NH Handbook of Income Distribution, volume 2B
A.B. Atkinson and F.J. Bourguignon (Eds.) Chapter 15

Inequality in Macroeconomics *

Vincenzo Quadrini
University of Southern California

José-Víctor Ríos-Rull
University of Minnesota
Federal Reserve Bank of Minneapolis
CAERP, CEPR, NBER

July 17, 2014

Abstract

We revise some of the main ways in which the study of aggregate performance of an economy overlaps with the study of inequality.

*We are grateful for the comments from many people over the years. Many people contributed direct comments about and calculations for this paper; They include the editors for this handbook, but also Makoto Nakajima, Mariacristina De Nardi, Josep Pijoan and Thomas Piketty. Other, such as David Wiczer and Moritz Kuhn, made even more direct contributions. We got invaluable research assistance from Kai Ding and Gero Dolfus, and comments by Sergio Salgado and Annaliina Soikkanen and editorial help from Joan Gieseke. We thank all of them. Ríos-Rull thanks the National Science Fundation for grant SES-1156228. The views expressed herein ore those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

Contents

In a handbook devoted to income distribution, a chapter devoted to macroeconomics should start by clarifying the role of macroeconomics. Two of the main concerns about macroeconomics are aggregation and general equilibrium. The first ensures that the various sections of the economy aggregate; that is, by adding the incomes, wealth, and other variables of all households, we obtain the economy-wide value of these variables. The second macroeconomic concern is general equilibrium, that is, how changes in any section of the economy propagate to other sections of the economy via implied adjustments in prices and tax rates that are necessary to clear markets and balance the government's budget constraint. Although a large body of macroeconomic research abstracts from distributional considerations among individuals and households, a significant strand of studies are also concerned about the interaction between distribution and aggregate outcomes. In this chapter we will explore the possible interactions between distribution and the aggregate dynamics of the economy.

Since Bertola [2000] (the macroeconomics chapter in Volume 1 of this handbook series), many changes have taken place in the way macroeconomists deal with income inequality. One important change is that interest has shifted ways from the study of the relation between inequality and long-term growth and has focused more on other aspects of macroeconomic performance. Perhaps the main reason for this change is that, in general, there is less concern about long-term growth. Now, the more popular view is that all advanced economies grow by about 2% annually. The main question is, what does it take for less developed countries to accelerate the process of development and join the group of rich countries? As a result of these changes, macroeconomists have two main concerns with regard to inequality. One is what determines the joint distribution of earnings (or labor income) and wealth, and the other is how the explicit account of empirically sound inequality shapes the answers to the standard questions in macroeconomics. The models typically used feature a large number of agents that differ in earnings, wealth, and, in some cases, other characteristics. Consequently, we find it convenient to separate the two main branches of macroeconomic studies in this vein: the branch primarily interested in understanding the sources or causes of inequality and the branch concerned with the consequences of inequality for the aggregate performance of the economy. Such a distinction is not always applicable, yet it provides a natural organization of the literature. We will also find it occasionally convenient to separate studies that are primarily interested in economic growth from those focusing on business cycles.

The outline of the chapter is as follows. We start in Section 1 with some facts on the United States income and wealth distribution that are relevant from a macroeconomic point of view. We look at both cross-sectional evidence and the changes observed in the last few decades. Although some of these facts are analyzed in more detail in other chapters of this handbook (for example, chapters 8, 9, and 10), it will be useful to summarize them here since they are the reference for some of the theories we will review in this chapter.

After summarizing the main empirical facts about income and wealth distribution, we take a look in Section 2 at how macroeconomists make sense of these facts. First, we show how the distribution of wealth is determined given an exogenous process for earnings. After reviewing the models used by macroeconomists to examine this question and their success in replicating the wealth distribution observed in the data, we turn to models of endogenous determination of earnings. We look at models of human capital investment to determine why some people are more successful than others, in the sense of earning higher labor incomes. Thus, we will look at earnings inequality not as a purely stochastic process (luck) but as an outcome of different mechanisms such as investment in human capital (for example, education) or the higher relative demand of certain skills (affecting the relative prices of certain skills compared with other skills). The section concludes with a look at how occupational choices can also determine labor earnings. While this section is concerned with inequality as a permanent or steady-state phenomenon, the occupational choice part is informed by the bad employment performance of the Great Recession so it includes business cycle aspects.

Next we turn to the dynamics of inequality. Section 3 studies how inequality may change both over the business cycle and over a longer horizon. Here we consider a simple model in which factor shares—of capital and labor—can change.

Section 4 deals with what is possibly one of the most exciting ways in which macroeconomics and inequality interact: the role of financial markets or, more specifically, financial frictions. We start by looking at how the ability to borrow shapes the income and wealth distribution (and the allocation efficiency) by reallocating investment funds to entrepreneurs that are efficient and reliable, but not always both. We then turn to how wealth inequality is shaped by borrowing ability even when the rate of return of savings is equated across households. First, we look at how the sheer ability to borrow shapes inequality, and then we consider endogenous theories of borrowing where financial frictions arise from the institutional environment. We also look at various extensions of these ideas where the frictions are endogenous. In addition to exploring the effects of financial frictions on inequality, we look at the long-term effects on the performance of the economy, including some issues that have become of concern to macroeconomists, such as implications for global imbalances.

In Section 5 we analyze how the political system interacts with inequality to yield different policies that have an impact on the aggregate performance of the economy. People have different views about the desirability of alternative economic policies that depend on the position of individuals in the economy-wide distribution of income. The aggregation of individual preferences leads to the choice of particular policies. As the distribution of income changes, so does the choice of policies, which in turn affect the aggregate performance of the economy.

Section 6 concludes the chapter with a global assessment of what may be behind some of the changes observed in the last few decades.

Finally a note of caution and a disclaimer. Throughout the chapter we make use of various theoretical models that, for expositional purpose, are kept simple. Although this makes the intuitions of the basic mechanisms easy to understand, it also implies that these models may not be completely suited to address quantitative questions. Therefore, even if we often illustrate the properties of the model quantitatively, we should be careful in interpreting the simulation numbers since they are often intended to provide a qualitative, rather than quantitative, assessment of the model. The disclaimer is about the necessary incompleteness of this chapter. As much as we have tried to provide a general presentation of the studies that deal with inequality in macroeconomics, covering all possible subjects is impossible. There are many topics that we do not review. For example, we exclude studies that introduce behavioral elements in macroeconomic analysis. In part, this is motivated by our limited expertise in these subjects. We have also avoided for the most part the study of inequality in developing countries. We have marginally touched issues like the impact of the rise in inequality on the U.S. economy, the macroeconomic causes for the rising inequality, and globalization and inequality among others. Perhaps, more importantly, we have only scratched the surface of how income inequality translates into consumption inequality which is what most economists think that is what really matters. The topic of income inequality is quite vast and different authors would write it quite differently, in fact, Thomas Piketty's chapter in this Volume includes some macro modelling of the wealth distribution with a very different flavor than this chapter.

1 Some facts on the income and wealth distribution

Here we outline some general features of the Lorenz curves for earnings and wealth and their correlation and persistence over a medium span of 5 to 10 years for both individuals and across generations. We draw data from Díaz-Giménez et al. [2011], Kuhn [2014], and Budría et al. [2001] for the United States.[1] The data come from the Survey of Consumer Finances (SCF).

[1] In Krueger et al. [2010] various macroeconomists study inequality in a variety of other countries with a similar way of looking at the data as we do in this chapter.

Although the facts are for the U.S. economy, they may apply in variying degrees to other countries. In general, the United States is a more extreme version of the other developed countries in the sense that it is characterized by higher inequality.

TABLE 1
Distribution of Earnings and Net Worth in the U.S. Economy Source Kuhn (2014)

	Bottom (%)			*Quintiles*					*Top (%)*			*All*
	0–1	1–5	5–10	1st	2nd	3rd	4th	5th	90–95	95–99	99–100	0–100
Shares of Total Sample sorted by earnings (%)												
Earnings	-0.1	0	0	-0.1	3.5	11	20.6	65.0	12.1	18.3	18.0	100
Income	0.8	0.4	0.9	6.5	8.5	10.5	17.8	56.7	10.3	16.5	15.9	100
Net worth	4.5	0.5	0.7	11.6	13	6.3	9.9	59.1	8.6	21.2	20.2	100
Shares of Total Sample sorted by net worth (%)												
Earnings	0.9	2.8	2.3	8.4	10.7	14.6	17.5	48.8	10.3	15.7	10.8	100.0
Income	0.8	2.6	2.2	8.4	10.0	13.9	17.6	50.1	10.2	15.3	12.2	100.0
Net worth	-0.3	-0.3	-0.1	-0.7	0.7	3.3	10.0	86.7	13.5	26.8	34.1	100.0

Data are from the 2010 Survey of Consumer Finances. Income includes all transfers including food stamps. Earnings are defined as the part of income earned by labor. Farm and business incomes are assigned split between labor (93.4%) and capital (6.6%) (a much lower capital share than in other years.)

