# 11

# Inequality and Development

> "Them that's got, shall get
> Them that's not, shall lose."
>
> — Billie Holiday, 1939.

## 11.1. Introduction

In the introduction to Chapter 10, we argued that an interest in inequality can have intrinsic or instrumental foundations. To many, economic equality is a worthy objective in itself. But even if you aren't one of those individuals, inequality may still be of interest to you, for *instrumental* reasons.

Economic growth affects income distribution, but income distribution also affects economic outcomes. Individuals hold endowments in different forms: raw labor, savings in the bank, skills via education, capital by way of a business, land, and so on. These endowments interact in the marketplace. They determine the supply and demand of different goods and services, and the prices at which these are bought and sold. Those prices will determine how much people earn, and the distribution of those earnings. They also determine how people consume and save. At the end of the day, endowments are updated — either accumulated or depleted. Then tomorrow comes around, and the whole thing starts up again. In this way, economic growth and economic inequality intertwine and evolve together, and this ever-recurrent process determines a host of economic, social and even psychological outcomes such as income, employment, poverty, social conflict, life-satisfaction, and so on. A useful schematic diagram for this is:

$$
\begin{tikzcd}
 & & \text{earnings} \arrow[dr] & \\
\text{endowments} \arrow[r] & \text{market prices} \arrow[ur] \arrow[dr] & & \text{endowments} \\
 & & \begin{array}{c}\text{consumption}\\ \text{and savings}\end{array} \arrow[ur] &
\end{tikzcd}
$$

To some degree, there is a parallel between theories of economic growth, studied earlier, and studies of economic inequality. This whole process of endowment accumulation or depletion looks like a mini-growth model for each household (or an individual and her descendants). Convergence and divergence, notions so often studied in growth models, address the distribution of income or wealth across *countries*, while economic inequality studies income or wealth distributions across *persons*. So it will not be surprising to find that old questions of convergence or divergence make a natural reappearance, cloaked in the language of inequality: do inequalities worsen over time or narrow? Do markets exacerbate past inequalities, or eradicate them, or so they simply replicate them in some neutral fashion? The evolution of economic inequality is a question of central importance, whether you are a philosopher or a policy-maker, or just a plain old economist.

But let's not take the parallel too far, at least with respect to traditional growth models. Interaction *across* countries is not central in those models. The interaction across individuals via the market place is a core feature of the inequality story.

## 11.2. The Evolution of Inequality: Macroeconomic Context

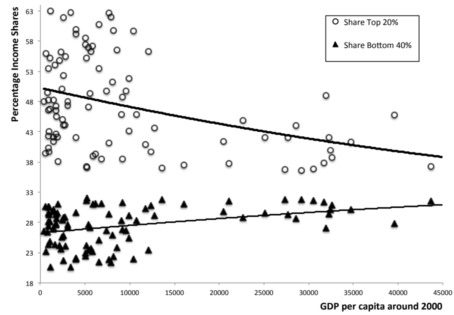
The great scholars of development economics — Paul Rosenstein-Rodan, Gunnar Myrdal, Albert Hirschmann, Arthur Lewis, and others — did their work in the mid-20th century. From the perspective of economic equality, these were generally hopeful times. The enormous inequalities of the first decades of the twentieth century were dying away in the United States and in Europe, as the postwar period brought a sustained, prosperous middle class into being. Overall, and especially after the Great Depression, inequality displayed a general downward trend into the 1970s. It is no surprise that economists writing during these times believed that sustained economic growth would automatically and spontaneously take care of the problem of distribution.

**11.2.1. Kuznets and the Inverted-U Hypothesis.** A leading economist in this tradition was Simon Kuznets. He is perhaps best known today for the famous "Inverted-U Hypothesis," presented in his 1954 Presidential Address to the American Economic Association. Kuznets (1955, 1963), as also Oshima (1962), went on to argue that economic progress, measured by per capita income, is initially accompanied by rising inequality, but that these disparities ultimately go away as the benefits of development permeate more widely. The resulting pattern has an inverted-U shape.[114] You can still see vestiges of the pattern in a cross-section over countries. Figure 11.1 is reproduced from Chapter 2. Both panels in that Figure plot the income shares of poorest 40% of the population (triangles) and the richest 20% of the population as we move over

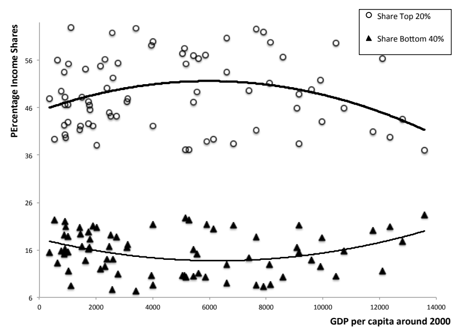
[114] The Kuznets hypothesis spawned a literature, initially based on attempts to establish it as a universal law or at least as a strong statistical regularity. Early examples of cross-section studies include Kravis (1960), Paukert (1973), Adelman and Morris (1973), Chenery and Syrquin (1975), Ahluwalia (1976), Ahluwalia, Carter, and Chenery (1979), Bacha (1979), Papanek and Kyn (1986), Bourguignon and Morrisson (1989, 1990), and Anand and Kanbur (1993a, b). But even during this early wave of research, there were doubts. The hypothesis is about the evolution of inequality in any *single* country over *time*, while what we see in Figure 11.1 — and in the data from which Kuznets drew his conclusions, as well as the cross-section studies — are different countries at similar moments in time. It requires a leap of faith to convert the latter into the former, the leap being that all countries not only have a Kuznets curve each, they have the *same* curve. Fields and Jakubson (1994) and Deininger and Squire (1996) were among the first authors in sounding these warnings. The coefficients on per-capita income (and its square) are either insignificant, or they have the wrong signs for an inverted-U, a not-surprising precursor of the great U-turn that we now see so clearly. As the 20th century came to a close and the new millenium began to unfold, it became clear that the inverted-U hypothesis was far from being the dominant law of inequality it was once held to be (Gallup 2012).

countries of varying per-capita income. Panel A suggests that on average, inequality appears to fall as we move from low-income countries to high-income countries. Panel B tells a somewhat different story by "stretching out" the horizontal axis of per-capita incomes to consider only those countries with under \$15,000 of income per-capita. Now the inverted-U can be seen among the poorest countries: inequality appears to *rise*, then fall over the cross-section.

Perhaps the clearest story that fits this hypothesis — one that Kuznets himself favored — is the agriculture-industry transition, where the higher-wage industrial sector initially generates inequality, while the later arrivals to that sector play a game of catch-up. Thus (so goes the argument) inequality is first low when everyone is in agriculture, and low again when the transition to industry has been completed, but is highest "in the middle," when some have moved to industry while the rest are still in agriculture. That transformation still casts some light on cross-country patterns. But agriculture and industry are no longer the only games in town.




**(a)** The full range of per-capita GDP




**(b)** Per-capita GDP below \$15,000

**Figure 11.1.** Income shares of poorest 40% and richest 20% in various countries. Source: The World Bank; `http://databank.worldbank.org`.

**11.2.2. The Great U-Turn.** Indeed, the individual experiences of countries over recent decades run distinctly counter to Figure 11.1. Figure 11.2, from the *World Inequality Database*, plots the income shares of the top 10% of income earners for selected countries over 1910–2010. The pattern it displays is robust, no matter whether you look at the top 1%, 5% or 10%. After the 1970s there has been a veritable upsurge in economic inequality. The trend is particularly visible in the United States and in many European countries, and the increase continues unabated. In the United States, top income shares are now as sizable as they were in the mayhem of the "Roaring Twenties," just before things fell apart in the Great Depression. But a similar story is true for the United Kingdom and for other European

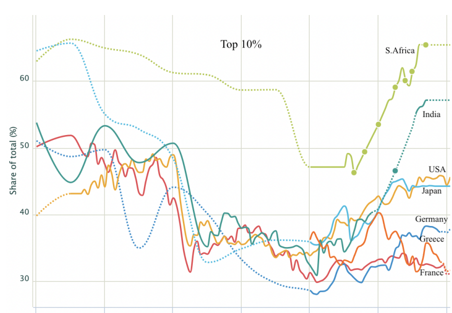



**Figure 11.2.** Income shares of top 10% in selected countries. Source: World Inequality Database.

countries, and developing countries are no exception either. Figures 11.1 and 11.2 appear to stand in stark contradiction to each other, but they are not. Patterns that appear over the cross-section of countries cannot be extrapolated to corresponding patterns for each country *over time*.

By way of sly reference to Kuznets, one could call this recent rise in inequality the great U-*turn*.[115] Clearly, we no longer live (if we ever did) in an optimistic universe in which income and wealth concentrations are happily eroded by aggregate economic progress. If anything, it is the other way around. The view that "growth will take care of inequality" has largely evaporated.

**11.2.3. Inverted-U's and U-Turns.** It's worth distinguishing between three different kinds of income growth. The first — and the most placid — consists of those changes that occur on an everyday basis: the steady accumulation of wealth, the acquisition of skills, ongoing gains in work productivity, and so on. Think of this as some steady sequence of 2 or 3% annual raises that you might receive at work, as well as gradual increases in your capital income stemming from the accumulation of wealth.

The second source of change is inherently disequalizing: some sector (such as engineering, software, or biotech) takes off, and there is a frenetic increase in demand for individuals with these skills. The economy as a whole registers growth, of course, but this growth is highly concentrated in a relatively small number of sectors. These growth spurts are intrinsically inequality-*creating*.

Finally, there are those changes that are "compensatory" to the second: as the growth spurt manifests itself in high incomes in some sectors, the incomes spread through the economy as demands for all sorts of other goods and services rise. Engineers buy houses, doctors buy cars, and even geeks go on vacations. In addition, more people acquire the skills that are currently in demand, tempering the rates of return to such skills and spreading the income gains more evenly through society.

