality through non-distortionary policies.

Also, the model shows that increasing flexibility in labor force reallocation may have positive or negative effects on steady-state growth, including wage growth. A particularly complex relationship between growth, labor market rigidities and inequality thus emerges.

4.5 Appendix to Chapter 4

4.5.1 Appendix 4A

We show here that the constant returns to scale function

$$Y_{ijt} = \tilde{A}_{it} X_{ijt}^{\alpha} L_{ijt}^{\beta} \left[\frac{G_t}{f(n)} \right]^{1-\alpha-\beta} \tag{4.55}$$

can indeed be rewritten as

$$Y_{it} = A_{it} X_{it}^{\alpha} L_{it}^{\beta} \tag{4.56}$$

and with a unique, constant and strictly defined output maximizing number of firms operating in all sectors. Equation (4.55) implies that the public good is an essential good in the production of final output. We assume, along the lines of Barro and Sala-i-Martin (1992), that the public good is subject to congestion. Instead of assuming that congestion is an increasing function of total output, we assume that congestion is a function of the total number of firms operating in the economy. The amount of "effective" public good is thus a decreasing function of the number of firms. In effect, an excessive number of firms using roads, asking for services from the public administration, etc. makes the effective amount available to each firm smaller. We make the very general assumption for the moment that the effective amount of public good is equal to $\frac{G_t}{f(n)}$.

It should be obvious from equation (4.55) that on the one hand, decreasing returns to scale in X and L in any sector acts as an incentive to decrease the size of the firms. There is a countervailing force, however, that necessarily limits the number of firms, as the amount of effective public good decreases with n. We proceed to show that, under certain conditions on $f(n)$, the output maximizing number of firms is equal in any sector and constant across time and independent of the level of public good G.

Suppose for a moment there is only one sector with production function $Y_{jt} = \tilde{A}_t X_{jt}^{\alpha} L_{jt}^{\beta} \left[\frac{G_t}{f(n)} \right]^{1-\alpha-\beta}$, where j is a firm index. If there are n firms of equal size, total output is given by

$$Y_t = nY_{it} = n\tilde{A}_t \left[\frac{X_t}{n}\right]^{\alpha} \left[\frac{L_t}{n}\right]^{\beta} \left[\frac{G_t}{f(n)}\right]^{1-\alpha-\beta} \tag{4.57}$$

where $X_t = nX_{jt}$ and $L_t = nL_{jt}$. This can be rewritten as

$$Y_t = \left[\frac{n}{f(n)}\right]^{1-\alpha-\beta} \tilde{A}_t G_t^{1-\alpha-\beta} X_t^{\alpha} L_t^{\beta} \tag{4.58}$$

Several cases are thus possible:

- $\frac{n}{f(n)}$ is everywhere decreasing in n. In such a case, the optimal number of firm is equal to 1.
- $\frac{n}{f(n)}$ is everywhere increasing in n. In such a case, the optimal firm size in infinitesimally small, and $n \to \infty$.
- $\frac{n}{f(n)}$ is first increasing in n then decreasing in n. In such a case, there exists a unique positive (greater than 1) output maximizing number of firms n^*. This number is constant across time and independent of G.
- $f(n) = n$. In such a case, the output maximizing number of firms in indeterminate.

Empirically, the third and fourth cases are the most relevant. We chose to focus on the third case, which implies a unique, constant and finite number of output maximizing firms. For example, $\frac{n}{f(n)}$ could take the following form: $\frac{n}{f(n)} = \frac{n}{1+\left[\frac{n}{\lambda}-\frac{1}{\lambda}\right]^{\psi}}$, with $\lambda > 1$ and $\psi > 1$. Graphically, we would have

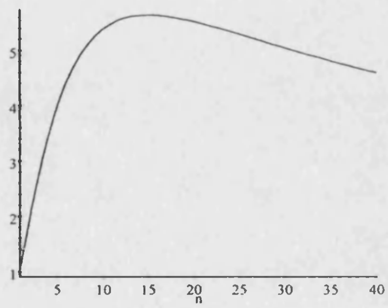


Figure 4.21: $\frac{n}{f(n)}$ with $\lambda = 10$, $\psi = 1.5$

With such a functional form, we assume that there is no congestion when only one firm operates, and that the congestion effect is low for a relatively small number of firms, but that it becomes significant once a certain threshold is attained. Given that n^* is a constant and independent of G, we can write the production function for the sector as a whole as $Y_t = \left[\frac{n^*}{f(n^*)}\right]^{1-\alpha-\beta} \tilde{A}_t G_t^{1-\alpha-\beta} X_t^\alpha L_t^\beta$. If we consider G_t as given in period t, we can rewrite this as

$$Y_t = A_t X_t^\alpha L_t^\beta \tag{4.59}$$

The same argument can be applied if we have a continuum of sectors i in $[0,1]$. In such a case, the production function becomes,

$$Y_{ijt} = \tilde{A}_{it} X_{ijt}^\alpha L_{ijt}^\beta \left[\frac{G_t}{f\left[\int_0^1 n_i di\right]} \right]^{1-\alpha-\beta} \tag{4.60}$$

where n_i is the number of firms in sector i. By symmetry of sectors, we have that $n_i^* = n_j^* = n^*$. All sectors thus have the same number of output maximizing firms. Following the same argument as for the one-sector case, we can thus rewrite the production function for sector i at time t as

$$Y_{it} = A_{it} X_{it}^\alpha L_{it}^\beta \tag{4.61}$$

where $A_{it} = \left[\frac{n^*}{f\left[\int_0^1 n_i^* di\right]} \right]^{1-\alpha-\beta} G_t^{1-\alpha-\beta} \tilde{A}_{it}$. If, for simplicity, we take G_t as constant over time , we have that A_{it} is just a scaled version of $\tilde{A}_{it}$.

4.5.2 Appendix 4B

Define the relative technology parameter a_{it} as

$$a_{it} = \frac{A_{it}}{A_t^{Max}} \tag{4.62}$$

By construction, relative technology parameters are distributed over $[0,1]$.

Technological progress is defined in such a way that a sector experiencing discovery jumps from A_{it} to A_t^{Max}. Let $F(A,t)$ be the cumulative distribution function of absolute technology parameters at time t. By definition, $F(A_0^{Max},0) = 1$.

The density to the left of A_0^{Max} falls at a rate equal to $\frac{dF\left(A_0^{Max},t\right)}{dt} = -F\left(A_0^{Max},t\right)\Psi_t$, where $\Psi_t = \lambda\frac{R_t}{\eta A_t^{Max\frac{1}{1-\alpha}}}$ is the Poisson arrival rate. On the steady-state, Ψ_t is a constant (implying that the absolute amount of resources devoted to R&D, R_t increases together with A_t^{Max}). We have thus the following differential equation with its associated initial condition:

$$\frac{dF\left(A_0^{Max},t\right)}{dt} = -F\left(A_0^{Max},t\right)\Psi \tag{4.63}$$

$$F\left(A_0^{Max},0\right) = 1 \tag{4.64}$$

Also, we have seen that $\frac{\dot{A}_t^{Max}}{A_t^{Max}} = \lambda\frac{R_t}{\eta A_t^{Max\frac{1}{1-\alpha}}}\ln\gamma$, which means that in steady-state

$$\frac{dA_t^{Max}}{dt} = A_t^{Max}\Psi\ln\gamma \tag{4.65}$$

The differential equation (4.63) has the following unique solution:

$$F\left(A_0^{Max},t\right) = e^{-\int\limits_0^t \Psi ds} \tag{4.66}$$

Equation (4.65), on the other hand, implies that

$$A_t^{Max} = A_0^{Max}e^{\ln\gamma\int\limits_0^t \Psi ds} \tag{4.67}$$

From this, we have that $\left[\frac{A_0^{Max}}{A_t^{Max}}\right]^{\frac{1}{\ln\gamma}} = e^{-\int\limits_0^t \Psi ds}$. Using equation (4.66), we find that

$$F\left(A_0^{Max},t\right) = \left[\frac{A_0^{Max}}{A_t^{Max}}\right]^{\frac{1}{\ln\gamma}} \tag{4.68}$$

which is obviously valid for any $s \leq t$, i.e. $F\left(A_s^{Max},t\right) = \left[\frac{A_s^{Max}}{A_t^{Max}}\right]^{\frac{1}{\ln\gamma}}$.

Since A_t^{Max} is a constant at time t, we can scale the random variable A_s^{Max} by A_t^{Max} and obtain that $F\left(\frac{A_s^{Max}}{A_t^{Max}}, t\right) = \left[\frac{A_s^{Max}}{A_t^{Max}}\right]^{\frac{1}{\ln \gamma}}$. The cumulative distribution function of relative technology parameters is thus given by

$$H(a) = a^{\frac{1}{\ln \gamma}}$$

Bibliography

[1] Aghion, Philippe and Peter Howitt (1992). "A Model of Growth through Creative Destruction." Econometrica, March 1992, 60 (2), pp. 323-51.

[2] Aghion, Philippe and Peter Howitt (1998). Endogenous Growth Theory. Cambridge, MA, MIT Press.

[3] Barro, Robert J. and Xavier Sala-i-Martin (1992). "Public Finance in Models of Economic Growth." Review of Economic Studies, October 1992, 59 (4), pp.645-661.

[4] Barro, Robert J. and Xavier Sala-i-Martin (1995). Economic Growth. New York, NY, McGraw-Hill.

[5] Bénabou, Roland (1996a). "Unequal Societies." C.V. Starr Center for Applied Economics, Economic Research Reports, May 1996, 96-17.

[6] Bénabou, Roland (1996b). "Inequality and Growth." C.V. Starr Center for Applied Economics, Economic Research Reports, July 1996, 96-22.

[7] Denison, E. F. (1967). Why Growth Rates Differ: Postwar Experiences in Nine Western Countries. Washington, D.C., Brookings.

[8] Dollar, David and Aart Kraay (2000). "Growth *Is* Good for the Poor." World Bank Working Paper, March 2000.

[9] Grossman Gene M. and Elhanan Helpman (1991a). Innovation and Growth in the Global Economy. Cambridge MA, MIT Press.

[10] Grossman Gene M. and Elhanan Helpman (1991b). "Quality Ladders in the Theory of Growth". Review of Economic Studies, 58, pp. 43-61.

[11] Kuznets, Simon (1955). "Economic Growth and Income Inequality." The American Economic Review, March 1955, XLV (1), pp. 1-28.

[12] Persson, Torsten and Guido Tabellini (1994). "Is Inequality Harmful for Growth?". The American Economic Review, June 1994, 84(3), pp. 600-21.

[13] Pissarides, Christopher A. (1990). Equilibrium Unemployment Theory. Oxford, Basil Blackwell.

[14] Poirson, Helene (2000). "The Impact of Intersectoral Labor Reallocation on Economic Growth". IMF Working Paper, WP/00/104.

[15] Robinson, S. (1971). "Sources of Growth in Less-Developed Countries". Quarterly Journal of Economics, 1971, pp. 379-401.

[16] Schumpeter, J. A. (1934). The Theory of Economic Development. Cambridge, MA, Harvard University Press.

[17] Squire, L. (1981). Employment Policy in Development Countries: a Survey of Issues and Evidence. New York, NY, Oxford University Press.