Table 1 shows the shares of earnings (the part of the income that can be attributed to labor), income (earnings plus capital income plus government transfers before taxes), and wealth (both financial and real assets, but not defined benefits pensions). See Díaz-Giménez et al. [2011] for details about the definition of these variables. As we can see, a large number of households have zero or negative earnings, almost two-thirds of all earnings come from the top quintile, and the top 1% receive almost 20% of all earnings. Our definition of earnings includes part of self-employment income that is imputed as labor income.[2] Since self-employment income can be negative, several households have zero or even negative earnings in our sample. Negative earnings contribute to a much higher Gini index compared to other measures provided in the literature, such as those that result from using earnings data from the Current Population Survey (CPS). However, correcting the CPS data by tax information could lead to Gini indexes that exceed 0.6 (see Alvaredo [2011]) and the discussion in chapter 9). Wealth is more concentrated than income, and the poorest quintile holds negative wealth. Furthermore, more than 85% of the wealth is held by the richest quintile and more than one-third of all wealth by the richest 1%. Table 2 shows a few measures

[2]In fact, due to the recession, the overall fraction of business income attributed to labor is larger in 2010 (93.4%) and in earlier waves of the SCF when it was around 85%. See Díaz-Giménez et al. [2011] for details on how to calculate such fraction.

of dispersion that are useful to keep in mind.

TABLE 2
Concentration and Skewness of the Distributions for 2010. Source: Kuhn [2014].

	Earnings	*Income*	*Wealth*
Coefficient of Variation	3.26	3.45	6.35
Variance of the Logs	1.41	0.92	4.65
Gini index	0.65	0.55	0.85
Top 1%/Lowest 40%	210	67	47,534
Location of Mean (%)	70	73	83
Mean/Median	1.85	1.70	6.42

The properties of the distribution of earnings, income, and wealth (total (Wea) and excluding housing (N-H-W)) have changed in the last few years. Tables 3, 4, and 5 show the values of a few measures of concentration for 1998, 2007 and 2010. For earnings, the Ginis, the Coefficient of Variations, the various ratios involving the median and the shares of top groups have all increased, most of them monotonically. For income the picture presented by the SCF is muddier. The Gini seems unchanged, with some measures indicating an increase in inequality and others a decrease. The same seems to have happened for wealth. While the Gini, the ratio of the 90th percentile to the median and of the mean to the median have all gone up, the shares of the top 10%, 1% and 0.1% have either remained stable or have gone down.

TABLE 3
Changes in Concentration. Source: Kuhn [2014].

	Ginis				Coeff. of Variation			
	Ear	Inc	Wea	N-H-W	Ear	Inc	Wea	N-H-W
1998	0.61	0.55	0.80	0.86	2.86	3.56	6.47	7.93
2007	0.64	0.57	0.82	0.88	3.60	4.32	6.01	7.59
2010	0.65	0.55	0.85	0.89	3.26	3.45	6.35	7.70

The modest evidence for an increase in inequality in income and wealth contrasts drastically with the picture reported by Piketty and Saez [2003], Piketty [2014], and Saez and Zucman [2014] who have used tax data. They have documented a big increase in inequality in the last in both income and wealth. Their evidence for income is direct as it comes straight from tax returns. The evidence for the evolution of wealth concentration in Saez and Zucman [2014] is indirect, it imputates the value of the assets that generate the reported capital income using the capitalization method with the rates of return for each class of assets that are obtained in the Flow of Funds. Yet it is quite persuasive. They suggest that the discrepancies that result between using the SCF and using tax data is due mostly to the top 0.1% of the top wealth holders. The SCF excludes the richest 400 households (the Forbes 400), and it is quite possible that even within income strata response rates to the SCF voluntary questionnaire vary by income. The two sets of data are complementary and the SCF is working hard to improve how it represents the very richest. There is likely to be a major update of the SCF that tries hard to improve in these dimensions. Hopefully, such improvements in the SCF will be available in the next few months and we will get a much better picture of the characteristics of the very rich.

TABLE 4
Changes of relevant ratios involving the Medians. Source Kuhn [2014].

	Median to 30th percentile				90th percentile to Median				Mean to Median			
	Ear	Inc	Wea	N-H-W	Ear	Inc	Wea	N-H-W	Ear	Inc	Wea	N-H-W
1998	2.80	1.71	4.00	4.54	3.18	2.87	6.88	12.56	1.57	1.62	3.95	7.66
2007	2.77	1.68	4.54	4.73	3.41	3.00	7.55	15.73	1.72	1.77	4.60	10.39
2010	3.30	1.64	5.24	4.11	3.79	3.10	12.37	23.33	1.85	1.71	6.42	13.18

Turning to consumption inequality, there are some doubts about whether it has also increased. Using data from the Consumer Expenditure Survey, Krueger and Perri [2006] have documented that consumption inequality has increased only slightly. However, Attanasio et al. [2004] and Aguiar and Bils [2011] claim that the increase in consumption inequality has been more significant. One of the reasons these studies reach different conclusions is because they use different survey data, Krueger and Perri [2006] use data in which consumption comes from survey collecting interviews, and Attanasio et al. [2004] use diary data. In addition, Aguiar and Bils [2011] have argued that the observed reduction in the quality of the consumption data in terms of how much of aggregate consumption it is recovered in the interviews is concentrated among goods that are mostly purchased by rich people and by among the high income groups and both features point to a larger increase in the underlying consumption inequality than what is obtained by using the data without especial adjustments.

TABLE 5
Changes in the Top Shares. Source: Kuhn [2014].

	Share of the Top 10%			Share of the Top 1%			Share of the Top 0.1%		
	Ear	Inc	Wea	Ear	Inc	Wea	Ear	Inc	Wea
1998	43.5	16.1	1.7	42.8	17.4	6.1	68.6	33.9	12.5
2007	47.0	18.7	1.9	46.9	21.0	7.8	71.4	33.6	12.4
2010	48.4	18.0	1.7	44.5	17.2	5.6	74.4	34.1	12.3

To summarize, over the last 10-20 years the evidence points to a sizable increase in inequality.

2 Modeling the sources of macro inequality

In this section we show what macroeconomics has to say about inequality. We start by exploring in Section 2.1 the implications of existing macro models for the distribution of wealth. Most of the models reviewed in this section start from the assumption that the process for earnings is exogenous: it can be stochastic, but it cannot be affected by individual decisions. In Section 2.1.3 we describe some of the theories in which the process for earnings is endogenous in the sense of being affected by individual choices. Since individual choices respond to policies, these models have interesting predictions about the impact of economic policies on the distribution of earnings.

2.1 Theories of wealth inequality given the process for earnings

We start this section by emphasizing the limits of the neoclassical growth model with infinitely lived agents and complete markets in predicting wealth inequality. After reviewing the prediction of the overlapping generations model, we analyze models with incomplete markets. As we will see, the consideration of market incompleteness allows for more precise predictions about the distribution of wealth for given processes of individual earnings.

2.1.1 The irrelevance of income and wealth inequality in the neoclassical model

The deterministic neoclassical growth model says very little about income and wealth inequality. Note that we mean the neoclassical growth model in its modern meaning of incorporating fully optimizing saving behavior.[3] In an important article by Chatterjee [1994], reiterated later by Caselli and Ventura [2000], it is shown that any initial distribution of wealth is essentially self-perpetuating. To see this, consider the typical problem of a household $i \in \{1, \cdots, \mathrm{I}\}$. Using recursive notation with primes denoting next period variables, the household's problem can be

[3] As such the standard analysis of Stiglitz [1969] does not apply, since there saving behavior is postulated not derived from first principles.

written as

$$
\begin{aligned}
v^i(a) &= \max_{c,a'} u^i(c) + \beta_i \, v^i(a'), && (1)\\
\text{subject to} \qquad & c + a' = a(1+r) + \epsilon_i w. && (2)
\end{aligned}
$$

Here, $u^i(c)$ is a standard utility function (differentiable, strictly concave), $\beta_i \in [0,1]$ is the discount factor, and ϵ_i is the household's endowment of efficient units of labor that we assume constant for now. The necessary condition for optimality is

$$
u_c^i(c_i) = \beta_i \, (1+r') \, u_c^i(c_i'), \tag{3}
$$

where $u_c^i(c_i)$ is the marginal utility of consumption. In a steady state, the allocation is constant over time, $c_i = c_i'$, and $r = r'$, which requires that the rate of return on savings is equal to the rate of time preference in every period, that is, $\beta_i = (1+r)^{-1}$. One implication is that, if households have interior first order conditions so that equation (3) is satisfied with equality, then $\beta_i = \beta$ for all i. Otherwise, some households would reduce their assets as much as they can until they reach some lower bound that depends on the borrowing ability.

Since the rate of return in the neoclassical growth model is given by the marginal productivity of capital, we have that

$$
\beta^{-1} = 1 + r = F_K(K,N) - \delta, \tag{4}
$$

where K is aggregate capital, $N = \sum_i \epsilon_i$ is the aggregate effective labor (hours worked weighted by their efficiency), F is the production function, and δ is the constant rate of capital depreciation. In the neoclassical growth model physical capital is the only form of wealth, so the following has to hold:

$$
K = \sum_i a_i, \tag{5}
$$

where a_i are the assets held by household i. Note that these last three equations are the only ones imposed by the theory. It turns out that any distribution of wealth $\{a_i\}_{i=1}^{I}$ that satisfies equations (4)-(5) is a steady state of this economy in which each individual household i consumes its income, $c_i = ar + \epsilon_i w$. This is the sense in which the theory poses no constraints whatsoever on the distribution of a. Note that this is true no matter how efficiency units of labor (and hence earnings) are distributed across households. Non-separability between consumption and leisure does not change this finding.

Small details qualify the behavior of the system outside a steady state. Under constant relative risk aversion (CRRA) preferences, equation (3) can be written as $\left(\frac{c_i'}{c_i}\right)^{\sigma} = \beta_i(1+r')$, where $1/\sigma$

is the intertemporal elasticity of substitution. Depending on the joint distribution of earnings and wealth, the evolution of the wealth distribution is dictated by this equation and the budget constraint.