At any point in time, it is likely that some combination of all three phenomena is at work. An inverted-U would be the outcome if it is more likely that disequalizing changes occur at low levels of income, whereas compensatory changes occur at higher levels of income. And indeed, as already discussed, there was historical reason to believe that this was the implacable order of things, because of the agriculture-industry structural transformation. It inspired Kuznets to formulate his hypothesis.

But Kuznets was writing in the mid-twentieth century, when all within his experiential ken was agriculture and industry and not much else. But there's more to life now than agriculture and industry. Yes, that was one of the greatest structural transformations. But we are now in the throes of other great transformations. The IT revolution brought about another seismic shift, and a displacement of unskilled labor that is still not over. When the dust settles, IT too will have created a rise in inequality, followed by a Kuznets-like adjustment as job-seekers across generations struggle to deal with the creation of new occupational niches, and the disappearance of others.

In turn, IT has made other transformations possible, such as decentralized service industries (for transport or housing), a revolution in finance, highly data-intensive

[115] The nomenclature isn't mine; it has been used — primarily in the United States by several commentators and authors — to describe rising inequality.

biotechnology and health sectors, and massive online platforms for social interaction. Each such transformation can be matched to a sequence of disequalizing and compensatory changes. Each creates its own inequalities, as the lucky or farsighted individuals already in the beneficiary sector experience an upsurge in their incomes. That inequality then serves as an impetus to reallocation, as individuals (or their progeny) in the “lagging" sectors attempt to relocate to the growing sector. Whether or not that relocation can occur will depend on how quickly the new generation can adjust. Whether or not that relocation is actually *successful* will depend on the next tsunami of unevenness and where it hits, and so it goes. The complexity of, and variation in these paths can leave simplistic theories such as the inverted-U hypothesis without much explanatory power at all. At the same time, the idea of uneven and compensating variations is extremely important if we are to keep track of background macroeconomic trends as we proceed to more microeconomic considerations, which we shall now do.

## 11.3. The Evolution of Inequality: Microeconomic Concepts

What aspects of individual behavior and market structure are relevant for a detailed study of inequality and its evolution? A little sensible accounting will take us a long way. Some of that accounting can be done by exploiting the equations that resemble the growth model of Chapter 3, except that now we will use them for individuals or households rather than the entire economy. (These parallels will take us part of the way, but occasionally we will need to abandon them; I will tell you when that happens.)

**11.3.1. A Taxonomy For Sources of Inequality.** Here is some simple taxonomy to organize our thoughts:

1. *Savings.* All other things equal, differences in savings rates across individuals will cause their relative incomes to change over time. If savings rates rise with incomes, there will be continued divergence. See equation (11.1) below.

2. *Rates of Return.* Variations in the rate of return to capital across people will affect economic disparities. For instance, if richer individuals earn higher rates of return on capital, inequality will increase over time. See equation (11.3) below.

3. *Occupational Choice and Entrepreneurship.* Income inequality will affect access to capital markets, thereby influencing the connection between individual wealth and human capital, including skilled occupational choice or entrepreneurship; all these can be captured as special cases by the production function $f$, as in equation (11.4) below.

4. *Demand.* Income distributions will affect the pattern and composition of demand for products, and therefore individual incomes; these will be captured by the derived returns to various kinds of labor, as in equation (11.4) below.

5. *Politics and Policies.* Income distributions will affect the choice of taxes on labor and capital income, thereby modifying the links between wealth and economic outcomes in equation (11.2) below.

In what follows we shall study these issues, but doing them all together will drive you (and me) crazy. As always, we will find it convenient to shut down additional concerns while focusing on one or two issues at a time. Items (1)–(3) will be the subject of this chapter, while we leave (4) and (5) for a later chapter.

**11.3.2. Individual Accumulation.** We pull some of the ideas above into a simple enveloping framework. Denote the wealth or capital of a person — call her Asha — by $k$, her income by $y$, and her consumption at any date by $c$. Then Asha's wealth at date $t$ must evolve according to the equation

$$k(t+1) = y(t) - c(t) + k(t), \tag{11.1}$$

because $y(t) - c(t)$ is savings, and added to existing wealth $k(t)$ it gives us new wealth $k(t+1)$. The similarity to the growth equation (3.5) in Chapter 3 should be obvious, though we do not regard here the savings rate as a parameter but are interested in how it might be *chosen*.[116] That interest summarizes item 1 above.

Now, Asha's income $y$ evolves as well, and in general that evolution could depend on her current wealth, or some overall macro-state of the economy, call it $\theta(t)$, over which Asha has no control. We write this as:

$$y(t) = f(k(t), \theta(t)), \tag{11.2}$$

where $f$ is catch-all notation for the individual's income-producing opportunities at date $t$. A macro-state is anything that influences Asha's ability to convert her capital (physical or human or both) into income, including the forces of supply and demand economy-wide that determine wage rates to different skills or the rate of return to capital. Suppressing the macro-state momentarily, $f$ could be a standard production function of the sort studied in the growth model of Chapter 3; e.g., $y(t) = Ak(t)^{\alpha}$. But it could stand for other things as well. For instance, if Asha *only* earns income on her capital at rate $r$ on her wealth $k(t)$ (a true capitalist, our Asha!), we could write

$$f(k(t)) = rk(t). \tag{11.3}$$

This equation adds on item 2 to item 1 in the taxonomy of the previous section. As students of inequality, we would be interested the determinants of $r$, and indeed in how $r$ might vary over different individuals, If Asha and another individual, say Madhavi, have access to same equations (11.1)–(11.3) but they have varying savings rates or rates of return to their capital investments, then Asha and Madhavi's wealths could well separate over time or persist in their differences, generating or maintaining economic inequality between them.

A third interpretation of $f$ is very different. Think of each date $t$ as a lifetime, so that each individual is now a sequence of individuals (a parent-child chain), or a *dynasty*, and $k(t)$ could be the dynasty's *human capital* in each generation. Under that interpretation, $f$ could represent the labor-market reward for skills; for instance:

$$\begin{aligned} f(k(t)) &= \underline{w} \text{ for } k(t) < \bar{x} \\ &= \bar{w} \text{ for } k(t) > \bar{x}; \end{aligned} \tag{11.4}$$

or in words, each descendant in the Asha-dynasty earns an "unskilled wage" if she has human capital below some "entry threshold" $\bar{x}$, and a "skilled wage" if her human capital crosses that threshold. The interpretation is that each parent of generation $t$ in the Asha-dynasty would choose to skill her child by incurring the cost $\bar{x}$, and depending on her choices, the child would earn either the low- or the high-skilled wage during her lifetime. But it isn't just human capital. A similar specification holds for entrepreneurship, or the setting up of a private business. In this interpretation, $\bar{x}$

[116] Another obvious difference: we are applying that equation to an individual or a household, not to an economy. In particular, there is no corresponding notion of wealth depreciation ($\delta = 0$).

would stand for the setup cost of a firm, and the difference between $\bar{w}$ and $\underline{w}$ would now be viewed as the additional profits accruing to the business if entrepreneurs such as Asha and Madhavi were to incur those setup costs. Whatever the specific interpretation might be, the Asha and Madhavi dynasties could face the very same *equations* but end up making different *choices*, perhaps because of some pre-existing wealth inequality that they had to begin with, so that they have different access to these different occupational niches.

In these settings, which we might think of broadly as *occupational choice* (see Item 3 in the taxonomy), wages or rates of return or educational thresholds could additionally evolve or change with macroeconomic conditions, such as those we've described in Section 11.2. Or they could change with the choices of other dynasties living in the same economy as Asha and Madhavi, and facing similar equations of occupational choice and wealth evolution. Asha's or Madhavi's income-generating opportunities might look exogenous to *them*, but they are truly endogenous outcomes in an economy with many Asha-Madhavi-like actors. For instance, extend (11.4) to read:

$$\begin{aligned} f(k(t), \theta(t)) &= \underline{w}(\theta(t)) \text{ for } k(t) < \bar{x} \\ &= \bar{w}(\theta(t)) \text{ for } k(t) > \bar{x}; \end{aligned}$$

The macro parameter $\theta(t)$ might incorporate the supply of labor by other individuals, which will affect the skilled and unskilled wages. Or — if we did the same with equation (11.3) — it could incorporate the overall demand for credit, which will affect the rate of return on Asha's savings. This is quite different from the standard growth model of Chapter 3, in which "Asha" or "Madhavi" are countries developing over time, each with its own exogenous technology.

There is no particular reason to have just two thresholds and skills, as depicted in equation (11.4). There could be a plethora of occupational and entrepreneurial choices, each with its entry threshold and with its wages and returns. These returns will depend in turn on the composition of demand for goods and services in the economy, via their derived demand for factors of production. That ties into item 4 of our taxonomy. And last but not least — *all* these returns are subject to political influences and policies. Think of returns to capital, the panoply of skilled and unskilled wages, the revenues from running a business — all of it is as influenced by policy as it is by markets. That is where item 5 in our taxonomy appears. Now we see why our little equation system in this Section can be so useful in organizing our thoughts.

## 11.4. Savings Rates and Rates of Return to Capital

In early 2016, a story went viral: apparently, just 62 of the richest persons in the world have more wealth than *half the entire population* (around 3.6 billion).[117] At one level, this isn't surprising: if you have any money at all in your bank account, and no debt to pay off, you will have more wealth than the combined wealth of about half of all people living in one of the richest countries in the world, the United States. The reason is simple: the bottom half of the United States — defined by wealth — has about zero net wealth.[118] From this perspective, our viral story is interesting but not terribly

[117] "An Economy for the 1%," Oxfam Briefing Paper, 2016.