Chapter 5

Winner-Take-All Markets, Human Capital, Inequality and Growth

5.1 Introduction

Striking back in 1999, George Lucas added an estimated $400 million in earnings to his (financial) empire. In another stars' war, Oprah Winfrey, the Queen Amidala of reality shows, earned a staggering $150 million both in 1999 and 2000. At the same time, 23 people in the U.S.A. had fortunes in excess of $10 billion, and 13 young Americans (under 40) had accumulated fortunes in excess of $1 billion, all of them as "wiz-kids" in new-economy sectors[1].

Meanwhile, President Clinton had to fight hard to raise the Federal minimum wage in 1996 and 1997 to $4.75 and $5.15 per hour. Such increases, however, only marginally compensated the erosion of the real minimum wage initiated in the 1970s. In real terms, the minimum wage reached $10,900 per annum (1998 dollars) on the basis of a 40-hour week, lower than in 1956 and 30 percent lower than in 1968 (Figure 5.1).

[1] Estimates are from Forbes "Celebrity 100" and "400 U.S. Richest" and Fortune "America's Forty Richest under 40."

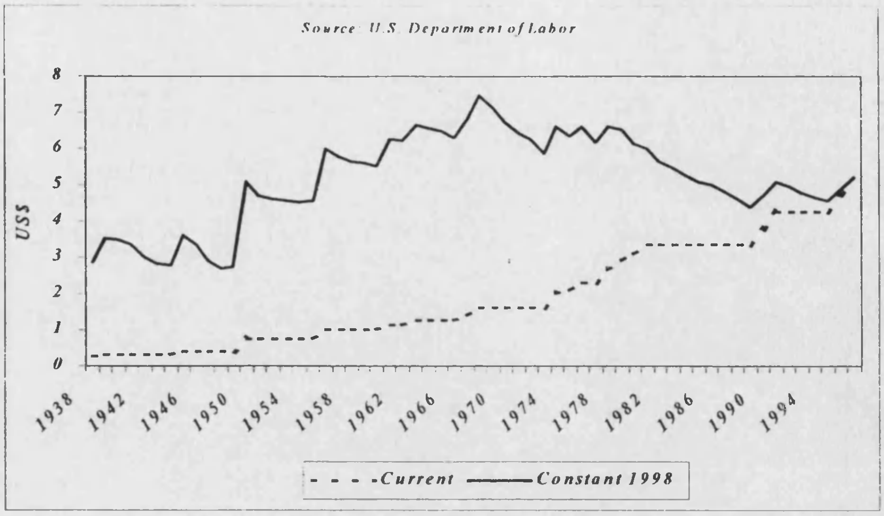


Figure 5.1: Real US minimum wage

On a less anecdotal note, the US Census Bureau reports that the top quintile of households appropriated 49.2 percent of aggregate income in 1998, up from just under 44 percent in 1980. More strikingly still, the top 5 percent of households earned 21.4 percent of aggregate income in 1998, up from 15.8 percent in 1980, while the bottom quintile earned a miserable 3.6 percent in 1998, down from 4.3 percent in 1980.

The same data reveal that the top 40 percent of households earned just under 6 times as much, on average, than the bottom 40 percent of households in 1998, compared with just under 5 times as much in the early 1970s[2]. There has thus been relatively little change in relative earnings positions among these two broadly defined categories of the population (Figure 5.2).

A more marked trend has occurred when the top 20 percent of households are compared with the bottom 20 percent. In such a case, the top households earned on average just under 10 times as much than the bottom 20 percent in the early 1970s, rising to just under 14 times as much in 1998.

The most important trend, however, occurred in the income concentration among the top 5 percent of households. While those earned on average just under 15 times

[2]The US Census Bureau reports a break in the series in 1993. Data are thus not fully comparable. In 1992 (pre-break), however, the highest 40 percent of households earned 5.4 times as much on average than the lowest 40 percent.

as much as the bottom 20 percent in the early 1970s, they earned 24 times as much in 1998.

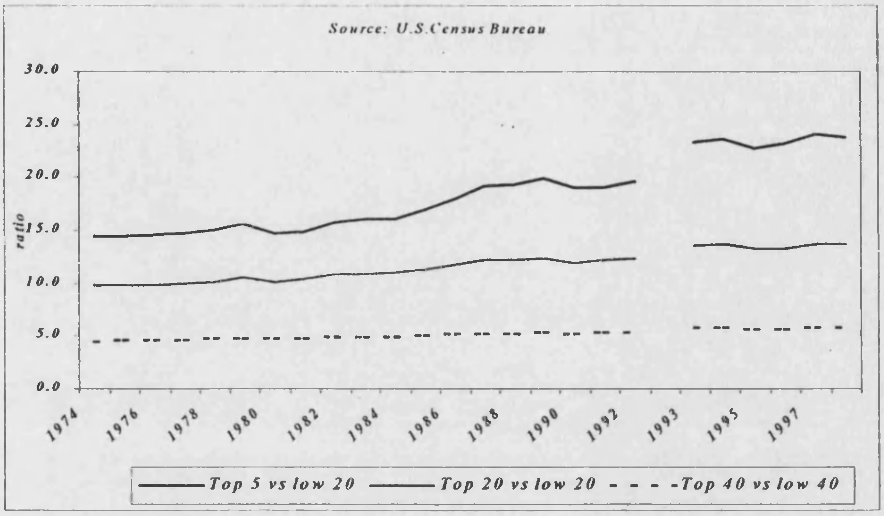


Figure 5.2: Relative average earnings

The rise in concentration of income toward the top is even more radical if one considers the very top (1 percent or less) income earners. Feenberg and Poterba (2000) calculate that the top 0.5 percent of households accounted for just over 6 percent of aggregate adjusted gross income in 1970, almost *doubling* to 11.3 percent in 1995 (Figure 5.3).

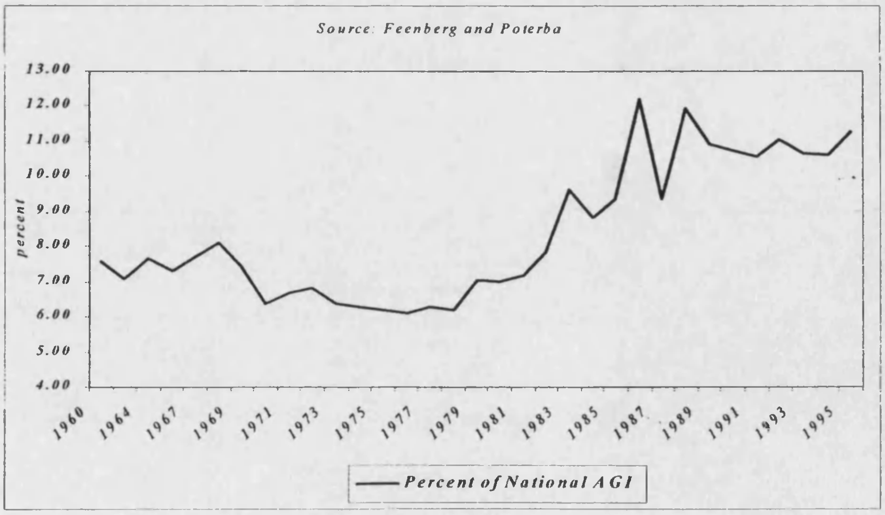


Figure 5.3: Top 0.5 percent of households as percentage of national AGI

While much research has been devoted to understanding the roots of the rising income inequality in the U.S. (and elsewhere) over the past decades, most of the literature has focused on explaining the rising skilled/unskilled wage premium, principally by invoking skill-biased technological change[3]. We argue here that this strand of the literature is unable to explain the rising concentration of income and wealth among a minute percentage of the population, however. Instead, we maintain that the increasing prevalence of winner-take-all (WTA) markets could be one of the main causes of such rising inequality.

We begin this chapter by characterizing WTA markets. We then proceed to model the agents' investment decision in human capital in such WTA markets, in parallel with the incentives to invest in education. This allows us to shed new light on the incentives and growth effects of redistributive taxation, whereby extreme income earners are subject to a tax on part of their income.

5.2 The Winner-Take(s)-(It)-All Phenomenon

Abba (1980) powerfully highlighted, if quite informally at least melodiously and successfully, the existence of WTA phenomena. They were participating themselves, in fact, in one of the most glamourous and obvious WTA markets: popular music. In a more academic, but still rather informal fashion, Frank and Cook (1995) analyze the mechanisms underlying the emergence of WTA markets in American society, and their consequences for economic efficiency, inequality and social ethics. The subtitle of their book, "*How more and more Americans compete for ever fewer and bigger prizes, encouraging economic waste, income inequality, and an impoverished cultural life*", is quite revealing of their analysis.

Frank and Cook, in essence, estimate that WTA markets generate wasteful investments as a result of perverse private incentives to agents responding to competition in peculiarly structured markets. Their book, however, consists essentially of an enumeration of examples of WTA markets, and fails to provide any formal modelling.

[3] See, among others, the studies by Murphy and Welch (1992), Gottschalk (1993, 1997), Autor, Katz and Krueger (1998), Machin and Van Reenen (1998), Acemoglu (1998) and Aghion and Howitt (1998).

Efforts to formalize some WTA markets have been carried out elsewhere. Most notable are papers by Rosen (1981) and Rosen and Sanderson (2000), which consider, theoretically and empirically, respectively, the "economics of superstar" and labor markets in professional sports. Although valuable, these papers are quite specific and do not consider WTA markets globally.

A different strand of the literature has attempted to explain why risk-averse agents choose to participate in lotteries and gambling (fair or unfair). Simultaneous participation in lotteries and insurance purchasing indeed constitutes a puzzle for expected utility maximization theory. Friedman and Savage (1948) first attempted to explain such behavior by proposing a class of utility function compatible with risk aversion and local preference for risk. Conlisk (1993) works along the lines of Friedman and Savage to explain why globally risk-averse agents engage into gambling activities involving small stakes.

The following section underlines the essence of WTA markets and shows how and why they have become more prevalent in the recent past. Typical examples of WTA phenomena are highlighted.

5.2.1 Characteristics and Typical WTA Markets

As is clear from their name, WTA markets are characterized by the fact that of all the participants in the "market" or "game", only the winner gains a prize. The other key feature of WTA markets is the marginal difference in effort, quality, skill, or other key characteristics that exists between the winner(s) and the losers. The difference between winning and loosing, although potentially extreme in terms of payoff, may thus be partly attributable to luck or other external factors outside the participants' control.

Athletes in general compete in what can be considered as the ultimate WTA markets. A tiny minority of aspiring professional athletes ever make it to the top, by definition of competitive sport, and the distribution of payoffs, however large they may be overall, is extremely skewed towards the winners[4]. Also, whether one ends up as a top-earning

[4]Michael Jordan earned around $80 million annually in the last years of his career as an active NBA

professional depends, aside from talent, dedication and effort, significantly on luck: the luck of avoiding injury, finding the proper trainer and infrastructure, starting the sport appropriate to one's potential talent, etc.