What other possibilities does the neoclassical growth model or its variants offer? Not many. Consider heterogeneity in the per period utility function. We have already noted that this does not change any steady-state consideration. Outside the steady state, the model just takes the initial wealth distribution and uses the first order conditions and the budget constraints to propagate the wealth distribution into the future, essentially dispersing or concentrating the wealth distribution without much endogenous action on the part of the model.

What about stochastic versions of these economies? With complete markets, all idiosyncratic uncertainty disappears (it is insured away), whereas the aggregate uncertainty is borne by those who are more willing to bear it. If such ability to bear the risk is increasing in wealth, then the model could generate some redistribution in response to aggregate shocks. But abstracting from aggregate uncertainty, we will see that the irrelevance result no longer applies when markets are incomplete (and agents continue to face idiosyncratic shocks). Before exploring the implications of incomplete markets, however, we briefly review the overlapping generations model.

2.1.2 Overlapping generations models and wealth inequality

In overlapping generations models, new households are born every period and live up to a certain number of periods J (they may also die earlier with some probability).[4] In what follows, we abstract from differences among households in any given age cohort and assume that the heterogeneity is only between cohorts. Households in age cohort j have earnings ϵ_j, which we take as exogenous. This specification can accommodate retirement and, with some extra work, government-provided social security (see Section 2.2 for theories of the determination of age-

[4] Sometimes the literature uses the term "overlapping generations model" for environments in which new agents are born every period and die with some probability at any point in the future. We refer to this particular environment as the Blanchard-Yaari model (Blanchard [1985] and Yaari [1965]), which is very similar mathematically to the infinitely lived model.

specific earnings). In a steady state, households solve the following problem:

$$\max_{\{c_j, a_{j+1}\}_{j=1}^J} \quad \sum_{j=1}^{J} \beta_j \, u(c_j), \tag{6}$$

$$\text{subject to} \quad c_j + a_{j+1} = a_j \, (1 + r) + w \, \epsilon_j, \tag{7}$$

$$a_1 = 0, \tag{8}$$

$$a_{J+1} \geq 0. \tag{9}$$

Here, β_j is the specific weight that households place in the age-j utility. Note that households are born with no assets and cannot die with debts. Steady-state factor prices are r and w. The solution of the problem includes age-specific consumptions, c_j, and asset holdings, a_j, that satisfy the Euler equation

$$u_c(c_j) = \frac{\beta_{j+1}}{\beta_j} \, (1 + r) \, u_c(c_{j+1}). \tag{10}$$

Steady-state factor prices are equal to the marginal productivities of a neoclassical production function with respect to aggregate capital, $K = \sum_{j=1}^{J} A_j$, and labor, $N = \sum_{j=1}^{J} \epsilon_j$. We are using capital A_j to denote the assets of households of age j of which there are many, making it an aggregate variable which explains the use of capital letters.

When mapping these models to the data, we calibrate the earnings profile to have an inverse U shape as in the data (even after including social security payments). If we pose a constant discount rate, that is, we substitute β_j with β^j in equation (6), as most researchers do, the model generates wealth holdings with an inverse U shape that typically peaks a little beyond sixty years of age. From that point on, the model predicts a slow but certain depletion of assets until death. Since in equilibrium household wealth has to add up to capital, households have to save during their finite lifetime to accumulate the whole capital stock. Although the prediction of the overlapping generations model in terms of lifetime wealth is broadly consistent with the data, the prediction for lifetime consumption is not. The strong incentive to save, together with the Euler equation, implies that $c_{j+1} > c_j$ for all j. In the data, however, consumption is also hump shaped. Various approaches are proposed in the literature to get around this shortcoming. They include demographic shifters, non-separable leisure in the utility function (Auerbach and Kotlikoff [1987], Ríos-Rull [1996]), existence of both durable goods and incomplete financial markets (Fernandez-Villaverde and Krueger [2010]), borrowing constraints and low rates of return (Gourinchas and Parker [2002]), and others.

With stochastic mortality, the model produces identical predictions as long as there is a market for annuities (which are available even if scarcely used). To see why, consider the probability of surviving between ages j and $j+1$, which we denote by ϕ_j. The survival probability multiplies the discount factor $\frac{\beta_{j+1}}{\beta_j}$, capturing the fact that the household gets utility only if alive.[5] Fairly priced (i.e., issued at zero expected cost) annuities imply that households save by purchasing them, and one unit of savings today yields $\frac{1}{\phi_j}(1+r)$ units of the good tomorrow if the household survives and zero otherwise. Clearly, this asset dominates non-contingent investment of savings and the budget constraint (7) becomes

$$c_j + \frac{a_{j+1}}{\phi_j} = a_j\,(1+r) + w. \tag{11}$$

It can be verified that with these modifications to discounting and the budget constraint, we obtain the same first order conditions as in (10).

If we assume that there are no annuities, as in Hansen and Imrohoroglu [2008], we have to make some assumption about the allocation of the assets left by the deceased households. There are various options. One possibility is to assume that any household is like a pharaoh and assets are buried with their owners. The predictions of the model change a little relative to the basic model because there is now a smaller amount of total wealth due to the lower rate of return tilting the allocation toward young ages. Other options include the assumption that there is a 100% estate tax (with implications identical to that of the pharaoh model except for the use of public revenues) or that the assets of the deceased go to those in a certain age group. If the assets are distributed equally among the households of certain age groups, the wealth distribution will present a hike at the age at which households inherit, which is not a feature of the data. A more attractive alternative that has not been directly explored is to build direct links between a dead household and a randomly chosen younger household that inherits the assets. In this case, there will be limited within-cohort inequality that results from differences in the timing of the death and the wealth of the ancestors.

What about versions of overlapping generations economies with aggregate shocks? With aggregate shocks, even if there are markets for one-period-ahead state-contingent assets, there could be incomplete insurance because households that are not alive cannot insure each other. The answers depend first on the size of the shocks. For (small) business cycle type shocks, there are no great differences between the allocations implied by complete or incomplete markets. Ríos-Rull [1996] and Ríos-Rull [1994] find that the allocations are almost identical with and without typical business cycle shocks. Larger and persistent shocks are a different matter. For example,

[5] In this model, there is nothing that the household can do to affect survival, so the relative value of being alive or dead is irrelevant. With a CRRA utility function with more curvature than log, the utility is negative and, implicitly, our formulation seems to indicate that the household would rather be dead than alive.

Krueger and Kubler [2006] study the role of social security in reducing market incompleteness across generations and do not find large effects. Glover et al. [2011] study the redistributional implications of the (most recent) recession and find that the loss of output and consequent drop in the price of assets affect the old generations more than the young ones. The intuition for this result is that the recent crisis has been associated with large drops in asset prices, including housing, and old generations own more assets than the young.

If markets for the insurance of idiosyncratic risks are not present and households can save only by holding non-contingent assets, the situation changes dramatically and the model has very tight predictions. We will see this in the next section.

2.1.3 Stationary theories of earnings and wealth inequality

When households do not have access to insurance against shocks, the accumulation of riskless assets acts as a mechanism that allows households to smooth consumption—saving in good times when earnings are above the mean and dissaving in bad times. This means that in environments in which households are subject to uninsurable risks, those that have been lucky and have enjoyed good realizations of the shocks are wealthier than those that faced adverse realizations. This type of ex post inequality has been widely studied in models in which the risk was on endowments or earnings and agents could save only in the form of non-state-contingent assets. The basic theory was first developed in Bewley [1977] and the general equilibrium and quantitative properties were studied later by İmrohoroğlu [1989], Huggett [1993], and Aiyagari [1994]. These ideas have important applications such as those in Carroll [1997] and Gourinchas and Parker [2002].

Successive studies have extended these models to improve the ability to generate greater wealth inequality. Among these approaches are the addition of special earning risks (Castañeda et al. [2003]), entrepreneurial risks (Quadrini [2000], Cagetti and De Nardi [2006], Angeletos [2007], Buera [2009]), endogenous accumulation of human capital (Terajima [2006]), and stochastic discounting (Krusell and Smith Jr. [1998]). Since in these models inequality is endogenous, the degree of wealth concentration can be affected by policies. This opened the way to studies that investigate the importance of taxation policies for wealth inequality. Example are Díaz-Giménez and Pijoan-Mas [2011], Cagetti and De Nardi [2009] and Benhabib et al. [2011].

We start by reviewing how to pose the process for earnings (Section 2.1.3.1), and then we describe the main features of the Aiyagari model (Section 2.1.3.2).

2.1.3.1 Stochastic representation of earnings A large body of literature tries to provide a parsimonious representation of the stochastic processes for wages or earnings. This literature uses panel data to estimate a univariate process for labor income or earnings, sometimes at the level of individual earners and sometimes at the household level (which is more in line with the data used in Tables 1 and 2).[6] (See, for example, Guvenen [2009] or Guvenen and Kuruşçu [2010, 2012].)

One important feature to take into account, as we will see below, is that the most common data sets do not include the very rich. The Survey of Consumer Finances (SCF) is designed to provide a better picture of the rich but, unfortunately, it has no panel dimension and therefore cannot be used to separate individual effects from shocks and other interesting property that affects the most appropriate representation of earnings as a stochastic process. A comparison between the properties of the cross section in both data sets gives an idea of the differences in the sample. Recent work using either tax data (Atkinson et al. [2011] and DeBacker et al. [2011]) or social security data (Guvenen et al. [2012]) looks very promising in terms of including both the very top earners and information about the persistence of their earnings.

2.1.3.2 The Aiyagari (1994) Model Consider an economy populated by many, in fact a continuum, of infinitely lived agents that can be of finitely many types $i \in I$. They are subject to shocks that cannot be insured. Without loss of generality, we pose that there are finitely many possible realizations of the shock $m \in M$ and that it follows a Markov chain with (possibly type-specific) transition matrix $\Gamma^i_{m,m'}$. For compactness, we write Γ to denote a block diagonal matrix in which each block is $\Gamma^i_{m,m'}$. For the most part, the shock refers to the agents' endowment of efficiency units of labor, so we denote the shock by $s \in S = \{s^1, s^2, \cdots, s^M\}$.