[118] That probably includes a lot of rich people who are debt-leveraged. If one defines the bottom half by income — and looks at the wealth of the bottom half of income earners — the story has a bit more content.

informative. And yet, it does alert us to something: that *differential* savings across different economic classes could account for a significant part of the rising inequality that we see.

**11.4.1. Savings Rates and Income.** Think of how savings rates might vary with income. At very low levels, considerations of *minimum subsistence* are paramount, and the minimum needs of food, clothing and shelter overwhelmingly dictate current expenditure. Savings could be zero or even negative when debt is taken into account.

As income increases, a progressively greater fraction of additional income can be put away for the future, so that the savings rate begins to rise. For no group of people is this truer than those out of poverty, yet some distance away from the pleasures enjoyed by the rich. This group includes individuals and families who both aspire to a better economic life, and can also act on those aspirations. Such individuals typically save large fractions of their income.

What happens at still higher levels of income is not entirely clear. We now enter the realm of the rich and ultra-rich in developing countries. Eager to attain the desirable consumption seen worldwide, their own consumptions could be pushed up. While they still save more in absolute terms, their *rate* of savings may well be lower. Against this argument is the observation that very high levels of income lead to extremely high rates of saving, as *the very act of accumulation* may become an end in itself.

All in all, we would expect savings rates to climb through much of the income distribution, perhaps moving around as we pass through the higher income categories. Exact estimates are notoriously hard to come by. One important reason is that we (the statisticians or researchers) find it hard to measure "permanent income." Someone may have just received a windfall gain and might be saving a large fraction of that windfall (or blowing it) because it is a transitory amount. Or someone with a negative shock to income may be maintaining high consumption in the anticipation that this will all even out with time.[119] What we're after in contrast, is the proclivity to save out of anticipated, permanent income. Table 11.1 shows us some estimates from the United States for different income quintiles in the population, as well as for the top 5% and top 1%. The issue of transitory income is avoided here by looking at multi-year income averages, or by instrumenting for income using durable-goods (vehicle) consumption. Rates of savings out of permanent income appear to systematically climb with permanent income, with the richest 1% saving over half their income.

**11.4.2. Savings Rate Variations and Observed Inequality.** Focus here just on financial capital accumulation, so that equation (11.3) is relevant: we're presuming that all income is capital income. Combine equations (11.1) and (11.3) to get an intuitive equation for wealth evolution:

$$k(t+1) = (y(t) - c(t)) + k(t) = sy(t) + k(t) = (1 + sr)k(t), \tag{11.5}$$

which in words reads: "I save a fraction $s$ of my capital income (which is all my income), and add it to old wealth to get new wealth." Our focus on those "pure accumulators" who live *entirely* off capital income will give us a good sense of how $s$ and $r$ affects the growth of wealth. (Indeed, authors such as Piketty 2014 have argued that the upper

[119] For the original discussion of such issues, see Friedman (19xx).

| | 6-Yr Income Average | Instrumented By Vehicle Consumption |
|---|---|---|
| Quintile 1 | 1.4 | 2.8 |
| Quintile 2 | 9.0 | 14.0 |
| Quintile 3 | 11.1 | 13.4 |
| Quintile 4 | 17.3 | 17.3 |
| Quintile 5 | 23.6 | 28.6 |
| Top 5% | 37.2 | 50.5 |
| Top 1% | 51.2 | 35.6 |

**Table 11.1.** Average Savings Rates at Different Relative Incomes. Source: Dynan, Skinner and Zeldes (2004).

income groups do indeed have predominant fractions of financial capital in their overall portfolio.) Could this breed of capital-income-wielding high earners explain the data on inequality? Let's play with the rise in inequality that we've seen for the top income groups over 1970–2010. Our data for developing countries is limited, so we illustrate our approach using the United States and the United Kingdom.

Say that the average rate of growth in the economy is $g$, which is around 2% for the US and UK. The rich, on the other hand, are presumably growing according to equation (11.5), so that if the initial share of the rich (at some fictitious date 0) is $x(0)$, then $t$ periods later it will be

$$x(t) = x_0 \left( \frac{1+sr}{1+g} \right)^t .$$

The point is that this equation allows us to back out $r$ if we know $s$ and $\{x(t)\}$:

$$r = \frac{[x(t)/x(0)]^{1/t}(1+g) - 1}{s}. \tag{11.6}$$

As an illustration, consider the top 10% of income earners in the United States. The *Top Incomes Database* reports that they took in a share of about a third of total income in 1970, which has since climbed to a whopping 47% by 2000. That is, $x_{1970} = 1/3$ which rises to $x_{2000} = 47/100$ thirty years later. What annual rate of return must they have earned to explain this increase? Fortunately, we have Table 11.1 to guide us. I will use a savings rate of 35% for this group, which is probably a bit on the optimistic side. Substituting that value as well as $g = 2\%$ in (11.6), we can back out the following value of $r$: 9.7% per annum, net of inflation, every year, from 1970–2000.

We can perform similar calculations for the top 1% in the United States. Once again, the *Top Incomes Database* reports $x_{1980} = 8/100$, rising to $x_{2005} = 18/100$ a quarter of a century later. I use the Dynan *et al* (2004) estimate of $s = 51\%$ for this group, which backs out an impressive real annual rate of return of 10.5% for this group.[120] For the top 10% of the income distribution in Europe over 1980–2010, the implied rate is lower

[120] Thanks to the *Top Incomes Database*, one can perform similar calculations for the top 0.1% in the United States: their share climbed from $x_{1980} = 2.2/100$ to $x_{2007} = 8/100$. Savings estimates for this group are unavailable but if they save half their income, then the implied rate of return on capital is $r = 14.4\%$, whereas if they save 3/4 of their income, then $r = 9.6\%$.

but still an extremely large 7.5% per year, while for the top 1% in the United Kingdom (1980–2005) the corresponding number is 11.4%.

**11.4.3. The Implied Rates of Return: A Discussion.** Are these implied rates convincing? If you kept your money in the United States stock market over the last century, you would have probably made (inflation adjusted, and including dividends) between 6% and 7% on your money. In Europe the corresponding return would be quite a bit smaller, certainly under 5%.[121] In the 1970s and the 2000s you would have made far less, in the 1980s and 1990s you would have made quite a bit more. To pull off 10% a year or more, over 1970–2010 would have been a feat of unusual wizardry.

That said, remember that this is the top 10% of the population we are looking at. That is a sub-population which has made higher rates of return than the population average.[122] In that light, the numbers are not surprising. The more interesting question is how such high rates of return are sustained.

One obvious consideration is, of course, risk appetite: the less rich may be less able to bear the enormous fluctuations that equity ownership could entail. But the so-called "equity premium" is already built into the average 6-7% available on the market: we are discussing here a return that is around 50% higher than that. Attitudes to risk cannot fully explain these returns.

Then there is the question of information. Someone who has more money in financial ventures is naturally more incentivized to gather information about financial sectors and markets. Whether they are systematically successful in doing so is, of course, another matter. On this matter, the direct data from developed countries appears to be inconclusive, on average (see Benhabib and Bisin 2016 for a summary of the evidence). Appendix 1 contains a model of incentivized information-gathering.

Information aside, however, the major explanation for return differentials is that such returns are unavailable to the not-rich, and that they arise not on the stock market but elsewhere. That can happen from two sources. First, the real gains have come in high rates of return to private, unincorporated businesses which do not appear on the stock market. This is an extremely important consideration for developing countries, where entire avenues of business are left relatively wide open and exploitable because of the general absence of capital markets or deep pockets to finance those businesses.

The second source is skilled labor incomes, or returns to human capital. As in the case of private, unincorporated business, there is no way to share in the returns to such activities by holding publicly traded stocks. You will need to acquire those skills yourself by making — or by being *able* to make — the right educational and occupational choices. If capital markets do not readily fund those activities, and we shall see below that they do not, you or your parents will need to fund them. Without the initial wealth to do so, you could simply be out of luck.

---

[121] Table 1 in the *Credit Suisse Investment Returns Yearbook* (2014) lists rates of return for several developed countries over the 20th century. The United States comes in at 6.59%, while the corresponding number for the United Kingdom is 5.33%.

[122] Also note that "the top 10%" does not represent the same people over time. New, successful, individuals enter that decile, while the relative failures leave. That also biases upward any rate of return we calculate using simply the incomes of the highest decile.

Both the returns to entrepreneurship and to human capital are described well by an equation such as (11.4). In the former case, $x$ is to be viewed as a setup cost, and in the latter case as a threshold for acquiring skills. It is no surprise that these two items are precisely those that appear in any story of uneven growth. They feature in the disequalizing changes that initially accompany a structural transformation: agriculture-industry, the IT revolution, the rise of the service economy, or globalization. In each case there is a huge rise in the premium to entrepreneurship or human capital. That premium will later be competed away as others enter. But as already discussed, the *ongoing* battle between uneven change and later catch-up will never end.

## 11.5. Occupational Choice

Highly unequal societies can significantly distort investment decisions away from optimality. Meet Gilberto. He could be a budding physicist or have enormous business potential, but if he is poor, his poverty will prevent him from exploiting these talents. After all, setting up a business requires upfront investments, and doing a degree in physics takes money.

This sort of explanation sounds trite, but it isn't. One might ask, for instance: why can't Gilberto approach a bank for a loan to finance his business or his education? But if you are thinking of college loans in a developing country or fund-raising on Kickstarter, think again. Even in the most developed of regions, affordable *outside* finance is the exception rather than the rule. It is far easier to convince Gilberto's parents rather than a bank that Gilberto has potential. That generates a fundamental source of difference between Gilberto and his equally talented friend but with richer parents. Upfront funding or self-funding is a dominant means of financing new businesses or education.

