United Nations Development Programme
Regional Bureau for Latin America and the Caribbean

Research for Public Policy
Inclusive Development

ID-10-2009

Markets and the Dynamics of Inequality: Theoretical Perspectives♦

Jaime Kahhat *

Key words: inequality, poverty, intergenerational transmission, capital market
JEL Codes: D31, D63, I32, O12, E24

♦ Document prepared for UNDP Project on "Markets, the State and the Dynamics of Inequality: How to Advance Inclusive Growth" (Co-funded by RBLAC and BDP-Poverty Group).

* Universidad de Alicante.

Please cite this document as: Kahhat, J. (2009) "Markets and the Dynamics of Inequality: Theoretical Perspectives", Research for Public Policy, Inclusive Development, ID-10-2009, RBLAC-UNDP, New York

Markets and the Dynamics of Inequality: Theoretical Perspectives

Jaime Kahhat
Universidad de Alicante

1 Introduction

In its *2006 World Development Report* the World Bank argues that there are two broad sets of reasons to believe that inequality can be detrimental to long-run growth and development. First, high levels of economic and political inequality often lead to the creation of poor institutional arrangements that favor the interests of powerful elites. By protecting personal and property rights selectively, poor institutions are likely to restrain social mobility and reduce the investment and innovation opportunities of the majority of the population. Second, a highly unequal distribution of wealth may exacerbate the efficiency loss originating from market imperfections, in particular capital market imperfections. When capital markets are less than perfect, resources do not flow where expected returns are higher and the poor may be credit constrained regardless of the expected returns on their investments. By increasing the number of credit constrained agents in the economy, higher levels of wealth inequality usually lead to more inefficient allocations of resources.

The political and economic links between unequal opportunities and inefficient outcomes affect the performance of the economy in the short run but also tend to reproduce the initial pattern of inequality over time. For example, children of wealthy parents are usually better nourished, receive more and better education, and are healthier than children born in poor families. As a result, children born in poor families usually earn less throughout their lives than children born in wealthy families, reproducing the initial pattern of economic inequality. Moreover, since greater economic power often leads to greater political power, the poor will continue to have less voice in the political process. The wealthy will continue to be able to alter institutions to better serve their interest while the poor will continue to be unable to influence the institutional arrangements that hinder their ability to accumulate wealth. In this way, a highly unequal initial distribution of wealth and power may become inimical to long-run growth and development.

The previous arguments suggest that the functional aspects of inequality are more severe in poor developing countries without deeply rooted democratic institutions and well

developed financial structures than in their wealthier counterparts. Therefore, a better understanding of the political and economic mechanisms that tend to reproduce unequal opportunities over time is central to the design of policies aimed at promoting economic prosperity and development. This article will review the main theories examining the economic mechanisms through which inequality and development are interconnected. These theories explore the economic forces that shape the evolution of the distribution of wealth. At the same time, they explain the way in which wealth inequality influences economic performance. A central question asked by most of this literature is: Do historical inequalities affect the long-run performance of an economy? In this article we will provide a detailed discussion of the main theoretical approaches attempting to answer this question.

The article is organized as follows. Section 2 reviews the main theories examining the market forces that shape the intergenerational evolution of inequality. Benabou (1996) refers to these theories as the capital market imperfections approach. In Section 3 we examine the relationship between inequality and development implied by the transitional dynamics of the theories in the capital market imperfections approach. Section 4 discusses the recent theoretical literature on the effect of technological change on inequality and Section 5 concludes.

2 The Capital Market Imperfections Approach

Does a market economy display an intrinsic tendency to reduce differences in wealth and income across households, or does it exacerbate those differences? Is there a mechanism inherent in the workings of a market economy that can generate inequality? One point of view argues that competitive markets tend to reduce differences in wealth and income across households. If inequality persists in the long run it is only because idiosyncratic shocks create differences in earnings within each generation. Inequality would eventually vanish if these random shocks were not present. An alternative view argues that the natural tendency of a market economy is to generate inequality, even if all agents are ex-ante identical. In this view, inequality is a result of the way in which returns to different occupations are determined by competitive markets. In between these two views lies a third

view that argues that a market economy displays neither an intrinsic tendency to generate inequality nor an intrinsic tendency to reduce differences in wealth across households. Whether competitive markets reduce or exacerbate the initial level of inequality depends on the underlying characteristics of the economy. Mookherjee and Ray (2006) refer to these alternative points of view as the *equalization*, *disequalization*, and *neutrality* views, respectively. These alternative views provide radically different accounts of the role played by market forces on the intergenerational transmission of inequality. Consequently, the policy implications derived from them will vary significantly. This section reviews the main theories in each of these views and provides simple models to illustrate their main arguments[1].

The theories discussed in this section take as historically given an initial distribution of wealth and then proceed to determine how market forces affect the evolution of this distribution. A central question they address in the process is: Do historical inequalities matter for economic development? In other words, does the initial distribution of wealth affect long-run economic performance? A general idea, common to all the literature cited here, is that a comprehensive understanding of the functional aspects of inequality requires taking into account the role of capital market imperfections in the intergenerational transmission of inequality. When capital markets are perfect, investment decisions depend exclusively on the expected return on investments and on the market price of capital, adjusted for the extra risk. Economic agents can always borrow what they need and their investment decisions are not influenced by the way in which aggregate wealth is distributed across households. Therefore, the distribution of wealth does not affect economic performance. Furthermore, wealth inequality in one generation does not create unequal opportunities for the next generation. Markets, however, are not perfect and intergenerational transfers do play a significant role in the transmission of inequality and the investment decisions of economic agents. Parental investments in children's human capital (e.g., investments in nutrition, education, and health), parental investments in children's social capital (e.g., maintaining or improving a family's social status, reputation,

[1] The technical details of each of these models and the proofs of some of the main results are provided in the Appendix.

and connections), and financial bequests generate a playing field that is far from being level. When capital markets are less than perfect, intergenerational transfers translate the wealth inequality in one generation into unequal opportunities for the next generation, reducing the degree of social mobility and reproducing the initial pattern of inequality in the short run. Moreover, these unequal opportunities affect individual investment decisions and may lead to inefficient economic outcomes. In these cases, policy interventions aimed at reducing inequality can improve the short-run efficiency of the economy. Whether unequal opportunities in one generation have persistent effects or not is one source of disagreement between the equalization, neutrality, and disequalization views.

2.1 The Equalization View

In its inception, the equalization view emerged as a reformulation of the neoclassical theory of economic growth. One of the main predictions of the neoclassical theory of growth is the long-run convergence of per capita incomes across countries. With a concave production function, the return on the capital stock should be higher in poor countries, where capital is scarce, than in rich countries. The higher return should encourage a higher rate of investment implying that poor countries will grow faster than rich ones, reducing the income gap between them. In this way, poor countries would eventually catch up with their wealthier counterparts by gradually accumulating capital over time. A similar argument lies at the core of early theories in the equalization view, with convergence of wealth occurring at the family level instead of at the country level (e.g., Champernowne, 1953; Becker and Tomes 1979, 1986; Loury, 1981). The difference is that the tendency towards convergence predicted by the neoclassical theory of growth is partly offset in the equalization view by idiosyncratic shocks to the income of individuals. Most theories of income distribution rely on ongoing shocks to abilities or opportunities as a natural source of intergenerational social mobility[2]. However, neither in the neutrality view nor in the disequalization view do these shocks acquire the prominent role they play in the equalization view. The logic of convergence is such an integral part of the equalization view that inequality can only persist

[2] These shocks may represent, among other things, the outcomes of risky investments or, alternatively, the random assignment of individual attributes such as innate ability. Becker and Tomes (1979) refer to the former as "market luck" and to the latter as "endowment luck".

in the long run as a result of random shocks to the income of individuals. This point is emphasized by Aghion and Bolton (1992, p. 603) who argue that "[a]ny attempt at explaining intergenerational mobility as well as the endogenous generation of inequality must include uncertainty in individual incomes."

The logic behind the central predictions of early theories in the equalization view can be succinctly stated by using a simple model. Consider an economy with a continuum of identical individuals who live for one period and have a single offspring. Individuals start their lives with an initial wealth, say *wi*, inherited from their parents and choose an allocation of investments to maximize their expected end-of-period wealth. At the end of their lives they consume a fraction $1-\beta$ of their end-of-period wealth and bequeath the rest to their children. These bequests become the initial wealth of the next generation. There are two investment opportunities in this economy. There is a safe asset that yields a low constant return of $r > 1$ per unit of capital invested. In addition, individuals can undertake a risky entrepreneurial project that yields a return of $\alpha f(k)$ when total investment is k. The production function $f(\cdot)$ is strictly increasing and strictly concave in k. The term α represents an idiosyncratic shock independently and identically distributed for all agents on the interval [0, 2] with $E[\alpha] = 1$. The optimal capital investment in the entrepreneurial project, k^*, is given by the equality of the expected marginal productivity of capital and the return on the safe asset, $f'(k^*) = r$. Credit and insurance markets are missing and individuals' investment decisions are wealth constrained[3]. Therefore, those individuals for whom $w_i \leq k*$ will invest all their wealth in their projects and those for whom $w_i > k^*$ will invest k^* in their projects and $w_i - k^*$ in the safe asset.

Society can be divided into two social classes: the poor, whose inherited wealth is below k^*, and the rich, whose inherited wealth is above k^*. It follows from the concavity of the production function that wealth inequality will lead to an inefficient allocation of investments across families. The allocation of investments is inefficient because the expected marginal return on capital investments is greater for poor entrepreneurs than for

[3] The assumption of missing credit and insurance markets implies an extreme form of capital market imperfection. Our results would not change substantially if a less severe form of capital market imperfection is used as long as some poor agents are credit constrained and individuals cannot insure away all idiosyncratic risk. For simplicity, we will continue to use the assumption of missing capital markets throughout this section.

rich ones. This implies that, for a given average wealth in the economy, the greater the level of inequality, the lower the rate of growth. Two conclusions follow immediately from this result. First, a one time redistribution of wealth from the rich to the poor can improve efficiency, increasing future aggregate wealth and the rate of growth in the short run. Redistribution increases future aggregate wealth by reallocating resources from activities with a low return, investments in the safe asset, to activities with a higher expected return, the entrepreneurial projects of poor agents. The extent of the efficiency gains from redistribution depends on the average wealth in the economy. Given a distribution of wealth the efficiency gains from redistribution decrease as the average wealth in the economy increases. Indeed, the decentralized allocation of investments will be efficient if the economy is rich enough such that everyone can afford the optimal level of investment. Second, since the expected marginal return on capital investments is higher for poor entrepreneurs, the average wealth of poor families should grow faster than the average wealth of rich families. This is a weaker form of convergence than that predicted by the neoclassical theory of growth. In particular, this result does not imply that, at some point in the future, everyone's wealth will converge to a common long-run average wealth. Inequality persists in the long run because idiosyncratic shocks generate earnings inequality within each generation. However, long-run inequality will be independent of the initial level of inequality in the economy. To make this point clear, we need to look at the steady state of the model.

Consider the evolution of a single family's wealth. In a world without uncertainty in which α is constant, we could easily trace the wealth of a single family by looking at the optimal allocation of investments in each generation. In any period, given an initial wealth for the current generation, the optimal allocation of investments determines the end-of-period wealth and the bequest left to the next generation. In this way, we can generate the path followed by a family's bequests by using the bequest left by one generation as the initial wealth of the following generation. If we trace the path followed by these bequests we could determine the family's long-run wealth. In this case, the concavity of the production function implies that the wealth of each family converges to a common long-run wealth and inequality eventually vanishes. However, once we allow α to be random, the

intergenerational evolution of a family's wealth becomes a stochastic process. In any period, given an initial wealth for the current generation, the optimal allocation of investments and the distribution of idiosyncratic shocks determine a probability distribution over the set of possible end-of-period wealth. Since individuals bequeath a constant fraction of their end-of-period wealth, this defines a probability distribution over the set of possible bequests for the next generation. Using the distribution over bequests of one generation as the distribution over the set of possible initial wealth for the next generation we could generate the sequence of probability distributions over bequests for each generation. Under some standard assumptions about the production function and the distribution of idiosyncratic shocks, this sequence of distributions converges to a unique limiting distribution irrespective of the family's initial wealth[4]. Since all families are identical, except for their initial wealth, we can interpret this limiting distribution as the long-run distribution of wealth for the entire population.

The properties of the long-run wealth distribution have profound implications for our understanding of the intergenerational transmission of inequality and for the design of policies aimed at reducing inequality. First, the distribution is unique and globally stable. In other words, starting from any initial distribution of wealth the economy will eventually converge to the same long-run wealth distribution. This implies that historical inequalities do not play a role in determining long-run macroeconomic performance. Furthermore, this also implies that the efficiency gains from a one time redistribution of wealth are short lived. Irrespective of the way wealth was distributed, or redistributed, in the past the economy will eventually converge to its unique long-run distribution. Second, the long-run wealth distribution is ergodic, so there are no poverty traps in the economy. There is a positive probability that a poor family may become rich in finite time and vice-versa. This implies that the average wealth of all families should be equal if computed over sufficiently long time horizons. Finally, inequality within each generation and social mobility across

[4] The stochastic process followed by the wealth of a family is linear. Therefore, we can use standard techniques to prove existence, uniqueness, and global stability of a limiting distribution. The fact that the transition rule for the wealth of a family does not depend on the distribution of wealth for the entire population at any point in time is crucial for this and other results. This characteristic is shared by almost all theories in the equalization view.

generations are inextricably intertwined. In the long run, the level of inequality and the degree of social mobility are jointly determined by the distribution of the idiosyncratic shocks or, as Becker and Tomes (1979) define it, the "inequality in luck". A society with substantial inequality in luck will experience substantial wealth inequality but a large degree of social mobility. The close relation between inequality and social mobility predicted by these models raises one serious question from a normative point of view. As Loury (1981, p. 851) points out, "[t]he basic problem for a normative analysis of inequality raised by this joint determination is that the concept of a 'more equal' income distribution is not easily defined. For instance, how is one to compare a situation in which there is only a slight degree of inequality among families in any given generation but no mobility within families across generations, with a circumstance in which there is substantial intragenerational income dispersion but also a large degree of intergenerational mobility? Which situation evidences less inequality?"

It should be clear from the previous example that the prediction of convergence in early theories in the equalization view is not a result of market forces reducing differences in wealth across households. More recent developments in the equalization view provide a rationale for market forces in the process of convergence. Greenwood and Jovanovic (1990) focus on the endogenous development of financial institutions and their effect on the performance of the economy. Financial intermediation promotes growth by collecting and processing information that allows an efficient allocation of investments. Furthermore, financial intermediaries provide a safer return to their investors by pooling risks across a large number of investors. However, building a financial superstructure capable of providing these services is costly. Consequently, financial superstructures tend to form gradually as an economy grows. In early an intermediate stage of development, the scope of financial markets is limited and only a small part of the population can invest through financial intermediaries. Market forces promoting convergence of income only emerge during the final stages of development when a large part of the population has enough wealth to invest in financial markets and an extensive network for financial intermediation emerges. As more individuals invest in financial markets, income inequality begins to decline and the distribution of wealth converges to an invariant distribution.

Other theories in the equalization view provide a more detailed analysis of the role played by capital markets in the process of convergence of wealth across households. For example, Banerjee and Newman (1991) emphasize the role of insurance markets as the driving force behind the convergence of wealth. They argue that, due to moral hazard in production, poor entrepreneurs will bear less absolute risk than rich entrepreneurs[5]. Consequently, the poor are more likely to undertake risky entrepreneurial projects with higher expected returns on investments. This implies that, on average, the wealth of the poor will grow faster than the wealth of the rich leading to the conclusion of convergence of wealth across households. One shortcoming of Banerjee and Newman's argument is that it relies on the rather unrealistic prediction that the poor have better access to insurance markets than the rich. In practice, potential entrepreneurs with assets that can be used as collateral tend to have better access to capital markets than those without assets. Aghion and Bolton (1997) look instead at the way credit markets influence the evolution of inequality. They argue that the opportunity cost of capital is higher for the poor than for the rich due to limited liability constraints in credit contracts[6]. In early stages of development, when aggregate wealth in the economy is low, the interest rate is high and poor individuals are discouraged from borrowing. However, if entrepreneurs can accumulate capital sufficiently fast then, as the economy grows, aggregate savings will increase pushing down the interest rate. As the interest rate decreases, the opportunity cost of capital for the poor decreases, allowing them to undertake more profitable investments. Therefore, on average, the wealth of all families converges to a common wealth in the long run irrespective of the initial distribution of wealth[7].

[5] This result is a consequence of the particular form of capital market imperfection in the model. Since the marginal disutility of effort increases with the wealth of an agent, insurance markets will require rich entrepreneurs to bear more risk than the poor to induce them to supply the optimal amount of effort.

[6] Since the rich own more assets that can be used as collateral on loans than the poor, the rich have more incentives to put high effort in their entrepreneurial projects. Therefore, the interest rate on loans will be higher for the poor to compensate lenders for the lower probability of repayment.

[7] It is worth mentioning that this convergence result relies heavily on some strong restrictions on the parameters of the model so that capital accumulates sufficiently fast in equilibrium. When these conditions are not met the model gives rise to a multiplicity of steady states, some involving persistent inequality and others involving convergence of wealth across families. Therefore, in a strict sense, this theory belongs in the neutrality view.

Galor and Moav (2006) provide an alternative mechanism through which market forces can promote convergence of wealth across households. They argue that the increasing importance of human capital in production led to a socioeconomic transformation that eliminated the historical differences between capitalists and workers. When the aggregate stock of physical capital in the economy is low, physical capital is the prime engine of growth and society is divided between rich capitalists and poor workers. However, as the aggregate stock of physical capital grows, capital-skill complementarity implies that the accumulation of human capital becomes a necessary condition to sustain the rate of return on physical capital. Since the accumulation of human capital is subject to decreasing returns at the individual level, the aggregate stock of human capital increases faster with the universal provision of education. Therefore, capital-skill complementarity made it beneficial for capitalists to support publicly financed education. This self-interested change in the attitude of capitalists, they argue, was a significant force behind the social and political reforms in Western Europe during the 19th century that cause the demise of the class structure. Galor and Moav's argument rests on the assumption that physical capital accumulation increased the demand for more educated workers during the 19th century due to capital-skill complementarity. There are two problems with this premise. First, although there is no detailed econometric evidence, historical evidence suggests that technological advances during the 19th century were de-skilling (or skill replacing), reducing the demand for more educated workers (Goldin and Katz, 1998; Acemoglu, 2002). Second, recent empirical evidence suggests that the demand for skills is primarily driven by technology-skill complementarity, not capital-skill complementarity. In the United States, for example, the main determinant of the demand for more educated workers in the last decades is not equipment capital, but high-tech capital (Autor, Katz, and Krueger, 1998). This evidence suggests that technological progress, not physical capital accumulation, is responsible for the increase in the demand for more educated workers.