The difference between winners and losers is often minute as well, and sometimes even quite subjective. While Mary Lou Retton became an American icon after her Olympic titles in gymnastics in 1984, subsequently earning millions of dollars in endorsement contracts, her vice-champion rapidly fell in oblivion without reaping any significant financial benefits: who remembers Ecaterina Szabo, Kathy Johnson or Julianne McNamara? The difference between those gymnasts, however, cannot be argued to have been more than marginal, and was certainly not perfectly objectively measurable, hence partly attributable to luck[5]. Also, the investment, sacrifices, and effort consented must have been similar.

Competitive sport is only one among the WTA markets with extreme payoff for the winners, yet marginal differences between winners and losers, and where luck (chance) plays a central role. Show-business in general is the other most visible WTA market.

While the list of highly-paid Hollywood actors may seem long, it pales in comparison to the number of professionals in the performing arts struggling to make a living. As the Bureau of Labor Statistics' Occupational Outlook Handbook reports, "*most actors struggle for a toehold in the profession and pick up parts wherever they can. (...) Median annual earnings of actors, directors, and producers were $27,400 in 1998. (...) In reality, earnings for most actors are low because employment is so erratic. Screen Actors Guild reports that the average income its members earn from acting is less than $5,000 a year.*"

Extreme earnings differentials are also to be found among singers and musicians. Technological changes in the reproduction of sound and images have given artists a global reach that could not be thought of before. The number of highly paid performers

player, and before his recent second come-back. The average annual salary in the NBA in 1998 hovered around $2.5 million. Michael Schumacher earned about $50 million in 2000 as a Formula-One driver.

[5]In such instance, say the luck of being charismatic and having a pretty face. In other instances, it could be the luck of being from the appropriate country or having an artistic style fitting a certain mold, ...

pales in comparison to the number struggling to make a living out of singing, acting, etc. It is nevertheless difficult to argue that in terms of talent or effort, the winners are such a huge distance away from an endless list of aspirant stars.

The key distinguishing features of such WTA markets, beyond the winner-take-all nature itself, are thus twofold. First, marginal differences in quality or effort can translate in extreme differences in payoffs. Second, the payoff (outside option) of the losers are usually very low. We will henceforth characterize these markets as "full WTA".

However important and visible the "glamorous" WTA markets described above may be, the extreme concentration of income in the U.S. and its increase over the past decades also find their root in another "strand" of WTA market with some characteristics that distinguish them from the ones just described.

Fortune's "*America's Forty Richest under 40*" list consists almost exclusively of executives of companies in computer and internet related sectors. Thirteen of them held assets in excess of $1 billion in 2000. Forbes' 2000 list of "*400 U.S. Richest*" was still spearheaded by Bill Gates with assets of $63 billion. The total wealth of these 400 people reached $1.2 trillion, equivalent to about 10 percent of the NYSE total market value at the end of 2000. Elsewhere, it is estimated that 2.5 million people in the U.S. hold liquid financial assets in excess of $1 million, representing total assets of $14.3 trillion[6].

America's wealthiests consist, in the end, mostly of highly paid executives, lawyers or investment bankers, more than "glamorous" stars. The extreme income of these people, we argue, also results to a large extent from their participation in a type of WTA market, where marginal differences in talent or effort may translate into extreme differences in payoffs.

The difference between a successful Disney manager and Michael Eisner may indeed be rather slim, whether in terms of education, effort on the job or managerial talent. The difference in earnings, however, is nothing but extreme. The same marginal difference often holds between a successful investment banker and a partner at Goldman Sachs, a successful lawyer and a partner in a top practice in New York, ...

[6]Merril Lynch and Gemini Consulting (2000).

As in the case of full WTA markets, chance and/or peculiar market structures play a key role in determining the outcome of who wins and how large the difference in payoff between a winner and a loser is. Also, marginal differences in talent or effort can translate in extreme differences in payoffs in what we will henceforth characterize as "partial WTA" markets.

A crucial difference between full and partial WTA markets resides in the outside option of the losers. While in full WTA markets, the payoff to the losers tends to be very low, losers in partial WTA markets tend to receive decent payoffs in any case. Losers in such markets are indeed managers that never made it as CEO, lawyers that could not make it to partner in a major practice, computer scientists that did not turn into Bill Gates or Lawrence Ellison, etc.

5.3 Modelling

No full general model is developed. Instead, we try to formalize some of the phenomena observed in WTA markets: why people take part, how they make their decisions to invest in such markets as opposed to education, and the consequences for inequality, aggregate investment in human capital and growth.

5.3.1 Full WTA Markets

Why do people invest in full WTA markets? Is the potential financial payoff sufficiently large to entice agents to invest despite the very small odds of winning? Is it sufficient to compensate for the partial "randomness" of the outcome? Is it sufficient to justify large investments without clear knowledge of the actual probability of winning[7]? It is not obvious at first sight why participation in full WTA markets is empirically so important and generalized. Having answered these questions, it would be interesting as well to investigate whether there should be a link between individual wealth and the (relative) amount of investment in full WTA markets. Casual observation does indeed seem to

[7] The "type" of an agent in full WTA markets is often only revealed after a long period of significant investment: it is difficult to properly assess the probability of success early on.

indicate that poorer families tend to invest relatively more in full WTA markets than richer families, in contrast to investment in education.

A Simple Modelling Tool

To build some intuition and concentrate on the very core issues of the problem, suppose the following setting. A number N of agents indexed by i live for two periods (1,2) and receive an endowment $A_{i,t}$ in both periods. For simplicity and without loss of generality, we postulate that $A_{i,1} = A_{i,2} = A_i$. Agents differ in their individual endowments, which are distributed in $\left[A_l, A^h\right]$. The endowment in period 1 can either be consumed or invested in a full WTA market. The amount W_i invested in the full WTA market provides the agent with a chance of winning a very large payoff S^E in period 2, where by assumption $S^E >> A_i$.

The probability of winning in a full WTA market obviously depends on one's investment, on the total number of participants, and on the others' efforts (investment). To keep the framework as simple as possible, we will simply assume that the probability of winning, denoted as the probability of the binary variable X taking value 1, is given by

$$P(X = 0|W_i) = 1 - \left(1 - e^{-\gamma W_i}\right) p \tag{5.1}$$

$$P(X = 1|W_i) = \left(1 - e^{-\gamma W_i}\right) p \tag{5.2}$$

where γ and $0 < p < 1$ are parameters. This functional form implies that we ignore the effect of others' efforts on an individual's probability of winning. By assumption as well, there are decreasing returns to the investment W_i: the probability of winning is capped at p regardless of the amount W_i invested, and the marginal "productivity" of W_i in increasing the probability of winning is falling in W_i.

In order to focus on the WTA market investment choice, we abstract from the possibility of saving. Assuming $A_{i,1} = A_{i,2} = A_i$ implies zero saving if the subjective discount rate is equal to the interest rate in any case. Consumption in period 2 can thus take two values, depending on whether $X = 0$ (losing) or $X = 1$ (winning):

$$C_{i,2}^{0} = A_i \quad (5.3)$$
$$C_{i,2}^{1} = A_i + S^E \quad (5.4)$$

Agents maximize expected lifetime utility subject to their budget constraint:

$$\max_{W_i} U(C_{i,1}) + \beta E[U(C_{i,2})] \quad (5.5)$$

$$s.t. \quad C_{i,1} = A_i - W_i \quad (5.6)$$
$$E(C_{i,2}) = A_i + E(X) S^E \quad (5.7)$$

This can be rewritten as

$$\max_{W_i} U(A_i - W_i) + \beta \left\{ \left[1 - \left(1 - e^{-\gamma W_i}\right) p\right] U(A_i) + \left(1 - e^{-\gamma W_i}\right) p U\left(A_i + S^E\right) \right\} \quad (5.8)$$

Optimization thus requires that

$$U'(C_{i,1}) = \beta\gamma e^{-\gamma W_i} p \left[U\left(A_i + S^E\right) - U(A_i)\right] \quad (5.9)$$

Equation (5.9) is a rather unconventional Euler equation. The marginal utility of consumption in period 1 is indeed traded off against a marginal increment in the probability of obtaining a non-marginal jump in utility in period 2 from $U(A_i)$ to $U\left(A_i + S^E\right)$. Utility of consumption in period 2 does not appear in marginal terms, but as a difference in utility *levels* between two possible states of the world.

Why Do People Invest in Full WTA Markets at All?

A quick calibration allows us to draw some important conclusions about participation in full WTA markets. Assume a CES utility function $U(C) = \frac{C^{1-\theta}-1}{1-\theta}$, where θ is the reciprocal of the intertemporal elasticity of substitution ($\theta = \frac{1}{\sigma}$).

Empirical studies do not agree on the magnitude of the intertemporal elasticity of substitution σ. Hall (1988) finds evidence "*supporting the strong conclusion that the elasticity is unlikely to be much above 0.1, and may well be zero.*" Beaudry and van Wincoop (1996), on the other hand, find "*evidence indicating that the IES is significantly different from zero and probably close to one*", by using a framework allowing for the existence of rule-of-thumb consumers who consume their current income.

Although empirical studies do not provide definite results, the intertemporal elasticity of substitution is low in all likelihood. Relevant values of θ are thus likely to be (well) above 1, and perhaps as high as 10 or more, reflecting both a low intertemporal elasticity of substitution and a rapidly falling marginal utility of consumption.

A simple calibration can then be done. Assume that endowments are included in $[A_l, A^h]$ and that the winning "prize" of the WTA markets is 100 times the highest endowment A^h. This merely assumes that the prize is "big" for any agent in the economy. We assume also that $\gamma = \frac{5}{A_l}$, which is just a way to postulate that if the agent with least resources invests his entire endowment of period 1 in the WTA market, he obtains a probability of winning that is very close to the maximum achievable, p. A reasonable assumption for p seems to be that $p < 0.01$ and with the actual value probably significantly lower than 1 percent. "Overshooting" the actual value of p obviously makes investment in the WTA market more attractive.

Using parameter values $\theta = 2$, $p = 0.01$, $A_l = 5$, $A^h = 50$ and $\beta = 0.98$, it is easy to verify that there exists no strictly positive value W_i^* satisfying equation (5.9). In other words, the full WTA market is *never* attractive enough to justify *any* investment at all, even for the agents with highest endowments.

A robustness analysis shows that a positive equilibrium level of investment W_i^* can be obtained only for non-realistically low values of θ (say $\theta = 0.5$), which implies an extremely high rate of intertemporal substitution and that the marginal utility of consumption falls very slowly. It is thus not possible to reconcile empirical observations of large investments in full WTA markets, particularly by poorer agents, with a simple expected utility maximization model. This failure is similar to the failure of the expected utility model with risk-averse agents to explain people's participation in lotteries (fair or unfair). This comparison suggests a more comprehensive approach that might explain agents' participation in full WTA markets.

A More Realistic View

Conlisk (1993), observes that standard expected utility theory, treating potential financial gain as the sole motive of gambling, is unable to explain the simultaneous purchase of lotteries and insurance. His paper then shows that if one accepts the assumption that gambling provides some utility of itself, however small, risk-averse behavior (purchasing insurance) can be observed together with risk-loving behavior by a given optimizing agent.