Households do not care for leisure and assess consumption streams through a per period utility function $u^i(c)$ with intertemporal discount factor β_i. The utility function and the discounting may be type specific.

We start by considering the most primitive financial structure in which households have access to saving only in one-period noncontingent assets, and they cannot borrow. To map the model to a real economy and to consider its empirical implications, we build the model economy on top of a neoclassical growth model with exogenous labor supply. With a Cobb-Douglas production function, the prices of capital (rental rate of capital) and labor (wage) depend only on the capital-labor ratio. Since the aggregate labor supply is constant, prices depend only on aggregate capital

[6] The process could also be for wages, but given the low variability of individual variance of hours worked for primary earners, wages and earnings have similar properties.

K, and we can express them as $r(K)$ and $w(K)$.

We consider only steady-state equilibria in which households face a constant interest rate r and a constant wage w per efficiency unit of labor. This approach is common in these types of studies because it greatly simplifies the computational burden. In fact, by focusing on steady states, when we solve the individual problem we can ignore the evolution of the aggregate states and we only need to keep track of the individual states (household's type i, asset position a, and realization of the idiosyncratic shock m). Of course, by doing so we have to exclude from the analysis changes that affect the whole economy such as aggregate productivity shocks or structural changes. The consideration of aggregate and recurrent shocks represents a major computational complication (see, for example, Krusell and Smith Jr. [1998]). However, if we restrict ourselves to explore the implications of a one-time completely unexpected shock (a somewhat oxymoronic term) or structural change, then the computation remains tractable. For simplicity of exposition, we limit the analysis here to steady-state comparisons with the caveat that in the real economy, the distribution will take a long time to converge to a new steady state.

The household's problem can be written as

$$v^i(m, a; K) \;=\; \max_{c,a'} \; u^i(c) + \beta_i \sum_{m'} \Gamma^i_{m,m'} \, v^i(m', a'; K), \qquad \text{s.t.} \tag{12}$$

$$c + a' = w \, s^m + a(1 + r), \tag{13}$$

$$a' \geq 0, \tag{14}$$

where the superscript i denotes the household's type. Since this does not vary over time, we wrote it outside the arguments of the value function. The first order condition is given by

$$u^i_c(c) \geq \beta_i \, (1 + r) \sum_{m'} \Gamma_{m,m'} \, u^i_c(c'), \qquad \text{with equality if } a'_i > 0. \tag{15}$$

Standard results show that this problem is well behaved and the solution is given by a function $a'_i(m, a; K)$. Moreover, when $\beta_i^{-1} < (1 + r)$, it is easy to show that for all i and m, there is a level of wealth $\overline{a}$ such that $a'_i(m, \overline{a}) < \overline{a}$. This means that there is a maximum level of wealth accumulated by an individual household. Thus, the set of possible asset holdings is the compact set $\mathcal{A} = [0, \overline{a}]$.

To describe the economy, we could use a list of households with their types, shocks, and assets along with their names, but it is easier to use a measure x. This measure tells us how agents have certain characteristics in the space (i, m, a). Then, aggregate capital, which is just the sum of the assets of all households, can be written as

$$K = \int a \ \ dx. \tag{16}$$

The measure x gives us all the information that we need. For example, the total amount of efficiency units of labor or aggregate labor input is equal to

$$N = \int s^m \ \ dx, \tag{17}$$

and the variances of both wealth and earnings are

$$\sigma_k^2 = \int (a - K)^2 \ dx, \tag{18}$$

$$\sigma_N^2 = \int (s^m - N)^2 \ dx. \tag{19}$$

To calculate the Gini index for wealth, we need to compute the Lorenz curve and then calculate its integral. Note that any point of the Lorenez curve, for example, its value at 0.99 denoted by $\ell_{0.99}$, is one minus the share of wealth held by the richest 1%. To compute $\ell_{0.99}$, we start by finding the threshold of wealth that separates the richest 1% from the rest of households. Once we have found the threshold, we compute the wealth held by those households with wealth above the threshold relative to total wealth. The Gini index is simply twice the area between the Lorenz curve and the triangle below the diagonal between zero and 1. (See Figure 4, for instance.) Other inequality statistics are also readily obtained from x, including those pertaining to the joint distribution of earnings and wealth and their intertemporal persistence.

The Aiyagari model has unique predictions about wealth and income inequality for any specification of the process of earnings. Therefore, the determination of the properties of the earning process becomes the central issue in the application of this model to the data. Should we think of people as being all ex ante equal in the sense that there is only one i type and they differ only in the realization of the shock? Or should we think of people consisting of different ex ante types? In either case, how do we determine which process to use? We now turn to this issue.

As we have seen in Section 1, the distribution of wealth in the United States is highly skewed, with about one-third of all the wealth in the hands of a mere 1% of households. How did those households become so wealthy? In order to become rich, households need both motive and opportunity. The reason for the opportunity is clear: at some point, the households had to have high enough earnings to be able to save and accumulate high levels of wealth. Motive is also important: why should households save rather than consume if they are impatient? If high earnings are not going to be around forever, then prudent households would want to save for the bad times that are likely to lie ahead. The issue is whether motives and opportunities are big enough to generate the wealth concentration observed in the United States.

To choose the actual parameterization of the earnings process for the model, we first need a Markovian process for earnings. If the focus is on the U.S. economy, one possibility is to specify a process estimated using the Panel Study of Income Dynamics (PSID). This is what Aiyagari [1994] did in his seminal paper, which relied on existing empirical studies such as Abowd and Card [1987] and Heaton and Lucas [1996].

The results are disappointing. The first row of Table 6 shows the shares of wealth of key groups in the U.S. data, and the second row displays those same shares as predicted by a model in which the earning process is calibrated using PSID data. The red line in Figure 4a shows the associated Lorenz curve. There is very little inequality compared with the inequality observed in the U.S. economy. The shares of wealth of the top 1% and the top quintile generated by the model are 4% and 27%, respectively, whereas in the United States these shares are 34% and 87%. The use of alternative estimates of the earning process, such as those provided by Storesletten et al. [2001], improves the performance of the model but only marginally.

Table 6

Concentration of Wealth of Various Economies (U.S. data Source Kuhn (2014))

Economy	1st	2nd	3rd	4th	5th	Top 10%	Top 5%	Top 1%	Gini
2010 U.S. Data	-0.7	0.7	3.3	10.0	86.7	74.4	60.9	34.1	0.85
PSID Ear-Pers	3.7	10.1	17.0	25.1	44.1	26.7	15.5	4.0	0.41
SCF Ear-Wea	0.0	0.0	0.3	4.5	95.2	78.3	53.2	14.7	0.87
Emp-Unemp	10.4	16.2	19.6	23.4	30.4	16.8	9.1	2.3	0.19
Stochastic β	1.7	6.5	12.5	21.1	59.3	40.3	25.3	6.7	0.56

Aiyagari [1994] used the PSID Ear-Pers calibration, Castañeda et al. [2003] used the SCF Ear-Wea calibration, and Krusell and Smith Jr. [1998] used both the Emp-Unemp and the Stochastic β calibrations.

The model fails to replicate the high concentration of wealth observed in the data for many possible reasons. One obvious explanation is that the model misses important pieces; for example, the model ignores life cycle heterogeneity with all the demographic complications of actual lives. Or it ignores the permanent characteristics of people as well as education and human capital acquisition. It also ignores the fact that lives are affected by many other types of shocks such as health or unforseen expenditures.

Castañeda et al. [2003] take a different approach and argue that the reason for the failure is the misrepresentation of the process of earnings. The PSID sample does not include very rich households. A comparison of its data with that of the SCF in which the emphasis is on wealthy people shows a large mismatch. The PSID does a better job at including people outside the top 10% of income earners and asset holders, but it is not appropriate to capture the dynamic properties of the incomes earned by the top of the distribution. Based on this observation, Castañeda et al. [2003] propose to ignore the PSID and focus instead in the specification of a process for earnings where the cross-sectional dispersion is similar to the SCF but its persistence is engineered so that it replicates the main features of the wealth inequality.

The third row of Table 6 displays the wealth distribution of an economy where the earning process has been calibrated following the above criteria. As we can see from the table, the model replicates quite well (by construction) the empirical data. Comparing the two processes for earnings (at least in the parsimonious representation used in Díaz et al. [2003]) is very useful. The version of their economy designed to replicate the properties of the original Aiyagari economy has three values for earnings that are essentially symmetric, as are the persistence properties of the process: the earnings of agents in the middle and top thirds are 1.28 times and 1.63 times, respectively, the earnings of those of the bottom third. Households in both, the top and the bottom thirds, have a one-third probability of moving out of their current situation by the next period. This society is very equal. Moreover, encountering bad luck, that is, being sent to the bottom third, is not really that bad. Clearly, because there are few motives to save money in this society, households soon stop doing so and consume all of their income.

The process that replicates the U.S. wealth distribution is extremely different from the process just described. The bottom of the society is now almost *half* of all households. Moreover, once at the bottom, less than 1% of these households move up each year. Households that make up the middle class, almost one-half of the population, earn 5 times more than those at the bottom and have an equal chance (1%) of moving up or down. Only 6% of the households are in the top earnings group, and their earnings are huge: 47 times of those of the poor and 9 times those of the middle class. More than 8% of these households will move down in a given year. Although these particular values are somewhat arbitrary, they give an accurate sense of how extreme both motives and apportunities have to be in order to induce the U.S. wealth disparities in a model in which agents differ only in the realization of a common earnings process.