In summary, the theories in the equalization view predict that starting from any initial distribution of wealth the economy will converge to a unique invariant wealth distribution. This long-run wealth distribution is ergodic, so there are no poverty traps in the economy. Furthermore, social mobility across generations and inequality within each generation

depend on the distribution of random shocks, i.e., on the inequality in luck. A society with substantial inequality in luck will experience substantial wealth inequality but a large degree of social mobility. These conclusions have rather strong implications for the way in which we understand the effect of a market economy on the intergenerational transmission of inequality. In particular, they imply that historical inequalities vanish over time because competitive markets tend to reduce differences in wealth across families. If inequality persists in the long run it is only because idiosyncratic shocks create differences in earnings within each generation. In the absence of these shocks the differences in wealth will eventually vanish. Finally, there is consensus within this view that temporary policy interventions, such as a one time redistribution of wealth, cannot have permanent effects on inequality or economic performance. Irrespective of the way in which wealth is redistributed, the economy will eventually converge to the unique invariant wealth distribution, leaving inequality unaltered in the long run. However, there is disagreement on the effect of permanent redistributive policies. Some theories predict that a permanent redistribution may increase efficiency by allowing the economy to move to a better allocation of investments and may enhance welfare by reducing the income uncertainty of future generations (e.g., Loury, 1981; Aghion and Bolton, 1997). Other theories predict that a permanent redistribution may lead to a new long-run equilibrium with a lower average income and greater inequality because investments are discouraged by the reduction in aftertax rates of return (e.g., Becker and Tomes, 1979, 1986). In between these extremes there is a third point of view which argues that a permanent redistribution of wealth will have no effect on inequality or welfare because the welfare enhancing insurance effect of a redistribution is completely offset by its negative incentive effect (e.g., Banerjee and Newman, 1991).

2.2 The Neutrality View

The main premise in the equalization view is that even if poor families are hindered by credit constraints in the short run, they could still catch up with wealthier families in the long run by gradually accumulating wealth over time. As a result, the initial wealth distribution affects economic performance in the short run, but not in the long run. In contrast, the neutrality view argues that if there are indivisibilities in the initial level of

investment necessary to acquire human or physical capital then poor families may never accumulate enough wealth to meet such a threshold. In this case, the initial wealth distribution affects economic performance both in the short and in the long run. Multiple steady states are possible, some involving persistent inequality and others involving convergence in wealth across families. Furthermore, the initial distribution of wealth will determine the long-run equilibrium of the economy. Consequently, historical inequalities may persist generation after generation.

The following model illustrates the main macroeconomic effects of indivisible investments. Consider an economy with a continuum of identical individuals who live for one period and have a single offspring. Individuals start their lives with an initial wealth, say w_i, inherited from their parents and choose an allocation of investments to maximize their end-of-period wealth. At the end of their lives they consume a fraction $1 - \beta$ of their end-of-period wealth and bequeath the rest to their children. These bequests become the initial wealth of the next generation. There are two investment opportunities in this economy. There is a safe asset that yields a low constant return of $r > 1$ per unit of capital invested. In addition to the safe asset, individuals can invest in an entrepreneurial project. Entrepreneurs can use one of two technologies, a cottage technology or an industrial technology. The cottage technology yields a return of $f(k)$ when total investment is k. The function $f(\cdot)$ is a strictly increasing and strictly concave production function. The industrial technology yields a return of $Af(k)$ only if $k \geq \underline{k}$, where $A > 1$[8]. The optimal capital investments in cottage and industrial production, k^* and k^{**} respectively, are given by the equality of the marginal productivity of capital and the return on the safe asset, i.e., $f'(k^*) = r$ and $Af'(k^{**}) = r$. Since industrial production yields a higher return than cottage production, each individual prefers to use the industrial technology. However, credit markets are missing and only those individuals whose initial wealth is above $\underline{k}$ are able to become industrial entrepreneurs. In the short run, society can be divided into four social classes: the lower class, whose inherited wealth is below k^*, the lower-middle class, whose inherited wealth is between k^* and $\underline{k}$, the upper-

[8] Note that there are no idiosyncratic shocks in this model. As we will see below, the models in the neutrality view do not require the presence of idiosyncratic shocks to generate persistent inequality. Nevertheless, some models do incorporate idiosyncratic shocks to generate steady state intergenerational social mobility.

middle class, whose inherited wealth is between $\underline{k}$ and k^{**}, and the upper class, whose inherited wealth is above k^{**}. Members of the lower class invest all their wealth in cottage production, those in the lower-middle class invest k^* in cottage production and $w_i - k^*$ in the safe asset, those in the upper-middle class invest all their wealth in industrial production while those in the upper class invest k^{**} in industrial production and $w_i - k^{**}$ in the safe asset.

Does inequality in the distribution of wealth lead to an inefficient allocation of investments across families? The answer was clear in the model developed in Section 2.1, given the average wealth in the economy, greater inequality led to a less efficient allocation of investments reducing the rate of growth. In this model, however, the relation between inequality and economic growth is nonmonotonic and greater inequality can have a positive, negative, or no effect on efficiency and the rate of growth. For example, consider a poor economy in which average wealth, $\bar{w}$, is below the minimum investment required to use the industrial technology, say $\bar{w} \in (k^*, \underline{k})$. If there is perfect equality in the distribution of wealth then $w_i = \bar{w}$ for each individual in the economy and everyone will invest k^* in cottage production and $\bar{w} - k^*$ in the safe asset. Consider now the following experiment: divide the population into two groups of the same size, reduce the wealth of each individual in the first group by $\tilde{w}$ and increase the wealth of each individual in the second group by the same amount. There is more inequality in the new distribution of wealth than in the initial distribution but the average wealth in the economy is the same in both cases. Whether the increase in inequality leads to a less efficient allocation of investments and a lower rate of growth depends on the size of $\tilde{w}$. If $\tilde{w}$ is sufficiently small such that $\bar{w} - \tilde{w} \geq k^*$ and $\bar{w} - \tilde{w} < k$ then the new allocation of investments is identical to the initial allocation, i.e., individuals will invest k^* in cottage production and the rest of their wealth in the safe asset. In this case, there is no efficiency loss and greater inequality will have no effect on the rate of growth. For larger values of $\tilde{w}$ such that $\bar{w} - \tilde{w} < k^*$ and $\bar{w} + \tilde{w} < \underline{k}$, the allocation of investments becomes less efficient. Those who lose from the increase in inequality will invest all their wealth, $\bar{w} - \tilde{w} < k^*$, in cottage production and those who gain invest k^* in cottage production and the rest of their wealth in the safe asset. Greater inequality leads to a lower rate of growth in this case because resources are transferred

from the entrepreneurial projects of those who lose from the increase in inequality to activities with a lower return, investments in the safe asset. Finally, for even larger values of $\tilde{w}$ such that $\bar{w} - \tilde{w} < k^*$ and $\bar{w} + \tilde{w} \in [\underline{k}, k^{**}]$ the new allocation of investments may be more efficient than the initial allocation. Those who lose from the increase in inequality will have even less wealth to invest in their cottage factories but those who gain will have enough wealth to become industrial entrepreneurs. If the yields from industrial production are sufficiently high then greater inequality will lead to a more efficient allocation of investments and a higher rate of growth[9].

The previous results change if the economy is sufficiently rich such that average wealth is above the minimum investment required to use the industrial technology. For example, if $\bar{w} \in [\underline{k}, k^{**}]$ then perfect equality in the distribution of wealth leads to an optimal allocation of investments across families. Each individual becomes an industrial entrepreneur and the marginal returns on capital investments are the same for all individuals. In this case, any increase in inequality will lead to a less efficient allocation of investments. Moreover, the greater the level of inequality, the lower the rate of growth in the economy[10].

What are the long-run properties of this economy? In the short run lower and lower-middle class families can only invest in cottage production while their wealthier counterparts become industrial entrepreneurs. The concavity of the production function implies that the wealth of a family investing in cottage production generation after generation will converge to a wealth w^* in the long run while the wealth of a family investing in industrial production generation after generation converges to $w^{**} > w^*$. Whether or not families who initially invest in cottage production can accumulate enough wealth over time to become industrial entrepreneurs in the future depends on the returns on the cottage technology which, in turn,

[9] Output per capita increases if $Af(\bar{w} + \tilde{w}) - [f(k^*) + (\bar{w} - k^*)r] > f(k^*) + (\bar{w} - k^*)r - f(\bar{w} - \tilde{w})$, which is true for sufficiently large values of A. It is worth noting that even larger increases in inequality, say a value of $\tilde{w}$ such that $\bar{w} - \tilde{w} < k^*$ and $\bar{w} + \tilde{w} > k^{**}$, may reduce the rate of growth.

[10] Note that perfect equality also leads to an optimal allocation of investments if the economy is even richer such that $\bar{w} > k^{**}$. However, in this case a sufficiently small increase in inequality will have no effect on the allocation of investments.

determine the relation between $\underline{k}$ and w^*. If the returns on the cottage technology are sufficiently high such that $\beta w^* \geq \underline{k}$ then lower and lower-middle class families will eventually be able to use the more efficient industrial technology. Thus, the wealth of each family converges to w^{**} and inequality vanishes in the long run. However, if the returns on the cottage technology are low and $\beta w^* < \underline{k}$ then lower and lower-middle class families will never accumulate enough wealth to become industrial entrepreneurs. In the former case the model predicts convergence in wealth across individuals, as do the theories in the equalization view, and in the latter case the model displays the multiplicity of steady states and the hysteresis (or history dependence) characteristic of the theories in the neutrality view.

When $\beta w^* < \underline{k}$ society is divided into two social classes in the long run, an upper class of industrial entrepreneurs whose wealth is equal to w^{**} and a lower class of cottage factory owners whose wealth is equal to w^*. Thus, a steady state for this economy is fully characterized by the fraction of industrial entrepreneurs in the population, say Λ. This economy has a connected continuum of steady states which contains one steady state that involves perfect equality in the distribution of wealth and a continuum of steady states that involve persistent inequality. Moreover, the initial distribution of wealth completely determines the development path of the economy. If the fraction of the population with a wealth above $\underline{k}$ in the initial distribution of wealth is equal to $\tilde{\lambda}$ then the steady state fraction of industrial entrepreneurs in the population will be $\Lambda = \tilde{\lambda}$. This implies that if there is perfect equality in the initial distribution of wealth then the economy converges to a steady state in which aggregate wealth is evenly distributed across individuals. Persistent inequality can only emerge if there is inequality in the initial distribution of wealth. Note that the initial distribution of wealth not only determines the long-run wealth distribution but also the long-run income per capita. Given an initial average wealth in the economy, societies with a larger initial fraction of industrial entrepreneurs in the population will have higher levels of income per capita in the long run. Finally, note that the existence of a connected continuum of steady states implies that any change in the initial distribution of wealth that alters $\tilde{\lambda}$ will lead the economy to a different steady state. It follows that

temporary policy interventions, such as a one time redistribution that increases $\tilde{\lambda}$, will reduce inequality and increase income per capita permanently[11].

The previous discussion implies that policies aimed at promoting economic growth in poor economies, where average wealth is below the minimum investment required to undertake industrial production, should allow some initial inequality in the distribution of wealth. In a poor economy, perfect equality in the distribution of wealth leads to an inefficient allocation of investments both in the short and in the long run. In the short run, each individual in the economy invests in cottage production while the more efficient industrial technology is not used. As a result, the economy converges to a steady state in which nobody is engage in industrial production, i.e., $\Lambda = 0$. Some degree of inequality may promote growth in the short run by allowing some individuals to become industrial entrepreneurs. As the economy grows and the upper and upper-middle classes accumulate wealth, policy interventions can increase the rate of growth by gradually redistributing wealth from the upper and upper-middle classes to the lower-middle class, and perhaps the most prosperous families in the lower class. Once the economy is rich enough such that average wealth is above the minimum investment required to undertake industrial production then the best way to advance economic efficiency and growth is to promote equity. In a rich economy, an egalitarian distribution of wealth leads to an efficient allocation of investments in the short and in the long run. In the short run, each individual becomes an industrial entrepreneur and the marginal returns on capital investments are the same for all individuals. As a result, the economy converges to the Pareto-optimal steady state in which everyone is engaged in industrial production, i.e., $\Lambda = 1$. Note that this efficiency-based development strategy may not necessarily benefit the poorest members of society in the short run. The objective of redistribution in this case is to maximize the productivity of investments in the economy by increasing the amount of capital invested in industrial production. In other words, the objective is to maximize the number of families investing in industrial production at every step of the development process. Since lower-middle class families need less additional wealth to become industrial entrepreneurs than

[11] Galor and Zeira (1993) were the first to formally prove these results in a model in which investments in physical capital are replaced by investments in human capital.

poor families, an efficiency-based redistribution may involve transfers to the lower-middle class, but not necessarily to the lower class.

So far we have assumed that there are no market interactions in the economy. Introducing market interactions (e.g., interactions in labor or credit markets) substantially complicates the analysis. The joint determination of the wealth distribution and market prices, such as wages and the interest rate, produces an interactive framework in which both influence each other. When capital markets are less than perfect, the distribution of wealth influences market prices through its effect on the investments decisions of the agents in the economy. At the same time, market prices determine the earnings of economic agents, influencing the distribution of wealth. When we take into account the general equilibrium interactions between inequality and market prices, multiple locally isolated steady states can emerge. This means that small changes in the initial distribution of wealth can move the economy towards a development path leading to a steady state with substantially different levels of inequality and income per capita[12].

An interesting version of this argument is the one developed by Banerjee and Newman (1993) who analyze agents' occupational choices when the wage rate is endogenously determined. The presence of capital market imperfections in addition to indivisible investments imply that wealthy agents choose between becoming entrepreneurs and being self-employed while poor agents choose between working for an entrepreneur and remaining idle and subsist. The middle class, whose inherited wealth is enough to be self-employed but not sufficient to become entrepreneurs, choose to be self-employed. In principle, poor families may be able to accumulate enough wealth over time to be self-employed. However, their ability to accumulate wealth depends on the wage rate, which is determined by the ratio of poor to wealthy families in the economy. Banerjee and Newman show that for some sets of parameter values the economy has only two locally isolated

[12] In contrast, in models with a connected continuum of steady states, small changes in the initial wealth distribution will move the economy to a nearby steady state with similar levels of inequality and income per capita.

steady states with large basins of attraction[13]. In one steady state the ratio of poor to wealthy families is low, labor is relatively scarce and the equilibrium wage rate is high. The high wage rate allows for a high degree of upward mobility for the lower class keeping the labor supply low, reinforcing the high equilibrium wage rate. Although there are three social classes, large upper and middle classes which are mostly self-employed and a small lower class of workers who receive high wages, there is also substantial intergenerational mobility. In contrast, in the second steady state the ratio of poor to wealthy is high, labor is relatively abundant and the equilibrium wage rate is low. The low wage rate makes it impossible for lower-class families to accumulate enough wealth to be self-employed in the future, preserving the high supply of labor. In turn, the persistently high supply of labor reinforces the low equilibrium wage. This is a steady state with limited intergenerational mobility and three well defined social classes, a small upper class of wealthy entrepreneurs, a middle class of self-employed families, and a large lower class of poor families who either work for low wages or remain idle and subsist. The same class structure is perpetuated generation after generation because there is little or no upward mobility for the lower class. Banerjee and Newman show that societies with an unequal initial distribution of wealth are more likely to converge to the low-wage low-mobility steady state than egalitarian societies.

Piketty (1997) uses similar arguments to show that the joint determination of the wealth distribution and the interest rate may lead to multiple self-sustaining steady states. He considers an economy in which every agent is a self-employed entrepreneur. Credit markets are incomplete and poor entrepreneurs, who need large loans to make the optimal investment, may be credit-rationed. Whether the poor are credit-rationed or not depends on the equilibrium interest rate. In particular, the set of credit-rationed agents in the economy

[13] The dynamical system derived from this model is too complex to provide a complete characterization of the behavior of the economy. Banerjee and Newman restrict their analysis to two cases, *the cottage and the factory* and *prosperity and stagnation*, which correspond to two particular sets of parameter values. The following discussion refers to steady states in the perturbed version of *the cottage and the factory*. See Ghatak and Jiang (2002) for a complete characterization of the transitional dynamics in a simpler version of the model.

increases as the interest rate increases[14]. Therefore, steady states associated with higher interest rates have higher levels of inequality and lower levels of income per capita and social mobility. Intuitively, in poor economies, where capital is scarce, the equilibrium interest rate will be high. The high interest rate implies that a large fraction of the population will be credit-rationed, making capital accumulation more difficult and keeping the economy poor. In contrast, if the economy is rich then the equilibrium interest rate will be low. In this case there is little or no credit-rationing and this in turn speeds up the process of capital accumulation keeping the economy rich.

Other theories in the neutrality view examine alternative sources of multiplicities and hysteresis. For example, Durlauf (1996) focuses on the process of economic stratification resulting from endogenous neighborhood formation. The basic framework relies on the premise that education is locally financed within each neighborhood with proportional income taxes and that human capital formation exhibits decreasing average costs. Neighborhoods cannot borrow to finance their investments in education and increases in the level of education provided within a neighborhood require fixed indivisible investments. Due to the assumption of proportional taxes, income heterogeneity within neighborhoods involves redistribution from the rich to the poor, creating incentives for wealthy families to form small homogeneous neighborhoods. On the other hand, decreasing average costs in human capital formation promotes the creation of large neighborhoods, implying income heterogeneity within neighborhoods. Whether society is characterized by small homogeneous or large heterogeneous neighborhoods depends on the distribution of income. A highly unequal society promotes economic stratification of neighborhoods leading to persistent inequality. The initial pattern of inequality is reproduced generation after generation because children in wealthier communities receive a higher level of education than children in poorer communities. On the other hand, an egalitarian society promotes the creation of large heterogeneous neighborhoods. Since all children in a neighborhood

[14] Piketty assumes that credit markets are incomplete due to limited liability constraints in credit contracts. This implies that an entrepreneur's incentive to supply high effort decreases as the interest rate increases. Therefore, as the interest rate increases, lenders will reduce the size of loans to give borrowers the proper incentives to supply high effort, increasing the fraction of credit-rationed entrepreneurs in the economy.

receive the same human capital investment, heterogeneous neighborhoods promote convergence of wealth across families.