**11.5.1. Imperfect Access to Capital Markets.** Even if a bank is convinced of Gilberto's potential, it may worry about the possibility of a default, either involuntary (when things really don't work out) or strategic (when things do work out, but Gilberto defaults anyway). That is why a borrower is typically screened for his or her ability to repay, as well as for past dealings, which signal not only the ability but the *willingness* to repay. Therefore, all other things being equal, a borrower who can put up collateral is more likely to get a loan, compared to a borrower who can't. Collateral not only (partially) covers the lender in the event of a default, but it also reduces the borrower's proclivity to default in the first place.

Here's an example. Say Gilberto wants to start a business, and has personal assets worth R\$100,000. The business entails setting up a small factory, which involves a startup cost of R\$200,000. The business will hire fifty workers, who will be paid R\$5,000 each, and produce and sell widgets for a total revenue of R\$500,000. Suppose that the lifetime of the business is one year, and that after this the loan must be repaid.

Gilberto approaches a bank. He puts up his assets as collateral. The interest rate on the loan is 10%. If he does not repay, then suppose for dramatic effect that Gilberto flees the country,[123] but that his assets (R\$100,000) with the bank are seized, and so are half the variable profits he made during the year (= $(1/2) \times \text{R\$}250,000$, or R\$125,000). There

[123] This isn't entirely fictional: in the 2008 financial crisis, abandoned cars were left at Dubai airport by frantic failed businessmen seeking to flee the country, and presumably their creditors. See, for example, `http://www.nytimes.com/2009/02/12/world/middleeast/12dubai.html`.

| Items | 100K Collateral Repay | 100K Collateral Default | 20K Collateral Repay | 20K Collateral Default |
|---|---|---|---|---|
| Direct payment | 220,000 | 0 | 220,000 | 0 |
| Collateral loss/credit | -110,000 | 0 | -22,000 | 0 |
| Fine/jailtime (exp. value) | 0 | 60,000 | 0 | 60,000 |
| Seizure of profits (exp. value) | 0 | 125,000 | 0 | 125,000 |
| **Total** | 110,000 | 185,000 | 198,000 | 185,000 |

**Table 11.2.** Repayment and Default Payouts For Different Collaterals

is also a chance that he will be arrested and fined or even imprisoned. The expected cost of this punishment (including the probability of capture) is, say, R\$120,000. The flip side: if Gilberto defaults, he pockets the outstanding loan plus interest.

Table 11.2 records the costs of repayment and default. The first two columns deal with the case in which collateral advanced is R\$100,000. The repayment column shows that Gilberto pays back 220,000 but receives a credit of 110,000 for his collateral plus imputed interest. If he defaults, he pays nothing, but loses the collateral, and faces expected penalties on fines and profit-seizure as shown. Clearly, the costs of default outweigh the benefits, so that Gilberto will repay the loan. What if Gilberto had only R\$20,000 to put up as collateral? In that case, by going through the same balance sheet (see last two columns of Table 11.2), it is easy to see that the costs of default are now smaller than the cost of repayment. So Gilberto will default. The *same* person is transformed from a circumspect complier to a defiant defaulter, not because of a sudden personality transplant, but because of collateral!

Let's generalize the essential features of this example with the help of some elementary algebra. Let Gilberto's startup cost be $S$, and suppose that his business consists of hiring $\ell$ industrial workers at wage $w$ to produce an output $f(\ell)$, where $f$ is a production function that just depends on labor. Then profit is $[f(\ell) - w\ell]$.

Suppose that Gilberto puts all his wealth $W$ up as collateral with a bank, borrows $S$, makes his money, and now the time comes for him to repay $S(1+r)$. He could try defaulting on the loan. (More precisely, this is a strategy that the *bank* imagines Gilberto might try.) Of course, he will lose the collateral, now worth $W(1+r)$. He also faces capture and punishment with equivalent money value $F$, in addition to the loss of a fraction $\lambda$ of the profits from his business.[124] Therefore, the bank reasons that Gilberto *will* honor the loan if

$$S(1+r) \leq W(1+r) + F + \lambda[f(\ell) - w(t)\ell],$$

and rearranging this, we obtain the requirement

$$W \geq S - \frac{F + \lambda[f(\ell) - w(t)\ell]}{1+r}. \tag{11.7}$$

So the bank will only advance the loan if Gilberto's initial wealth is "high enough," in the sense captured by inequality (11.7). If initial wealth is lower, he cannot *credibly* convince the bank that the loan will be repaid.

[124] The fact that $\lambda$ is only a fraction captures the fact that you may not be caught for sure, and even if you are, it may not be possible for the lending authority to seize *all* your profits.

The smaller are the values of $F$ (the expected cost of imprisonment) and $\lambda$ (the fraction of Gilberto's profits appropriated), the more stringent is the requirement of initial wealth. That makes sense: if it is very difficult to catch an offender, all that the bank has left to go on is the collateral that was put up in the first place. Market conditions such as the going wage rate also determine access to the credit market. If wages are relatively low, the profits from entrepreneurship are high, and it's easier to get a loan to go into business (the borrower has more to lose under a default).

This completes our simple description of an imperfect credit market, which is just a direct extension of our earlier example. It is, of course, a caricature of real-world calculations, but it isn't a bad one. For instance, we might argue that there are other costs of default, including a loss of future reputation, but there is nothing to prevent us from monetizing these costs as well and including them in the preceding calculations.[125] Perhaps it will then be harder to default, but the qualitative message of the example is unaltered: *credit markets might be shut down for individuals who have relatively small amounts of collateral*. This is true because these individuals cannot credibly convince their creditors that they will not default on their debt obligations.

This shutdown has little or nothing to do with the intrinsic characteristics of borrowers. They could be just as honest as anyone else, but no bank or lender will bet hard cash on it. With the ability to post collateral, those considerations change. In Chapter 19, we will return to these questions and study credit markets in detail.

**11.5.2. Labor Market Equilibrium With Imperfect Capital Markets.** We are now going to marry this capital market scenario to a wider setting which explicitly includes the labor market.[126] Consider an economy with just three occupations: subsistence producer, worker, and entrepreneur. Subsistence producers produce some fixed amount $z$ with their labor. A worker can earn a wage $w$. (The endogenous determination of $w$ will be a central part of our story, so keep an eye on it.) An entrepreneur hires workers, but her business requires startup capital. Each entrepreneur must contend with the capital market in exactly the way that I described in Section 11.5.

I want you to think of all individuals as identical to begin with, *except* for their starting wealth, which we take to be unequally distributed. This initial distribution of wealth, along with the going market wage, yields the fraction of the population that is shut out of entrepreneurship. The reason is that for each wage $w$, inequality (11.7) gives us the minimum wealth level required for access to credit. The fraction shut out from credit is just the fraction with wealth below this threshold. It's as simple as that. We know, moreover, that this minimum rises with the wage rate. (For — as already discussed — once the wage rate rises, there is less profit to be made, and therefore less to lose from a default. Banks therefore demand higher collateral.) It follows that the

[125] Likewise, there are adjustments that can be made for innate honesty, in case you find this example alarmingly cynical. As described, I have not allowed for any qualms of conscience on the borrower's part, but as long as individuals are moved to *some* extent by the economic considerations described here, a variant of this example can easily be constructed to incorporate (a certain degree of) honesty. More importantly, it is not so much the question of a *particular* borrower's honesty but the bank's assessment of the average honesty over *all* borrowers.

[126] While the model I use to make these points is largely of my own devising, it draws its inspiration from a sizable literature on occupational choice and economic inequality: Banerjee and Newman (1993), Galor and Zeira (1993), Ljungqvist (1993), Ray and Streufert (1993), Freeman (1996), Ghatak and Jiang (2002), Mookherjee and Ray (2002, 2003, 2010), and Ray (1990, 2006).

share of the population with access to entrepreneurship loans will generally fall as the wage rate climbs.

The individuals thus excluded from entrepreneurship must choose between subsistence and market labor. That choice depends on the wage rate. Consult the upper panel of Figure 11.3. Wages less than $z$, the subsistence income, will generate a zero supply of workers. At $w = z$, there is a jump in labor supply, because all the subsistence choices now get altered in favor of joining the labor market. For higher wages, the labor supply steadily increases, as more and more people get shut out of entrepreneurship and switch their occupational choice to labor. This process continues until we reach a high enough wage, call it $\bar{w}$, such that the profit from running a business becomes exactly the same as labor income.[127] After *this* point, it should be clear that everyone, whether they can be entrepreneurs or not, will jump into the labor market. If wages exceed $\bar{w}$, labor income exceeds profit income, so no one will want to be an entrepreneur. What emerges, then, is a supply curve of labor all right, but a rather nonstandard one. Its elasticity or slope, for instance, is determined by the going distribution of wealth and workings of the credit market. (In standard models, the slope of the labor supply curve typically mirrors labor-leisure preferences among the population.)