Krusell and Smith Jr. [1998] pursue a very different approach to get a suitable wealth distribution. Instead of tracking the behavior of earnings, they pose a simple employment/unemployment process to generate earnings inequality, and they assume that the discount rate of individual agents is also stochastic. Therefore, in addition to the idiosyncratic shock to earnings, they consider a second idiosyncratic shock to the discount rate. The earnings process alone generates almost no inequality because the only way to get rich is through remaining employed but without earning more than other employed workers. In the extension with stochastic discounting, they assume that β can take three values, $\{0.9858, 0.9894, 0.9930\}$, with a symmetric distribution that satisfies the following properties: (i) the average duration of the extremes is 50 years (so it lasts the length of the adult life of a person); (ii) the transition from extreme to extreme requires a spell in the middle; and (iii) the (stationary) size of the middle group is 80%. Interestingly, the model with stochastic discounting yields disarmingly similar inequality indexes as in the data (see the last row of Table 6).

Various papers used life-cycle models with idiosyncratic risk to study wealth inequality. An important early contribution is Huggett [1996]. De Nardi [2004], Cagetti and De Nardi [2006], and Cagetti and De Nardi [2009] study the role of bequest, estate taxation and entrepreneurship in shaping the wealth distribution.

2.2 Theories of earnings inequality

So far we have described how macroeconomists think of the distribution of wealth *given* the distribution of earnings. But what about the distribution of earnings itself? Where is it coming from? In general, we can think of the differences in earnings as resulting from a combination of heterogeneity in (i) innate abilities or ex ante luck that persists for the lifetime of the agent; (ii) ex post luck due to the realization of shocks that are not under the control of the agent; (iii) effort or occupational choice; and (iv) investment in human capital. In the model considered in the previous section, the heterogeneity in earnings was only a consequence of innate heterogeneity (captured by the agents' type) and ex post luck (captured by the Markov process for skills). In this section we make the earnings endogenous by allowing for the optimal choice of effort and investment in human capital. An important consequence of endogenizing the earning process is that the distribution of earnings can be affected by several factors including financial market development (which, for example, facilitates access to the financing of investment in human capital) and taxation policies (which, for example, affect the marginal decision of effort and investment in human capital).

Next we briefly describe three aspects of models with endogenous earnings: models in which earnings are the result of explicit choices of either learning by doing or learning by not doing, including education (Section 2.2.1); models in which not all types of labor are perfect substitutes

(Section 2.2.3); and models with occupational choices in which agents decide which occupation to take (Section 2.2.5).

2.2.1 Human capital investments

A common approach to endogenizing the process for earnings is the assumption that human capital is endogenous and depends on the individual investment chosen by agents. If the return from the investment is stochastic, then agents will be characterized ex post by different levels of human capital and, therefore, unequal earnings. An interesting feature of this set up is that it generates a positive relation between the aggregate performance of the economy and the degree of inequality. More specifically, higher investment in human capital leads to higher income or growth or both, but also to higher inequality because investment amplifies the impact of idiosyncratic shocks.

We illustrate this point with a simple model without taking a stand on the issue of what type of investment yields higher human capital. The investment can be either the result of time and hence forgone output or leisure, or the result of effort that generates a disutility, or the result of investment in goods. In this sense and at this level of abstraction, it accommodates both learning by direct investment in schooling or more general human capital investment such as that pioneered by Ben-Porath [1967]. It also captures the key mechanisms formalized in more recent studies such as Guvenen et al. [2009], Manuelli and Seshadri [2010], and Huggett et al. [2011]. For an extensive discussion of the life cycle human capital model see von Weizsàcker [1993].

Consider an economy with a continuum of risk-neutral workers, each characterized by human capital h. Production, which in this simple model corresponds to earnings, is equal to the worker's human capital h. Individual human capital can be enhanced with investment captured by the variable y. Investment is costly in terms of either utility or output (given risk neutrality, they are essentially the same). We assume that the cost take the form $\frac{\alpha y^2 h}{2}$ and human capital evolves according to

$$h' = h(1 + y\varepsilon'),$$

where ε' is an i.i.d. random variable with $\mathbb{E}\varepsilon' = \bar{\varepsilon}$. To simplify notation, we denote by $g(y, \varepsilon') = 1 + y\varepsilon'$ the gross growth rate of human capital.

Since the outcome of the investment is stochastic, the model generates a complex distribution of human capital among workers. In the long run, the distribution will be degenerate since at the individual level h follows a random walk. To make the distribution stationary and keep the model simple, we assume that workers die with probability λ in each period and are replaced by the same mass of newborn workers. To allow for ex ante heterogeneity or innate abilities, we also assume that newborn agents are heterogeneous in initial human capital. In particular, there are I types of newborn agents indexed by $i \in \{1, ..., \mathrm{I}\}$, each of size x_0^i and with initial human capital h_0^i. The initial distribution of newborn agents satisfies $\sum_i x_0^i = \lambda$.

Because of the linearity assumption, it will be convenient to normalize by h the optimization problem solved by a worker. We can then write the problem recursively as

$$\omega = \max_y \left\{ 1 - \frac{\alpha y^2}{2} + \beta(1-\lambda)\mathbb{E}\left[g(y, \varepsilon')\, \omega' \right] \right\}, \tag{20}$$

where ω is the expected lifetime utility normalized by human capital h. The non-normalized lifetime utility is ωh. Of course, the linearity of the accumulation function is crucial here. If the new human capital was a Cobb-Douglas function of old human capital, as in the Ben-Porath [1967] model, the analysis would be more complex analytically.

The first order condition gives

$$\alpha\, y = \beta\, (1-\lambda)\, \bar{\varepsilon}\, \omega', \tag{21}$$

where $\bar{\varepsilon}$ is the average value of the stochastic variable ε.

Since the first order condition is independent of h, the investment variable y is constant over time, which in turn implies that the normalized lifetime utility for the worker, ω, is constant. Therefore, y and ω can be determined by the two equations that define the value for the worker and the optimal investment, that is,

$$\omega = 1 - \frac{\alpha y^2}{2} + \beta(1-\lambda)(1 + y\, \bar{\varepsilon})\, \omega, \tag{22}$$

$$\alpha y = \beta(1-\lambda)\, \omega\, \bar{\varepsilon}. \tag{23}$$

Given the distribution of human capital for newborn workers x_0 and the investment variable y, we can determine the economy-wide distribution of human capital (equal to the distribution of earnings) and compute a cross-sectional index of inequality. We focus on the square of the coefficient of variation, that is,

$$\text{Inequality index} \equiv \frac{\mathrm{Var}(h)}{\mathrm{Ave}(h)^2},$$

which can be calculated exactly in a steady-state equilibrium.

Before we do this, note that the mass or measure of agents of age $j+1$ is given by $\sum_i x_j^i = \sum_i x_0^i \, (1-\lambda)^j$ and the average human capital is equal to

$$\text{Ave}(h) = \sum_i \; x_0^i \; \sum_{j=0}^{\infty} (1-\lambda)^j \; \mathbb{E}_j \; h_j^i. \tag{24}$$

The index j denotes the age of the worker and i the cohort of newly born agents with human capital h_0^i. The population size of newborn agents of type i is x_0^i, and the total mass of newborn agents is $\sum_i x_0^i = \lambda$. Since workers survive with probability $1-\lambda$, the fraction who are still alive after j periods is $(1-\lambda)^j$.

The cross-sectional variance of h is calculated using the formula

$$\text{Var}(h) = \sum_i x_0^i \sum_{j=0}^{\infty} (1-\lambda)^j \;\; \mathbb{E}_j \left[h_j^i - \text{Ave}(h) \right]^2, \tag{25}$$

which has an interpretation similar to the formula used to compute the average h. Of course, for the variance to be finite, we have to impose some parameter restrictions. In particular, we need to impose that the death probability λ is sufficiently large and the return on human capital accumulation $\mathbb{E}[\varepsilon']$ is not too big.

Using (24) and (25), Appendix A shows that the average human capital and the inequality index take the forms

$$\text{Ave}(h) \;\; = \;\; \frac{\lambda \; \bar{h}_0}{1-(1-\lambda) \; \mathbb{E} \; [g(y,\varepsilon)]}, \tag{26}$$

$$\text{Inequality index} \;\; = \;\; \left[\frac{\sum_i x_0^i (h_0^i)^2}{\bar{h}_0^2} \right] \frac{\left[1-(1-\lambda)\,\mathbb{E}g(y,\varepsilon)\right]^2}{1-(1-\lambda)\,\mathbb{E}g(y,\varepsilon)^2} - 1, \tag{27}$$

where $\bar{h}_0$ is the aggregate human capital of newborn agents.

We can see from equation (26) that the average human capital and, therefore, aggregate output are strictly increasing in the investment variable y. This is intuitive given the structure of the model.

As far as the inequality index is concerned, equation (27) shows that this results from the product of two terms. The first term in parentheses captures the ex ante inequality, that is, the distribution of human capital at birth. If all agents are born with the same human capital, this term is 1. However, if the initial endowment is heterogeneous (heterogeneity in innate abilities), then this term is bigger than 1. The second term in parentheses captures the inequality generated by investment. It is easy to show that this term, and therefore, the inequality index, are strictly increasing in y. Since the average value of h is also strictly increasing in y, we have established that there is a positive relation between macroeconomic performance and inequality. The intuition for this dependence is simple. If $y = 0$, human capital for all workers will be equal to h_0^i and the inequality index is fully determined by the ex ante heterogeneity. As y becomes positive, inequality increases for two reasons. First, since the growth rate $g(y, \varepsilon)$ is stochastic, human capital will differ *within* the same age-cohort of workers. Second, since each age-cohort experiences growth, the average human capital will differ *between* different age-cohorts.[7] Both mechanisms are amplified by the growth rate of human capital, which increases in the investment y.