Finally, Mani (2001) examines another potential source of multiple steady states, the link between inequality and the patterns of demand for goods. Income determines not only the level of consumption but also the composition of the basket of goods consumed by a household. As an individual's wealth increases, the share of basic goods in consumption decreases while the share of higher quality and more sophisticated goods increases. Thus, the distribution of wealth determines the overall pattern of demand for goods in society. Since the production of more sophisticated goods usually requires workers with higher levels of skills, the pattern of demand for goods determines the demand for different types of labor and so influences the wages for different degrees of skilled labor. Acquiring a high level of skills may be prohibitively costly for poor families. Nevertheless, if the return to medium-skilled labor is sufficiently high then the medium-skills sector can serve as a bridge for poor families to catch up with wealthier families. The return to medium-skilled labor, however, depends on the distribution of wealth. Mani argues that a highly unequal distribution of wealth implies a low demand for goods that use medium-skilled labor and a low return for medium-skilled workers. Consequently, medium-skilled workers will not be able to leave a sufficiently large bequest to allow their children to become high-skilled workers, preserving the initially high level of inequality and reinforcing the low return to medium-skilled workers. A highly egalitarian distribution of wealth, in contrast, implies a robust demand for goods that use medium-skilled labor and a high return to medium-skilled workers. In turn, the high return to medium skilled workers allows for a high degree of upward mobility, preserving the level of inequality low and reinforcing the high return to medium-skilled workers.

To summarize, the theories in the neutrality view predict the coexistence of multiple steady states, some involving persistent inequality and others involving convergence in wealth across households. Steady states with more unequal distributions of wealth are usually associated with lower levels of income per capita and lower degrees of social mobility. The specific steady state the economy converges to depends on the initial distribution of wealth.

Typically, economies with lower levels of inequality in the initial distribution of wealth converge to steady states with less inequality, more social mobility, and a higher income per capita. These conclusions imply that a market economy displays neither an intrinsic tendency to generate inequality nor an intrinsic tendency to reduce differences in wealth across families. A market economy may reproduce or even exacerbate the initial level of inequality. Nevertheless, in the absence of idiosyncratic shocks, if there is perfect equality in the initial distribution of wealth then equality will be preserved in the long run. In this sense, these are models of exogenous inequality. Inequality does not arise through an internal mechanism of the models. Persistent inequality is only possible if there is some exogenous source of inequality to begin with. Finally, the multiplicity of steady states and the hysteresis characteristic of the theories in the neutrality view imply that temporary policy interventions may have permanent effects. A one time redistribution of wealth can improve the productive efficiency of the economy in the short run by expanding the investment opportunities of the poor and middle class who need to borrow to invest. This short-run efficiency improvement may be all that is required to move the economy to an alternative path of development leading to a steady state with a higher income per capita and a higher degree of social mobility. The rationale for redistribution in this context is to equalize investment opportunities allowing more agents to make productive investments. Therefore, the beneficiaries of this efficiency-enhancing redistribution are not necessarily the poorest in the economy but those agents who are more likely to make these investments.

2.3 The Disequalization View

The prediction of convergence of wealth across families in the equalization view stands in sharp contrast to the multiplicity of steady states and hysteresis in the neutrality view. Nevertheless, despite these differences, the equalization and neutrality views share one common conclusion, namely, that there is no intrinsic tendency in a market economy to generate inequality. This is the main difference between the disequalization view and the previous theories. According to the disequalization view, the emergence of inequality is a natural consequence of market forces promoting separation in investment choices. If different professions, requiring the acquisition of different levels of skills, are necessary in production then the returns on investments in different professions will adjust such that all

professions are supplied in equilibrium. In particular, individuals in professions that involve higher training costs must earn higher wages. Hence, even if all individuals are ex-ante identical, there must be inequality in earnings. Furthermore, these theories predict that competitive markets will tend to over compensate individuals in higher paying professions strengthening earnings inequalities over time and generating inequality in welfare for future generations.

A simple model will be useful to illustrate the logic of the disequalization view. The following is a simple version of the model developed in Ray (1990, 2006). Consider an economy with a continuum of individuals who live for one period and have a single offspring. Individuals spend all their time working and allocate their labor earnings between consumption and investments in children's education. These educational investments are the only type of intergenerational transfers in this economy. For simplicity, assume that investments in children's education are driven by parental concerns about children's income. In particular, preferences are given by $U = u(c) + w_c$, where c is consumption, w_c is child's income, and $u(\cdot)$ is a strictly increasing, strictly concave utility function. Credit markets are missing and individuals are constrained to use their own labor earnings to finance their consumption and investments in education. The children of parents who invest in education become skilled workers while the children of parents who do not invest remain unskilled. Investments in education are indivisible and the cost of providing a child with skills is equal to x. Firms use a constant returns to scale technology which combines skilled and unskilled workers to produce a single consumption good. Skilled and unskilled labor are complements in production and the production function exhibits diminishing returns in its two inputs. The labor market is competitive and the equilibrium wages for skilled and unskilled workers equal their respective marginal products. This implies that the wages for skilled and unskilled workers as well as aggregate output are completely determined by the ratio of skilled workers in the economy. Thus, a steady state

for this economy is fully characterized by its ratio of skilled workers. Most theories in the disequalization view are reformulations of this basic framework[15].

The main predictions of the theories in the disequalization view can be easily understood by looking at the properties that must be satisfied by any steady state in this economy. First, there cannot be a steady state involving perfect equality in earnings, any steady state must exhibit some degree of earnings inequality. Perfect equality in earnings requires a sufficiently large ratio of skilled workers in the economy. However, if there is no earnings differential to compensate for educational expenditures then parents will have no incentive to educate their children. Therefore, a sufficiently high ratio of skilled workers such that the wages for skilled and unskilled workers are the same cannot be sustained as a steady state. This implies that inequality will endogenously emerge and persist in the long run even if there is perfect equality in the initial distribution of wealth. Second, there cannot be social mobility in any steady state, i.e., skilled workers invest in the education of their children while unskilled workers choose to keep their children uneducated[16]. This result follows from the concavity of the utility function. Since capital markets are missing, individuals have to reduce consumption to finance their educational expenditures. With a strictly concave utility function, the loss of utility from forgone consumption is higher for unskilled workers, who have a lower income, than for skilled workers. Hence if it is optimal for an unskilled worker to invest in education then it must also be optimal for a skilled worker to do so. Since the ratio of skilled workers is constant in any steady state, it must be true that only skilled workers invest in education. Finally, in any steady state the utility of skilled workers must be higher than the utility of unskilled workers. In other words, there is inequality in welfare as well as earnings. Earnings differentials in steady state overcompensate for educational expenditures allowing skilled families to enjoy a higher consumption than unskilled families[17].

[15] See, for example, Ray (1990, 2006), Ljungqvist (1993), Owen and Weil (1998), Maoz and Moav (1999), Mookherjee and Ray (2003), and Mookherjee and Napel (2007).

[16] The lack of social mobility in this model is clearly unrealistic. One way to introduce some intergenerational mobility is to assume heterogeneity among individuals due to idiosyncratic shocks (see, for example, Owen and Weil, 1998; Maoz and Moav, 1999; and Mookherjee and Napel, 2007).

[17] Freeman (1996) reaches similar conclusions in a model in which indivisible investments in education are replaced by an increasing returns to scale technology.

Are there multiple steady states or will the economy converges to a unique steady state in the long run? Furthermore, are steady states efficient or inefficient? The role of history in shaping an economy's long-run economic performance depends on the answer to the first question while the scope for efficiency-based policy interventions depends on the answer to the second. It turns out that the answers to these questions are not independent from each other; inefficient steady states are possible only if there are multiple steady states. The existence of multiple steady states, in turn, depends on the richness of the occupational structure in the economy. If the number of occupations in the economy is finite then there are multiple steady states. For example, when there are just two occupations, skilled and unskilled, Ray (1990, 2006) proves the existence of a continuum of steady states. Besides this continuum of steady states, other steady states with smaller ratios of skilled workers are also possible. Thus, the set of steady states need not be connected. Steady states can be ranked in terms of overall inequality and aggregate output. As we move from steady states with low ratios of skilled workers to steady states with higher ratios of skilled workers, aggregate output increases and the difference between the wages of skilled and unskilled workers decreases. Thus, higher steady-state ratios of skilled workers are associated with lower degrees of inequality and higher levels of income per capita. In contrast, if there is a continuum of occupations requiring continuously increasing indivisible investments then there exists a unique steady state. For example, Mookherjee and Ray (2003) prove that if there is a continuum of distinct levels of skills and a continuous training costs function that determines the cost of acquiring different levels of skills then the steady state is unique.

Mookherjee and Ray (2003) also provide a characterization of the constrained efficiency of steady states. A steady state is efficient if there is no other feasible allocation such that each individual in each generation is at least indifferent between the two and some of them are strictly better off. This implies that overinvestment in education, but not underinvestment, is consistent with efficiency. When the number of occupations in the economy is finite and multiple steady states emerge, a continuum of inefficient steady states coexists with a continuum of efficient ones. Mookherjee and Ray show the existence of a threshold such that steady states with a ratio of skilled workers below the threshold are inefficient while

steady states with a ratio of skilled workers above the threshold are efficient. The threshold is determined by the optimal level of investment in education. Thus, any ratio of skilled workers below the threshold implies underinvestment in education while any ratio above it implies overinvestment. The multiplicity of steady states implies that historical inequalities will affect long-run economic performance. Moreover, the existence of inefficient steady states implies that temporary policy interventions can raise long-run per capita income while reducing inequality[18].

When there is a continuum of occupations, the unique steady state turns out to be efficient. Moreover, in this steady state investments in education are optimal, i.e., there is no overinvestment. These results imply that there is no scope for efficiency enhancing redistributive policies in the case of a continuum of occupations[19]. Mookherjee and Ray (2003, p. 372) conclude that "efficiency-based arguments for interventionist policies must rely on the existence of investment indivisibilities, analogous to arguments concerning history dependence."

Our analysis of the disequalization view so far suggests that persistent inequality is an inevitable consequence of market forces promoting separation in investment choices. This result is partly due to the assumption that investments in education are the only type of intergenerational transfers. Ljungqvist (1993) shows that when financial bequests are allowed, steady states that involve perfect equality in wealth and welfare are possible[20]. Financial bequests may generate steady states with perfect equality because they can complement educational expenditures in a way that may compensate for earnings differentials in the long run. In a steady state with wealth equality, the earnings differential

[18] It is important to note that these results are sensitive to the introduction of idiosyncratic shocks into the model. Mookherjee and Napel (2007) show that the set of steady states shrinks dramatically when heterogeneity is introduced in order to generate steady-state mobility. Therefore, long-run economic performance becomes less history dependent. Furthermore, the effect of temporary policy interventions on long-run macroeconomic outcomes is considerably reduced.

[19] Mookherjee and Ray (2003) do not prove that the global stability of the unique steady state. Therefore, it is not possible to determine the effect of temporary policy interventions when the economy is not in its unique steady state.

[20] To be sure, Ljunqvist considers an economy populated by infinitely lived agents, not an overlapping generations model. However, similar results would obtain in an overlapping generations model with dynastic utility functions in which financial bequests play the role of savings.

is just enough to make both skilled and unskilled workers indifferent between investing in their children's education or leaving them a financial bequest. In any generation there is inequality in labor earnings but not necessarily wealth or welfare inequality. Steady states with perfect equality have higher ratios of skilled workers and higher levels of income per capita than steady states involving persistent inequality.

In summary, the theories in the disequalization view predict the emergence of inequality as a natural consequence of market forces promoting separation in investment choices. The existence of multiple steady states and the possibility of efficiency enhancing redistributive policies in this view rely on the existence of indivisibilities in investment opportunities. When multiple steady states emerge, steady states with more unequal distributions of wealth are associated with lower levels of income per capita due to underinvestment in education. As a result, steady states with highly unequal distributions of wealth are inefficient and redistributive policies can improve both equity and efficiency. Moreover, as in the neutrality view, the existence of multiple steady states implies that temporary policy interventions can increase income per capita while reducing inequality in the long run. In contrast, if investment opportunities are perfectly divisible then a unique steady state emerges. In this case, the unique steady state is efficient so redistributive policies cannot be grounded on efficiency arguments. Finally, although these theories predict the emergence and persistence of inequality in earned incomes they do not preclude the possibility of a steady state with perfect equality in the distribution of wealth and welfare. If the initial aggregate wealth is evenly distributed then financial bequests can compensate for earnings inequality generating a steady state with perfect equality in wealth and welfare.

2.4 Equalization, Neutrality, or Disequalization?

The equalization view enjoyed a dominant status in the economic literature on the evolution of inequality until the 1990's. Although this dominant status was challenged by the emergence of the theories in the neutrality and disequalization views, the equalization view remained as the mainstream view in economics. The empirical evidence accumulated during the 1970's and 1980's supported its central prediction of convergence of income across households. Becker and Tomes (1986) surveyed a series of studies comparing

parent's and son's incomes and found a low level of intergenerational persistence of economic status. In particular, the correlation between parents' and sons' earnings in these studies averaged 0.15 for the United States, suggesting a high level of intergenerational mobility. Based on this evidence Becker and Tomes (1986, p. S32) concluded that "[a]side from families victimized by discrimination, regression to the mean in earnings in the United States and other rich countries appears to be rapid .... Almost all earnings advantages and disadvantages of ancestors are wiped out in three generations." However, more recent empirical studies show that the estimates of low correlation between the incomes of parents and their sons found in previous studies were the result of various types of measurement errors. When controlled for measurement errors, the correlation between parents' and sons' incomes in the United States may be up to three times the average of the studies surveyed by Becker and Tomes, i.e., an estimated correlation somewhere between 0.4 and 0.45 (see, e.g., Solon, 1992, 1999; Zimmerman, 1992; and Bowles and Gintis, 2002). Mazumder (2005) argues that even these recent empirical studies underestimate the correlation between parent's and son's incomes due to transitory fluctuations in earnings. He finds that this correlation is around 0.6 in the United States[21]. This evidence suggests a substantially higher level of intergenerational persistence of economic status than was previously thought to be the case. Furthermore, Hertz (2005) finds that the degree of intergenerational persistence of economic status in the United States is particularly high at the top and at the bottom of the income distribution. For example, 22.9 percent of children with parents in the top decile of the income distribution remained there as adults and 40.7 percent of them remained in the top quintile. At the other extreme of the income distribution, only 1.3 percent of the children with parents in the bottom decile moved to the top decile and 3.7 percent moved to the top quintile while 31.2 percent of them remained in the lowest decile and 50.7 percent of them remained in the lowest quintile. These findings suggest that parental income and wealth are important determinants of the economic status of children.

So, does a market economy display an intrinsic tendency to reduce differences in wealth and income across households, or does it exacerbate those differences? The persistence of

[21] Dearden, Machin, and Reed (1997) report a similar result for the United Kingdom. They estimate a correlation of 0.57.

inequality in earned incomes predicted by the neutrality and disequalization views is consistent with the high level of intergenerational persistence of economic status found in recent empirical studies. Nevertheless, this empirical evidence does not necessarily imply that we should discard the equalization view in favor of either the neutrality or the disequalization views. A high level of intergenerational persistence of economic status does not preclude the possibility of convergence of income predicted by the equalization view, although this convergence may take place at a substantially slower pace than previously thought. Moreover, we still do not completely understand the mechanisms that determine the intergenerational transmission of economic status. As Bowles and Gintis (2002, p. 4-5) put it, "the transmission of economic success across generations remains something of a black box ... The fundamental problem is not that we are measuring the right variables poorly, but that we are missing some of the important variables entirely."

An alternative empirical test of the predictions of the three views is to provide an answer to the question: Does inequality hinder economic growth? While the equalization view argues that current inequality has no effect on long-run economic growth, the neutrality and disequalization views predict that more unequal societies tend to have lower levels of income per capita in the long run. However, the search for empirical evidence to answer this seemingly simple question has proved to be both difficult and controversial. The availability and quality of data on inequality has been one of the central problems for empirical analyses of the relation between inequality and growth. Most empirical analyses use the distribution of income to measure inequality and data on income inequality has traditionally suffered from large measurement errors. Moreover, data derived from surveys usually differ on the time surveys are conducted, levels of coverage (e.g., nationwide, urban), definitions of income (e.g., gross or net income), and definitions of recipient units (e.g., households, economically active persons, individuals). These differences significantly reduce cross-country comparability of the data[22].

[22] New data sets on inequality, in particular the Deininger and Squire (1996) data set and the 2007 World Income Inequality Database compiled by the United Nations, are a considerable improvement over previous data sets. Nevertheless, the extent of measurements errors remains significant.

Even if we ignore measurement and comparability problems of the data on income inequality, we are faced with the problem of choosing between the contradictory results that have plagued this literature. Early attempts to measure the effect of inequality on growth usually proceeded by adding inequality as an independent variable in a standard growth regression. In these early studies, the ordinary least squares (OLS) estimates of the effect of inequality on growth are consistently negative and usually significant[23]. Li and Zou (1998) and Forbes (2000) criticize these early studies arguing that omitted country-specific characteristics biased the OLS estimates of the effect of inequality on growth. Using fixed effects estimators, both studies find a positive and significant effect of inequality on growth[24]. However, Barro (2000) argues that the fixed effects estimates exacerbate the biases due to measurement errors because gini coefficients tend to be relatively stable over time and variations across countries are more important than variations across time. Barro uses a three-stage least squares estimator which takes the country-specific characteristics as random and finds that inequality has no significant effect on growth. So, does inequality hinder economic growth after all? Both the negative OLS estimates and the positive fixed effects estimates of the effect of inequality on growth appear to be robust. Indeed, Forbes (2000, p. 871) argues that the "estimates of a short-run positive relationship between inequality and growth within a given country do not directly contradict the previously reported long-run negative relationship across countries." Thus, the empirical evidence on the relation between inequality and growth at the country level does not allow to unequivocally determine the effect of inequality on economic growth[25].

[23] See, for example, Alesina and Rodrik (1994), Persson and Tabellini (1994), Alesina and Perotti (1996), Perotti (1996), and Deininger and Squire (1998). See Benabou (1996) and Perotti (1996) for extensive surveys of this literature.

[24] The differencing implicit in fixed effects estimation eliminates the country-specific fixed effects allowing the usual interpretation of the coefficients on initial inequality in the regression equation.