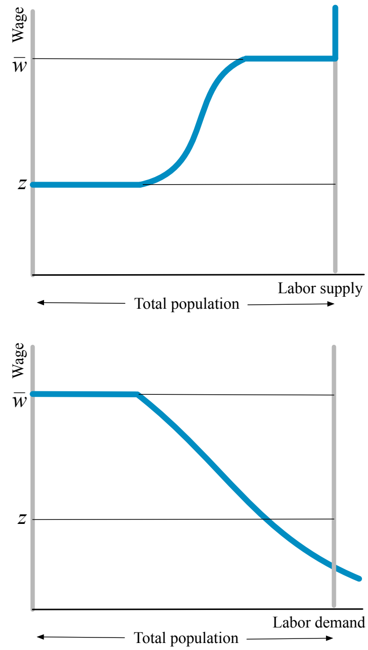



**Figure 11.3.** Labor supply and demand curves. Note the rightward jumps at $z$ and at $\bar{w}$.

We now turn to the *demand* for labor. Start with a high wage that exceeds $\bar{w}$. Obviously, at such wages there is no demand for labor at all, because no one wants to be an entrepreneur. Moving down to $\bar{w}$, we see a sudden rightward jump in the demand for labor as people now enter entrepreneurship. (This mirrors the rightward jump in the supply curve of labor at $\bar{w}$, though the lengths of the two jumps need not be the same.) Thereafter, as the wage falls, the demand for labor steadily rises, capturing the fact that more individuals have access to the credit market with lower wages, *and* the fact that each employer will expand $\ell$ as the wage rate falls. Panel B of Figure 11.3 summarizes labor demand.

We can now put the two curves in Figure 11.3 together to determine *equilibrium wages*. Note well just how the prevailing distribution of wealth feeds into the choice of occupations and therefore determines the shape of these supply and demand curves, and consequently the wage rate. The three possible outcomes are described in the three panels of Figure 11.4. Panel (a) shows what happens if the distribution of income is highly unequal, *or* if the economy is extremely poor. Then there is a large number of individuals with very low wealth. This situation has the effect of creating a sizable

[127]This wage $\bar{w}$ solves $[f(\bar{\ell}) - \bar{w}\bar{\ell}] - rS = \bar{w}$, where $\bar{\ell}$ is the profit-maximizing choice of employment at wage $\bar{w}$. Note that we debit the cost of loan service $rS$ from business profits. Of course, we are assuming that the subsistence level $z$ is less than $\bar{w}$; otherwise there would be no industrial sector in the model.

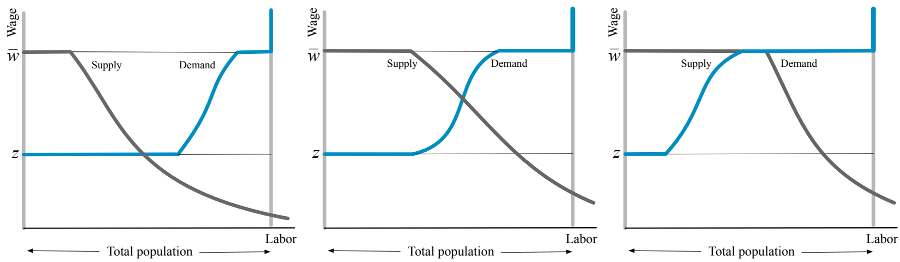



**Figure 11.4.** Equilibrium wages under varying degrees of inequality

supply of labor at any wage exceeding subsistence levels, simply because there are a greater number of individuals barred from entrepreneurship. For exactly the same reason, the demand curve for labor is low, at any wage rate. The result is an intersection of the two curves at the minimum subsistence wage $z$. For the lucky few who are entrepreneurs, however, profits (and so income and wealth) are high.

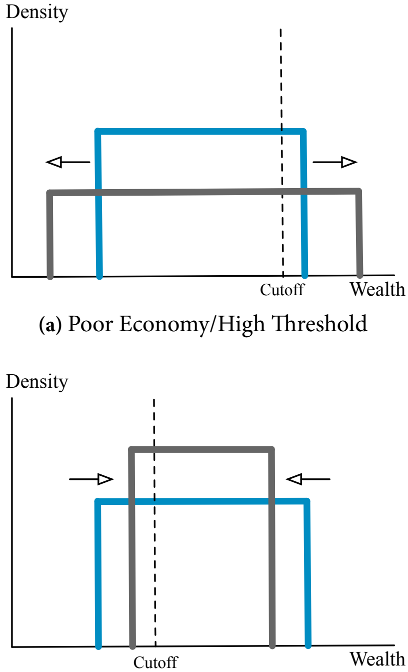
Figure 11.4b displays an intermediate situation of moderate inequality or average wealth, where a sizable fraction of the population is shut out of entrepreneurship, while another sizable fraction is not. Labor demand and supply curves intersect at some wage rate that lies between subsistence $z$ and the high wage $\bar{w}$. Finally, at the opposite extreme, if there is a great deal of equality, *or* if the economy is very rich, relatively few people are barred from entrepreneurship. In general, therefore, individuals will only enter the labor market when wages are high enough to provide an attractive alternative to entrepreneurship. Consequently, the supply curve of labor shifts inward, and the demand curve shifts outward, leading to an equilibrium wage of $\bar{w}$. Note that in such a situation, wages and profits are equalized; see Figure 11.4c.

**11.5.3. Endemic Inefficiency.** This simple model captures well the inefficiency of an economic equilibrium in which capital markets are imperfect, and also has something to say about the connections between inequality and inefficiency. We take as our efficiency criterion the maximization of net output in the non-subsistence sector. Appendix 2, entitled "The Magic of Markets," develops this idea in detail. We only use the essential argument here, which is to note that in any efficient allocation, an entrepreneur's profit net of setup costs must equal the working wage. For say the former is higher, then it would pay to move an individual from workerhood to entrepreneurship, where her contribution is larger, which shows that the original allocation is inefficient. Similarly, if the former is smaller, the opposite movement would be called for — from entrepreneurship to workerhood.

It should be obvious that there can never be an inefficient market outcome in which the second change above is called for. The reason is that a potential entrepreneur could make that change *on her own*, with no need for loans. But the first change is impeded by the potential difficulty of accessing the capital market. Return to Panel A of Figure 11.4 in which industrial wages are reduced to the subsistence alternative. In this situation there are some individuals in the subsistence sector. What if a fraction of these individuals could have become entrepreneurs? They would then have generated profits for themselves, which would exceed their subsistence level of income *and* in the process would have pulled more workers into the industrial sector. That scenario creates an unambiguous efficiency improvement. In fact, some section of the population is made better off while no one else is made worse off.

In the "intermediate" regime described by Figure 11.4b, inefficiency continues to persist. Because the equilibrium wage is lower than the wage $\bar{w}$ at which entrepreneur profits and worker wages are equalized, it is socially efficient for more individuals to become entrepreneurs. Worker incomes would also rise, because of the resulting upward pressure on wages. Existing employers of labor would be worse off but nevertheless, aggregate national income must climb with this change. Why doesn't the market permit these improvements to organically arise? The reason is that the improvements require additional access to credit, and such access is hindered because of the distribution of wealth. That creates inefficiency in the economy as a whole. Even if we do not care about distributional matters *per se*, the resulting inefficiency might still matter to us. Or at least, it should.

It is only in Figure 11.4c that entrepreneurial access is so plentiful that wages rise to the level of $\bar{w}$. In this case, further easing of the credit market serves no function at all: the outcome is efficient to begin with. For more details on this case, see the Appendix.




**(a)** Poor Economy/High Threshold

**(b)** Richer Economy/Low Threshold

**Figure 11.5.** Redistribution and efficiency

**11.5.4. Inequality and Efficiency.** Would redistribution ameliorate or exacerbate this inefficiency? The answer is complex, and depends on the going distribution of wealth, as well as its average level. Use Figure 11.5 to accompany the discussion that follows. Both panels depict a pair of uniform density function for the wealth distribution, one in blue for the initial distribution, and one in grey following a change in that distribution. Each panel also shows the going wealth threshold needed for access to the capital market, evaluated at the initial equilibrium wage before the change.

Consider an extremely poor society, in which most individuals have wealth that lie below the cutoff required to access the credit market (Figure 11.5a). Then a *disequalization* of wealth — while painful for the relatively worse-off — can create a situation in which at least some higher fraction of individuals can access the capital market, or have access to self-funding. This is a terrible prospect and one can only hope that the resulting gain in employment might compensate for some of this loss. A dramatic way of thinking about it is to visualize how ancient monuments such as the Great Pyramids or the Taj Mahal were built. With an equal distribution of wealth, those monuments would simply not have been built. As much as we revere these architectural wonders today, how do we even begin to evaluate the now-distant but terrible inequities that made such enterprises possible?

In societies with higher levels of average wealth, the opposite is true. A progressive redistribution of that wealth would permit a larger number of individuals to chase their entrepreneurial dreams, taking wealth away from those far above the minimum required to enter into productive business, and redistributing that wealth to individuals just below the cutoff. Now *equality* and efficiency go hand in hand, presumably a more palatable prospect. A disequalization of wealth means that the uniform distribution spreads out even further; an equalization compresses that distribution. Panel A shows that if the wealth threshold is high relative to the wealth distribution (which is just another way or saying that the society is relatively poor), then disequalizations generate greater entrepreneurship. Just the opposite occurs when society is wealthy on average relative to the minimum wealth threshold; see Panel (b).

**11.5.5. Inequality Begets Inequality.** Our discussion so far elides an important point, which is that the wealth distribution in the society is itself an outcome of the occupational choices made in the market. When we incorporate this observation, it reveals an intrinsic tendency for inequality to beget itself. Look again at Figure 11.4a. Its outcome is generated by the fact that the majority of individuals are shut out from access to credit, so that the labor market is flooded from the supply side and is tightfisted on the demand side. *Such an outcome goes precisely toward reinforcing the inequalities that we started with*. People earning subsistence wages are unable to acquire wealth, while wealthy entrepreneurs make high profits off the fact that labor is cheap. The next period's wealth distribution therefore tends to replicate the wealth distribution that led to this state of affairs in the first place.

Thus high inequality not only gives rise to inefficient outcomes, it tends to replicate itself, prolonging the inefficiency. This persistent disparity does not arise from any deep-rooted difference, genetic or nurtured, among individuals. It stems from the fact that the poor are shut out of projects that yield high rates of return, and so locked into poverty. There is no "convergence" to attentuate these wealth disparities over time. Billie Holiday, quoted at the start of this chapter, was not off the mark at all.