Using this model, we can analyze how changes that have an impact on the incentives to invest in human capital affect macroeconomic performance and inequality simultaneously. An example is a change in income taxes.

Suppose that the government taxes income at rate τ. The equilibrium conditions (22) and (23) become

$$\omega = 1 - \tau - \frac{\alpha y^2}{2} + \beta(1-\lambda)\,(1 + y\,\bar{\varepsilon})\omega, \tag{28}$$

$$\alpha y = \beta(1-\lambda)\,\omega\,\bar{\varepsilon}. \tag{29}$$

A bit of algebra shows that y is strictly decreasing in τ. Effectively, the tax reduces the value of human capital, ω, which in turn must be associated with a reduction in y (see equation (29)). Then, we can see from equations (26) and (27) that higher taxes reduce inequality but also reduce the average human capital. This mechanism captures, in stylized form, the idea of Guvenen et al. [2009] used to explain cross-country wage inequality. They argue that higher taxation of labor accounts for the wage compression and lower productivity in Europe relative to United States. Note also that the effects of higher labor taxation in the short run would differ from those in the long run in environments like this. In this particular model, taxation has no short-run disincentive effects (they would exist, however, if leisure were valued). Taxation does have long-run effects because agents would invest less in human capital. Empirical studies based only on short-run data would miss these effects.

[7] This is in addition to the differences in initial human capital among the I types of newborn agents.

2.2.2 Human capital investment versus learning by doing

The stylized model considered in the previous section can easily be extended to include learning by doing. To do so, we can simply interpret the variable y as the fraction of time spent investing (in human capital) and $1-y$ the fraction of time spent producing. Output is produced according to the function $h(1-\alpha y^2/2)$, which is strictly decreasing and concave in the time spent investing. The equation determining the evolution of human capital becomes

$$h' = h(1 + y\varepsilon') + \chi(1 - y).$$

The first term captures the time spent investing, while the second results from learning by producing. The analysis conducted so far extends trivially to this case. In particular, the two equations (22) and (23) become

$$\begin{aligned} \omega &= 1 - \frac{\alpha y^2}{2} + \beta(1-\lambda)\Big[(1 + y\bar{\varepsilon}) + \chi(1-y)\Big]\omega, \qquad (30) \\ \alpha y &= \beta(1-\lambda)\omega(\bar{\varepsilon} - \chi). \qquad (31) \end{aligned}$$

A further extension is to assume that the return from learning by doing is stochastic, that is, χ is a stochastic variable. Also, we could consider the special case in which the evolution of human capital is determined only by learning by doing. This case is obtained by setting $\varepsilon = 0$. These extensions do not change the basic properties of the model illustrated in the previous subsection, including the analysis of the short- and long-run effects of labor income taxation.

2.2.3 Prices of Skills

So far we have presented a model in which there is only one type of human capital or skills. Individuals have different levels of human capital and, therefore, earn different incomes. In reality, different types of skills are combined together with physical capital to produce goods and services. If those skills are not additive, they have a relative price that may be changing, implying that the distribution of income also depends on those relative prices, which in turn depend on the relative supplies and demands of the various skills.

To fix these ideas, suppose that there are three types of agents according to their skill types,

H_1, H_2, and H_3. Production takes place through the technology

$$(H_1 + AH_2)^\theta H_3^{1-\theta}.$$

Assuming that markets are competitive, the prices of the three types of skills are equal to their marginal productivities, that is,

$$\begin{aligned}
W_1 &= \theta(H_1 + AH_2)^{\theta-1} H_3^{1-\theta}, \\
W_2 &= \theta A(H_1 + AH_2)^{\theta-1} H_3^{1-\theta}, \\
W_3 &= (1-\theta)(H_1 + AH_2)^{\theta} H_3^{-\theta}.
\end{aligned}$$

In this example, the relative prices for the three types of skills depend on three factors: (i) the relative supplies of the skill types; (ii) the parameter A determining the productivity of H_2 relative to H_1; (iii) the parameter θ determining the relative productivity between the aggregation of H_1 and H_2 on one side and H_3 on the other. For example, an increase in the parameter A, keeping constant the relative supplies of the three skills, increases the productivity of H_2 and H_3 but reduces the marginal productivity of H_1. This changes the distribution of income between the three groups. The change in A could be the result of particular technological progress. As we will see in Section 3, a similar idea has been used by Krusell et al. [2000] to explain the increase in the skill premium observed in the United States since 1980.

2.2.4 Search and inequality

Where does workers' luck come from? Some economists think it is from the arbitrariness of the process that matches workers to jobs. The idea is that some firms are better than others, and these firms end up paying more for essentially identical workers. The argument relies on two considerations. The first is that certain frictions make it difficult for firms to get a worker. The second is that wages depend on the characteristics of both workers and firms.

We can discuss these ideas with the help of the basic labor market model (see Pissarides [1990]) in which firms are created through the random matching of job vacancies and unemployed workers. Workers have linear utility $\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \ c_t$. Risk neutrality implies that the interest rate is constant and equal to $r = 1/\beta - 1$.

A firm is created by paying a cost κ_0 that entails a draw of a productivity level z from the distribution $F(z)$. After the initial draw, z stays constant over time. Then the firm has to post a vacancy at cost κ_1. If matched with an unemployed worker, the firm produces output z starting in the next period until the match is separated, which happens exogenously with probability λ. The firm can use only one worker. The number of newly formed matches is determined by the function $M(v,u)$, where v is the number of vacancies and u is the number of unemployed workers. The probability that a vacancy is filled is $q = M(v,u)/v$, and the probability that an unemployed worker finds occupation is $p = M(v,u)/u$. The second ingredient of this model is that wages are determined through Nash bargaining, where we denote by η the bargaining power of workers. A worker attached to a firm with productivity z is paid the wage $w(z)$ and the firm earns $z - w(z)$.

The value of a firm that has a worker can be written recursively as

$$J^1(z) \;\;=\;\; \Big\{ z - w(z) + \beta(1-\lambda)\; J^1(z) \Big\}, \tag{32}$$

which implies that the value is $J^1(z) = \frac{z-w(z)}{1-\beta(1-\lambda)}$. A newly created firm has value

$$J^0(z) \;\;=\;\; \Big\{ -\kappa_1 + \beta(1-\lambda)\, \Big[qJ^1(z) + (1-q)J^0(z) \Big] \Big\}, \tag{33}$$

which can also be expressed as $J^0(z) = \frac{-\kappa_1 + \beta(1-\lambda)\frac{(z-w(z))q}{1-\beta(1-\lambda)}}{1-\beta(1-\lambda)(1-q)}$.

The value of a worker employed by a firm with productivity z is

$$W(z) = w(z) + \beta\Big[(1-\lambda)W(z) + \lambda U\Big], \tag{34}$$

where U is the value if the worker does not have a job. Such value is given by

$$U = \overline{u} + \beta\left\{ p \int \; W(z)\; F(dz) + (1-p)U \right\}, \tag{35}$$

where $\overline{u}$ is the flow utility for the unemployed worker.

To derive the bargaining problem, let's define the following functions:

$$\widehat{J}(z,w) \;\;=\;\; z - w + \beta(1-\lambda)J^1(z), \tag{36}$$

$$\widehat{W}(z,w) \;\;=\;\; w + \beta\, \Big[(1-\lambda)W(z) + \lambda U\Big]. \tag{37}$$

These functions are, respectively, the value of a firm and the value of an employed worker, given an arbitrary wage w paid in the current period and future wages determined by the function $w(z)$.

The actual wage function $w(z)$ is the solution to the problem

$$\max_{w} \left[\widehat{J}(z,w) - J^0(z)\right]^{1-\eta} \left[\widehat{W}(z) - U\right]^{\eta}. \tag{38}$$

Notice that the terms inside the brackets describe, respectively, what the firm and the worker would lose if they do not reach an agreement and break the match. Parameter η captures the bargaining power of workers. To reach an equilibrium, a couple of additional conditions are needed. One is free entry of firms, that is, the expected value of creating a firm, $\int J^0(z)\, F(dz)$, equals its cost, κ_0. To get a steady-state equilibrium, total firm creation has to be sufficient to create enough vacancies to replace the jobs of workers that join unemployment from job separation.

It is easy to see that the wage is an increasing function of z. In this fashion, a theory of wage inequality can arise from the sheer luck of matching with a very productive firm, even though there is nothing inherently different between two workers in different firms. Hornstein et al. [2011] proposed a new method to assess the quantitative importance of the wage dispersion induced by search frictions and found that it is very small. In fact, the actual dispersion is 20 times larger than the dispersion generated by the type of search frictions described here. To understand this finding, think of an intermediate step between a worker being matched with a firm and before the actual bargaining process takes place. In this step, the worker could forecast what the wage will be and could potentially choose whether to take the job or keep searching. The minimum wage makes the worker indifferent between accepting the job and continuing searching. Such a wage can be compared with the average wage that workers get. Hornstein et al. [2011] found that for empirically sound values of the parameters, the difference between the minimum and the average wages generated by search frictions was tiny.