[25] Banerjee and Duflo (2003) provide a possible explanation for these contradictory results. They argue that there is no theoretical basis for assuming that the relation between inequality and growth can be captured by a linear regression like the ones used in previous studies. Using a random effects estimator to measure the effect of inequality on growth in a nonlinear equation, they find that the relation between changes in inequality and changes in the rate of growth has the shape of an inverted-U, i.e., changes in inequality in either direction reduce the rate of growth. Banerjee and Duflo (2003, p. 268) conclude that the nonlinearities in the relationship between changes in inequality and changes in growth rates "explain why different variants of the basic linear model (OLS, fixed effects, random effects) have generated very different conclusions: In many cases, it turns out that the differences arise out of giving different structural interpretations to the same reduced form evidence."

In conclusion, the empirical evidence is not strong enough to select one view over the other two. Thus, we will not delve deeper into the empirical literature. Instead, we will focus our attention on the key differences in underlying assumptions that distinguish the equalization, neutrality, and disequalization views. In addition, we will also analyze the role played by capital market imperfections in each of these views.

2.4.1 Convexities, Non-Convexities, and Complementarities

Which are the key differences that distinguish the equalization, neutrality, and disequalization views? All the theories in each of the three views acknowledge the importance of capital market imperfections in understanding the functional aspects of inequality. Indeed, most of these theories make very similar assumptions about the fundamental characteristics of markets. Thus, the source of disagreement between the three views must lie somewhere else. As we will see below, disagreement arises from assumptions about the technologies available in the economy. In turn, these assumptions implicitly define the set of occupational choices (i.e., the set of entrepreneurial activities and professions) available to the agents in the economy.

Most theories in the equalization view rely on the assumption of a convex technology as the driving force behind the process of convergence of earnings and wealth across households[26]. There are two characteristics of a convex technology that allow poor families to eventually catch up with their wealthier counterparts. First, a convex technology implicitly requires all inputs, including physical and human capital, to be perfectly divisible. This implies that the same set of occupations is available to all agents in the economy, irrespective of their wealth. Due to credit constraints, poor agents may end up investing less than their wealthier counterparts. Nevertheless, they have access to the same entrepreneurial activities and professions as the wealthy. Second, a convex technology explicitly requires decreasing marginal returns on investments in physical and human

[26] Two exceptions to this rule are Banerjee and Newman (1991) and Aghion and Bolton (1997). In both cases technology is not convex due to indivisibilities of investments. As mentioned before, the prediction of convergence in Banerjee and Newman's model is a consequence of the rather unrealistic assumption that the poor have better access to insurance markets than the rich. In the case of Aghion and Bolton's model, the prediction of convergence is only valid for a restricted set of parameter values. For other parameter values the predictions of the model are similar to the predictions of the theories in the neutrality view.

capital. If capital markets are less than perfect then the wealthy will invest more than the poor. As a consequence, the marginal return on investments should be higher for the poor than for the wealthy. Under standard assumptions about preferences, this difference in marginal returns on investments implies that the wealth of poor families will grow faster than the wealth of rich families. This result leads to the conclusion of convergence of wealth and income in the long run.

In contrast, most theories in the neutrality view are based on the assumption of technological non-convexities due to indivisibilities of investment opportunities[27]. Indivisibilities of investments in human capital arise if, for instance, the practice of some professions requires a minimum number of years of education or if the returns to education increase only after completing some minimal level of formal education (e.g., secondary or college education). The direct costs of primary and secondary education are usually low in most countries. However, the opportunity cost of secondary education, as measured by forgone earnings while attending school, can be a substantial burden for poor families. Similarly, indivisibilities of investments in physical capital arise if undertaking an entrepreneurial project requires an initial setup cost or if a business is productive only if investment crosses certain threshold. In any of these examples the accumulation of physical or human capital requires an initial indivisible investment. When capital markets are less than perfect, access to occupations requiring high initial investments may be prohibitively costly for poor families. In other words, the set of occupational choices available to the poor is smaller than the one available to the rich. If occupations with the highest returns require sufficiently high initial indivisible investments then poor families may never accumulate enough wealth to meet such a threshold. As a consequence, poor families may

[27] One exception to this rule is Piketty (1997) who considers an economy where self-employed entrepreneurs use a standard concave production function. However, Piketty assumes that output depends on an entrepreneur's effort which is a discrete variable that can only take two values, high and low. This assumption implies that, in equilibrium, there are only two investment options, a high-capital high-effort investment and a low-capital low-effort investment. Thus, for all practical purposes, technology behaves as if an initial indivisible investment is required to access the entrepreneurial activity with the highest return. In particular, as in any model with indivisible investments, there is a threshold such that poor individuals whose wealth is below the threshold can only make the low-capital low-effort investment.

be excluded from the most productive occupations generation after generation and historical inequalities may persist or even expand over time.

The theories in the disequalization view, on the other hand, rely on the assumption that diverse occupations, requiring different initial indivisible investments, complement each other in production. The assumption of complementarities in production has two main implications that are essential to understand the predictions of the disequalization view. First, it implies that different levels of human capital correspond to different occupations. In other words, not all human capital is homogeneous and different occupations are not perfect substitutes for each other. For instance, neither nurses are perfect substitutes for surgeons in the operation room nor construction workers are perfect substitutes for architects in designing a building. As a consequence, the rate of return on a unit of human capital need not be the same across occupations. Second, it implies that some occupational diversity is essential in production. Since different occupations are not perfect substitutes and the production process involves a variety of tasks requiring different skills, some minimal occupational diversity is necessary. For example, both nurses and surgeons are necessary to perform surgery and both construction workers and architects are necessary to build a building. If diverse occupations requiring different initial indivisible investments are essential to production then the returns on investments in different occupations must adjust such that all occupations are supplied in equilibrium. Therefore, individuals in occupations that require higher indivisible investments must earn higher wages. In other words, the emergence of earnings inequality is a necessary outcome of market forces promoting separation of investment choices. Moreover, under standard assumptions about preferences, competitive markets will overcompensate individuals in higher paying occupations strengthening earnings inequalities over time and generating inequality in welfare for future generations.

It is important to notice that the presence of indivisible investments per se is not a sufficient condition for the existence of multiple steady states. For instance, each occupation in Mookherjee and Ray (2003) requires its own specific indivisible investment. Nevertheless, the economy has a unique steady state. The existence of multiple steady states and,

therefore, the importance of history in shaping the process of economic development are related to a more fundamental characteristic of the economy, the richness of the occupational structure (i.e., the diversity of entrepreneurial activities and professions requiring different initial indivisible investments) in the economy. Multiple steady states arise in the neutrality view and in most theories in the disequalization view because physical or human are not perfectly divisible and the occupational structure in the economy is not perfectly rich.

If physical or human capital are not perfectly divisible and the occupational structure in the economy is not perfectly rich then the set of initial indivisible investments required by different occupations in the economy is not connected. When the set of indivisible investments is not connected, there are a finite number of thresholds such that: (*i*) the set of occupations between two consecutive thresholds is not empty and is the same for all levels of wealth, and (*ii*) individuals only gain access to a discrete number of new occupations with higher returns when their wealth reaches the next threshold. In other words, the set of occupational choices available to the individuals in the economy is not a continuous function of their wealth. These discontinuities can lead to multiple steady states if the returns on the set of occupations between two consecutive thresholds are not high enough to allow families to accumulate enough wealth to reach the next threshold. In this case some families may be trapped in the same occupations generation after generation and may never catch up with wealthier families who have access to investment opportunities with higher returns. This point was made clear in the model developed in Section 2.2 in which there is one threshold, $\underline{k}$, which determines access to the more efficient industrial technology. If the return on the cottage technology is sufficiently high then the steady state is unique. Poor families eventually accumulate enough wealth to become industrial entrepreneurs and inequality vanishes in the long run. On the other hand, if the return on the cottage technology is low then multiple steady states emerge and history determines the long-run fate of the economy. Poor families never accumulate enough wealth to use the industrial technology and are trapped in cottage production generation after generation while wealthy families use the more efficient industrial technology. In contrast, if either physical and human capital are perfectly divisible or if there is a continuum of occupations

requiring continuously increasing indivisible investments then the set of occupational choices is a continuous function of wealth[28]. In turn, the continuity of the set of occupational choices implies that families are not trapped in occupations with low returns on investments and the steady state is unique. Hence, the role of history in shaping the process of economic development depends on the richness of the set of investment opportunities in the economy. Mookherjee and Ray (2006) emphasize this point in a model encompassing each of the three views as special cases[29].

Mookherjee and Ray (2006) extend the framework developed inMookherjee and Ray (2003) to allow for financial bequests and the use of physical capital in production. Mookherjee and Ray argue that two attributes of occupational diversity are central to understand the differences between the three views: the *span* and the *richness* of the occupational structure in the economy. The span condition refers to the range of indivisible investments across occupations while the richness condition refers to the diversity of occupations requiring different indivisible investments. The uniqueness or multiplicity of steady states depends on the richness of the occupational structure in the economy. If the occupational structure in the economy is not perfectly rich then the set of indivisible investments is not connected and multiple steady states arise. However, if there is a continuum of occupations requiring continuously increasing indivisible investments then there exists a unique steady state. Therefore, historical inequalities do not matter for economic development when the occupational structure in the economy is perfectly rich. Whether the market is equalizing or disequalizing depends on the span of the occupational structure in the economy. As in Ljungqvist (1993), steady states that involve perfect equality may arise in this model because financial bequests complement educational investments in a way that may compensate for earnings differentials in the long run. However, whether or not financial bequests can compensate for earnings differentials depends on whether or not the returns on human and physical capital investments are the

[28] In the former case the set of indivisible investments is empty, the economy behaves as if there is only one occupation and, therefore, the set of occupational choices is trivially continuous since it is the same for all levels of wealth. In the latter case the set of indivisible investments is an interval and the set of occupational choices becomes a continuously increasing function of wealth.

[29] See Matsuyama (2000, 2006) for alternative models that include each of the three views as special cases.

same. Mookherjee and Ray prove that the returns on human capital investments equal the returns on physical capital investments only for occupations requiring low indivisible investments. For occupations requiring sufficiently high indivisible investments the returns on human capital investments strictly exceed the returns on physical capital investments. Therefore, if the span is sufficiently wide (i.e., if the range of indivisible investments is sufficiently large) then every steady state must involve persistent inequality across families. Steady states that involve perfect equality are only possible when the span is sufficiently narrow. In other words, when the span of occupational choices is sufficiently wide then markets are *disequalizing* and all steady states must exhibit inequality in earnings and welfare. When the span is narrow, the richness of the occupational structure distinguishes the equalization view from the neutrality view. If the occupational structure of the economy is perfectly rich then *equalizing* forces dominate and the steady state is unique. However, if the occupational structure of the economy is not perfectly rich then markets are neutral towards inequality and multiple steady states arise.

2.4.2 The Role of Capital Market Imperfections

Throughout this section we argued that capital market imperfections play a central role in the predictions of the equalization, neutrality, and disequalization views. We will conclude this section by examining the role of capital market imperfections in more detail[30]. Our first three results deal with optimal individual behavior when capital markets are perfect and apply to all the theories in each of the three views. Our last result refers to the predictions of the simple models developed in Sections 2.1 and 2.2.

Consider first the role of imperfections in insurance markets. Most theories of income distribution rely on idiosyncratic shocks to abilities or opportunities as a natural source of steady-state intergenerational mobility. The underlying assumption is that individuals cannot insure away all idiosyncratic risk due to imperfections in insurance markets. Hence, idiosyncratic shocks become a source of variation in incomes across individuals. Now

[30] It is important to note that the empirical evidence seems to support the assumption of capital market imperfections. For example, after an extensive survey of the literature the World Bank (2006, p. 101-102) concludes that "markets in developing countries are highly imperfect, and those who do not have enough wealth or social status tend to underinvest."

suppose that insurance markets work perfectly. If individuals are risk averse then they would insure themselves perfectly against uncertainty about the outcome of risky entrepreneurial projects. If, in addition, individuals are altruistic toward their children then they would insure their children perfectly against uncertainty about their innate abilities. As a consequence, idiosyncratic shocks have no effect on the income of individuals. Therefore, if individuals are risk averse and altruistic toward their children, we have the following result.

> **Result 1**: If insurance markets are perfect then idiosyncratic shocks have no effect on intragenerational inequality or intergenerational mobility.

The main implication of this result is that, once the economy reaches its steady state, there will be no intergenerational mobility and the distribution of wealth will be invariant over time. In other words, identical families will remain identical and the ranking of families will be preserved generation after generation. Since there is no role for idiosyncratic shocks in a world in which insurance markets are perfect, we will assume that there are no idiosyncratic shocks in the remainder of this section.

Consider now the role of imperfections in credit markets[31]. When credit markets are less than perfect, access to credit is usually conditioned on the borrower's wealth. For example, interest rates may be lower for individuals who can provide collateral assets or, in some cases, access to credit may be rationed or even denied to those who cannot provide collateral assets. In the former case the opportunity cost of capital is higher for the poor than for the wealthy while in the latter the poor are credit constrained. In either case the poor will end up investing less than the wealthy even if they have the same investment opportunities. Thus, the wealth distribution will influence the allocation of investments in the economy. In contrast, in a world in which credit markets work perfectly, investment decisions depend exclusively on the return on investments and on the market price of

[31] In what follows, we will assume that credit and human capital markets are one and the same. Although the sources of market imperfections are fundamentally different in each case, this simplifying assumption does not affect our results.

capital. Economic agents can always borrow what they need and the allocation of investments is not influenced by the way in which aggregate wealth is distributed across households. Thus, the next result follows.

> **Result 2**: If credit markets are perfect then the distribution of wealth does not affect the allocation of investments in the economy.

Two conclusions follow immediately from this result. First, wealth inequality does not lead to an inefficient allocation of investments across families. In a perfect credit market there is a single interest rate for borrowers and lenders. Since optimal investment decisions imply equality between the marginal return on investments and the interest rate, the marginal return on investments should be the same for all individuals in the economy. Therefore, the allocation of investments will be efficient. Second, regardless of their wealth, individuals will choose those investment opportunities that yield the highest possible returns. This continues to be the case even if those investment opportunities require an initial indivisible investment. This last conclusion leads us to the following corollary.

> **Result 3**: Suppose that the accumulation of human or physical capital requires an initial indivisible investment. If credit markets are perfect then the distribution of wealth does not affect the allocation of investments in the economy.

Recall that the presence of indivisibilities of investment opportunities is what distinguishes the theories in the neutrality view from those in the equalization view. The predictions of the neutrality view diverge from those of the equalization view because, in a world in which credit markets are less than perfect, poor families may never accumulate enough wealth to invest in entrepreneurial activities or professions requiring a sufficiently high initial indivisible investment. As a result, historical inequalities may persist over time. This corollary implies that if credit markets work perfectly then the predictions of the neutrality view will be similar to the predictions of the equalization view.

Our first result established the invariance of the wealth distribution once the economy reaches a steady state. Our next result provides a characterization of the transitional dynamics of the wealth distribution in the model developed in Section 2.1. Note that Result 3 implies that these transitional dynamics also characterize the evolution of inequality in the model developed in Section 2.2. Consider the following variation of the model developed in Section 2.1. First, assume that there are no idiosyncratic shocks, i.e., $\alpha = 1$ for all families. Second, instead of a safe asset assume that there is a credit market in which, at the beginning of each period, individuals can borrow as much as they want at an interest rate r_b or lend as much as they want at an interest rate r_l. All loans are repaid at the end of each period[32]. Say that the credit market is perfect if $r_b = r_l \equiv r_t$, for all t, and r_t is determined endogenously each period by equating the aggregate supply with the aggregate demand for loans. Then we have the following result.

> **Result 4**: If credit markets are perfect then the initial distribution of wealth is preserved as the economy converges to its steady state.

The intuition behind this result is simple. If credit markets are perfect then the marginal return on investments should be the same for all individuals in the economy. Since individuals bequeath a constant fraction of their end-of-period wealth, the wealth of all families in the economy should grow at the same rate. Therefore, the initial distribution of wealth is reproduced generation after generation. In other words, if credit markets are perfect then historical inequalities are reproduced over time. Chaterjee (1994) shows that a similar result holds true in a standard neoclassical model with perfectly competitive capital markets. Chatterjee proves that if preferences are quasi-homothetic then the ranking of families is preserved and the distribution of wealth becomes invariant over time. Thus, Chaterjee (1994, p. 99) concludes that "[s]omewhat surprisingly, imperfections in the capital market could, in the long run, improve the distribution of wealth. In particular, in a

[32] Note that if contract enforcement is difficult and lenders have to keep track of borrowers to make sure their loans get repaid then it is possible that $r_b > r_l$. In this case the predictions of this model are similar to the predictions of the model developed in Section 2.1. In particular, the marginal return on investments for the poor who need to borrow to invest will be equal to r_b while the marginal return for the rich will be equal to r_l. Thus, inequality will lead to an inefficient allocation of investments across families. Moreover, since $r_b > r_l$ the wealth of poor families will grow faster than the wealth of rich families.

world where equity and credit markets are absent, the long-run distribution of wealth is perfectly equal since all agents have access to the same technology and individually converge to the same long-run capital stock."

We have not discussed the effect of introducing a perfectly competitive credit market on the predictions of the model developed in Section 2.3. As we have already seen, introducing a credit market is a simple task in the models developed in Sections 2.1 and 2.2. Unfortunately, introducing a credit market in the model developed in Section 2.3 requires a level of analytical complexity that goes beyond the competence of this survey[33]. Thus, we cannot provide a formal analysis of the changes in the predictions of this model. Nevertheless, there is some room for conjecture. The main predictions of the model in Section 2.3 are that any steady state must involve some degree of inequality in earned incomes and welfare. It seems most plausible to assume that any steady state must involve some degree of inequality in earnings even if credit markets are perfect. However, the question of whether or not any steady state must involve some degree of inequality in welfare is too complex to be answered outside a formal theoretical framework. In the disequalization view, inequality in earned incomes is a necessary outcome of market forces promoting separation in investment choices. Since all occupations must be supplied in equilibrium, earnings differentials must, at least, *compensate* individuals for their educational expenditures. Otherwise, some occupations will not be supplied. This line of reasoning continues to be true whether or not credit markets are perfect. Welfare inequality, on the other hand, requires earnings differentials to *overcompensate* individuals for their educational expenditures. When credit markets are missing, this situation can be sustained as a steady state because educational expenditures are financed through forgone consumption. Since the utility function is strictly concave and unskilled workers earn lower wages than skilled workers, the welfare loss from forgone consumption is higher for unskilled than for skilled workers. As a result, unskilled workers may have no incentive to invest in their children's education even if earnings differentials overcompensate skilled

[33] To be more precise, a meaningful credit market requires allowing for some form of financial bequests, i.e., intergenerational transfers of credit market contractual obligations. Leaving aside the question of whether or not these intergenerational transfers are legally viable (see Loury, 1981), the problem is that allowing for any type of financial bequests is precisely what complicates the analysis (see Mookherjee and Ray, 2006).

workers for their educational expenditures. This last result may change when credit markets are perfect. If loan repayments can be delegated to future generations then a skilled-unskilled wage differential that overcompensates for educational expenses may induce some unskilled workers to educate their children. In this case the wage differential will be eroded by the higher supply of skilled labor. Whether or not this process would lead to a steady state in which earnings differentials exactly compensate for educational expenses and welfare inequality vanishes is not clear.