But: had we started with low inequality, that may *also* have been self-perpetuating. Consider, for instance, the situation depicted in Figure 11.4b. In this case all economic agents earn the same, and as time passes, there is no reason for this state of affairs to change (unless something else is different, such as rates of savings across individuals). Appendix 3 contains a simple algebraic description of these multiple steady states, using a variant of our model that emphasizes human capital accumulation.

As in some of our earlier material on development traps, this model is in the spirit of a "frozen accident." It tells us little about how a history of high inequality comes about in the first place, but suggests that a history of high inequality may persist into the indefinite future, carrying with it inefficiencies in production. The very same economy may exhibit different levels of output and investment if its history were to change to one of low initial inequality. But there isn't a magic wand to change history.

**11.5.6. Sustained Growth and Setup Costs.** If you have been following this argument closely, you might raise a natural objection at this point. Look, you might say, all these problems occur only in the here and now. *Over time*, people will save and their wealth will increase. Sooner or later everybody will be free of the credit constraint, because they will all have sufficient collateral to be entrepreneurs if they so wish. Thus after some time, everything should look like Figure 11.4c. The inefficiency you speak of is only temporary, so what is all the fuss about?

This is a good question. The answer lies in thinking about just what constitutes the startup cost that we've so blithely blackboxed with the label $S$. Presumably, the startup cost of a business includes the purchase of plant and equipment — physical capital, in other words. If we go beyond the very simple model of this section, we also see a role for startup *human* capital: skilled technicians, researchers, scientists, trained managers, and so on. All of this goes into $S$. If we begin to think about the economy as it runs over time, surely these startup costs will change as overall wealth changes. For instance, we would expect the costs that are denominated in terms of human capital to rise along with national wealth: the wages of scientists and engineers will rise. Thus startup costs are endogenous in an extended view of this model and will presumably increase as wealth is accumulated. The whole question then turns on how the *ratio* of startup costs to wealth changes over time. For instance, if wealth is accumulated faster than startup costs increase, your objection would indeed be correct: the inefficiency is only an ephemeral one. If this is not the case—if startups keep pace with wealth accumulation—then these inefficiencies can persist into the indefinite future.[128]

**11.5.7. A Reinterpretation Based on Human Capital.** The analysis so far illustrates a general principle that is of widespread applicability. *Inequality has a built-in tendency to beget inefficiency, because it does not permit people at the lower end of the wealth or income scale to fully exploit their capabilities*. We've illustrated this by the inability of a section of the population to become entrepreneurs. But that is only one example. Replace "entrepreneurship" by "human capital," and a more general point emerges. There is a parallel between educational loans, and loans to set up a business.

Actually, in the case of education, matters are possibly worse, because unlike the alienable assets of a business, human capital cannot be seized and transferred to a creditor in the event of default. Thus human capital cannot be put up as collateral, whereas a house or business can be pledged, at least to some extent, as collateral in the event of failure to repay. It follows that the constraints on human capital loans are even more severe, dollar for dollar. As Glenn Loury (1981) has observed,

> Early childhood investments in nutrition or preschool education are fundamentally income constrained. Nor should we expect a competitive loan market to completely eliminate the dispersion in expected rates of return to training across families. Legally, poor parents will not be able to constrain their children to honor debts incurred on their behalf. Nor will the newly-matured children of wealthy families be able to attach the (human) assets of their less well-off counterparts, should the latter decide for whatever reasons not to repay their loans.

Loury was writing about the U.S. economy, and so was Arthur Okun (1975) when he judged the constrained accumulation of human capital to be "one of the most serious inefficiencies of the American economy today." Consider the same phenomenon magnified severalfold for developing countries.

[128]Thus further research on this topic will have to study the composition of startup costs and how various components are affected by the development process. See Mookherjee and Ray (2003) and Rigolini (2004).

Thus the poor have to fund educational choices out of retained earnings, wealth, or abstention from currently productive work. Because they are poor, the marginal cost of doing so may be prohibitively high. If a wealthier person were to loan a poor person money for education, an economy-wide improvement in efficiency would be created. However, the perceived difficulties of loan repayment make such an efficiency-enhancing move extremely problematic.

Finally, inequalities in education feed back and reinforce the initial differences in inequality. This part of the story is also analogous to the model of the previous section. Multiple development paths can result: one characterized by high inequality, low levels of education, and inefficient market outcomes; the other characterized by low inequality, widespread education, and equalization of the rates of return to education across various groups in society, which enhances efficiency. Appendix 3 to this chapter formalizes such a model. It has much in common with the story of entrepreneurship studied in this chapter, but takes a step further in endogenously solving for the distribution of income, along with the market outcome.

## Appendix 1: Investing in Investment

In this Appendix, we write down a model of "investing in investment," in which a person — call him Scrooge — expends systematic effort to increase the rate of return on his assets. The idea is simple: someone with more financial wealth will spend more effort finding good rates of return on it, which will tell us that the *rate* of return on wealth could be increasing in wealth. (But there will be a small twist.) Suppose, then, that Scrooge derives pleasure from a lifetime utility function of the form

$$\sum_{t=0}^{\infty} \delta^t c(t)^{1-\sigma},$$

where $0 < \sigma < 1$, so that Scrooge has diminishing marginal utility of consumption at any date. If you are more comfortable with abstraction, you can take this to be just any strictly concave utility function $u(c)$ of consumption without changing any results below.

Scrooge has two sources of wealth at any date $t$. First, capital: if $F_{t-1}$ represents his financial assets from last period holdings and $r_{t-1}$ the rate of return earned, then his current assets are $(1+r_{t-1})F_{t-1}$. Second, labor: suppose that Scrooge earns an income of $w$ times the fraction of his day he spends on the labor market. The remaining fraction is spent hunting for good returns on the stock market. (It is possible — though unlikely — that Scrooge does other things with his time, such as sleeping or going to the movies, but we'll ignore those there.) Then Scrooge's *consumption* today is given by

$$c(t) = (1 + r_{t-1})F_{t-1} + w(1 - e(t)) - F(t), \tag{11.8}$$

where $F(t)$ represents the assets laid away for tomorrow, $1 - e(t)$ is the fraction of labor time, and $e(t)$ is the fraction of time spent "investing in investment," so:

$$r(t) = \beta e(t). \tag{11.9}$$

That is, each hour spent researching the market gets him a bit more of a rate of return. Of course, this is all an abstraction: it isn't that some deterministic rate of return is obtained for sure, without any uncertainty at all. But nothing will change if we extend the model to include stochastic returns. Also, the linear form I have assumed here is completely unimportant. You could replace it with any concave function $r(t) = g(e(t))$ without changing a single result.

Now this looks like a complicated problem. At every date Scrooge has two instruments at his disposal, both of which affect the allocation of consumption over time: $F(t)$, which determines how much he saves for the future, and $e(t)$, which affects his rate of return on $F(t)$. If we just focus on these two variables at date $t$ (keeping all other instruments fixed) and think of how they affect the allocation of consumption across today and tomorrow, then we can get some necessary conditions for an optimum allocation.[129] First, if Scrooge adjusts $F(t)$ to equate marginal benefits over time, then we get the condition

$$\delta^t(1-\sigma)c(t)^{-\sigma} = \delta^{t+1}(1+r(t))(1-\sigma)c(t+1)^{-\sigma}.$$

The left-hand side is the marginal utility cost today from increasing $F(t)$ by a tiny bit. But that augments consumption tomorrow, and brings additional utility equal to the right-hand side of the above equation. Canceling common terms and moving things around a bit, we see that

$$\left(\frac{c(t+1)}{c(t)}\right)^{\sigma} = \delta(1+r(t)). \tag{11.10}$$

Likewise, an increase in $e(t)$ affects consumption today (it lowers labor income), but it translates into a higher rate of return tomorrow and so more consumption. Equating marginal costs and benefits over today and tomorrow:

$$\delta^t(1-\sigma)c(t)^{-\sigma}w = \delta^{t+1}\beta F(t)(1-\sigma)c(t+1)^{-\sigma},$$

which simplifies after some elementary manipulation to the condition

$$\left(\frac{c(t+1)}{c(t)}\right)^{\sigma} = \delta\frac{F(t)}{w}\beta. \tag{11.11}$$

Combine Equations (11.10) and (11.11) to see that

$$1+r(t) = \frac{F(t)}{w}\beta \tag{11.12}$$

for all dates $t$. Equation (11.12) gives us the following proposition:

Proposition 11.1. *Individuals with a higher ratio of financial wealth to labor income earn a higher rate of return, and so grow faster.*

Notice that the proposition does *not* say that a richer person has a higher rate of return, but that someone who a higher *ratio* of financial wealth to labor income earns a higher rate of return. That happens because Scrooge uses his *own* effort to carry out research on rates of return. The opportunity cost of that wealth is his labor income, so that a person with high labor income may not expend that effort even if he is asset-rich. It would all depend on the dimension along with he is relatively *richer*.