This suggest a similar approach to the WTA issue. It seems rather obvious that individuals derive utility from participation in full WTA markets, in addition to the potential financial gain. Children do enjoy playing sports, acting or singing, and may have a strong preference for these activities as opposed to schooling. We will thus assume that individuals directly derive utility $V(W)$ from their investment W in the WTA market.

Beyond earning high salaries and having an activity that provides utility by itself, winners in WTA markets also enjoy an additional payoff, which is not frequently recognized or modeled by economists, but which can be very important: status. Friedman and Savage (1948) first pointed out that an individual's utility function may depend on his own wealth relative to that of others in society. Similarly Becker, Murphy and Werning (2000) argue that status is an important element in consumers' preferences and that status and income interact in some key ways in the utility function. They assume that utility is defined as $U = u(c, s)$, with $u_c > 0$, $u_s > 0$, $u_{cc} < 0$. Their key assumption is that $u_{cs} > 0$, i.e. that a rise in status s increases the marginal utility of consumption c.

Because a gain in status (becoming a "star") is such an important element of full WTA markets, we follow Becker et al. in assuming that the marginal utility of consumption is increasing in status. For simplicity, we assume that status can only have two states, 0 and 1, depending upon whether one is a loser or a winner, respectively, in the WTA market. A practical way to capture the hypothesis is to assume that the utility function of a winner $U^{w}(C)$ has the property that $U_{c}^{w} > U_{c}$ for any given level of consumption. Or, more specifically, that

$$\begin{aligned} U(C) &= \frac{C^{1-\theta}-1}{1-\theta} \quad \text{in period 1 and if } X=0 \text{ (loser in period 2)} && (5.10) \\ U^{w}(C) &= \frac{C^{1-\theta_w}-1}{1-\theta_w} \quad \text{in period 2 if } X=1 \text{ (winner in period 2)} && (5.11) \\ \theta_w &< \theta && (5.12) \end{aligned}$$

Integrating these two assumptions, we have that agents seek to maximize the following expected lifetime utility:

$$\max_{W_i} U(C_{i,1}) + \varepsilon V(W_i) + \beta E\left[U(C_{i,2})\right] \tag{5.13}$$

$$\begin{aligned} s.t. \qquad C_{i,1} &= A_i - W_i && (5.14) \\ E(C_{i,2}) &= A_i + E(X)\, S^{E} && (5.15) \end{aligned}$$

where U and V are standard concave utility functions and ε is a parameter representing the importance of the utility derived directly from participation in the WTA market.

Optimization thus requires that

$$U'(C_{i,1}) = \varepsilon V'(W_i) + \beta \frac{\partial P(X=1|W_i)}{\partial W_i}\left[U^{w}(A_i+S^{E}) - U(A_i)\right] \tag{5.16}$$

where $\frac{\partial P(X=1|W_i)}{\partial W_i} = \gamma e^{-\gamma W_i} p$ is the "marginal productivity" of W_i in increasing the probability of winning in the WTA market.

In his decision to invest in the WTA market, the agent considers several elements.

On the one hand, there is a trade off between the present utility cost of investing and the direct utility generated by participation in the WTA market. On the other hand, there is a trade off between the present utility cost of investing and the marginal return of that investment in terms of gain in the probability of experiencing a discrete jump in utility in the second period.

It is clear from the assumptions that the LHS of equation (5.16) is continuously increasing in W_i (given A_i) and that the RHS is continuously decreasing in W_i, as it is the sum of the marginal utility of W_i and a term consisting of the product of the marginal productivity of W_i in increasing the probability of winning (decreasing in W_i by assumption) and a constant. A strictly positive and unique optimal level of investment is thus guaranteed if the two schedules cross. Whether they cross is not obvious at first sight.

If participation in the WTA market provides utility independently from consumption, however, every agent will *always* invest a strictly positive amount, regardless of endowment[8]. This would be the case even for small ε because $\lim_{W \to 0} V'(W) = \infty$.[9] In a way, deriving utility from participation in the WTA market allows agents to diversify their "consumption basket", an opportunity they would not pass, regardless of the size of the expected financial payoff in period 2. Equation (5.16) thus determines a unique optimal amount W_i^* given endowment A_i.

It is not possible in general to solve explicitly for the optimal level of investment in the WTA market W_i^*, even if one assumes simple utility functions (say quadratic or logarithmic). Equation (5.16) does define an implicit function between A_i and W_i^*, however. It is thus possible to study the properties of $\frac{dW_i^*}{dA_i}$. Differentiating equation (5.16), we have that

$$\frac{dW_i^*}{dA_i} = \frac{-\frac{\partial[U'(A_i-W_i)]}{\partial A_i} + \beta \frac{\partial P(X=1|W_i)}{\partial W_i} \frac{\partial[U^w(A_i+S^E)-U(A_i)]}{\partial A_i}}{\frac{\partial[U'(A_i-W_i)]}{\partial W_i} - \varepsilon \frac{\partial[V'(W_i)]}{\partial W_i} - \beta \frac{\partial^2 P(X=1|W_i)}{\partial W_i^2} \left[U^w(A_i + S^E) - U(A_i)\right]} \tag{5.17}$$

[8]This result is similar to Conlisk (1993), who elaborates and provides additional proofs in the setting of lottery purchases.

[9]We allow for ε to be small, but not so small that $\varepsilon \to 0$ and $\varepsilon \lim_{W \to 0} V'(W) \neq \infty$.

The sign of $\frac{dW_i^*}{dA_i}$ and hence the shape of the implicitly defined function $W_i^* = h(A_i)$ can be determined by looking at the sign of each element on the RHS of equation (5.17). The first term of the denominator is positive because we assume that the utility function U is concave in C_1. Similarly, because the utility function V and the function $P(X = 1|W_i)$ are both concave in W_i by assumption, the second and third terms are negative, making the whole denominator positive in all cases.

The first term of the numerator is negative by assumption as well. The sign of the second term depends on the sign of $\frac{\partial[U^w(A_i+S^E)-U(A_i)]}{\partial A_i}$, the other part of the term being positive. By assumption, $U^w(\cdot)$ is "less concave" than $U(\cdot)$, but the level of the argument in the function is augmented by S^E in the former and not in the latter. The functions $U^w(A_i + S^E)$ and $U(A_i)$ are thus plotted as follows:

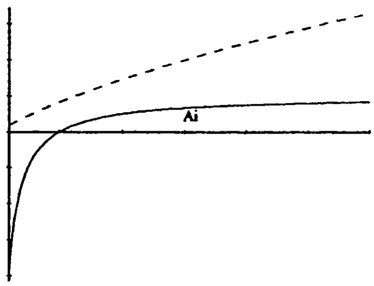


Figure 5.4: $U^w(A_i + S^E)$ (dots) and $U(A_i)$

By assumption, $\lim_{A_i \to 0} U'(A_i) = \infty$. Because $S^E > 0$, however, $\lim_{A_i \to 0} U^{w'}(A_i + S^E) = c \neq \infty$. It must then be that in some initial range of A_i, the marginal utility of an extra unit of endowment is larger for a loser (with utility function U) than for a winner (with utility function U^w), despite the assumption that U is "more concave" than U^w. This latter fact, however, also means that at a certain level (call it A_i^c), the marginal utilities of an extra unit of endowment are equal for both loser and winner. For any level of endowment beyond A_i^c also, $\frac{\partial[U^w(A_i+S^E)]}{\partial A_i} > \frac{\partial[U(A_i)]}{\partial A_i}$. This implies that the difference $U^w(A_i + S^E) - U(A_i)$ falls from a local maximum at $A_i = 0$ to a global minimum at a strictly positive level A_i^c and then rises again. We can thus conclude that

$$\frac{\partial\left[U^w\left(A_i+S^E\right)-U\left(A_i\right)\right]}{\partial A_i} < 0 \quad \forall\ 0<A_i<A_i^c \tag{5.18}$$

$$\frac{\partial\left[U^w\left(A_i+S^E\right)-U\left(A_i\right)\right]}{\partial A_i} > 0 \quad \forall\ A_i>A_i^c \tag{5.19}$$

Using this result, it is possible to determine the shape of $\frac{dW_i^*}{dA_i}$. If $A_i > A_i^c$, we can conclude unequivocally that $\frac{dW_i^*}{dA_i} > 0$. If $0 < A_i < A_i^c$ on the other hand, we have that $\frac{dW_i^*}{dA_i} \gtrless 0$, depending on the assumptions on the utility functions and their parameters. The implicit function $W_i^* = h(A_i)$ can thus take two possible shapes as shown in Figures 5.5 and 5.6. In the first case, if the positive term dominates the negative one when $0 < A_i < A_i^c$, we have

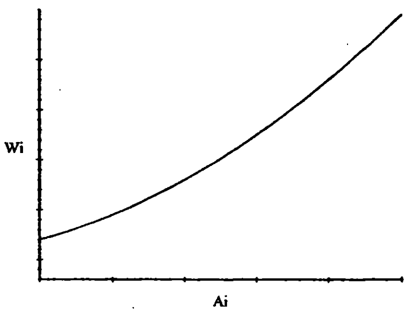


Figure 5.5: $W_i^* = h(A_i)$

In the second case, if the negative term dominates the positive one in a range between $[0, A_i^c]$, we have

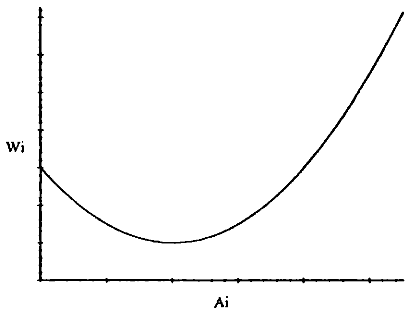


Figure 5.6: $W_i^* = h(A_i)$

The theory outlined above thus highlights two possible relationships between the level of endowment and the optimal amount of investment in the WTA market. The monotonically increasing function is perhaps what may seem most intuitive at first sight. The simple model does show, however, that the relationship between endowment and the level of investment may be more complex than it seems.

The second case indeed shows that the level of investment as a function of endowment may follow a J-curve, which means that, in a stylized way, poor agents are enticed to invest more in a risky activity than "middle-class" agents, despite having a higher marginal utility of consumption in the first period. The attractiveness of investing in the WTA market later increases again as the agent moves from "middle-class" to "upper-class". This, in a real-life example, would imply that those investing most in attempting to become sport or show-business stars, for example, should be found among the "lower-class" and the "upper-class".

Equation (5.16) points to a few more results about the agents' optimal investment strategy. A higher valuation of participation in the WTA market per se, independently of the potential financial gain in period 2 (i.e. a high value of ε), clearly implies a higher choice of investment W_i^*, ceteris paribus.

The importance of status in society, modeled here as a higher marginal utility of consumption for high-status people than low-status people, also influences the investment decision. Societies that give a very high status to winners provide more incentives for investment in WTA markets, and hence higher W_i^*, ceteris paribus. The size of the pure financial payoff of being a winner, S^E, is also positively correlated with investment W_i^*.

The simple modelling tool developed above allows us to derive some interesting results about participation in WTA markets. In particular, it shows that the "rich", and more surprisingly and importantly the "poor", may be those who choose to invest most in these types of risky markets with very uncertain but potentially large payoffs. This simplified framework, however, is limited in the sense that it focuses exclusively on a single investment choice, i.e. whether to invest in a WTA market, and how much. Thus, it cannot be used as is to explain why people may prefer to invest in full WTA markets *as opposed to* investing in other types of human capital, say education.