3 Inequality and Development

The previous section surveyed the main theories examining the economic mechanisms through which inequality and development are interconnected. In this section we analyze the transitional dynamics of some of these theories[34]. These transitional dynamics implicitly establish an intertemporal relation between inequality and development. Most of the literature on the relation between inequality and development revolves around the Kuznets (1955) inverted-U hypothesis. The intuition behind the Kuznets hypothesis can be better understood in the context of Lewis (1954) model of a dual economy. According to Lewis, the economy is divided into two sectors: a traditional rural agricultural sector and a modern urban industrial sector. The process of development is then identified with the growth of the urban industrial sector, i.e., with the processes of urbanization and industrialization. Industrialization is fueled by labor migrating from the rural to the urban sectors. The main premise behind the Kuznets hypothesis is that both average income and income inequality in the rural sector are lower than in the urban sector. In early stages of development, most of the population is in the rural sector and the economy's overall average income and degree of inequality are low. As the rural population begins to migrate to the urban sector, average income and income inequality increase. The economy's average income increases because migrants experience a rise in their income. However, since a larger fraction of the population is now in the high-inequality urban sector, income inequality also increases. Thus, during the early stages of development the relation between inequality and

[34] The analysis in this section is per force limited to those theories in which the transitional dynamics of the model can be clearly assessed.

development tends to be positive. Kuznets argues that once the early stages of development have passed, a variety of factors converge to reduce the degree of inequality in the urban sector[35]. Hence, as the size of the urban sector continues to grow, inequality begins to decline. Therefore, during the later stages of development the relation between inequality and development becomes negative. The previous arguments led Kuznets (1955, p. 18) to conjecture "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." In other words, the relation between inequality and development should follow an inverted-U pattern. This conjecture is known as the Kuznets inverted-U hypothesis or the Kuznets curve.

The empirical evidence on the Kuznets hypothesis is mixed[36]. The Kuznets hypothesis was widely accepted as an empirical regularity until the 1980's. Early cross-country studies found strong support for an inverted-U pattern in the relation between inequality and development, as measured by income per capita (see, e.g., Paukert, 1973; and Ahluwalia, 1976a, 1976b)[37]. However, the findings of subsequent cross-country studies were more controversial. Some of these studies do find empirical support for the Kuznets hypothesis but are more cautious in assessing the relevance of their results. For example, Papanek and Kyn (1986) find evidence in favor of the Kuznets hypothesis but argue that the evidence is not strong and may be weakening over time. Bourguignon and Morrison (1990) also find evidence in favor of the Kuznets hypothesis but note that this evidence largely vanishes with the addition of variables such as education or protection into the regression equation. Barro (2000), on the other hand, finds that the Kuznets hypothesis is a clear empirical

[35] For example, Kuznets argues that, after the early stages of urbanization and industrialization have passed, most of the urban population would be "native", i.e., born in cities. This new industrial labor force would be more experienced and efficient than previous immigrant workers. Moreover, a native urban population would be more able to organize and exercise greater political power than an urban population composed mostly of immigrants. These factors would raise the income share of urban workers, reducing the degree of inequality in the urban sector.

[36] See Kanbur (2000) for an extensive survey of the empirical literature on the Kuznets hypothesis.

[37] Although this is essentially an intertemporal relation, most studies use cross-country data due to lack of sufficient reliable longitudinal data on inequality for individual countries. We have already discussed the main problems with data on income inequality in Section 2.4.

regularity but that this hypothesis explains little of the variations in inequality across countries over time. Other studies show that the evidence in favor of the Kuznets hypothesis is not robust to the specification of the regression equation (e.g., Anand and Kanbur, 1993) or sample composition (e.g., Deininger and Squire, 1998). Perhaps the most damaging critique to cross-country evidence on the Kuznets hypothesis comes from the observation that income inequality varies significantly across countries but is relatively stable within countries over time. Using a sample for 49 developed and developing countries covering the period 1947-1994, Li, Squire, and Zou (1998) find that variation across countries explains around 90 percent of the variance in Gini coefficients while variation over time explains only a small percentage. They conclude that the determinants of inequality are significantly different across countries but relatively stable over time. These results suggest that evidence in favor of an inverted-U relationship between inequality and development in cross-country studies may be biased due to region-specific or country-specific characteristics. This possibility is tested by Deininger and Squire (1998) who find that support for the Kuznets hypothesis vanishes when either a dummy for Latin American countries or when country-specific dummy variables are included in the regression equation. In their view, there is no "unmovable universal law" governing the evolution of inequality and development. In contrast, Barro (2000) finds that the relation between inequality and development follows an inverted-U pattern even after a dummy for Latin American countries or country-specific dummy variables are included[38].

The evidence for individual countries is equally mixed. Historical data from a few developed countries, such as Great Britain and the United States, seems to support an inverted-U relationship between inequality and development. For example, Lindert (1986) and Lindert and Williamson (1985) find that, after a period of increased inequality during the Industrial Revolution era, the wealth distribution in England followed a tendency towards increased equality, especially since World War I. Indeed, Lindert and Williamson

[38] These contradictory results illustrate one of the shortcomings of the evidence in favor of the Kuznets hypothesis mentioned earlier, i.e., the results are not robust to the specification of the regression equation. While Barro (2000) follows Ahluwalia (1976b) and uses income per capita and income per capita squared as explanatory variables, Deininger and Squire (1998) follow Anand and Kanbur (1993) and use income per capita and the inverse of income per capita as explanatory variables.

(1985, p. 344) argue that the "British experience since 1688 looks like an excellent advertisement for the Kuznets curve." However, Deininger and Squire (1998) find little support for the Kuznets hypothesis in a sample of 49 developing and developed countries using data since the 1960's. They find no statistically significant relationship between inequality and development in 40 of the 49 countries in the sample. In 4 of the remaining 9 countries the data supports the presence of a direct-U, not an inverted-U, relationship. This leaves us with 5 out of 49 countries in which the evidence seems to support the Kuznets hypothesis. However, the evidence in favor of the Kuznets hypothesis is dubious in 2 of these countries. In Hungary a sudden increase in inequality coupled with a fall in income per capita creates the false impression of an inverted-U relation and in Brazil support for the Kuznets hypothesis vanishes by deleting one observation for 1960. In other words, the data supports the Kuznets hypothesis only in 3 of the 49 countries: Mexico, the Philippines, and Trinidad.

After this brief survey of the empirical literature we are driven to agree with the World Bank (2006, p.43-44) in concluding that "[t]here is still no consensus on a systematic relationship between the long-term growth processes of industrialization and urbanization and overall inequality." In contrast, the conclusions of the theoretical literature on the relation between inequality and development are more straightforward. The transitional dynamics discussed in this section can be classified into two types of processes: (*i*) Monotonic processes in which inequality either monotonically increases or monotonically decreases with development, and (*ii*) Kuznets processes in which the relation between inequality and development follows an inverted-U pattern.

3.1 Monotonic Processes

Perhaps the model in Section 2.1 is the simplest example of a monotonic process. In this model there is a unique and globally stable invariant distribution. Therefore, the level of inequality in the economy always converges to its steady-state level. When the initial level of inequality is higher than the steady-state level, the intergenerational evolution of inequality is driven by two features of the model: investment decisions are wealth constrained and the production function is strictly concave. As a result, the expected

marginal return on capital investments is greater for poor entrepreneurs than for rich ones. Since individuals bequeath a constant fraction of their end-of-period wealth to the next generation, on average, the wealth of poor families should grow faster than the wealth of rich families. Therefore, inequality monotonically decreases as the economy grows toward its steady state. When the initial level of inequality is lower than the steady-state level, the intergenerational evolution of inequality is driven mostly by idiosyncratic shocks. Since idiosyncratic shocks are a source of variation in income across individuals, inequality monotonically increases as the economy converges to its unique steady state. Analogous transitional dynamics describe the evolution of inequality and income per capita in Loury (1981) and Banerjee and Newman (1991). However, when the initial level of inequality is higher than the steady-state level, the driving force behind the evolution of inequality in Banerjee and Newman (1991) is not the concavity of the production function. Instead, they argue that insurance markets will require rich entrepreneurs to bear more absolute risk than poor entrepreneurs. As a consequence, the poor are more likely than the rich to undertake risky entrepreneurial projects with high expected returns. This, in turn, implies that, on average, the wealth of poor families should grow faster than the wealth of rich families.

Another example of a monotonic process is Ray (1990, 2006). The transitional dynamics in Ray's model are different from those in the previous models in two aspects. First, since there are multiple steady states, not all development paths lead to the same steady state. Second, convergence is not gradual but sudden when the initial level of inequality is lower than the lowest possible steady-state level of inequality. Steady states in this model are fully characterized by their ratio of skilled workers. Consider an initial distribution of wealth such that the initial ratio of skilled workers in the economy is not a steady state. If the initial ratio of skilled workers is smaller than the highest steady state (i.e., the steady state with the highest ratio of skilled workers) then the wage differential will be large enough to induce unskilled workers to educate their children. Therefore, the ratio of skilled workers increases over time and the economy gradually converges to the closest steady state with a ratio of skilled workers higher than the initial ratio. Inequality monotonically decreases along this development path because the wage differential decreases as the ratio of skilled workers increases. At the same time, since the initial ratio of skilled workers

implies underinvestment in education, income per capita increases along this development path. In other words, the relation between inequality and development is negative. On the other hand, if the initial ratio of skilled workers exceeds the highest steady state then the wage differential is not sufficient to compensate parents for their educational expenditures. Some skilled workers choose to leave their children uneducated and the economy converges to the highest steady state immediately. That is, there is a sudden increase in inequality immediately followed by stability.

An alternative type of monotonic process is described by the transitional dynamics of the model in Section 2.2. This model has a connected continuum of steady states which contains one steady state that involves perfect equality and a continuum of steady states that involve persistent inequality. If the initial level of inequality is sufficiently low then the economy always converges to the steady state with a perfectly egalitarian distribution of wealth, i.e., either all families invest in cottage production or all families invest in industrial production in the long run. Since investment decisions are wealth constrained and the production function is strictly concave, inequality monotonically decreases while income per capita increases as the economy converges to the steady state. For higher initial levels of inequality the economy always converges to a steady state that involves persistent inequality. Lowerand lower-middle-class families are trapped in cottage production generation after generation while upper-middle- and upper-class families invest in industrial production. Inequality and income per capita monotonically increase as the economy converges to a steady state with only two social classes: poor cottage factory owners and wealthy industrial entrepreneurs. In other words, the relation between inequality and development is positive as the middle class vanishes and society becomes completely polarized over time. Similar transitional dynamics describe the evolution of inequality and income per capita in Galor and Zeira (1993). Galor and Zeira consider an economy in which investments in physical capital are replaced by investments in human capital. When the initial level of inequality is low the economy converges to a perfectly egalitarian steady state in which either all families invest in education or all remain unskilled generation after generation. The difference with the previous model is that the relation between inequality and development is negative along paths leading to the steady

state in which all families invest in education but is positive along paths leading to the steady state in which all families remain unskilled[39]. When the initial level of inequality is sufficiently high, the relation between inequality and development is the same as in the model in Section 2.2. In the long run society is completely polarized and families are divided into two social classes: wealthy families who invest in education and poor families who remain unskilled generation after generation. Both inequality and income per capita increase as the economy converges to its steady state. Galor and Zeira prove the existence of a poverty trap threshold such that families whose initial wealth is above this threshold invest in education generation after generation. The earlier generations of families whose initial wealth is below the threshold may invest in education but their descendants will eventually become unskilled and will remain unskilled generation after generation. Hence, the development path of the economy is fully determined by the initial distribution of wealth.

3.2 Kuznets Processes

One of the earliest examples of a Kuznets process is Greenwood and Jovanovic (1990). They argue that the relation between inequality and development is heavily influenced by the development of financial markets. In early stages of development only few individuals with sufficient wealth can invest through financial intermediaries. At this stage, the role of financial markets is insignificant, the economy grows slowly and the economy's overall average income and level of inequality are low. When the economy enters an intermediate stage of development, a financial superstructure begins to grow as more individuals accumulate enough wealth to invest in financial markets. The efficiency enhancing properties of financial institutions accelerate economic growth. At the same time, the gap between the rich, who can afford to invest in financial markets, and the poor, who are left out, widens. As the economy reaches the final stages of development, an extensive structure for financial intermediation has emerged. Since most people get to enjoy the high returns on financial investments, inequality declines until the distribution of wealth converges to an invariant distribution. In contrast, Aghion and Bolton (1997) emphasize the role of credit

[39] In both cases inequality decreases as the economy converges to the steady state. However, in the former case income per capita increases while in the latter income per capita decreases.

markets on the relation between inequality and development. In early stages of development, when aggregate wealth in the economy is low, the interest rate is high and poor individuals are discouraged from borrowing. Only upper- and middle-class families share the high expected return on capital investments and, on average, their wealth grows faster than the wealth of lower-class families. Thus, in early stages of development inequality increases because the gap between the rich and the poor gets wider. However, if entrepreneurs can accumulate capital sufficiently fast then, as the economy grows, aggregate savings will increase pushing down the interest rate. As the interest rate decreases, the opportunity cost of capital for the poor decreases, allowing them to undertake more profitable investments. In the long run, most families undertake entrepreneurial projects and, on average, the wealth of all families converges to a common long-run wealth. In other words, the initial tendency towards increased inequality eventually stops and the gap between rich and poor families begins to narrow.

Galor and Moav (2006) provide an alternative example of a Kuznets process. Galor and Moav argue that capital-skill complementarities produced a change in the attitude of capitalists that led to improvements in publicly financed education. These improvements, in turn, lead to the demise of the class structure in 19th century Western Europe. During early stages of development the aggregate stock of physical capital is low and the return on physical capital investments is high. Since the working class remains mostly uneducated during this period, labor earnings are low. Hence, the wealth of capitalists grows faster than the wealth of workers widening the gap between the two social classes. As the aggregate stock of physical capital grows, capital-skill complementarity makes the accumulation of human capital essential to sustain the rate of return on physical capital. As a consequence, capitalists become more willing to support improvements in publicly financed education, increasing the human capital of the working class. As the human capital of the working class increases, their income share also increases and inequality falls. Thus, in early stages of development physical capital accumulation is the prime engine of growth and the relation between inequality and growth is positive. However, as the stock of physical capital grows, the relative importance of human capital accumulation increases, the initial

tendency toward greater inequality is reversed and the relation between inequality and development becomes negative.

Finally, consider the model in Banerjee and Newman (1993). This model has two steady states: a high-wage high-mobility steady state and a low-wage low-mobility steady state. Banerjee and Newman assume that the middle class always experiences substantial upward mobility but little downward mobility. The transitional dynamics for the upper and lower classes depend on the equilibrium wage rate along the development path. When the equilibrium wage rate is low, most of the downward mobility from the upper class is to the middle class while there is little upward mobility from the lower class to the top two classes. When the equilibrium wage rate is high, there is substantial upward mobility from the lower class to the middle class and downward mobility from the upper class to the lower class. Development paths leading to the high-wage high-mobility steady state imply that during the last stages of development the high-wage dynamics determine the evolution of the income distribution. In contrast, in development paths leading to the low-wage low-mobility steady state the evolution of the wealth distribution is always ruled by the low-wage dynamics. Nevertheless, the transitional dynamics leading to both steady states are similar during early stages of development. Consider an initial distribution of wealth such that most of the population is in the middle class and the ratio of poor to wealthy is high. Hence, labor is relatively abundant and the equilibrium wage rate is low. Since most of the population belongs to the middle class, the upward and downward flows of individuals from the middle class to the upper and lower classes outweighs the flows in the opposite directions and both the upper and lower classes grow. However, the upper class grows faster than the lower class because there is little social mobility between the top two classes and the bottom class. Therefore, the ratio of poor to wealthy increases over time. In early stages of development the ratio of poor to wealthy continues to be sufficiently high such that the equilibrium wage rate continues to be low. Since the upper and lower classes grow at the expense of the middle class, overall inequality increases as income per capita increases.

Once the early stages of development have passed, the transitional dynamics of the model depend on the initial size of the lower class. If the initial size of the lower class is sufficiently small then the increasing demand for labor will eventually exceed the supply and the equilibrium wage will be high. The high wage allows for substantial upward mobility for the lower class, reducing the fraction of potential entrepreneurs who can find workers. Most of the upper class chooses to be self-employed and downward mobility to the lower class falls. The size of the lower class continues to decline until the economy reaches the high-wage high-mobility steady state. Thus, during later stages of development inequality declines because the lower class receives higher wages and becomes smaller over time. However, for a slightly larger initial lower class the demand for labor will never exceed the supply and the economy converges to the low-wage low-mobility steady state. In this case, inequality continues to increase during later phases of development. Hence, this model can display both a Kuznets and a monotonic process. To be more precise, the transitional dynamics leading to the high-wage high-mobility steady state may be best described as a Kuznets process while those leading to the low-wage low-mobility steady state may be best described as a monotonic process.

4 Technological Change and Inequality

There is a long tradition in economics of theoretical research on the role of technological change in the process of economic growth. In comparison, theoretical research on the redistributive properties of new technologies is relatively new. Interest in the latter was ignited by recent trends in the wage and employment structures of most developed countries. These trends are well documented for the United States economy. Wage inequality was relatively stable in the United States until the 1970's when it started to rise at a fast pace (Acemoglu, 2002). The share of college graduates in total employment in 1996 was more than three times higher than in 1950 while the college premium increased by 25 percent during the same period (Autor, Katz, and Krueger, 1998). Moreover, increased demand for highly educated workers was predominantly a within-industry phenomenon, while labor reallocation from low-skill to high-skill industries was relatively small (Bound and Johnson, 1992; Katz and Murphy, 1992; Berman, Bound, and Griliches,

1994). Finally, the supply of highly educated workers rose substantially between 1940 and 1995. Autor, Katz, and Krueger (1998) document that the ratio of college equivalents to noncollege equivalents increased at an average rate of 2.35 percent a year between 1940 and 1970 and of 3.05 percent between 1970 and 1995[40].