2.2.5 Occupational choice and earnings inequality

Workers' choice of occupation has recently come to the fore as a source of income inequality. Income inequality may occur not only because workers accumulate different levels of human capital, but also because they work in different occupations. As evidenced by Kambourov and Manovskii [2009a], among others, human capital is largely occupation specific. We will discuss how occupation choices can directly affect the return to human capital and wage growth. In another vein, some workers' occupations may make them more sensitive to cyclical dynamics and unemployment. As in Wiczer [2013], the occupation specificity of human capital makes workers less flexible in response to specific shocks during the business cycle, and this generates inequality across occupations in terms of unemployment rates, unemployment duration, and earnings.

To see the pathways through which occupation choices may affect earnings inequality, consider a simple model with occupations indexed by $j = \{1, \dots, J\}$. Human capital is only imperfectly transferable between occupations. Therefore, the human capital of a worker with current human capital h in occupation j that switches to occupation ℓ becomes $h' = \omega(h, j, \ell)$.

This is the basic framework with which Kambourov and Manovskii [2009b] connected occupational mobility to wage inequality. Workers who remain in the same occupation experience the same wage growth. Workers who switch occupations lose human capital, that is, experience negative growth in earnings. Let's normalize $h = 1$ for experienced workers. When a new employee arrives, the human capital is $\omega_{j,\ell} = \omega(1, j, \ell)$, and it takes one period to become experienced. Let $g(j, \ell)$ be the probability that a worker switches from j to ℓ, and x_j is the measure of workers in occupation j. The variance of wages is

$$var(w) = \sum_j x_j \sum_\ell g(j, \ell) \Big(\omega_{j,\ell} - \mathbb{E}w\Big)^2.$$

Clearly, without switching occupations, the variance would be zero. But, occupational switching is not infrequent and has, in fact, been rising concurrently with the recent rise in earnings inequality—the probability of switching occupations rose by 19% from the 1970s to the 1990s.[8] Kambourov and Manovskii [2009b] connect the former to the latter, posing a common cause for both. If occupation-specific shocks are on the rise, they will affect wage inequality through two chanels. Directly, they will increase the dispersion in wages of those attached to an occupation, but shocks will also increase switching and create more wage inequality. Because these shocks are difficult to observe directly, Kambourov and Manovskii [2009b] use the occupational switching behavior of workers to inform the underlying process that would generate such behavior; as switching rose, the shocks must have also amplified. With this identification logic, occupational switching accounts for 30% of the overall rise in earnings inequality.

Unobserved, occupation-specific shocks are certainly not the only hypothesis for why workers move to different occupations. Several authors, e.g., Papageorgiou [2009] and Yamaguchi [2012], propose that workers' wages depend on occupation-specific match quality that is learned only through the course of the match. In these papers, earnings inequality is exacerbated by occupational mismatch, which slows the wage growth for some workers. On the other hand, aggregate factors like business cycle pressures and unemployment may also increase occupational switching. Indeed, unemployed workers are three times more likely than employed workers to

[8] This probability uses the 1970 Census occupations defintions at the three-digit level and PSID data

switch occupations. In this vein, we introduce a simple model with unemployment and search as motives to switch occupations.

The critical object to be determined in this environment is the stochastic finding rate, $\{p_{j,\ell}\}$, at which workers with occupation j find a job in occupation ℓ. Denote by $P = \{p_{j,\ell}\}_{j,\ell \in \{1,\cdots,J\}}$ the collection of all such finding rates. Denote by $g(j,\ell)$ the fraction of time spent searching for occupation ℓ by a worker with previous occupation j. We assume that looking for jobs in a particular area has decreasing returns at rate $\phi < 1$, so that not all workers from occupation j shift to the same type of new occupation ℓ. This simplification could be a stand-in for any number of more realistic elements of a model such as heterogeneous preferences. Clearly, the allocation of searching time determines the realized finding rate.

The characteristics of the equilibrium are going to depend on the type of wage-setting rule we use, but here we consider the simplest case in which workers earn their marginal product. Hence, the wage of a worker who has just switched occupations is $\omega_{j,\ell}$ and 1 for an experienced worker.

Let W be the value function for an employed worker and U for an unemployed worker. These functions are defined recursively as

$$
\begin{aligned}
W(\ell, P) &= \omega_{\ell,\ell} + \beta(1-\lambda)\mathbb{E}\, W(\ell, P') + \beta\lambda\, \mathbb{E} U(\ell, P'),\\
U(j, P) &= \max_{g(j,\ell)} \sum_{\ell=1,\ldots,J} g(j,\ell)^{\phi} p_{j,\ell}\Big(\omega_{j,\ell} + \beta\, \mathbb{E} W(\ell, P')\Big) + \left(1 - \sum_{\ell=1,\ldots,J} g(j,\ell)^{\phi} p_{j,\ell}\right) \times \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \Big(\bar{u} + \beta\, \mathbb{E} U(j, P')\Big).
\end{aligned}
$$

In this case, $g(j,\ell)$ is chosen so that, given wages and finding rates, the marginal return to search time satisfies

$$
\begin{aligned}
&g(j,j)^{\phi-1} p_{j,j} \left\{1 - \bar{u} + \beta\mathbb{E}\Big[W(j,P') - U(j,P')\Big]\right\}\\
&\qquad = g(j,\ell)^{\phi-1} p_{j,\ell} \left\{\omega_{j,\ell} - \bar{u} + \beta\mathbb{E}\Big[W(\ell,P') - U(j,P')\Big]\right\}, \qquad \forall \ell.
\end{aligned}
$$

To get a taste of the dynamics, suppose that $p_{j,j}$ falls. The indifference condition holds that search time toward this occupation falls so that $g(j,j)^{\phi-1}$ will rise. This increases earnings inequality through two channels: (1) the increase in the unemployment rate among type j workers and (2) more of the new matches go to different occupations, where they produce only $\omega_{j,\ell} < 1$.

The model we have presented has the essential elements of that explored in Wiczer [2013] and, in tying it to data, he shows that recessions often bring a correlated change in P that hurts some occupations much more than others. In such a case, workers from the affected occupations can be unemployed for very long durations and keep the level of unemployment high for a long time. To allow for this result, Wiczer [2013] builds upon Kambourov and Manovskii [2009b] by introducing search frictions, so that the job finding rate is endogeneous and affected by business cycle conditions. As in the case of a typical matching function, the finding rate is reduced by *congestion*.

To extend our framework to endogeneous matching frictions, following Wiczer [2013] let $p_{j,\ell} = p(z_\ell, g(j,\ell))$, where z_ℓ is a shock affecting hiring in occupation ℓ and the more workers looking for the same types of job lowers the finding rate, $\frac{\partial p}{\partial g} < 0$. Then, when $p_{j,j}$ falls, the probability of successfully switching occupations also falls. Whereas Kambourov and Manovskii [2009b] find their shocks to reconcile switching behavior, Wiczer [2013] maps his shocks to measured value-added by occupation. Looking directly at productivity allows Wiczer [2013] to address business cycles in which unemployment and earnings dispersion across occupations increases even though search frictions prevent workers from mass switches into new jobs.

How to identify "occupations" in the data is still an open question. Whereas Wiczer [2013] uses two-digit occupation codes, Carrillo-Tudela and Visschers [2013] take a similar model but with a finer definition of occupation that highlights the interaction between occupation-specific skills and other job characteristics such as location. Hence, the position of a machinist in Detroit may be even more volatile than machinists in general. Both papers generate significant volatility in unemployment and earnings over the business cycle beyond search models that abstract from occupational heterogeneity.

3 The dynamics of inequality

So far we have looked at how to build theories of inequality in earnings and wealth that aggregate into a macro model. Now we turn to the analysis of factors that affect the dynamics of inequality. We first consider in Section 3.1 changes in inequality that take place over the business cycle and in Section 3.2 we analyze the dynamics over a longer horizon.

3.1 Inequality and the business cycle

A well-established feature of the business cycle is that the labor share of income is highly counter-cyclical. As shown in Figure 1, the labor share in the U.S. economy tends to increase during recessions. The figure also shows a declining trend in the labor share since the early 1980s. To the extent that agents are heterogeneous in the sources of their income—that is, some agents earn primarily capital incomes while others earn primarily labor income—there is significant redistribution over the business cycle.

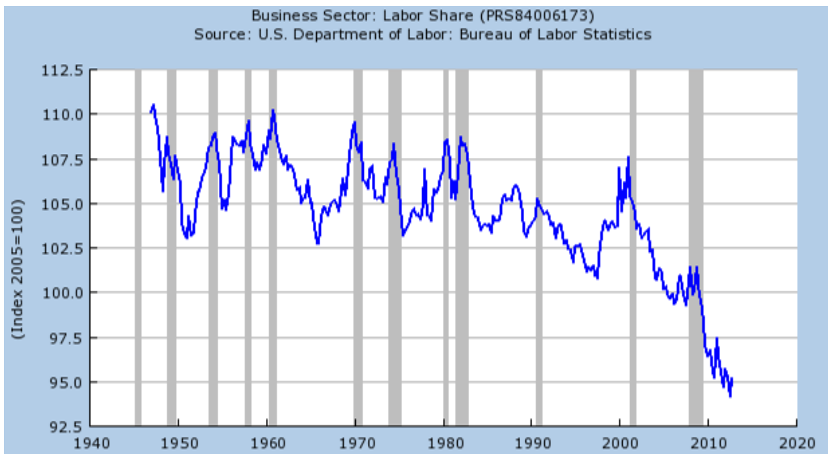


FIGURE 1

Labor share in the U.S. business sector as defined by the Bureau of Labor Statistics.

To capture the cyclical properties of the labor share, we have to deviate from the standard neoclassical model with a Cobb-Douglas production function, since in this model the labor share is constant. In this section, we review some models in which the compensation of workers is determined through bargaining between employers and workers. Since the bargaining strength of workers depends on macroeconomic conditions, this mechanism has the potential to generate a labor income share that changes over the business cycle.