5.3.2 Partial WTA Markets and Education

Investing time, effort and money to learn singing, acting, running or playing basketball in the hope of becoming a "star" in the field may seem risky enough given the objective odds of winning. In contrast, it seems that education offers a rather safe payoff structure.

One should recognize that education in many cases also offers the possibility to "play" in what we define as partial WTA markets. As we saw in the previous section, the majority of extreme income earners are to be found among CEOs, lawyers, investment bankers, etc. The key difference between full WTA markets and partial WTA markets, as highlighted earlier, resides in the outside option of the losers.

This section seeks to model individuals' choice of investment in education, which we characterize as a partial WTA market. The basic setting is similar to that used in the previous sections, in that we assume two-periods lived individuals maximizing expected lifetime utility.

At this stage, we need to introduce some assumptions about the return to education. While empirical studies differ on the exact specifications and parameters of the return to education, there is little doubt that education provides agents with a reasonable certainty of obtaining some kind of return, and that the payoff (wage) is increasing with the level of education. A high-school graduate can reasonably expect to find a better paid job than a high-school drop-out, but is also likely to be less paid than a college graduate, himself having lower wage expectations than someone with a graduate degree.

Such a relationship is obviously not fully empirically verified, but it is a useful and reasonable stylized fact that can serve as the basis for the assumption that education, in contrast to WTA types of human capital provides the individual with what will be denoted as *certainty return* $S^c(H_i)$. We assume that

$$S^c(H_i) = S^L + \left(1 - e^{-H_i^a b^{-a}}\right) \bar{S} \tag{5.20}$$

Equation (5.20) captures the assumption that education provides the individual with a non-random return that is an increasing function of the amount invested in period

1, H_i. Implicit in equation (5.20) is the assumption that the non-random return to education is somehow capped at $S^L + \bar{S}$, regardless of the amount H_i. This is not to say that the actual return to education is capped at that level, as we will also introduce a random, WTA-type, return to H_i later on. What this means is that a Master's graduate can expect "with certainty" a reasonably high wage, but that he cannot be certain to become one of the extremely highly paid managers.

Based on the amount invested in education H_i in period 1, an individual will thus earn with certainty a fraction $\left(1 - e^{-H_i^a b^{-a}}\right)$ of $\bar{S}$, where a and b are shape and scale parameters, respectively. Using this specification allows us to capture several possible assumptions on the marginal return to education by varying a. Assuming that $a \leq 1$ captures the assumption of a concave return function to investment H_i. Assuming that $a > 1$, on the other hand, allows to encompass a less traditional assumption about the marginal return to investment, which we argue is relevant in the case of education. If $a > 1$, the marginal return to education is increasing at first, reaches a maximum at a strictly positive level of H_i and falls subsequently, as depicted in Figures 5.7 to 5.10.

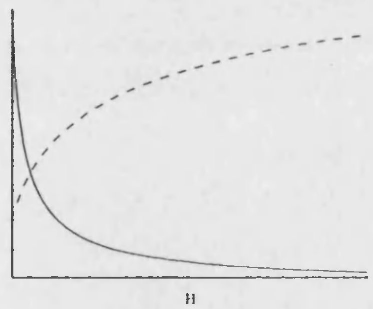


Fig. 5.7: Total (dots) and marg. ret., $a = 0.5$

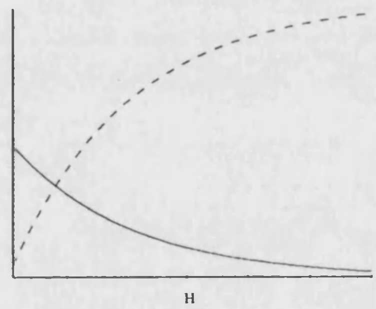


Fig. 5.8: Total (dots) and marg. ret., $a = 1$

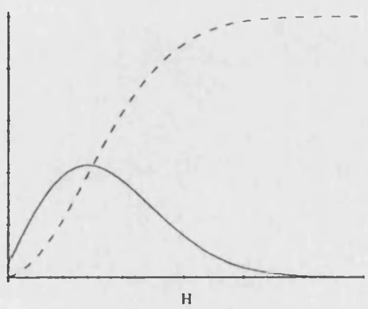


Fig. 5.9: Total (dots) and marg. ret., $a = 2$

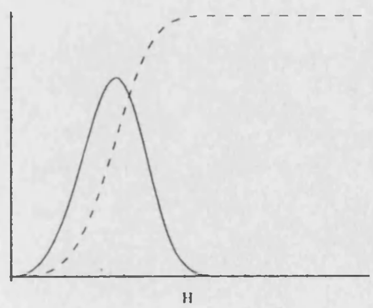


Fig. 5.10: Total (dots) and marg. ret., $a = 4$

A large proportion of the empirical work on the return to schooling postulates, following Mincer (1974), a linear relation between the logarithm of income Y and years of schooling S of the type $\log Y = a + bS + cX + dX^2 + \varepsilon$, where X is the number of years of experience. This linear relation in the logarithm of income implies that the *growth rate* of income for a marginal increase in schooling is constant and is equivalent to assuming that the absolute level of income is exponential in S, i.e. that the marginal return to schooling (in levels of income) is increasing everywhere in S.[10]

A significant amount of empirical work has tended to confirm this log-linear relationship (see Card, 1999). More recent empirical work has sought to test, however, whether the marginal return to education really is constant for any level of schooling. Hungerford and Solon (1987) and Belman and Heywood (1991) find evidence of "sheepskin effects" in the returns to education, whereby wages rise faster with an extra year of education if it leads to a certificate. More interestingly, Belzil and Hansen (2000) conclude that "*the null hypothesis that the local returns to schooling are constant is strongly rejected in favor of a specification where the local returns are estimated using 8 spline segments. (...) The local returns are very low until grade 11 (less than 1 percent per year), increase to 3.8 percent in grade 12 and exceed 10 percent only from grade 14 to grade 16.*"

New empirical work thus seems to point to non-constant returns to education, with low returns at low levels of schooling, peaking around college education before falling subsequently. Such a pattern intuitively reflects rather well the marginal return to education in developed economies. It seems indeed that obtaining a limited level of education offers little additional return with respect to fully unqualified labor. A critical minimum level of education seems to be necessary to ensure access to more qualified and better paid job.

In many ways, education also opens the door to participation in partial WTA markets, as we argued earlier. In order to capture this phenomenon, we assume that, in addition to the certainty return characterized above, education allows individuals to participate in a WTA market, which means that they may obtain a large payoff in addition to the certainty return, with some probability. Let this *uncertainty return* be modeled as

[10] The *level* of income is thus *convex* in schooling in Mincer's specification.

$$S^U(H_i) = XS^E \tag{5.21}$$

where S^E is large and X is a binary random variable as in the previous section.

One can reasonably think that low levels of education provide little chance of winning in education-based partial WTA markets. In order to become one of the top earning lawyers, doctors or executives, one generally needs to hold the appropriate degree. It would thus be reasonable to model the probability of winning similarly to the certainty return to education.

Instead, for simplicity and without weakening the results, we will assume that the probability of winning ($X = 1$) is as follows:

$$P(X = 1|H_i) = \frac{1}{M}\frac{H_i}{\bar{H}} \tag{5.22}$$

where $\bar{H}$ is the average level of investment in education and M is the number of participants.

The probability of winning is thus linear in H_i, which is not fully realistic but is a useful simplification that does not bias the results as it tends to overestimate the incentive to invest in education for the poor. Equation (5.22) implies that if everyone invests the same amount in education, all have a chance $\frac{1}{M}$ of winning.

Consumption in period 2 can take two values, depending on whether $X = 0$ (losing) or $X = 1$ (winning):

$$\begin{aligned} C_{i,2}^0 &= A_i + S^L + \left(1 - e^{-H_i^a b^{-a}}\right)\bar{S} & (5.23) \\ C_{i,2}^1 &= A_i + S^L + \left(1 - e^{-H_i^a b^{-a}}\right)\bar{S} + S^E & (5.24) \end{aligned}$$

As before, agents maximize expected lifetime utility subject to their budget constraint:

$$\max_{H_i} U(A_i - H_i) + \beta \left\{ \begin{array}{c} \left(1 - \frac{1}{M}\frac{H_i}{\bar{H}}\right) U\left[A_i + S^L + \left(1 - e^{-H_i^a b^{-a}}\right)\bar{S}\right] + \\ \left(\frac{1}{M}\frac{H_i}{\bar{H}}\right) U\left[A_i + S^L + \left(1 - e^{-H_i^a b^{-a}}\right)\bar{S} + S^E\right] \end{array} \right\} \tag{5.25}$$

Optimization requires that

$$U'(C_{i,1}) = \beta \left\{ \frac{\partial S^c(H_i)}{\partial H_i} E\left[U'(C_{i,2})\right] + \frac{\partial P(X=1|H_i)}{\partial H_i}\left[U(C_{i,2}^1) - U(C_{i,2}^0)\right] \right\} \tag{5.26}$$

where $\frac{\partial S^c(H_i)}{\partial H_i} = aH_i^{a-1}b^{-a}e^{-H_i^a b^{-a}}\ \bar{S}$ is the "marginal productivity" of H_i in increasing certainty return and $\frac{\partial P(X=1|H_i)}{\partial H_i} = \frac{1}{M}\frac{1}{\bar{H}}$ is the "marginal productivity" of H_i in increasing the probability of being a winner in the partial WTA market.

The second term on the RHS of Euler equation (5.26) is similar to that analyzed in the previous section. It implies that agents trade off marginal utility of consumption in period 1 against an increased probability in jumping from the utility *level* of a loser to that of a winner. Ignoring status effects analyzed above and given that $\frac{\partial P(X=1|H_i)}{\partial H_i}$ is a constant by assumption, the second term on the RHS of equation (5.26) is decreasing in H_i.

The first term on the RHS of equation (5.26) captures the traditional trade-off between marginal utility of consumption in period 1 and the gain in marginal utility in period 2 due to investment in education. By assumption, $E[U'(C_{i,2})]$ is decreasing in $C_{i,2}$ and hence in H_i. The marginal product of education in providing certainty return, $\frac{\partial S^c(H_i)}{\partial H_i}$, on the other hand, behaves differently according to the value of the parameter a and can lead to significantly different conclusions.

Case 1: Traditional Concavity, $a \leq 1$

The case $a \leq 1$ captures the traditional neo-classical assumption that the marginal return to investment is everywhere decreasing and respects the Inada conditions at the lower and upper limits. Under such an assumption, $\frac{\partial S^c(H_i)}{\partial H_i} E[U'(C_{i,2})]$ is everywhere decreasing in H_i and equation (5.26) determines a unique and strictly positive level of investment in education for any given level of endowment A_i, with the equilibrium H_i increasing in A_i. A strictly positive equilibrium always exists even though $\lim_{H_i \to 0} E\left[U'\left(C_{i,2}\right)\right] \neq \infty$, because when $a < 1$, $\lim_{H_i \to 0} \frac{\partial S^e(H_i)}{\partial H_i} = \infty$.