The trends for other OECD countries are similar to those observed in the United States. Most OECD countries experienced an increase in wage inequality by the middle of the 1980's. Nevertheless, the magnitude of the changes varies substantially between countries. Wage inequality increased sharply in some while it remained relatively stable in others. For example, during the 1980's wage inequality rose faster in the United Kingdom than in the United States while France and Germany experienced a more moderate rise in wage inequality (Machin, 1996; Aghion, 2002). The share of skilled workers in total wage costs and employment also increased in recent decades, with most skill upgrading occurring within industries (Berman, Bound, and Machin, 1998; Machin and Van Reenen, 1998; Berman and Machin, 2000). Between 1979 and 1992 unemployment in European OECD countries rose from 4.9 percent to 9.9 percent, with most of the unemployed being unskilled workers (Berman, Bound, and Machin, 1998). Finally, the supply of highly educated workers increased while the college premium showed no tendency to decline in most countries.

There is a growing consensus arguing that skill-biased technological change plays a central role in explaining these recent changes in labor markets. Autor, Katz, and Krueger (1998, p. 1170) provide a succinct summary of this view

> "The continued increase in the relative utilization of nonproduction workers and more-educated workers within detailed industries and within establishments in the United States, despite the rising relative wages of these groups during the 1980's and 1990's indicates strong within-industry and

[40] Autor, Katz, and Krueger define college equivalents as college graduates plus 0.5 times those with some college education and noncollege equivalents as those with a high school degree or less plus 0.5 times those with some college education.

within-establishment demand shifts favoring the more-educated that is often interpreted as reflecting skill-biased technological change."

The literature surveyed in this section provides a theoretical framework to understand the effect of technological change on the wage and employment structures of developed countries. The main premise is that new technologies favor skilled labor because skilled workers are more able than unskilled workers to deal with the changes involved in the adoption and operation of those technologies. Moreover, new technologies often replace tasks performed by unskilled workers with advanced machinery. As a result, the demand for skilled labor rises while the demand for unskilled labor plummets. This change in the pattern of demand for labor is responsible for exacerbating earnings inequality.

It is important to emphasize that the theories surveyed in this section are intrinsically different from the theories surveyed in Section 2 in two fundamental ways. First, these theories assume that all markets, including capital markets, are perfectly competitive. Second, these theories assume that parents are not altruistic toward their children and implicitly exclude the possibility of any type of intergenerational transfers. As a consequence, wealth inequality in one generation does not lead to unequal opportunities for the next generation. Moreover, for each generation the allocation of investments across families is efficient regardless of how wealth was distributed in the past. Therefore, there is no scope for efficiency-based policy interventions. Indeed, the theories surveyed in this section are not concerned with the normative or functional aspects of inequality. Instead, their focus is restricted to the analysis of the effect of technological changes on the intergenerational evolution of earnings inequality. These theories rely on idiosyncratic random shocks to ability to generate inequality within each generation and study how technological changes affect this ability-induced level of inequality[41].

One of the earliest theories linking technological change and inequality is Greenwood and Yarukoglu (1997). They consider an economy consisting of a large number of individuals

[41] Note that Result 1 in Section 2.4.2 does not apply in this context because there is no intergenerational altruism.

and a firm using skilled and unskilled labor to produce a single homogeneous good at a variety of plants. Plants differ in the technologies used in production and the firm can open and close plants as desired. The supply of skilled and unskilled labor depends on the occupational choices of individuals. Greenwood and Yarukoglu assume that the earnings of skilled workers increase with their ability and the cost of becoming skilled is a decreasing function of an individual's ability. Hence, there is a threshold such that in every generation only those individuals with ability levels above the threshold become skilled workers. Technological change increases total factor productivity and therefore has little effect on the relative demand for skilled labor. However, the adoption of new technologies involves a substantial learning cost and skilled labor has an advantage over unskilled labor at learning. Therefore, the demand for skilled labor increases during the initial phases of adoption and implementation of new technologies. Moreover, as the rate of technological progress increases, new technologies represent more radical departures from previous technologies and the cost of learning a new technology increases. Therefore, an acceleration in the rate of technological progress produces a sharp increase in the demand for skilled labor. In turn, the higher demand for skilled labor leads to an increase in the skill premium during the learning phase. Once the learning phase is over and workers become fully efficient at using the new technology, the firm substitutes relatively expensive skilled labor with more economical unskilled labor. This shift in labor demand drives down the skill premium. Thus, an acceleration in the rate of technological progress would lead to an increase in earnings inequality during the initial phases of adoption and implementation of the new technology and then to a fall in inequality as the new technology becomes established.

One implication of the assumptions in Greenwood and Yarukoglu (1997) is that all technologies yield the same return on the ability of skilled workers. Since there are only two types of labor in the economy (i.e., skilled and unskilled) and since all technologies require both types of labor as complementary inputs in production, in equilibrium the return on the ability of skilled workers must be the same across all technologies. In contrast, Galor and Tsiddon (1997) argue that new technologies require a labor force with higher levels of skills than previous technologies and, therefore, must yield a higher return on ability. In their view, the higher return on ability should lead to a reallocation of labor such that the

average ability of workers in technologically advanced sectors is higher than in sectors using older technologies. In turn, this reallocation of labor between new and old technologies is responsible for exacerbating earnings inequality.

Galor and Tsiddon (1997) consider an economy in which, in every period, two sectors produce a single homogeneous good. The two sectors differ in their production technologies, i.e., an old and a new technology. Whenever a newer technology is invented, the previously old technology is immediately abandoned. New technologies require a more educated labor force than previous technologies but yield a higher return on the ability of skilled workers. The earnings of individuals increase with their ability and parental specific human capital. The ability effect is stronger in the technologically advanced sector than in the sector using the old technology. The parental effect, on the other hand, is stronger for individuals who work in the same sector as their parents. Hence, the ability effect promotes intergenerational earnings mobility while the parental effect introduces some intergenerational persistence of economic status. The relative importance of the two effects and, therefore, the occupational choices of individuals depend on the pace of technological progress. Galor and Tsiddon identify periods of rapid technological progress with periods in which new technologies are invented. Since new technologies yield a higher return on ability, the ability effect becomes the dominant factor for most individuals and intergenerational mobility increases. Moreover, earnings inequality between high-skill and low-skill workers increases because the difference in the return on ability between the two sectors increases and because the average ability of high-skill workers increases while the average ability of low-skill workers decreases. In contrast, periods of slow technological progress are periods in which innovations make already existing technologies more accessible, reducing the need for and the return on ability. As a consequence, the parental effect is the dominating factor for most individuals and intergenerational mobility is attenuated. Earnings inequality decreases because the difference in the return on ability between the two sectors decreases and because the ability dispersion increases for high-skill and low-skill workers. Therefore, an acceleration in the rate of technological progress exacerbates earnings inequality but enhances earnings mobility while inequality declines but becomes more persistent as the rate of technological progress slows down.

Since technological change is labor augmenting in Galor and Tsiddon (1997), the ratio of capital to efficiency units of labor is the same across all technologies. Caselli (1999) argues instead that technological change increases total factor productivity and, therefore, the capital-labor ratio should be higher in firms using new technologies than in firms using older technologies. Moreover, since new technologies require a labor force with higher levels of skills than previous technologies, the capital endowment of high-skill workers must be higher than the capital endowment of low-skill workers. According to Caselli, these differences in capital-labor ratios between workers with different levels of skills are responsible for strengthening earnings inequality.

Caselli (1999) considers an economy in which, in every period, firms can choose from a menu of technologies available in the economy. New technologies are more productive than previous technologies but can only be operated by workers who have acquired the necessary skills[42]. When a new technology is introduced into the economy the effect on earnings inequality depends on the relation between the productivity gains of the new technology and the cost of learning the skills necessary to operate it. If the new technology brings sufficiently large productivity gains relative to the cost of learning it then all previous technologies are immediately abandoned. All individuals acquire the skills to operate the new technology and earnings inequality vanishes. If, instead, the new technology generates a moderate productivity gain relative to its learning cost then previous technologies are not immediately abandoned. Since individuals differ in cognitive ability, only those individuals with sufficiently low learning costs acquire the skills to operate the new technology while the rest acquire the skills to operate previous technologies. How is capital allocated between the different technologies in the economy? In equilibrium, the return on capital investments must be the same across all technologies because capital markets are perfectly competitive. Since technological change increases total factor productivity, this can only be achieved by increasing the capital-labor ratio of workers using the new technology relative to workers using previous technologies. In this case

[42] Caselli allows for skill-replacing (or de-skilling) as well as skill-biased technological changes. For ease of exposition we restrict attention to the skill-biased case.

earnings inequality increases with the introduction of the new technology because high-skill workers use a more productive technology and have a higher capital-labor ratio than low-skill workers[43]. Whether the initial increase in inequality eventually vanishes or persists in the long run depends on the behavior of the wage differential as the economy approaches its new steady state. As the economy grows toward its new steady state, the aggregate capital stock in the economy increases while the aggregate labor supply remains constant. This implies that capital-labor ratios must be increasing across all technologies. Since the returns on capital investments must be the same for all technologies, the wage differential between high-skill and low-skill workers must be increasing over time. Nevertheless, decreasing marginal returns on capital investments imply that the aggregate capital stock increases at a decreasing rate. Therefore, the wage differential increase at a decreasing rate until it eventually ceases to grow when the economy reaches its new steady state. As the wage differential increases, the proportion of the population acquiring the skills to operate the new technology increases. Hence, least efficient technologies are gradually abandoned as the economy approaches the new steady state. If the wage differential grows sufficiently fast then the new technology ultimately displaces all previous technologies. All individuals eventually acquire the skills to operate the new technology and earnings inequality vanishes in the long run. However, if the wage differential grows slowly then the economy may reach its new steady state before all previous technologies are completely abandoned. The labor market remains split into high-skill and low-skill workers and inequality persists in the long run.

The previous theories explain the increase in wage inequality *between* skilled and unskilled labor. However, one of the main characteristics of the recent changes in labor markets of developed countries is that most of the increase in wage inequality is explained by the increase in inequality *within* the groups of skilled and unskilled labor. Galor and Moav (2000) provide a theory that can account for the increase in inequality within as well as between the groups of skilled and unskilled labor.

[43] Furthermore, if the economy was in, or near to, its steady state prior to the introduction of the new technology then the increase in earnings inequality is accompanied by a decline in the wages of low-skill workers because their capital-labor ratio is lower than before.

Galor and Moav (2000) consider an economy in which skilled and unskilled labor are substitutes in production and technological progress leads to the immediate abandonment of old technologies. Since technological change increases total factor productivity, the level of technology has no effect on the relative demand for skilled labor. An acceleration in the rate of technological progress, on the other hand, increases the relative demand for skilled labor. The human capital and the earnings of individuals depend on their ability and the technological environment. Technological change has two opposite effects on individual earnings: an "erosion effect" and a "productivity effect". New technologies erode the human capital of workers and reduce their earnings because human capital is technology specific. Since cognitive ability speeds up the process of learning a new technology, ability mitigates the erosion effect. On the other hand, new technologies increase total factor productivity and, therefore, raise the earnings of skilled and unskilled workers. The influence of the erosion and productivity effects on earnings inequality depends on the rate of technological progress. When the rate of technological progress is constant, the relative demand for skilled labor is stable and the erosion effect is constant over time. Hence, the earnings of skilled and unskilled workers grow at the constant rate of technological progress due to the productivity effect and earnings inequality between and within groups remain stable. In contrast, an acceleration in the rate of technological progress accentuates the erosion and productivity effects because new technologies represent more radical departures from previous technologies. The rise in the relative demand for skilled labor in addition to the productivity effect produces a sharp increase in the return on the ability of skilled workers. In turn, the increase in the return on the ability of skilled workers increases the relative supply of skilled labor because it lowers the threshold level of ability above which individuals choose to become skilled workers. This implies that the average level of ability of skilled and unskilled workers declines. In equilibrium, the acceleration in the rate of technological progress produces an increase in earnings inequality between skilled and unskilled workers because the average earnings of skilled workers rises while the average earnings of unskilled workers may fall, despite the increase in the relative supply of skilled labor. The average earnings of skilled workers increase because the return on the ability of skilled workers increases. On the other hand, the average earnings of unskilled workers

may decline because both the relative demand for unskilled labor and the average level of ability of unskilled workers fall. Moreover, the acceleration in the rate of technological progress also increases earnings inequality within the groups of skilled and unskilled workers due to the erosion effect. The role of ability in mitigating the erosion of human capital is accentuated by the acceleration in the rate of technological progress. Therefore, the influence of ability on earnings increases for skilled and unskilled workers, increasing the dispersion in earnings within both groups of workers.

Galor and Tsiddon (1997), Greenwood and Yarukoglu (1997), Caselli (1999), and Galor and Moav (2000) study alternative channels through which technological change affects inequality. In this sense, their work should be considered as complementing each other rather than as competing theories. One shortcoming common to all these theories is that the direction of technological progress, i.e., the skill content of new technologies, is exogenously determined. Acemoglu (1998, 2002) and Krusell, Ohanian, Rios-Rull, and Violante (2000) provide alternative theories in which the skill content of new technologies is endogenously determined. Acemoglu argues that the direction of technological change depends on the relative profitability of developing different types of technologies. When the relative supply of skilled labor increases, the market for skilled-biased technologies expands making the development of such technologies more profitable. In contrast, new technologies are skill-replacing when the relative supply of unskilled labor increases. According to this theory, technological change in the United States has been skill-biased since the 1970's because the supply of skilled labor experienced a sharp increase in the late 1960's caused by the large supply of college graduates from the baby-boom generation. The increase in the relative supply of skilled labor expanded the size of the market for skill-biased technologies making the development of such technologies more profitable. Acceleration in the rate of growth of the demand for skilled labor in the 1970's is then explained by a change in the direction of technological progress, i.e., the development of ever more skilled-biased technologies. On the other hand, Krusell et al. argue that the increase in the stock of physical equipment resulting from the decline in the relative price of physical equipment goods in the post-war period led to an increase in the demand for skilled labor. Their premise is that physical equipment is more complementary to skilled

labor than to unskilled labor. The acceleration in rate of growth of the demand for skilled labor during the 1970's is then explained by the accelerated decline in the relative price of physical equipment goods since the early 1970's which led to an acceleration in the accumulation of physical equipment. Both theories provide compelling explanations of the trends discussed at the beginning of this section. However, it is difficult to distinguish between the predictions of these two theories and the predictions of the former set of theories based on the evidence of the last decades.

To summarize, the theories surveyed in this section argue that the pace of technological progress is uneven, accelerating at some times and slowing down at others. This implies an uneven growth in the demand for skilled labor[44]. Changes in wage inequality are then explained mostly by the changes in the pace of technological progress. Wage inequality increases in periods of accelerated skill-biased technological change and remains relatively stable in periods of slow technological progress[45]. According to this view, wage inequality was relatively stable before the 1970's because the rate of technological progress and the rate of growth of the supply of skilled labor were relatively stable. The sharp increase in wage inequality that started in the 1970's in some countries and 1980's in others was the consequence of an acceleration in the rate of growth of skill-biased technological change which led to an increase in the demand for skills. The most frequent explanation given for the acceleration in the rate of growth of skill-biased technological change is the widespread use of computers in the workplace in an era of rapid advances in information technology (see, among others, Krueger, 1993; Berman, Bound, and Griliches, 1994; Autor, Katz, and Krueger, 1998; Berman, Bound, and Machin, 1998). Indeed, Greenwood and Yorukoglu (1997) and Caselli (1999) argue that the 1970's marked the beginning of a "Third Industrial Revolution" fueled by advances in information technology.

[44] One exception is Acemoglu (1998, 2002) who argues that changes in the direction, not the speed, of technological progress influence the demand for skilled labor.

[45] An alternative view argues that skill-biased technological change takes place at a constant pace, implying a steady demand growth in favor of highly-skilled workers. Changes in wage inequality are then explained by the uneven growth in the supply of skilled labor. Wage inequality declines when the rate of growth of the supply of skilled labor exceeds the constant rate of skill-biased technological change and rises when the rate of growth of the supply of skilled labor falls behind this constant rate. See, for example, Katz and Murphy (1992) and Card and Lemieux (2001).

There are two main pieces of evidence from the United States economy that appear to support the hypothesis of an acceleration in the rate of growth of skill-biased technological change during the 1970's. The first piece of evidence is provided by Autor, Katz, and Krueger (1998). They show that between 1940 and 1970 the college premium in the United States decreased at an average rate of 0.11 percent a year while the supply of highly educated workers during the same period increased at an average rate of 2.35 percent a year. In contrast, between 1970 and 1995 the college premium increased at an average rate of 0.39 percent a year while the supply of highly educated workers increased at an average rate of 3.05 percent a year. If the rate of growth of the demand for skilled labor did not increase between 1940 and 1995 then the acceleration in the rate of growth of the supply of skilled labor between 1970 and 1995 should have caused the college premium to decrease between 1970 and 1995 even faster than it did during the previous thirty years. The rise in the college premium between 1970 and 1995 suggests an acceleration in the rate of growth of the demand for skilled labor.

The second piece of evidence in favor of an acceleration in skill bias is provided by Juhn, Murphy, and Pierce (1993). They show that within-group wage inequality remained stable or fell during the 1960's but increased sharply since the early 1970's in the United States[46]. There are three causes that can potentially explain the rise in within-group inequality during the 1970's: an increase in the return to unobserved characteristics of workers, an increase in the dispersion of those unobserved characteristics, or an increase in measurement errors. Juhn, Murphy, and Pierce interpret the trend toward increased within-group wage inequality as an increase in the return to unobserved characteristics and identify those unobserved characteristics with workers' ability[47]. They conclude that "[t]he differences in the timing of the increases in wage inequality within and between groups ... point to a rise

[46] Within-group, or residual, wage inequality is measured as the wage dispersion between observationally equivalent workers. In other words, within-group wage inequality is the residual of overall inequality after we control for observable characteristics of workers, such as years of education and years of work experience.

[47] This interpretation has been recently challenged by Lemieux (2006). Lemieux argues that there are serious problems of spurious growth in within-group inequality in the United States economy due to an increase in measurement errors and to changes in the composition of the labor force resulting from the aging of the baby boom generation and from increases in the average level of education of the labor force.

in the demand for skill that predates the recent rise in returns to education by about a decade" (Juhn, Murphy, and Pierce, 1993, p. 412). There are two reasons why Juhn, Murphy, and Pierce's conclusion is crucial for the hypothesis of an acceleration in the demand for skills. First, the idea of an acceleration in the demand for skilled labor during the 1970's is difficult to reconcile with the sharp decline in the college premium during that decade unless the return to other skills rose during the same period. Second, most of the rise in overall wage inequality during the 1970's and 1980's in the United States is accounted for by the rise in within-group wage inequality. Therefore, if the rise in within-group inequality was unrelated to the demand for skills then the acceleration in the demand for skills could only account for a small fraction of the rise in overall wage inequality.