But that changes if you can hire someone to manage your money. To capture that, let's change Equations (11.8) and (11.9) to

$$c(t) = (1+r_{t-1})F_{t-1} + w - z(t) - F(t), \tag{11.13}$$

[129] It can also be shown that under our assumptions, these necessary conditions are sufficient for an optimum, but we won't go into that here.

where

$$r(t) = \gamma z(t). \tag{11.14}$$

What's the interpretation? Now, instead of using his own labor time to research hot stocks, Scrooge is hiring an expert to do his research, and is paying the expert $z(t)$. That is also costly, of course, but it is costly in a different way. The amount he pays the expert depends on the *expert's* market income, but it does not depend on Scrooge's labor market capacity. Of course, this does not change the necessary condition (11.10), which comes from the choice of $F(t)$, but it does change the condition (11.11). When Scrooge adjusts $z(t)$ to equate marginal benefits today and tomorrow, we obtain the condition

$$\delta^t(1-\sigma)c(t)^{-\sigma} = \delta^{t+1}\gamma F(t)(1-\sigma)c(t+1)^{-\sigma},$$

or after simplification:

$$\left(\frac{c(t+1)}{c(t)}\right)^{\sigma} = \delta F(t)\gamma. \tag{11.15}$$

Equations (11.10) and (11.15) combine to tell us that

$$1 + r(t) = \gamma F(t), \tag{11.16}$$

which yields a different proposition:

PROPOSITION 11.2. *Individuals with higher financial wealth earn a higher rate of return, and so grow faster.*

Contrast the two propositions. Proposition 11.2 tells us that if a money manager is hired, then it's individuals with higher wealth that earn a higher rate of return on their investments, whereas according to Proposition 11.1, if own effort is used, the ratio of financial wealth to labor income is what matters. The two propositions help us understand which their contrasting implications is more likely to be applicable, under varying conditions.

## Appendix 2: The Magic of Markets

We show that the equilibrium of occupational choice into entrepreneurship or labor works beautifully *provided that* credit markets are perfect. Our course, our starting point is that they're not perfect, but we can often better understand the implications of a critical assumption by temporarily negating it. Besides, we will also derive the market wage explicitly in our exercise.

Index people on $[0,1]$, and each person $i$ in $[0,1]$ can become a *worker* or an *entrepreneurs*. To become a worker, you don't need any preparation: you walk into the labor market and start earning a wage. But there is a startup cost $S$ for becoming an entrepreneur. As soon as you pay this, you have access to a production function $f(\ell) = A\ell^{\alpha}$ (where $0 < \alpha < a$) and you can start making money from your business. How much do you make? Well, if entrepreneurs hire workers at wage $w$ (later, $w$ will equate supply and demand), then

$$\text{Profits} = A\ell^{\alpha} - w\ell - S(1+r),$$

where $r$ is the "opportunity rate of return" on the money you ploughed into setting up, or if it is someone else's money, it is the interest rate you've paid.

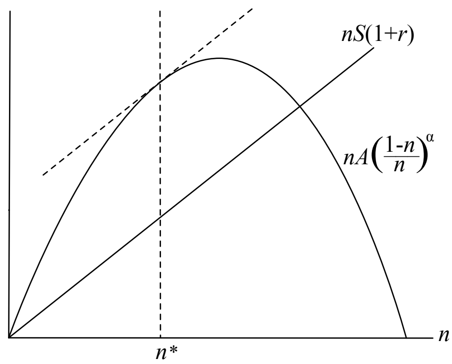



**Figure 11.6.** NET SOCIAL OUTPUT.

Our notion of efficiency is the *net output* accruing to society, which is

$$nA\left(\frac{1-n}{n}\right)^{\alpha} - nS(1+r), \tag{11.17}$$

where $n$ is the fraction of entrepreneurs (a number between 0 and 1).

Why is this net output? Well, if $n$ entrepreneurs are chosen, then there are $1-n$ workers available. They have to be allocated to the $n$ firms, but if the production function is concave, which it is (because $\alpha$ lies between 0 and 1), then the best way to do this is to allocate equally.[130] That is, each firm gets $(1-n)/n$ workers, so per-firm output multiplied by $n$, minus the setup cost, gives you the expression above. Figure 11.6 plots this social output as a function of $n$, which the omnipotent and benevolent social planner can freely choose. The curved line is the first term in equation (11.17). Its shape makes sense: when $n=0$, there are lots of workers but no firms (so no output), and when $n=1$ there are lots of firms but no workers (so again no output). In between output rises and then falls. The second term is the setup cost which is simply linear in $n$.

As Figure 11.6 shows, the way to maximize net output is to choose $n$ so that the marginal output benefit is equal to the marginal setup cost of another firm, which is $S(1+r)$. The former is the derivative of the output function with respect to $n$. Taking that derivative, the required condition is

$$A\left(\frac{1-n^*}{n^*}\right)^{\alpha} - \frac{\alpha}{n^*}A\left(\frac{1-n^*}{n^*}\right)^{\alpha-1} = S(1+r). \tag{11.18}$$

Can this efficient solution be achieved as a decentralized equilibrium? The answer is *yes*, if credit markets are perfect. For that equilibrium must satisfy two properties. First, because *anyone* has the option of becoming an entrepreneur (credit markets are perfect), it must be the case that entrepreneurial profits equal the wage; that is,

$$A(\ell)^{\alpha} - w\ell - S(1+r) = w. \tag{11.19}$$

[130] If two firms have unequal amount of labor, then transferring a unit of labor from the higher-employment firm to the lower-employment firm will create a net increase in output.

Next, all entrepreneurs make the same profit-maximization decision, so $\ell$ is the same for everyone, which means that $\ell = (1-n)/n$. Using this in (11.19),

$$A\left(\frac{1-n}{n}\right)^{\alpha} - w\frac{1-n}{n} - S(1+r) = w,$$

and moving terms around slightly,

$$A\left(\frac{1-n}{n}\right)^{\alpha} - w\frac{1}{n} = S(1+r). \tag{11.20}$$

The next condition comes from the fact that each entrepreneur must be incentivized to hire their $(1-n)/n$ units of labor. That is, the wage $w$ must equal marginal product with respect to $\ell$, evaluated at $(1-n)/n$:[131]

$$w = \alpha A\left(\frac{1-n}{n}\right)^{\alpha-1} \tag{11.21}$$

Substitute (11.21) into (11.20) to get

$$A\left(\frac{1-n}{n}\right)^{\alpha} - \frac{\alpha}{n}A\left(\frac{1-n}{n}\right)^{\alpha-1} = S(1+r). \tag{11.22}$$

Compare (11.18) with (11.22), they are the same. So $n$ equals $n^*$. This is the magic of markets: if they are frictionless and there are no externalities, they “implement" efficient outcomes. But these are big ifs. In particular, when credit markets are imperfect, the ability to access that market for setting up a firm will require some degree of self-funding, or collateral. The credit market will not be available to all, and as the this chapter shows, the overall outcome will generally no longer be efficient.

## Appendix 3: Steady States with Imperfect Capital Markets

In this appendix, we study a model of human capital which has very similar features to the labor market equilibrium studied in the main chapter. Indeed, we will go a bit further here and solve for the resulting wealth distribution in a steady state.

We adopt the simplest description of a dynastic household — at any date there is a single parent in each household with a single child. At the next date, the child grows up, replaces the parent, and now has a child of her own. Parent and child are linked together by altruism. Suppose that each parent can choose to educate their child by incurring an education cost; otherwise, the child grows up as an unskilled manual worker. We shall suppose that this is the *only* bequest that a parent can make to her child, and that no financial transfers are possible.[132]

Presumably, if there are few or no skilled people, the wage rate of an educated person is immensely high — that person would face enormous demand for her skills. On the other hand, if everyone is skilled, then *unskilled* tasks would command a premium, and the wage differential would vanish (or even turn negative, if educated people are

[131] Warning: do *not* differentiate with respect to $n$ to find marginal product, because that involves changing both the number of firms and the number of workers. No individual firm can do that. Instead, take the derivative of the production function $A(\ell)^{\alpha}$ with respect to $\ell$ and evaluate the answer at $\ell = (1-n)/n$, which is what we do in the main text.

[132] Elements of this model appear in Ljungqvist (1993), Freeman (1996) and Mookherjee and Ray (2003). The specific exposition follows Ray (2006). Mookherjee and Ray (2010) consider the more complex setting in which both financial transfers and educational investments are possible.

incapable of manual labor, which isn't a crazy thought at all). Specifically, denote skilled labor by $s$, unskilled labor by $u$, and introduce the production function

$$y = F(s,u)$$

which combines skilled and unskilled labor to produce an output. We will normalize our overall population to 1, with $n$ as the share getting skills, so that $s = n$ and $u = 1 - n$. This normalization won't matter at all if the production function is constant returns to scale in $s$ and $u$, which we shall assume. Wages in each occupation will equal their respective marginal products, and so depends on the share of skilled labor in the population just as we've described above. That is, if the skilled wage is $w_s$ and the unskilled wage $w_u$, these are both functions of the skill share $n$ in the population. Specifically, $w_s = w_s(n) = F_1(n, 1-n)$ and $w_u = w_u(n) = F_2(n, 1-n)$, where the subscripts stand for the partial derivatives of $F$ with respect to its first and second inputs. Our previous discussion makes it obvious that $w_s(n)$ is declining in $n$ while $w_u(n)$ is increasing in $n$, with $w_s(n) - w_u(n)$ extremely large when $n \simeq 0$, and close to zero or even negative when $n \simeq 1$.

Each person at $t$ derives utility from her consumption and the anticipated consumptions of all her descendants:

$$\sum_{s=t}^{\infty} \delta^{s-t} u(c_s),$$

where $0 < \delta < 1$ is the discount factor and $u$ is a one-period utility function with diminishing marginal utility. There is *no* capital market, and an educational investment, which costs an amount $X$, must be borne by the parent out of her own funds. Because there are no other bequests by assumption, those funds consist of the skilled wage $w_s$ if the parent is uneducated, or the unskilled wage $w_u$ if the parent is uneducated.