Case 2: Non-Convexities and Educational Poverty Trap, $a > 1$

The case $a > 1$ captures the idea that low levels of education provide relatively little certainty return. Optimal behavior as determined by Euler equation (5.26) is significantly different than in the "traditional" case: three possibilities arise, depending on the level of endowment A_i. These are easiest understood graphically. In all cases, the RHS of equation (5.26) describes a "bell curve" because $\lim_{H_i \to 0} \frac{\partial S^e(H_i)}{\partial H_i} E\left[U'\left(C_{i,2}\right)\right] = 0$ and $\frac{\partial S^e(H_i)}{\partial H_i}$ is increasing in an initial range then decreasing. The LHS is increasing in H_i by assumption on the utility function.

If an agent's endowment is low, equation (5.26) cannot be respected in any case as the two schedules never cross. The optimal decision is thus for the agent to undertake no investment in education whatsoever: maximum utility is reached with $H_i = 0$, $C_{i,1} = A_i$ and $C_{i,2} = A_i + S^L$. This can be seen from Figures 5.11 and 5.12.

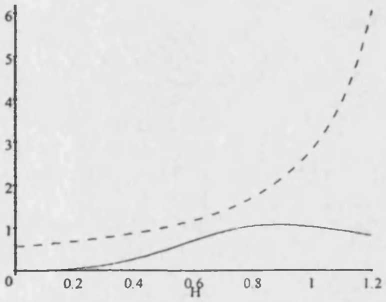


Fig. 5.11: RHS and LHS (dots), low A_i

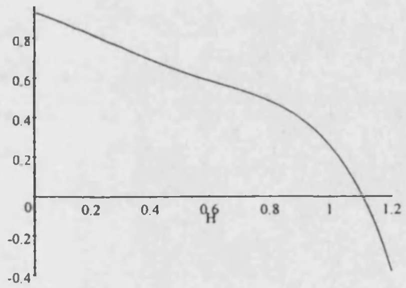


Fig. 5.12: Expected lifetime utility, low A_i

If an agent's endowment is high enough, the two schedules cross at *two* points, the lower crossing point H_i^{cl} and the upper crossing point H_i^{cu}, meaning that the Euler equation is respected twice and that there are two candidates for optimum (Figures 5.13 and 5.14).

The lower crossing point H_i^{cl} corresponds to a local *minimum* of expected lifetime utility, however. In the initial range of investment in education, the marginal return is low, which implies that the gain in marginal utility in period 2 is low as well. The gain in utility in period 2 through higher education, characterized by the area under RHS curve is lower than the loss in utility in period 1, characterized by the area under the LHS curve. Increasing education from $H_i = 0$ to H_i^{cl} thus generates a *net loss* in utility equivalent to the area comprised between the two curves in that range.

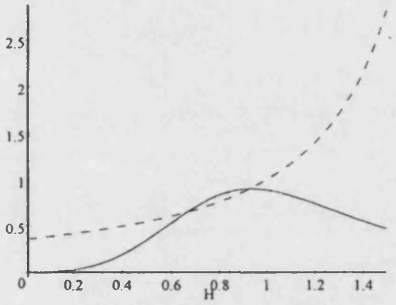


Fig. 5.13: RHS and LHS (dots), medium A_i

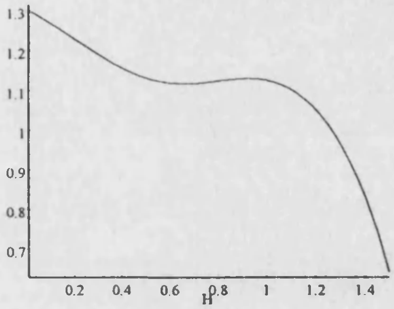
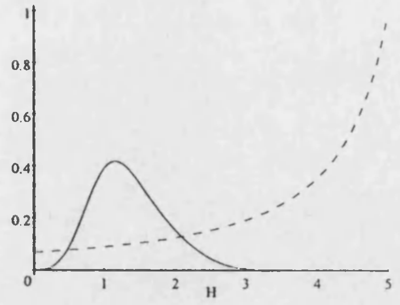


Fig. 5.14: Expected lifetime utility, medium A_i

Beyond H_i^{cl} and up to H_i^{cu}, the gain in utility in period 2 through higher education becomes higher than the loss in utility in period 1, implying a *net gain* in utility equivalent to the area comprised between the two curves in that range.

It should be obvious then that H_i^{cl} corresponds to a local minimum, while H_i^{cu} corresponds to a maximum, which could be either local or global. Whether it is global or local depends on whether the net loss associated with an increase in education from $H_i = 0$ to H_i^{cl} is bigger or smaller than the net gain associated with an increase in education from H_i^{cl} to H_i^{cu}.

If endowment A_i is at an intermediate level, the Euler equation is respected at two points, but H_i^{cu} corresponds to a local maximum. The global maximum is still reached at the corner solution with zero investment in education. If endowment A_i is high enough, on the other hand, H_i^{cu} corresponds to a global maximum (Figures 5.15 and 5.16).



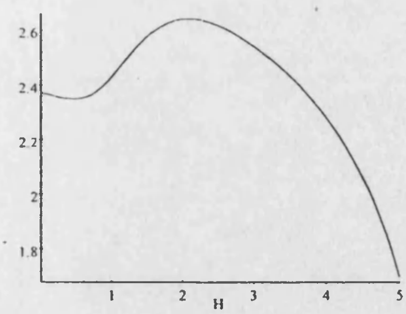


Fig. 5.15: RHS and LHS (dots), high A_i Fig. 5.16: Expected lifetime utility, high A_i

In general then, we have that agents invest nothing in education up until an endowment level A_i^{cu} for which the following condition is respected:

$$\int_0^{H_i^{cu}} \beta \left\{ \begin{array}{c} \frac{\partial S^c(H_i)}{\partial H_i} E\left[U'\left(C_{i,2}\right)\right] + \\ \frac{\partial P(X=1|H_i)}{\partial H_i}\left[U\left(C_{i,2}^1\right) - U\left(C_{i,2}^0\right)\right] \end{array} \right\} dH_i - \int_0^{H_i^{cu}} U'\left(C_{i,1}\right) dH_i = 0 \tag{5.27}$$

By integration, this condition can be rewritten in terms of the original expected lifetime utility maximization problem as

$$\left[U\left(A_i^{cu}\right) - U\left(A_i^{cu} - H_i^{cu}\right)\right] = \beta\left\{E\left[U\left(C_{i,2}\right)\right]_{|H_i=H_i^{cu}} - U\left[C_{i,2}\right]_{|H_i=0}\right\} \tag{5.28}$$

Equation (5.28) means that agents invest in education only from the level of endowment A_i^{cu} at which the loss of utility in period 1 as a result of the investment H_i^{cu} (RHS) is exactly equal to the discounted gain in expected utility in period 2.

Optimal investment behavior thus requires that

$$H_i^* = 0 \quad \text{if} \quad A_i < A_i^{cu} \tag{5.29}$$

$$H_i^* = H_i^{cu} \quad \text{if} \quad A_i > A_i^{cu} \tag{5.30}$$

This means that agents with low endowments are caught in an "educational poverty trap". Because the return to investment in education is low in an initial range, the poor are unable to devote to education the amount of resources necessary to reap the high rewards of the investment.

By implicit differentiation, we can use equation (5.26) to determine the response of H_i^{cu} to an increase in A_i. We have that

$$\frac{dH_i^{cu}}{dA_i} = \frac{-\frac{\partial[U'(A_i-H_i)]}{\partial A_i} + \beta \left\{ \begin{array}{c} \frac{\partial S^c(H_i)}{\partial H_i} \frac{\partial[E[U'(C_{i,2})]]}{\partial A_i}_{|H_i=H_i^{cu}} + \\ \frac{\partial P(X=1|H_i)}{\partial H_i} \frac{\partial[U(C_{i,2}^1)-U(C_{i,2}^0)]}{\partial A_i}_{|H_i=H_i^{cu}} \end{array} \right\}}{\frac{\partial[U'(A_i-H_i)]}{\partial H_i}_{|H_i=H_i^{cu}} - \beta \frac{\partial \left\{ \begin{array}{c} \frac{\partial S^c(H_i)}{\partial H_i} E\left[U'\left(C_{i,2}\right)\right] + \\ \frac{\partial P(X=1|H_i)}{\partial H_i} \left[U\left(C_{i,2}^1\right) - U\left(C_{i,2}^0\right)\right] \end{array} \right\}}{\partial H_i}_{|H_i=H_i^{cu}}} \tag{5.31}$$

The denominator of equation (5.31) is always positive. The first term is positive by assumption on the utility function. The second term can be either positive or negative, depending on the point H_i^{cu} at which the derivative is evaluated. If it is evaluated to the right of the maximum of the RHS of equation (5.26), the derivative is negative and the denominator is obviously positive. If it is evaluated to the left of the maximum of the RHS of equation (5.26), the derivative is positive, but also necessarily smaller than the first term of the denominator. This is so because H_i^{cu} is the second crossing point of a convex (LHS) and a concave (RHS) functions, which means that the difference RHS-LHS is decreasing in H_i from the left to reach zero at H_i^{cu}, and hence necessarily that the derivative of the LHS is larger than that of the RHS. It follows then that the denominator is always positive.

The first term of the numerator is negative by assumption on the utility function. The second term is also negative, however, which means that we cannot sign unequivocally the numerator. As we have seen earlier, however, it must be that no positive equilibrium H_i^* exists for an initial range of endowments. This means that, in order to have positive levels of investment in education H_i^* for some range of endowments at least, H_i^{cu} must be increasing in A_i. If this is true, then the numerator is positive, and the level of investment in education, when different from zero, is increasing in A_i.

We will assume that if this is true for some lower-middle range of endowment A_i (which it must be in order to have at least some agents investing in education), it is also true for higher ranges of endowment. This is most likely as $\frac{\partial S^c(H_i)}{\partial H_i}$ converges to zero for large H_i.

The equilibrium level of investment in education, H_i^*, is thus equal to zero in some initial range of endowment $[0, A_i^{cu}]$, jumps to a strictly positive level at A_i^{cu} and then increases continuously with A_i. This is confirmed by simulations, which additionally show that the increasing section of H_i^* is concave, which was to be expected as the marginal return to education converges to zero as H_i becomes large.

Agents with insufficiently high endowment are caught in the educational poverty trap. Because the return to education is high only for sufficiently large levels of investment, the poor are unable to mobilize enough resources to finance the critical level of education in the absence of financial markets. Non-convexities in the return to education thus reinforce ex-ante inequality among agents: the poor are bound to remain poor, while the rich are in a situation to reinforce their status in period 2 by investing large amounts in education. Escaping the educational poverty trap is not possible at all in the absence of a financial mechanism that allows agents with low endowments to invest sufficiently in education.

Public funding of education seems an obvious way to allow the poor to exit the educational poverty trap. One should note, however, that it may not be sufficient if the publicly provided amount of education (say compulsory "free" schooling) is below the threshold at which the return to education becomes really high. In such a case, poor agents may still optimally decide not to finance by themselves the incremental amount of education that does provide high return, particularly if the level of "free" publicly provided education already involved significant costs, say in terms of forgone earnings or non-publicly covered expenses in books, tutoring, etc.