The main problem with the hypothesis of an acceleration in the growth of skill-biased technological progress is that the period between 1970 and 1995 was a period of slow productivity growth relative to previous decades. Some theories predict a period of slower productivity growth resulting from rapid technological progress[48]. Nevertheless, the length of the period of slower total factor productivity growth in most OECD countries appears to contradict the hypothesis of an acceleration in the rate of technological progress since the 1970's. As Acemoglu (2002, p. 34) puts it, "[i]t is difficult to imagine how a new and radically more profitable technology will first lead to 25 years of substantially slower growth." Despite this shortcoming, the hypothesis of an acceleration in the growth of skill-biased technological change continues to be considered as one of the main causes of the changes in the wage and employment structures of developed countries.

4.1 Trade and Technological Change

An interesting empirical regularity in many developing countries is that changes in labor markets similar to those experienced in developed countries seem to coincide with episodes of trade liberalization. For example, the evidence for Latin American countries during the 1980's and 1990's suggests a pattern of changes in labor markets that resembles the

[48] For example, Greenwood and Yorukoglu (1997) argue that it takes time to learn how to use new technologies. Therefore, new technologies will slow down productivity growth during this learning period until firms and workers become fully efficient at using those technologies.

changes experienced in developed countries during the 1970's and 1980's. Wage inequality increased in most Latin American countries during the 1980's and 1990's (Sanchez-Paramo and Schady, 2003; Goldberg and Pavcnik, 2007). There is also strong evidence of an increase in the share of skilled workers in total employment and wage costs in several countries in the region (Robbins, 1996; Berman and Machin, 2000; Sanchez-Paramo and Schady, 2003). Most of the skill upgrading took place within industries, while there is little evidence of labor reallocation between industries (Berman and Machin, 2000). Moreover, the college premium increased in all countries for which there is available data (Cragg and Epelbaum, 1996; Attanasio, Goldberg, and Pavcnik, 2004; Goldberg and Pavcnik, 2007). However, Latin American countries did not experience a substantial increase in the supply of highly educated workers as the United States and other OECD countries (Attanasio and Szekely, 2000; Sanchez-Paramo and Schady, 2003). In most Latin American countries these changes in labor market conditions took place during periods of trade liberalization. The coincidence in the timing of trade reforms and changes in wage and employment structures is often interpreted as evidence in favor of trade-induced skill-biased technological change (Robbins, 1996; Berman and Machin, 2000; Sanchez-Paramo and Schady, 2003; Goldberg and Pavcnik, 2007). For example, Goldberg and Pavcnik (2007, p. 63) argue that "the most credible explanations for the distributional changes witnessed [in developing countries] in the past few decades would most likely involve interactions of trade openness with skilled-biased technological change."

One of the earliest studies to link trade liberalization and skill-biased technological change is Wood (1994) who argued that increased competition from abroad could lead to defensive skill-biased innovations. This idea is formalized by Thoenig and Verdier (2003) who argue that trade openness increases the threat of imitation of a firm's production process by outside competitors. Imitation, in turn, reduces the competitive advantage of a firm. Since simple unskilled-intensive technologies are easy to imitate, firms undertake *defensive skill-biased innovations* to lessen the increased threat of imitation after trade liberalization. According to this theory, skill-biased technological change should be more pronounced in industries that liberalized more. Acemoglu (2003) provides an alternative link between trade liberalization and skill-biased technological change. Acemoglu argues that the extent

of skill bias in technologies used in less developed countries after trade reforms depends on the relative supply of skilled labor in those countries. Trade opening may reduce the cost of skill-biased technologies in developing countries but will lead to an increase in imports of those technologies only in countries in which the use of skill-biased technologies is profitable. In other words, trade-induced skill-biased technological change is likely to occur in developing countries with a relatively high supply of skilled labor such that the market size for skill-biased technologies is sufficiently large. Although both theories provide persuasive arguments in favor of the hypothesis that skill-biased technological change has been an endogenous outcome of trade liberalization policies in Latin American countries, there is no conclusive empirical evidence to support these theories (see, for example, the discussion in Goldberg and Pavcnik, 2007).

5 Concluding Remarks

This article surveyed the theoretical literature examining the economic mechanisms that shape the intergenerational transmission of inequality. Several conclusions follow from the theories discussed throughout this article.

- First, greater equity in the distribution of wealth is not necessarily at odds with greater efficiency. Carefully designed policies aimed at promoting equality of opportunities can improve both welfare and economic performance by increasing productive investments.

- Second, promoting equality of opportunities may involve redistributions of wealth in some cases while in others it may require improving access to capital markets. More and better access to credit and insurance markets for the majority of the population may encourage productive investments and innovations leading to higher rates of growth and development.

- Third, by enhancing efficiency in the short run, temporary policy interventions may have positive long-term effects. More equality of opportunities in the present may allow a larger fraction of the population to accumulate enough wealth to undertake productive investments in the future. This process can generate a virtuous circle leading to greater equity and growth in the long run.

- Fourth, redistributive policies should target those individuals who are more likely to make productive investments. This implies that the poorest members of society may not necessarily benefit from these policies. Indeed, efficiency-based redistributions may not be the optimal policies to reduce poverty or inequality. Moreover, this also implies that allowing for some initial degree of inequality in low-income countries may be conducive to growth. Nevertheless, in middle-income countries with high levels of inequality, such as Latin American countries, any policy aimed at improving efficiency is likely to lead to a significantly more egalitarian distribution of wealth.

- Fifth, while more equality of opportunities may improve efficiency and equity both in the short and long run, any policy designed to improve equality of opportunities has to take into consideration the possible distortions in the structure of incentives in the economy. Policy interventions that ignore these effects may discourage investments and innovations, leading to a lower rate of growth.

Finally, it is important to highlight that the aim of policy interventions should be to promote equality of opportunities, not equality of outcomes. Differences in outcomes should not be the consequence of unequal opportunities. Nevertheless, equality of opportunities does not imply equality in outcomes. Differences in ability, effort, or simply pure luck can, and often do, generate unequal outcomes even under equal opportunities. Moreover, unequal outcomes usually play a central role in providing incentives to promote investments and innovations. By reducing these incentives, policy interventions aimed at promoting equality of outcomes may lead to lower growth and wealth in the long run.

Appendix

This appendix provides the main assumptions and proofs of some results of the models developed in Section 2. A complete characterization of the transitional dynamics of the model in Section 2.3 requires too many technical details and will be omitted here. The reader interested in the transitional dynamics of the model is referred to Ray (1990, 2006). The only difference between the two models is that Ray uses a dynastic utility function.

A1. The Equalization Model

Assumptions: (*i*) There is a unit mass of individuals who live for one period and have a single offspring, (*ii*) $\beta < 1/r$, (*iii*) $f'(k) > 0$ for all $k \geq 0$, $f''(k) > 0$ for all $k > 0$, $\lim_{k\to 0} f'(k) = \infty$, and $\lim_{k\to\infty} f'(k) = 0$, and (*iv*) idiosyncratic shocks are distributed on the interval [0, 2] independently and identically for all individuals. The distribution has a continuous and strictly positive density function g: $[0, 1] \to \mathbf{R}_+$.

Convergence: In any period t, the distribution of wealth across families may be conceived as a probability measure, say v_t, on $[0, \infty)$. We need to prove that there exists a unique invariant probability measure v^* on $[0, \infty)$ such that for any initial distribution of wealth v_0 the distribution of wealth in the economy eventually converges to v^*. First, note that the optimal allocation of investments for an individual with an inherited wealth equal to w_i is given by: if $w_i \leq k^*$ then invest w_i in the entrepreneurial project and if $w_i > k^*$ then invest k^* in the entrepreneurial project and $w_i - k^*$ in the safe asset. Now define $\overline{w}$ as follows:

$$\overline{w} = \begin{cases} \beta 2 f(\overline{w}) & \text{if } \beta 2 f(k^*) \leq k^* \\ \beta[2f(k^*) + r(\overline{w} - k^*) & \text{if } \beta 2 f(k^*) > k^* \end{cases}$$

Since $\beta < 1/r$ and $f(\cdot)$ is strictly concave, $\overline{w}$ is uniquely determined. Note that no family's wealth could persistently remain above $\overline{w}$. Hence, without loss of generality, we may restrict attention to probability measures v^* on the interval $[0, \overline{w}]$. For future reference, let

B denote the Borel sets of $[0, \bar{w}]$, P denote the set of probability measures on B, and μ denote the Lebesgue measure on B.

Since bequests are a constant fraction of parents' end-of-period wealth, we can treat the evolution of the wealth distribution in this economy as a Markoff process. Thus, the intergenerational evolution of the wealth distribution will be fully characterized once we have found the transition probability of this Markoff process. Let w denote the initial wealth of an individual and let w_0 denote the bequest left to the offspring. Given the optimal allocation of investments described above we have

$$w' = \begin{cases} \beta\alpha f(w) & \text{if } w \leq k^* \\ \beta[\alpha f(k^*) + r(w - k^*) & \text{if } w > k^* \end{cases}$$

Since α is random, w' is random. However, for any given value of α, w' is a continuous and strictly increasing function of w. Now, for any set $W \in B$, define the function h: $B \times \mathbf{R}_+ \rightarrow B$ as follows

$$h(W, w) \equiv \{\alpha \in [0, 2] | w' \in W\}$$

Hence, the probability that an individual with initial wealth w leaves a bequest in the set W is simply the probability that the individual has an ability endowment in the set $h(W, w)$. Then, the transition probability of the evolution of the wealth distribution is

$$Q(w, W) \equiv \int_{h(W,w)} g(\alpha)\mu(d\alpha)$$

Note that Q: $[0, \bar{w}] \times B \rightarrow [0, 1]$ is a well defined transition probability function, i.e., $Q(\cdot,\cdot)$ satisfies (*i*) $\forall w \in [0, \bar{w}]$, $Q(w,\cdot) \in P$ and (*ii*) $\forall W \in B$, $Q(\cdot, W)$ is a B-measurable function on $\mathbf{R}_+$. Moreover, note that for any initial wealth distribution $\nu_0 \in P$ we can generate the sequence of all future wealth distributions $\{\nu_t\}$ recursively as follows

$$v_t(W) = \int Q(w, W) v_{t-1}(dw)$$

The interpretation is that $v_t(W)$ is the fraction of the population with an initial wealth in the set W at time t. A steady-state wealth distribution is a measure $v^* \in P$ such that $\forall W \in B$

$$v^*(W) = \int Q(w, W) v^*(dw)$$

Thus, we need to prove that there exists a unique steady-state wealth distribution such that $\forall v_0 \in P$, $\lim_{t\to\infty} v_t = v^*$. This proof relies heavily on the properties of an operator associated to our transition probability function. In order to define this operator, let $C[0, \bar{w}]$ denote the set of bounded continuous real valued functions on $[0, \bar{w}]$. For any $\phi \in C[0, \bar{w}]$, define the operator $T\phi(\cdot)$ associated with our transition probability function as follows

$$T\phi(w) = \int \phi(w') Q(w, dw')$$

Since the distribution of idiosyncratic shocks has a continuous and strictly positive density function and since, for any α, w' is a continuous and strictly increasing function of w we have the following results: (*i*) $T\phi(\cdot)$ maps the set $C[0, \bar{w}]$ into itself and (*ii*) for every nondecreasing function $\phi \in C[0, \bar{w}]$, $T\phi(\cdot)$ is also nondecreasing. Finally, our assumptions allow us to proof that there is a $\tilde{w} \in [0, \bar{w}]$ such that every neighborhood of $\tilde{w}$ is entered, with positive probability and finite periodicity, infinitely often from any initial wealth level in $[0, \bar{w}]$ (this proof involves too many technical details and will be omitted here). With these results, we can use one of the many theorems in measure theory that establish the existence, uniqueness, and global stability of an invariant distribution under these conditions.

A2. The Neutrality Model

Assumptions: (*i*) There is a unit mass of individuals who live for one period and have a single offspring, (*ii*) $\beta < 1/r$, (*iii*) $f'(k) > 0$ for all $k \geq 0$, $f''(k) > 0$ for all $k > 0$, $\lim_{k\to 0} f'(k)=\infty$, and $\lim_{k\to\infty} f'(k)=0$, and (*iv*) $A > 1$, $\underline{k} > k^*$, and $Af'(\underline{k}) > r$.

Long-Run Properties of the Model: The optimal allocation of investments for an individual with an inherited wealth equal to w_i is given by: if $w_i \leq k^*$ then invest w_i in cottage production, if $w_i \in (k^*, \underline{k})$ then invest k^* in cottage production and $w_i - k^*$ in the safe asset, if $w_i \in [\underline{k}, k^{**}]$ then invest w_i in industrial production, and if $w_i > k^{**}$ then invest k^{**} in industrial production and $w_i - k^{**}$ in the safe asset. Define w^* as follows

$$w^* = \begin{cases} \beta f(w^*) & \text{if } \beta f(k^*) \leq k^* \\ \beta[f(k^*) + r(w^* - k^*) & \text{if } \beta f(k^*) > k^* \end{cases}.$$

Let w denote the initial wealth of an individual and let w' denote the bequest left to the offspring. Since $\beta < 1/r$ and $f(\cdot)$ is strictly concave, the following statement can be easily verified. Suppose that an individual invests in cottage production: (*i*) if $w > w^*$ then $w' < w$, and (*ii*) if $w < w^*$ then $w' > w$. In other words, the wealth of a family investing in cottage production generation after generation monotonically converges to w^* in the long run. Similarly, we can define w^{**} as

$$w^{**} = \begin{cases} \beta A f(w^{**}) & \text{if } \beta A f(k^{**}) \leq k^{**} \\ \beta[Af(k^{**}) + r(w^{**} - k^{**}) & \text{if } \beta A f(k^{**}) > k^{**} \end{cases}.$$

Thus, w^{**} is the long-run wealth of a family investing in industrial production generation after generation. Note that since $A > 1$ we have $w^{**} > w^*$.

The long-run properties of this model depend on the relation between $\underline{k}$ and w^*. If $\underline{k} \leq w^*$ then a family initially investing in cottage production will eventually accumulate enough wealth to invest in industrial production. In this case, the long-run properties of the model are indistinguishable from the long-run properties of the theories in the equalization view.

Namely, in the long run every family invests in industrial production, their wealth converges to w^{**}, and inequality vanishes.

On the other hand, if $\underline{k} > w^{*}$ then families whose initial wealth is below $\underline{k}$ will never accumulate enough wealth to use the industrial technology. In this case, society is divided into two social classes in the long run, an upper class of wealthy industrial entrepreneurs and a lower class of poor cottage factory owners. Moreover, the long-run socioeconomic status of a family is completely determined by its initial wealth. A family will invest in cottage production generation after generation and its wealth will converge to w^{*} if its initial wealth is below $\underline{k}$, otherwise the family will invest in industrial production generation after generation and its wealth will converge to w^{**}. Thus, the long-run levels of income per capita and inequality depend on the initial distribution of wealth, in particular on the fraction of families whose initial wealth is above $\underline{k}$.

To be more precise, let Λ denote the steady state fraction of industrial entrepreneurs in the population. Then there exist $\underline{\Lambda}$ and $\overline{\Lambda}$ such that $\overline{\Lambda} > \underline{\Lambda}$ and any $\Lambda \in [\underline{\Lambda}, \overline{\Lambda}]$ is a steady state. Note that in principle any $\Lambda \in [0, 1]$ can be sustained as a steady state of the model. However, without government interventions, the set of possible steady states is constrained by the initial average wealth in the economy, $\bar{w}$. If the economy is initially poor such that $\bar{w} < \underline{k}$ then $\underline{\Lambda} = 0$ and $\overline{\Lambda} < 1$. On the other hand, if the economy is initially rich such that $\bar{w} \geq \underline{k}$ then $\underline{\Lambda} > 0$ and $\overline{\Lambda} = 1$. Note that the set of steady states always includes a steady state with perfect equality in the distribution of wealth, either $\underline{\Lambda} = 0$ or $\overline{\Lambda} = 1$ is a steady state. The particular steady state the economy converges to depends on the fraction of families whose initial wealth is above $\underline{k}$. If $\tilde{\lambda}$ is the fraction of families whose initial wealth is greater than or equal to $\underline{k}$ then the steady state fraction of industrial entrepreneurs in the population is $\Lambda = \tilde{\lambda}$. Note that, given an initial average wealth in the economy, a more egalitarian distribution of wealth typically implies a higher $\tilde{\lambda}$. Thus, societies with higher levels of inequality in the initial distribution of wealth typically converge to more unequal steady states with lower levels of income per capita. Finally, note that if $\tilde{\lambda} = 0$ then $\Lambda = 0$ and if $\tilde{\lambda} = 1$ then $\Lambda = 1$. That is, if the initial distribution of wealth is sufficiently

egalitarian then the economy converges to a steady state with perfect equality in the distribution of wealth.

A3. The Disequalization Model

Assumptions (*i*) There is a unit mass of individuals who live for one period and have a single offspring, (*ii*) $u'(c) > 0$ for all $c \geq 0$, $u''(c) < 0$ for all $c > 0$, $\lim_{c\to 0} u'(c) = \infty$, and $\lim_{c\to\infty} u'(c) = 0$, (*iii*) if the production function is given by $Y_t = f(\lambda_t, 1 - \lambda_t)$, where λ_t represents the ratio of skilled workers in the population at time t, then $f_1(\lambda, 1 - \lambda) = f_2(\lambda, 1 - \lambda) > 0$ for all $\lambda \in [0, 1]$, $f_{11}(\lambda, 1 - \lambda) = f_{22}(\lambda, 1 - \lambda) < 0$ for all $\lambda \in (0, 1)$, $\lim_{\lambda\to 0} f_1(\lambda, 1-\lambda) = \lim_{\lambda\to 1} f_2(\lambda, 1-\lambda) = \infty$, and $\lim_{\lambda\to 1} f1(\lambda, 1-\lambda) = \lim\lambda\to 0\, f_2(\lambda, 1-\lambda) = 0$.