In general, the solution to such a model can be quite complicated (see Ray 2006 for the gory details), but we can focus on its *steady states*, in which a constant fraction $n$ are skilled, wages are constant at $w_s = w_s(n)$ and $w_u = w_u(n)$, and all parents keep replicating their skill status in their children. Let's think through how such a solution must work. First, it must be that no skilled person must prefer to de-skill their child; that is,

$$\frac{u(w_s - X)}{1-\delta} \geq u(w_s) + \frac{\delta}{1-\delta} u(w_u) \tag{11.23}$$

In similar vein, no unskilled parent must prefer to skill their child; that is,

$$\frac{u(w_u)}{1-\delta} \geq u(w_u - X) + \frac{\delta}{1-\delta} u(w_s - X). \tag{11.24}$$

Here are these two inequalities in words. Equation (11.23) has on its left hand side the overall utility of a skilled parent who skills her child: she incurs a cost $X$, which is therefore subtracted from her wage $w_s$. Thereafter, she anticipates (correctly so in this simple model) that her descendants will all educate *their* children, and so her utility is given by the infinite sum $u(w_s)[1 + \delta + \delta^2 + \ldots]$, which is just $u(w_s - X)/(1-\delta)$. On the right hand of (11.23) is the utility the same skilled parent would receive if she de-skills her child: she saves on the cost $X$ and therefore gets to consume the full skilled wage $w_s$, but thereafter knows that she will have uneducated descendants, the present utility value of which is given by $\delta u(w_u)/(1-\delta)$. The inequality in (11.23) states that *a skilled parent must willingly make the decision to keep her child skilled.*

In parallel fashion, the left hand side of (11.24) records the utility of an unskilled parent from leaving her child unskilled. (No education costs are incurred, but wages stay at the lower level $w_u$ for all time.) The right hand side records her utility were she to change her plan and educate her child: this would entail a current consumption of $w_u - X$, followed by a switch to perennial education by her descendants, which has present value $\delta u(w_s - X)/(1-\delta)$. The inequality (11.24) states that *an unskilled parent prefers, all things considered, to keep her child unskilled.*

Even at the level of simplicity of this model, that last sentence is so loaded that we must pause to consider exactly what it means. It is emphatically not a statement about some deviant preference on the part of the unskilled; in fact, *all* our parents have the same preferences. But the unskilled parent has a lower wage, and so the cost of education, while the same for all in money, is an impossible burden for her in *utils*. That first term on the right hand side of (11.24), $u(w_u - X)$, gives her unbearably low utility. With these thoughts in mind, we have:

Proposition 11.3. *Every $n$ such that $w_s = w_s(n)$ and $w_u = w_u(n)$ satisfy inequalities (11.23) and (11.24), or equivalently, every $n$ for which*

$$\underbrace{u(w_u) - u(w_u - X)}_{\text{Unskilled Cost}} \geq \underbrace{\frac{\delta}{1-\delta}\left[u(w_s - X) - u(w_u)\right]}_{\text{Future Benefit}} \geq \underbrace{u(w_s) - u(w_s - X)}_{\text{Skilled Cost}} \tag{11.25}$$

*must be a steady state.*

The double inequality (11.25) is just a compact rewriting of conditions (11.23) and (11.24). Its middle term can be interpreted as the present value of the benefit from educating your child. It is positive — it has to be — even if skilled individuals need to bear educational costs. Their skilled wage more than compensates for it. The left hand side is the current educational cost (measured in utils) for a unskilled parent. The inequality states that this cost is too high relative to the benefit conferred by the middle term. On the right hand side is the current educational cost (measured in utils) for a unskilled parent. The inequality states this time that the educational cost is lower than the benefit. The inequality (11.25) is like a sandwich, placing future educational benefits between the two utility costs of skilled and unskilled parents. *Any population share of skilled people which generates skilled and unskilled wages in a way that satisfies the sandwich inequality is a steady state.* That population share won't move over time.

Figure 11.7 brings out the implications of Proposition 11.3. The line marked "Unskilled Cost" graphs the utility cost of education for the unskilled — the expression $u(w_u) - u(w_u - X)$ in (11.25) — as a function of the skilled population share $n$. Notice that as $n$ goes up, the unskilled wage must rise (skilled labor being more plentiful), and therefore, by diminishing marginal utility, the utility cost of education goes down. So the "unskilled cost graph" is downward sloping. By a parallel argument, the line marked "Skilled Cost" is upward sloping. Notice that Unskilled Cost always exceeds Skilled Cost, at least as long as $w_s > w_u$. If the population share of skilled labor is ever so high that $w_s = w_u$, the two utility costs of education are equal. That is the case at $n_1$.

The third line in the figure, marked "Future Benefit," is also downward sloping. For as $n$ goes up, the skilled wage falls while the unskilled wage rises, thereby narrowing the benefits of acquiring an education. This line hits zero at a point $n_2$ where $w_s - X = w_u$.

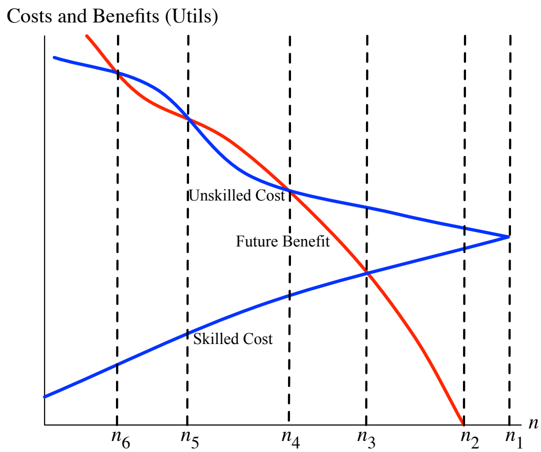



**Figure 11.7.** Steady States in the Human Capital Model.

It is smaller than $n_1$, where $w_s = w_u$. This point $n_2$ is a distinguished value and we will return to it below.

By Proposition 11.3, we know that every value of $n$ such that "Future Benefit" is sandwiched between "Unskilled Cost" and "Skilled Cost" is a possible steady state. Even a cursory look at Figure 11.7 tells us that:

(i) A steady state exists. Looking from right to left along the diagram, the first such steady state is $n_3$, at which Future Benefit just equals Skilled Cost (which in turn is lower that Unskilled Cost), so a sandwich just starts there.

(ii) There are many, many steady states — an infinity of them in our model![133] In the diagram, every skill share in the union of the intervals $[n_6, n_5]$ and $[n_4, n_3]$ is a steady state, and no share outside this set is.

(iii) *Every* steady state involves persistent inequality, not just in lifetime gross incomes but also in lifetime incomes net of the cost of education. Notice that the minimum inequality in a steady state is at the point $n_3$, at which $w_w - X > w_u$; the two are equal only at the higher skill share $n_2$ which is not a steady state.

A more careful look at Figure 11.7 tells us something more about the connection between inequality and inefficiency. Just as in the previous section, imagine that a benevolent planner were to choose $n$ to maximize net output in the economy. She would then choose the share of skilled people to maximize

$$F(n, 1-n) - nX,$$

where we recall that $F$ is the aggregate production function of the economy. This net output can be depicted exactly as we did for labor markets in Figure 11.6. The curve there can now be interpreted as the function $F(n, 1-n)$, while the straight line is just $nX$. The gap between the two, which is net output, is maximized at the point $n^*$ such that the slope of the curve equals the slope of the straight line. Taking derivatives in

[133] This large set of steady states is a consequence of our "two-occupation" model. As the number of distinct occupations rises, the measure of steady states can be shown to progressively shrink; see Mookherjee and Ray (2003).

the expression above with respect to $n$, this means that

$$F_1(n^*, 1-n^*) - F_2(n^*, 1-n^*) = X, \tag{11.26}$$

where the subscripts denote partial derivatives as already noted. But these partial derivatives are just marginal products, so if $n^*$ were to be the market skill share, then the corresponding skilled wage $w_s^*$ equals $f_1(n^*, 1-n^*)$, while the unskilled wage, $w_u^*$ equals $f_2(n^*, 1-n^*)$. It follows from (11.26) that

$$w_s^* - X = w_u^*,$$

but that means that $n^*$ is precisely our earlier distinguished point $n_2$! We already know that $n_2$ is above the highest possible steady state, so that gives us two more pieces of information from Figure 11.7:

(iv) Every steady state is strictly smaller than $n_2 = n^*$ and so is inefficient. Moreover, as we move from left to right over steady states in Figure 11.7, efficiency (or net output) rises, while at the same time the wage differential falls.

That net output rises as we move toward $n^*$ from below follows from eyeballing Figure 11.6. That the wage differential falls at the same time can be seen by studying Figure 11.7. Indeed, it is possible to show that the Lorenz curve of net consumptions in steady state is strictly more equal as we move left to right across the set of steady states. In short, inequality is negatively correlated with economic efficiency, a point that we could not make in the labor market model, because the inequality there was taken to be exogenous. Here, it is just as much an outcome, jointly determined with the skill share in the population.

We make a final remark on the human capital model, which is that our labor market model is closer to it than appears at first glance. To see this, reinterpret "skilled labor" as "entrepreneur" and "unskilled labor" as "worker," and let $X = S(1+r)$ stand for the setup cost including interest. Then we have exactly our entrepreneurial model, with the additional constraint that all such firms are family enterprises funded by parental money, and there are no credit markets at all. But of course, the model can be easily extended to incorporate the sort of limited access to credit that we have in the labor market model.[134]

[134]To make the argument complete, we must ask, what is the "production function" $F(s,u)$ for aggregate output in the entrepreneurship model? The answer is that $F(s,u) = sf\left(\frac{u}{s}\right)$. Notice that $F_2(s,u) = f'\left(\frac{u}{s}\right)$, which is precisely the wage in the entrepreneurship model, while $F_1(s,u) = f\left(\frac{u}{s}\right) - \frac{u}{s}f'\left(\frac{u}{s}\right) = f\left(\frac{u}{s}\right) - \frac{u}{s}w$, which is the profit to an entrepreneur, or her "wage."