5.3.3 Human Capital Investment Choices

How do agents choose between investing in full WTA markets and education? Does initial wealth play a role? We have tried in the previous sections to provide a framework to advance some answers to these questions. While it may not be fully realistic to characterize the investment decision as an either/or choice, it is a plausible simplification in that dedication (in terms of time and effort) is necessary for either type of investment, which are difficult to combine in reality.

It appears clearly from the previous two sections that poor agents may be "forced" out of the education choice. In a way, they are pushed into gambling for a very uncertain but high potential financial payoff in the future by investing in an activity that provides some utility in itself.

Agents with higher endowments, on the other hand, are in a position to invest sufficiently large amounts in education so as to obtain significant certainty returns in the future, which makes the investment worthwhile. They also enter the education-based partial WTA market, which gives them a probability of becoming one of the extremely rich.

The investment choices optimally made by agents in this setting are thus such that ex-ante inequality is reinforced ex-post. While the "extremely rich" in period 2 are drawn from the ex-ante poor as well as the ex-ante rich, the vast majority of the ex-ante poor are likely to remain so because they are pushed to invest in risky activities, while the ex-ante rich are in a position to ensure themselves a relatively high income through education.

Paradoxically, one can see that the incentive for the poor to invest in education rather than in the full WTA skill is lower if the minimum wage or unskilled wage, S^L is high. The return S^L is indeed the "outside option" of a loser in the full WTA market (with education $H_i = 0$). The higher the utility of the outside option, the more attractive the gamble is.

5.4 Inequality, WTA Markets, Human Capital and Growth

The rising inequality observed since the 1970s is most often explained in terms of skill-biased technological change, along the lines of Acemoglu (1998). According to this argument, technological progress has generated a shift in relative demand in favor of skilled labor, which has increased the wage premium relative to unskilled labor, despite the increased supply of skilled labor.

While this line of argument is quite convincing in order to explain the increasing wage differential between labor of distinct skill levels, it does not offer much when it comes to explain the sharp rise in income or wealth concentration among the very top of the distribution (top 5 percent or 1 percent). Yet, this has also been an important phenomenon of the 1980s and 90s, which has significantly contributed to rising overall inequality.

Frank and Cook (1995) show informally how WTA markets have become more and more prevalent in the U.S. over the past decades. Our framework seeks to understand why people actually decide to invest in such markets as opposed to investing in education. If one accepts that WTA markets have become more prevalent over the past decades, and that the poor tend to be disproportionately attracted by investment in such markets, a mechanism for rising income inequality at the top is in place. The poor will indeed remain so with a very high probability, while the increased prevalence of WTA markets will generate more winners and more concentration of earnings/wealth at the top.

Lucas (1988) advanced that economic growth is driven by the accumulation of human capital. Galor and Zeira (1993) and Aghion and Howitt (1998) later proposed the idea that inequality may be harmful to growth in the absence of financial markets. Their argument rests on the assumption that the marginal return to investment is decreasing at the individual level and that it is thus productivity-enhancing to distribute any given amount of total investment equally across all individuals. A more equal society would thus generate more growth, they argue, as a result of a better distribution of global investment across individuals.

The next sections consider two issues. We first look at whether inequality can be reduced through taxation on the extreme returns (S^E) without having excessive disincentive effects on the accumulation of human capital. We then briefly look at whether redistribution could generate an increase in the aggregate amount invested in education, and hence raise growth.

5.4.1 Redistributive Taxation and Incentives

Suppose that a proportional tax t is levied on the extreme return S^E. What is its effect on the agents' decision to invest in the full WTA market (W) or in education (H)?

In the case of the full WTA market, Euler equation (5.16) becomes

$$U'(C_{i,1}) = \varepsilon V'(W_i) + \beta \frac{\partial P(X=1|W_i)}{\partial W_i} \left[U^w\left(A_i + (1-t)S^E\right) - U(A_i)\right] \tag{5.32}$$

Differentiating implicitly, the effect of taxation on the equilibrium investment in the full WTA market, W_i^*, is given by

$$\frac{dW_i^*}{dt} = -\frac{\beta \frac{\partial P(X=1|W_i)}{\partial W_i} S^E U^{w'}(\cdot)}{\frac{\partial[U'(A_i-W_i)]}{\partial W_i} - \varepsilon \frac{\partial[V'(W_i)]}{\partial W_i} - \beta \frac{\partial P(X=1|W_i)}{\partial W_i}\left[U^w\left(A_i + (1-t)S^E\right) - U(A_i)\right]} < 0 \tag{5.33}$$

Increasing taxation thus reduces the incentive to invest in the full WTA market. Note, however, that investment in W_i occurs partly as a result of the direct utility derived from the activity. At the limit, even if the financial payoff of being a winner is fully taxed ($t = 1$), agents thus continue to invest a strictly positive amount in W_i.

By implicit differentiation, we can calculate that the effect of taxation on the equilibrium investment in education, H_i^*, is given by

$$\frac{dH_i^*}{dt} = \frac{-\beta\left\{\frac{\partial S^c(H_i)}{\partial H_i} P(X=1|H_i) S^E U''\left(C_{i,2}^1\right) + \frac{\partial P(X=1|H_i)}{\partial H_i} S^E U'\left(C_{i,2}^1\right)\right\}}{\frac{\partial[U'(A_i-H_i)]}{\partial H_i} - \beta \frac{\partial\left\{\begin{array}{c}\frac{\partial S^c(H_i)}{\partial H_i} E\left[U'(C_{i,2})\right] + \\ \frac{\partial P(X=1|H_i)}{\partial H_i}\left[U\left(C_{i,2}^1\right) - U\left(C_{i,2}^0\right)\right]\end{array}\right\}}{\partial H_i}} \tag{5.34}$$

where $C_{i,2}^1 = A_i + S^L + \left(1 - e^{-H_i^a b^{-a}}\right)\bar{S} + (1-t)\, S^E$.

We have already explained why the denominator in equation (5.34) is always positive. The numerator, on the other hand, could be either positive or negative, as the first term is always negative, while the second term is always positive. This means that $\frac{dH_i^*}{dt} \lesseqgtr 0$, i.e. a rise in the tax rate t could *increase* as well as decrease the equilibrium level of investment in education H_i^* (assuming that this level is strictly positive), which may seem paradoxical since taxation reduces the (expected) payoff to the investment.

Two forces are at play to explain this result. On the one hand, taxing extreme income S^E does indeed reduce the incentive to invest in education as the financial payoff for the winners in education-based partial WTA market is reduced. The extra level of utility that is gained by winning ($U\left[C_{i,2}^1\right] - U\left[C_{i,2}^0\right]$) is lower as a result of the tax, which lowers the incentive to invest in education in order to "play" in the WTA game. This is captured by the second term in the denominator of the RHS of equation (5.34).

On the other hand, taxing extreme income S^E also increases the expected marginal utility in period 2, $E\left[U'\left(C_{i,2}\right)\right]$. Because the "certainty return" to education is valued in terms of the expected marginal utility it brings in period 2, raising the latter obviously makes the investment in education for "certainty return" reasons more attractive. This is captured by the first term in the denominator of the RHS of equation (5.34).

Whether increasing the tax rate t ends up reducing or raising the equilibrium level of investment in education H_i^* thus depends on which of the two effects dominates the other. This will obviously depend on the values of the parameters, as well as the level of endowment A_i at which the impact of the tax is evaluated. The most important conclusion one has to draw from this is that taxing extreme income S^E does not necessarily have a negative incentive effect on the accumulation of capital H, and that if the effect is negative, it is likely to be very small as the "certainty return" to investment in education is unaffected by the tax.

5.4.2 Endowment Inequality and Growth

The peculiar structure of full or partial WTA markets is thus such that taxation on extreme returns seems possible without introducing major disincentive effects on the accumulation of human capital. It appears possible then, to put in place a redistributive mechanism financed from taxation on extreme income whereby the redistribution of income allows (some of) the poor to exit the educational poverty trap, without much reducing the level of investment in education through taxation. This could lead, in aggregate, to a higher level of education, and hence higher output.

A redistributive tax on extreme income, beyond having a direct impact on inequality could thus also have a sizeable impact by allowing some segment of the poor to exit the educational poverty trap. What the formal analysis of the previous sections shows, however, is that redistribution needs to be large enough if it is to be effective in taking people out of the educational poverty trap. A small amount of redistribution (say through cash payments) may indeed be insufficient to provide the incentive for poorer people to invest in education rather than in a full WTA skill. The amount of redistribution necessary would depend not only on a person's endowment, but also on the "amount" of education necessary to reach a high marginal payoff.

It might be better then to focus redistribution on a sub-group of the population at first in order to pull that group out of the educational poverty trap, rather than immediately spread the entire proceeds of the tax on attempting -but failing for lack of resources- to pull the entire "poor" group out of the trap. This dynamic setting is not modelled here, but if pulling a "household" out of the educational poverty trap can be perpetuated, it might be more efficient to focus redistribution on successive sub-groups over time, so as to progressively pull the entire population out of the trap. Attempting to do it all at once may indeed be impossible for lack of resources.

5.5 Conclusion

The spectacular rise in inequality in the U.S. over the past decades remains partly a theoretical puzzle. Skill-biased technological change arguments are theoretically appealing and elegant, and have definite empirical relevance. As we have argued, however, they only offer a partial explanation as they cannot tell us why income has tended to concentrate so much at the very top of the distribution, even though this concentration is in large part responsible for the increase in inequality over the past decades.

We have argued that the rise of WTA markets is an important reason of such a concentration of income. This chapter focuses on the reasons why agents might want to focus their investment in human capital on such markets instead of education, by making a few assumptions summarizing key characteristics of WTA markets.

By focusing on the investment decision, we show that the "poor" may in fact be attracted relatively more than the "middle class" in investing in a very risky type of human capital. Given the peculiar shape of returns to education, agents with low endowments may be caught in an educational poverty trap, which means that they end-up focusing their human capital investment on a very risky type of capital that provides direct utility, but that yields a very unlikely financial payoff, even though potentially extreme.

By doing so, ex-ante poverty is reinforced ex-post. We show that, given the incentive structure of investing in education, taxing the extreme payoff of winners in WTA markets may have very small disincentive effects. In fact, taxation of the extreme payoff may, in theory, even generate an extra incentive to invest in education.

Under such circumstances, it appears possible to introduce taxation on extreme income with the view of reducing inequality, pulling the poor out of the education poverty trap and increasing aggregate investment in education (growth). It emerges, however, that attempting to pull too many people too quickly out of the poverty trap may prove unproductive. Instead, focusing limited resources on a more restricted number of people sequentially may be more useful.

Bibliography

[1] Acemoglu, Daron (1998). "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." Quarterly Journal of Economics, vol. 113, 4, pp. 1055-1089.

[2] Aghion, Philippe and Peter Howitt (1998). Endogenous Growth Theory. Cambridge, MA, MIT Press.

[3] Autor, David H., Lawrence F. Katz and Alan B. Krueger (1998). "Computing Inequality: Have Computers Changed the Labor Market?" Quarterly Journal of Economics, vol. 113, 4, pp. 1169-1213.