Properties of the Steady States: The labor market is competitive and the equilibrium wages for skilled and unskilled workers, w_s and w_u, equal their respective marginal products, i.e.,

$$w_s = f_1(\lambda, 1 - \lambda) \quad \text{and} \quad w_u = f_2(\lambda, 1 - \lambda).$$

Note that the wages for skilled and unskilled workers as well as aggregate output are completely determined by the ratio of skilled workers in the economy. Thus, a steady state for this economy is fully characterized by its ratio of skilled workers, say Λ. Next, note that our assumptions about the production function imply that there exists $\tilde{\lambda}$ such that $w_s = f_1(\tilde{\lambda}, 1 - \tilde{\lambda}) = f_2(\tilde{\lambda}, 1 - \tilde{\lambda}) = w_u$. However, as we will see below, in any steady state it must be the case that $\Lambda \in (0, \tilde{\lambda})$.

In Section 2.3 we argued that any steady state in this model must exhibit some degree of earnings inequality, no social mobility, and welfare inequality between skilled and unskilled workers. In what follows, we will proof each of these statements.

Proposition 1 *Every steady state must exhibit some degree of earnings inequality.*

Proof: Let Λ be the steady-state ratio of skilled workers in the economy and let $\overline{w}_s$ and $\overline{w}_u$ be the resulting wages for skilled and unskilled workers, respectively. Perfect equality of earnings arises if and only if: either $\Lambda = 0$, or $\Lambda = 1$, or $\Lambda = \tilde{\lambda}$. Note that neither $\Lambda = 0$ nor $\Lambda = 1$ can be a steady state of this economy. In the former case the wage for skilled workers will be extremely high providing sufficient incentives for at least some families to educate their children while in the latter case the wage for unskilled workers will be higher than the wage for skilled workers and at least some families would be better off leaving their children uneducated. Moreover, since any $\Lambda > \tilde{\lambda}$ implies that the wage for unskilled workers will be higher than the wage for skilled workers it follows that $\Lambda \leq \tilde{\lambda}$. Therefore, we only need to prove that $\Lambda = \tilde{\lambda}$ cannot be sustained as a steady state.

Suppose, contrary to what we are required to prove, that $\Lambda = \tilde{\lambda}$. Note that $\Lambda = \tilde{\lambda}$ implies that in each generation a fraction $\tilde{\lambda}$ of parents invest in their children's education. An individual with labor earnings $\overline{w}_j$, for $j = u, s$, will invest in his/her child's education if and only if

$$u(\overline{w}_j - x) + \overline{w}_s \geq u(\overline{w}_j) + \overline{w}_u$$

Or equivalent

$$\overline{w}_s - \overline{w}_u \geq u(\overline{w}_j) - u(\overline{w}_j - x)$$

Since $u(\cdot)$ is strictly increasing we have $u(\overline{w}_j) - u(\overline{w}_j - x) > 0$. Therefore, a parent will invest in his/her child's education if and only if $\overline{w}_s - \overline{w}_u > 0$. However, $\Lambda = \tilde{\lambda}$ implies that $\overline{w}_s = \overline{w}_u$. A contradiction. Therefore, it must be the case that $\Lambda \in (0\,\tilde{\lambda})$ and $\overline{w}_s > \overline{w}_u$, i.e., every steady state must exhibit some degree of earnings inequality.

Proposition 2 *There is no social mobility in steady state.*

Proof: Unskilled workers will invest in education if and only if

$$\overline{w}_s - \overline{w}_u \geq u(\overline{w}_u) - u(\overline{w}_u - x)$$

and skilled workers will invest if and only if

$$\overline{w}_s - \overline{w}_u \geq u(\overline{w}_s) - u(\overline{w}_s - x)$$

Since $u(\cdot)$ is strictly concave and $\overline{w}_s > \overline{w}_u$,

$$u(\overline{w}_u) - u(\overline{w}_u - x) > u(\overline{w}_s) - u(\overline{w}_s - x)$$

Therefore, if it is optimal for an unskilled worker to invest in education then it is also optimal for a skilled worker to do so. Since the ratio of skilled workers is constant in any steady state, it must be the case that

$$u(w_u) - u(w_u - x) > w_s - w_u \geq u(w_s) - u(w_s - x),$$

i.e., only skilled workers invest in education.

Proposition 3 *In any steady state the utility of a skilled worker is higher than the utility of an unskilled worker.*

Proof: Since skilled workers invest in education it must be true that

$$u(\overline{w}_s - x) + \overline{w}_s \geq u(\overline{w}_s) + \overline{w}_u$$

On the other hand, since skilled workers earn higher wages than unskilled workers, ws > wu, it must be true that

$$u(\overline{w}_s) + \overline{w}_u > u(\overline{w}_u) + \overline{w}_u$$

These two inequalities imply

$$u(\overline{w}_s - x) + \overline{w}_s > u(\overline{w}_u) + \overline{w}_u,$$

i.e., the utility of a skilled worker is strictly higher than the utility of an unskilled worker.

References

[1] Acemoglu, Daron (1998) "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." *Quarterly Journal of Economics,* 113(4), pp. 1055-89.

[2] Acemoglu, Daron (2002) "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature*, 40(1), pp. 7-72.

[3] Acemoglu, Daron (2003) "Patterns of Skill Premia." *Review of Economic Studies*, 70(2), pp. 199-230.

[4] Aghion, Philippe (2002) "Schumpeterian Growth Theory and the Dynamics of Income Inequality*." Econometrica*, 70(3), pp. 855-82.

[5] Aghion, Philippe and Bolton, Patrick (1992) "Distribution and Growth in Models of Imperfect Capital Markets." *European Economic Review,* 36(2-3), pp. 603-11.

[6] Aghion, Philippe and Bolton, Patrick (1997) "A Trickle-Down Theory of Growth and Development with Debt Overhang." *Review of Economic Studies*, 64(2), pp. 151-72.

[7] Ahluwalia, Montek S. (1976a) "Income Distribution and Development." *American Economic Review*, 66(2), pp. 128-35.

[8] Ahluwalia, Montek S. (1976b) "Inequality, Poverty and Development." *Journal of Development Economics*, 3(4), pp. 307-42.

[9] Alesina, Alberto and Perotti, Roberto (1996) "Income Distribution, Political Instability and Investment." *European Economic Review*, 40(6), pp. 1203-28.

[10] Alesina, Alberto and Rodrik, Dani (1994) "Distributive Politics and Economic Growth." *Quarterly Journal of Economics*, 109(2), pp. 465-90.

[11] Anand, Sudhir and Kanbur, S.M.R. (1993) "The Kuznets Process and the Inequality-Development Relationship." *Journal of Development Economics*, 40(1), pp. 25-72.

[12] Attanasio, Orazio and Székely, Miguel (2000) "Household Saving in East Asia and Latin America: Inequality Demographics and All That." In B. Pleskovic and N. Stern (eds.), *Annual World Bank Conference on Development Economics*. Washington, D.C.: World Bank.

[13] Autor, David H.; Katz, Lawrence F. and Krueger, Alan B. (1998) "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics*, 113(4), pp. 1169-213.

[14] Banerjee, Abhijit V. and Duflo, Esther (2003) "Inequality and Growth: What Can the Data Say?" *Journal of Economic Growth*, 8(3), pp. 267-99.

[15] Banerjee, Abhijit V. and Newman, Andrew F. (1991) "Risk-Bearing and the Theory of Income Distribution." *Review of Economic Studies*, 58(2), pp. 211-35.

[16] Banerjee, Abhijit V. and Newman, Andrew F. (1993) "Occupational Choice and the Process of Development." *Journal of Political Economy*, 101(2), pp. 274-98.

[17] Barro, Robert (2000) "Inequality and Growth in a Panel of Countries." *Journal of Economic Growth*, 5(1), pp. 5-32.

[18] Becker, Gary and Tomes, Nigel (1979) "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility" *Journal of Political Economy*, 87(6), pp. 1153- 89.

[19] Becker, Gary and Tomes, Nigel (1986) "Human Capital and the Rise and Fall of Families*" Journal of Labor Economics*, 4(3), pp. S1-S39.

[20] Benabou, Roland (1996) "Inequality and Growth." In Ben Bernanke and Julio J. Rotemberg (eds.), *NBER Macroeconomics Annual 1996*. Cambridge, MA: MIT Press.

[21] Berman, Ely; Bound, John and Griliches, Zvi (1994) "Changes in the Demand for Skilled Labor within U.S. Manufacturing Industries: Evidence from the Annual Survey of Manufacturers." *Quarterly Journal of Economics*, 109(2), pp. 365-7.

[22] Berman, Ely; Bound, John and Machin, Stephen (1998) "Implications of Skill-Biased Technological Change: International Evidence." *Quarterly Journal of Economics*, 113(4), pp. 1245-80.

[23] Berman, Ely and Machin, Stephen (2000) "Skill-Biased Technology Transfer: Evidence of Factor Biased Technological Change in Developing Countries" Mimeo, Department of Economics, Boston University.

[24] Bound, John and Johnson, George (1992) "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations." *American Economic Review*, 83(3), pp. 371-92.

[25] Bourguignon, Francois and Morrison, Christian (1990) "Income Distribution, Development and Foreign Trade: A Cross-Sectional Analysis." *European Economic Review*, 34(6), pp. 1113-32.

[26] Bowles, Samuel and Gintis, Herbert (2002) "The Inheritance of Inequality." *Journal of Economic Perspective*s, 16(3), pp. 3-30.

[27] Card, David and Lemieux, Thomas (2001) "Can falling Supply Explain the Rising Return to College for Younger Men." *Quarterly Journal of Economics*, 116(2), pp. 705-46.

[28] Caselli, Francesco (1999) "Technological Revolutions." *American Economic Review*, 89(1), pp. 78-102.

[29] Champernowne, David G. (1953) "A Model of Income Distribution." *Economic Journal*, 63(250), pp. 318-51.

[30] Cragg, Michael Ian and Epelbaum, Mario (1996) "Why Has Wage Dispersion Grown in Mexico? Is It the Incidence of Reforms or the Growing Demand for Skills?" *Journal of Development Economics*, 51(1), pp. 96-116.

[31] Dearden, Lorraine; Machin, Stephen and Reed, Howard (1997) "Intergenerational Mobility in Britain." *Economic Journal*, 107(440), pp. 47-66.

[32] Deininger, Klaus and Squire, Lyn (1996) "A New Data Set Measuring Income Inequality." *World Bank Economic Review*, 10(3), pp. 565-91.

[33] Deininger, Klaus and Squire, Lyn (1998) "New Ways of Looking at Old Issues: Inequality and Growth." *Journal of Development Economics*, 57(2), pp. 259-87.

[34] Durlauf, Steven N. (1996) "A Theory of Persistent Income Inequality." *Journal of Economic Growth*, 1(1), pp. 75-93.

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

[36] Freeman, Scott (1996) "Equilibrium Income Inequality among Identical Agents." *Journal of Political Economy*, 104(5), pp. 1047-64.

[37] Galor, Oded and Moav, Omer (2000) "Ability Biased Technological Transition, Wage Inequality and Economic Growth." *Quarterly Journal of Economics*, 115(2), pp. 469-98.

[38] Galor, Oded and Moav, Omer (2006) "Das Human Capital: A Theory of the Demise of the Class Structure." *Review of Economic Studies*, 73(1), pp. 85-117.

[39] Galor, Oded and Tsiddon, Daniel (1997) "Technological Progress, Mobility, and Economic Growth." *American Economic Review*, 87(3), pp. 363-82.

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

[41] Ghatak, Maitreesh and Jiang, Neville N. (2002) "A Simple Model of Inequality, Occupational Choice and Development." *Journal of Development Economics*, 69(1), 205-226.

[42] Goldberg, Pinelopi K and Pavcnik, Nina (2007) "Distributional Effects of Globalization in Developing Countries." *Journal of Economic Literature*, 45(1), pp. 39-82.

[43] Goldin, Claudia and Katz, Lawrence F. (1998) "The Origins of Technology-Skill Complementarity." *Quarterly Journal of Economics*, 113(3), pp. 693-732.

[44] Greenwood, Jeremy and Jovanovic, Boyan (1990) "Financial Development, Growth and the Distribution of Income." *Journal of Political Economy*, 98(5), pp. 1076-107.

[45] Greenwood, Jeremy and Yorukoglu, Mehmet (1997) "1974." *Carnegie-Rochester Conference Series on Public Policy*, 46, pp. 49-95.

[46] Hertz, Thomas (2005) "Rags, Riches and Race: The Intergenerational Economic Mobility of Black and White Families in the United States." In Samuel Bowles, Herbert Gintis and Melissa Osborne (eds.), *Unequal Chances: Family Background and Economic Success*. Princeton, NJ: Princeton University Press.

[47] Juhn, Chinui; Murphy, Kevin M. and Pierce, Brooks (1993) "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy*, 101(3), pp. 410-42.

[48] Kanbur, Rabi (2000) "Income Distribution and Development." In Anthony B. Atkinson and Francois Bourguignon (eds.), *Handbook of Income Distribution. Amsterdam: North Holland.*

[49] Katz, Lawrence F. and Murphy, Kevin M. (1992) "Changes in RelativeWages, 1963-1987: Supply and Demand Factors*." Quarterly Journal of Economics*, 107(1), pp. 35-78.

[50] Krueger, Alan B. (1993) "How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989." *Quarterly Journal of Economics*, 108(1), pp. 33-60.

[51] Krusell, Per; Ohanian, Lee E.; Rios-Rull, Jose-Victor and Violante, Giovanni L. (2000) "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis." *Econometrica*, 68(5), pp. 1029-53.

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

[53] Lemieux, Thomas (2006) "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *American Economic Review*, 96(3), pp. 461-98.

[54] Lewis, Arthur (1954) "Economic Development with Unlimited Supplies of Labor." Manchester School of Economic and Social Studies, 22(2), pp. 139-91.

[55] Li, Hongyi and Zou, Heng-fu (1998) "Income Inequality Is Not Harmful for Growth: Theory and Evidence." *Review of Development Economics*, 2(3), pp. 318-34.

[56] Li, Hongyi; Squire, Lyn and Zou, Heng-fu (1998) "Explaining International and Intertemporal Variations in Income Inequality." *Economic Journal*, 108(446), pp. 26-43.

[57] Lindert, Peter H. (1986) "Unequal English Wealth since 1670." *Journal of Political Economy*, 94(6), pp. 1127-62.

[58] Lindert, Peter H. and Williamson, Jeffrey G. (1985) "Growth, Equality and History." *Explorations in Economic History*, 22(4), pp. 341-77.

[59] Ljungqvist, Lars (1993) "Economic Underdevelopment: The Case of a Missing Market for Human Capital." *Journal of Development Economics*, 40(2), pp. 219-39.

[60] Loury, Glenn (1981) "Intergenerational Transfers and the Distribution of Earnings." *Econometrica*, 49(4), pp. 843-67.

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

[62] Machin, Stephen and Van Reenen, John (1998) "Technology and Changes in Skill Structure: Evidence from Seven OECD Countries." *Quarterly Journal of Economics*, 113(4), pp. 1215-44.

[63] Mani, Anandi (2001) "Income Distribution and the Demand Constraint." *Journal of Economic Growth*, 6(2), pp. 107-33.

[64] Maoz, Yishay D. and Moav, Omer (1999) "Intergenerational Mobility and the Process of Development." *The Economic Journal*, 109(458), pp. 677-97.

[65] Matsuyama, Kiminori (2000) "Endogenous Inequality." *Review of Economic Studies*, 67(4), pp. 743-59.

[66] Matsuyama, Kiminori (2006) "Emergent Class Structure." *International Economic Review*, 47(2), pp. 327-60.

[67] Mazumder, Bhashkar (2005) "The Apple Falls Even Closer to the Tree thanWe Thought: New and Revised Estimates of the Intergenerational Inheritance of Earnings." In Samuel Bowles, Herbert Gintis and Melissa Osborne (eds.), *Unequal Chances: Family Background and Economic Success.* Princeton, NJ: Princeton University Press.

[68] Mookherjee, Dilip and Napel, Stefan (2007) "Intergenerational Mobility and Macroeconomic History Dependence." *Journal of Economic Theory*, 137(1), pp. 49-78.

[69] Mookherjee, Dilip and Ray, Debraj (2003) "Persistent Inequality." *Review of Economic Studies*, 70(2), pp. 369-93.

[70] Mookherjee, Dilip and Ray, Debraj (2006) "Occupational Diversity and Endogenous

Inequality", Mimeo, Department of Economics, Boston University.

[71] Owen, Ann L. andWeil, David N. (1998) "Intergenerational EarningsMobility, Inequality, and Growth." *Journal of Monetary Economics*, 4(1), pp. 71-104.

[72] Papanek, Gustav F. and Kyn, Oldrich (1986) "The Effect on Income Distribution of Development, the Growth Rate and Economic Strategy." *Journal of Development Economics*, 23(1), pp. 55-65.

[73] Paukert, Felix (1973) "Income Distribution at Different Levels of Development: A Survey of the Evidence." *International Labour Review*, 108(97), pp. 97-125.

[74] Perotti, Roberto (1996) "Growth, Income Distribution, and Democracy: What the Data Say." *Journal of Economic Growth*, 1(2), pp. 149-87.

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

[76] Piketty, Thomas (1997) "The Dynamics of the Wealth Distribution and the Interest Rate with Credit Rationing." *Review of Economic Studies*, 64(2), pp. 173-89.

[77] Ray, Debraj (1990) "Income Distribution and Macroeconomic Behavior." Mimeo, Department of Economics, New York University

[78] Ray, Debraj (2006) "On the Dynamics of Inequality." *Economic Theory*, 29(2), pp. 291-306.

[79] Robbins, Donald J. (1996) "Evidence on Trade and Wages in the Developing World." OECD Technical Paper No. 119.

[80] Sanchez-Paramo, Carolina and Schady, Norbert (2003) "Off and Running? Technology, Trade, and the Rising Demand for Skilled Workers in Latin America" World Bank Policy Research Working Paper Series No. 3015.

[81] Solon, Gary R. (1992) "Intergenerational Income Mobility in the United States." *American Economic Review*, 82(3), pp. 393-408.

[82] Solon, Gary R. (1999) "Intergenerational Mobility in the Labor Market." In Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics*, Volume 3A. Amsterdam: North Holland.

[83] Thoenig, Mathias and Verdier, Thierry (2003) "A Theory of Defensive Skill-Biased Innovation and Globalization." *American Economic Review*, 93(3), pp. 709-28.

[84] Venieris, Yiannis P. and Gupta, Dipak K. (1986) "Income Distribution and Sociopolitical Instability as Determinants of Savings: A Cross-Sectional Model." *Journal of Political Economy*, 94(4), pp. 873-83.

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

[86] World Bank (2006) World Development Report 2006: Equity and Development. New York: Oxford University Press.

[87] Zimmerman, David J. (1992) "Regression toward Mediocrity in Economic Stature." *American Economic Review*, 82(3), pp. 409-29.