[4] Beaudry, Paùl and Eric van Wincoop (1996). "The Intertemporal Elasticity of Substitution: an Exploration Using a US Panel of State Data." Economica, 1996, 63, pp. 495-512.

[5] Becker, Gary S., Kevin M. Murphy and Ivan Werning (2000). "Status, Lotteries and Inequality." Mimeo, University of Chicago.

[6] Belman, Dale and John S. Eywood (1991). "Sheepskin Effects in the Return to Education: an Examination of Women and Minorities." The Review of Economics and Statistics, 73, pp. 720-724.

[7] Belzil, Christian and Jorgen Hansen (2000). "Unobserved Ability and the Return to Schooling." Mimeo, Concordia University, Montreal.

[8] Bureau of Labor Statistics, Occupational Outlook Handbook, 2000.

[9] Card, David (1999). "The Causal Effect of Education on Earnings." Handbook of Labor Economics, Card and Ashenfelter editors, North-Holland.

[10] Clotfelter, Charles T. and Philipp J. Cook (1987). "Implicit Taxation in Lottery Finance." National Tax Journal, vol. XL, 4, December 1987, pp. 533-546.

[11] Clotfelter, Charles T. and Philipp J. Cook (1990). "On the Economics of State Lotteries." Journal of Economic Perspectives, vol. 4, 4, Fall 1990, pp. 105-119.

[12] Clotfelter, Charles T. and Philipp J. Cook (1993). "The Peculiar Scale Economies of Lotto." American Economic Review, vol. 83, 3, June 1993, pp. 634-643.

[13] Conlisk, John (1993). "The Utility of Gambling." Journal of Risk and Uncertainty, 6, pp. 255-275.

[14] Feenberg, Daniel R. and James M. Poterba (2000). "Taxing the Rich. The Income and Tax Share of Very High-Income Households, 1960-1995". AEA Papers and Proceedings, May 2000, pp. 264-270.

[15] Frank, Robert H. and Philip J. Cook (1995). The Winner-Take-All Society. The Free Press.

[16] Friedman, Milton and Leonard J. Savage (1948). "Utility Analysis of Choices Involving Risk." Journal of Political Economy, vol. 56, 4, pp. 279-304.

[17] Galor, Oded and Joseph Zeira (1993). "Income Distribution and Macroeconomics." Review of Economic Studies, vol. 60, 1, pp. 35-52.

[18] Goolsbee, Austan (2000a). "Taxes, High-Income Executives, and the Perils of Revenue Estimation in the New Economy." AEA Papers and Proceedings, May 2000, pp. 271-275.

[19] Goolsbee, Austan (2000b). "What Happens When You Tax the Rich? Evidence from Executive Compensation." Journal of Political Economy, vol. 108, 2, pp. 352-378.

[20] Gottschalk, Peter (1993). "Changes in Inequality of Family Income in Seven Industrialized Countries." American Economic Review, vol. 83, 2, pp. 136-142.

[21] Gottschalk, Peter (1997). "Inequality, Income Growth, and Mobility: the Basic Facts." Journal of Economic Perspectives, vol. 11, 2, pp. 21-40.

[22] Gregory, Nathaniel (1980). "Relative Wealth and Risk Taking: A Short Note on the Friedman-Savage Utility Function." Journal of Political Economy, 1980, vol. 88, 6, pp.1226-30.

[23] Hall, Robert E. (1988). "Intertemporal Substitution in Consumption." Journal of Political Economy, 1988, vol 96, 2, pp. 339-357.

[24] Hartley, Roger and Lisa Farrell (1998). "Can Expected Utility Theory Explain Gambling?", Mimeo, Keele University.

[25] Hungerford, Thomas and Gary Solon (1987). "Sheepskin Effects in the Returns to Education." The Review of Economics and Statistics, 69, pp. 175-177.

[26] Machin, Stephen and John Van Reenen (1998). "Technology and Changes in Skill Structure: Evidence from Seven OECD Countries." Quarterly Journal of Economics, vol. 113, 4, pp. 1215-1244.

[27] Merril Lynch and Gemini Consulting (2000). "World Wealth Report 2000."

[28] Mincer, Jacob (1974). Schooling, Experience and Earnings, New York, NY, Columbia University Press.

[29] Murphy, Kevin M. and Finis Welch (1992). "The Structure of Wages." Quarterly Journal of Economics, vol. 107, 1, pp.285-326.

[30] Psacharopoulos, George (1994). "Returns to Investment in Education: a Global Update." World Development, 1994, vol. 22, 9, pp. 1325-1343.

[31] Rose, Nancy L. and Catherine Wolfram (2000). "Executive Compensation. Has the "Million-Dollar Cap" Affected CEO Pay?" AEA Papers and Proceedings, May 2000, pp. 197-202.

[32] Rosen, Sherwin (1981). "The Economics of Superstars." American Economic Review, vol. 71, 5, pp. 845-858.

[33] Rosen, Sherwin (1997). "Manufactured Inequality." Journal of Labor Economics, 1997, vol. 15, 2, pp. 189-196.

[34] Rosen, Sherwin and Allen Sanderson (2000). "Labor Markets in Professional Sports." NBER Working Paper 7573.

[35] Sadler, Michael A. (2000). "Escaping Poverty: Risk-Taking and Endogenous Inequality in a Model of Equilibrium Growth." Review of Economic Dynamics, 2000, 3, pp. 704-725.

[36] Schoemaker, Paul J. H. (1982). "The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations." Journal of Economic Literature, vol. XX, June 1982, pp. 529-563.

[37] US Census Bureau (2000). "The Changing Shape of the Nation's Income Distribution, 1947-1998." US Department of Commerce, US Census Bureau.

[38] Walker, Ian (1998). "The Economics of Lotteries." Economic Policy, 27, December 1998, pp. 359-392.

Chapter 6

Conclusion

The links from growth to inequality and inequality to growth are manyfold, complex and particularly important for reasons related to poverty and social justice in particular, but also efficiency and welfare maximization. The complexity and breadth of the issues involved clearly prevent all-encompassing analyses. Instead, we focused on a few specific questions here, which has allowed us to derive some new theoretical results with normative implications.

Focusing on developing countries at first, we sought to determine how within-country inequality and international convergence (cross-country inequality) would behave in a dual economy subject to "globalization", narrowly defined as an opening of the economy to capital flows from abroad. In particular, we wanted to address the following questions: what determines the size and timing of capital flows from abroad? Why do certain countries attract so much more capital than others? Who benefits most, domestically, from capital flows, the "rich" or the "poor"?

The model elaborated in Chapter 3 allows us to answer these questions. The dual economy generates within-country inequality. While the rich operate, by assumption, in the sector that is susceptible to attract capital from abroad, we show that they do not disproportionately benefit from opening the economy to capital flows. "Globalization" in fact does not widen the relative gap between rich and poor at first, even though the absolute wage gap increases. This rather counter-intuitive result (the rich operate with more capital as a result of the opening up of the economy and hence their labor productivity increases) is explained by the complementarity in "output" between the two sectors of the economy, which is absent in the traditional approach in the line of Kuznets. An expansion of overall productivity in the "rich" sector thus also benefits the "poor" sector.

While "globalization" does not widen relative inequality at first, we show that it doubly benefits the poor in that it immediately raises their productivity and, most importantly, in that it allows the gap between rich and poor to close at an accelerated pace. While the gap would eventually close in any case in the framework of the model, opening up allows faster reduction in within-country inequality.

Opening up the economy to capital flows also benefits rich and poor alike by allowing a faster pace of international convergence, i.e. a faster reduction in cross-country inequality. The duality of the economy allows us to show, however, why convergence is not immediate despite international capital flows. It also underlines that the amount and timing of capital flows from abroad ultimately hinges upon domestic factors. Of particular importance are the development of good domestic infrastructures in terms of education and health, basic transportation and other services, and hence domestic saving. In other words, foreign investment (capital flows in general) follows domestic investment rather than precedes it.

The dual economy model of Chapter 3 is a useful characterization of a typical developing country, but it does not properly represent the structure of a developed economy, and it generates the conclusion that within-country inequality converges to zero in the long run. This, obviously, we do not observe in even the most developed economies. While it appears empirically that the more developed economies tend to have lower levels of inequality than lesser developed countries, inequality remains significant. There appears thus to be some kind of "steady-state" level of inequality.

The model of Chapter 4 seeks to investigate the roots of such an equilibrium level of inequality. Using an endogenous growth model, we typify the "developed" economy as one with a multiplicity (a continuum) of sectors, each mobilizing a share of the labor force. Growth is endogenized by allowing monopolistic production of intermediate goods of particular vintages and hence generating an incentive for R&D.

We then show how the labor force is permanently being reallocated across sectors, according to the level of technology attained in any sector at any given time, relative to all other sectors. Introducing frictions in the labor reallocation process, we show that equilibrium inequality emerges. It turns out then, that inequality is essentially determined by the extent of frictions in the labor reallocation process.

Such frictions, we show, influence steady-state growth in non-trivial ways. In fact, sub-optimal allocation of labor across sectors (an allocation that fails to equalize marginal productivities) may generate higher steady-state growth than an optimal allocation in the sense of equalized marginal productivities.

As a result, we show that the relationship between inequality and growth can be particularly complex. Different countries may exhibit different relationships depending on their characterizing parameters, and even a given country may or may not face a trade-off between equilibrium growth and inequality depending upon where it finds itself relative to the growth-maximizing amount of labor reallocation rigidities.

In Chapter 4, we thus draw a particularly nuanced view of the growth/inequality nexus through the introduction of a third parameter: frictions in labor force reallocation across sectors. This allows us also to shed some new light on another segment of the literature, that seeks to explain growth patterns through sectoral shifts in the labor force.

The prevalence of inequality even within the most developed countries led us to seek to explain the causes of such equilibrium inequality. There has also been, however, a tendency for inequality to increase again in many of the most developed economies over the past decades. In Chapter 5, we look critically at the traditional explanations of this phenomenon as they have been put forward in the literature. We underline that such explanations are unable to capture an important issue at the root of the recent increase in inequality: the extreme concentration of income at the top.

We then look at one potential explanation for such a concentration: the rising prevalence of winner-take-all (WTA) markets. Instead of modeling such markets in details, we seek to understand why agents may wish to participate in them. In particular, we show

how the poor may be disproportionately drawn to invest in such markets as opposed to investing in education, hence reinforcing ex-ante inequality.

We show that expected utility theory relating investment decisions exclusively to expected financial gains is unable to explain participation in typical WTA markets. Given the payoff structure and incentives to invest in WTA markets and the peculiar payoff structure of education, we explain why the poor may be drawn into investing in WTA markets rather than in education, and more so than the rich. This phenomenon, apart from perpetuating inequality, also has potential effects on growth, inasmuch as one accepts that growth is driven by the accumulation of certain types of human capital.

Analyzing the payoff structure of WTA markets, we also show that it may be possible to tax extreme income earners without affecting incentives to invest very much. There might thus be room for a redistributive tax that would not excessively discourage investment, while at the same time allow the poor to be taken out of poverty sustainably and positively affect growth.