We start in Section 3.1.1 by using the search and matching model developed in Section 2.2.4 modified to allow for the study of the determination of the labor share. The modification associates the creation of firms with actual investors, which gives an explicit separation of labor and capital income. We look at two versions of this model: a simple one and another in which investors use both debt and external financing with bonds. An important property of this model is that shocks that affect the bargaining position of workers also affect the distribution of income as well as the macroeconomic impact of these shocks. We will consider two types of shocks: standard productivity shocks and shocks that affect access to credit. Then in Section 3.1.2, we will review the financial accelerator model in which the distribution is also interconnected with the business cycle. We will conclude this section by discussing the ability of these models to replicate the empirical properties of the data beyond the contemporaneous correlation.

3.1.1 The determination of factor shares: Productivity shocks, bargaining power shocks, and financial shocks

Consider a version of the search and matching model described in Section 2.2.4 (Pissarides [1987]) in which the owners of firms—investors—are distinct from workers, but in which productivity is stochastic and common to all firms. Therefore, z is the same across firms and changes stochastically over time. We will focus on the distribution of income between investors and workers. Both types of agents have the same utility $\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t\, c_t$.

As before, a firm is created when a posted vacancy is filled by an unemployed worker. A new firm produces output until the match is destroyed exogenously, which happens with probability λ, but now the level of output varies over time. The number of matches is determined by the matching function $M(v,u)$, where v is vacancies and u unemployed workers. The probability that a vacancy is filled is $q = M(v,u)/v$, and the probability that an unemployed worker finds a job is $p = M(v,u)/u$. Wages are determined through Nash bargaining, where we denote by η the bargaining power of workers. We also consider the possibility that the bargaining power η may be stochastic. With these stochastic terms we rewrite the value functions for investors as

$$J^1(z,\eta) = z - w(z,\eta) + \beta(1-\lambda)\,\mathbb{E}J^1(z',\eta'|z,\eta), \tag{39}$$

$$J^0(z,\eta) = -\kappa_1 + \beta(1-\lambda)\,\Big[q\mathbb{E}J^1(z',\eta'|z,\eta) + (1-q)\mathbb{E}J^0(z',\eta'|z,\eta)\Big], \tag{40}$$

and for workers

$$W(z,\eta) = w(z,\eta) + \beta\,\mathbb{E}\Big[(1-\lambda)W(z',\eta') + \lambda U(z',\eta')\Big], \tag{41}$$

$$U(z,\eta) = \overline{u} + \beta\,\mathbb{E}\Big[pW(z',\eta') + (1-p)U(z',\eta')\Big]. \tag{42}$$

After some rearrangement, the values for the firm and the worker can be written as

$$J^1(z,\eta) - J^0(z,\eta) = (1-\eta)S(z,\eta), \tag{43}$$

$$W(z,\eta) - U(z,\eta) = \eta\, S(z,\eta), \tag{44}$$

where $S(z,\eta) = J^1(z,\eta) - J^0(z,\eta) + W(z,\eta) - U(z,\eta)$ is the bargaining surplus that is split between the contractual parties proportionally to their relative bargaining power. The surplus function can be written recursively as

$$S(z,\eta) = z - \overline{u} + (1-\lambda)\,\beta\,\mathbb{E}S(z',\eta') - \eta\,\beta\,p\,\mathbb{E}S(z'\eta'). \tag{45}$$

Using the free entry condition $\kappa_1 = q\beta\mathbb{E}S(z',\eta')$, the sharing rules (43) and (44), and the functions (39), (41), and (42), we can derive the following expression for the wage:

$$w(z,\eta) = (1-\eta)\bar{u} + \eta z + \frac{\eta\, p\, \kappa_1}{q}.$$

Shocks to Productivity Figure 2 plots the impulse responses of employment and the investor's share of income to a positive productivity shock under the heading baseline model. An economic boom is characterized by a larger share of income going to investors. However, the quantitative effects in terms of income distribution and employment are not large. The weak employment response is a well-known property of the matching model (see Costain and Reiter [2008] or Shimer [2005]). What is interesting is that the inability of the model to generate large employment fluctuations is related to the inability of the model to generate large movements in the distribution of income. Since wages respond too quickly to productivity, the share of income going to investors increases only slightly. As a result, the incentive to create new vacancies does not increase much. However, if wages would respond less, the increase in the share of income going to investors and the increase in employment would be bigger.

Recognizing the direct link between distribution and employment, several authors have proposed some mechanisms to generate smoother responses of wages and, therefore, larger fluctuations in income shares. Here we summarize three of these approaches. The first approach, proposed by Hagedorn and Manovskii [2008], is to assume that the flow utility received by workers when unemployed is not much smaller than the flow utility from working. In terms of the model, this is obtained by choosing a large value for the parameter $\bar{u}$, that is, the flow utility in the unemployment state. Although many consider this assumption implausible, the paper illustrates how this feature could bring the model closer to the data. The second approach proposed by Gertler and Trigari [2009] is to assume that wages are sticky. To illustrate these two cases, we first assign a higher value for the parameter $\bar{u}$ so that the replacement rate from unemployment is 95%. The impulse responses for this case, plotted in Figure 2, are labeled "High unemp. value." The figure also plots the impulse responses when the wage is exogenously fixed at the steady state with flexible wages (an extreme case of wage rigidity). As can be seen, both assumptions generate much higher volatility of employment and income distribution between investors and workers.

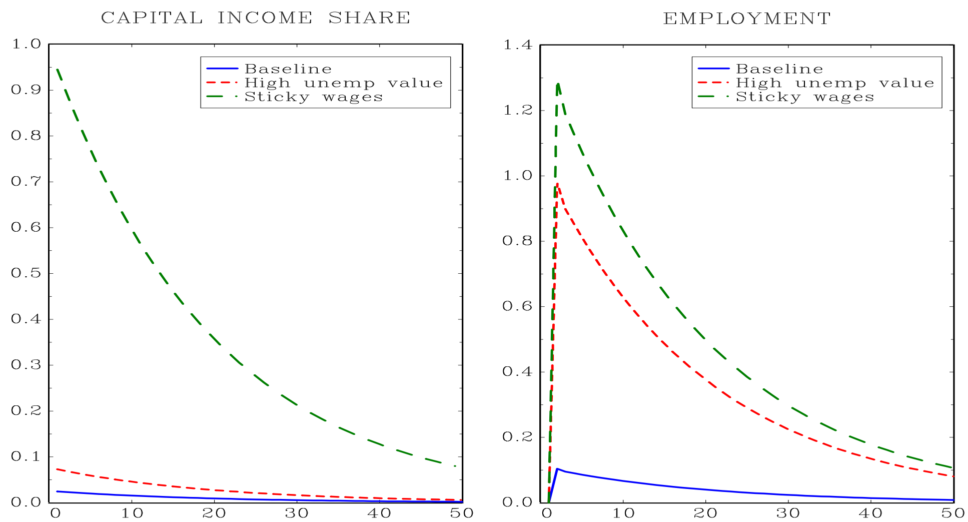


FIGURE 2
Impulse response to productivity shock. The common parameters to all versions of the model are $\beta = 0.985$, $\alpha = 0.5$, $\mathbb{E}z = 1$, $\rho_z = 0.95$, $\sigma_z = 0.01$. The remaining parameters $\overline{u}$, κ, λ and A are chosen to achieve the following steady state targets: a replacement rate of unemployment of 50% (95% in the model with high unemployment value), 10% unemployment rate, 93% probability of filling a vacancy, 70% probability of finding a job. The resulting values are $\overline{u} = 0.473$ (0.944 in the model with high unemployment value), $\kappa = 0.316$ (0.034 in the model with high unemployment value), $\lambda = 0.103$ and $A = 0.807$.

This shows that inequality and macroeconomic volatility are closely interconnected: more volatile income distribution over the business cycle is associated with greater macroeconomic volatility. A third approach is explored in Duras [2013]. The idea is that in periods with high productivity or output, the cost for workers to break the match is higher than in normal times. This weakens workers' bargaining position, alleviating the upward pressure on wages when productivity rises.

Shocks to Bargaining Power Although it has been customary to assume that macroeconomic fluctuations are driven by productivity shocks, economic disturbances have many other possible sources. Here we summarize the effects of shocks that impact directly on the distribution of income, in particular, shocks that directly affect the bargaining power of workers, that is, the bargaining share η. When η decreases, a larger share of income will go to investors, increasing income inequality. At the same time, as investors appropriate a larger share of the surplus, they have a higher incentive to hire workers, thereby inducing a macroeconomic expansion. A similar approach has been studied by Ríos-Rull and Santaeulàlia-Llopis [2010] in the context of a neoclassical model.

3.1.1.1 Financial Shocks The next step is to show that similar effects to those generated by shocks to bargaining power can be generated by the expansion and contraction of financial markets. The presentation of this case follows Monacelli et al. [2011].

Consider another slight modification of the search and matching model presented earlier, where we allow firms to borrow at the gross rate r. Borrowing, however, is subject to the constraint $b' \leq \phi \mathbb{E} J'(b')$, where ϕ is stochastic. This variable captures the possible changes to the tightness of credit markets.

The firm enters the period with debt b. Given the new debt b' and the wage w, the dividends paid to investors are $d = z - w + b'/R - b$, where $R = (1+r)/(1-\lambda)$ is the gross interest rate paid by the firm conditional on survival. We assume that in the event of exit, the firm defaults on the outstanding debt. Anticipating this, the lender charges the gross interest rate $R = (1+r)/(1-\lambda)$ so that the expected return from the loan is r. Notice that investors are shareholders and bondholders at the same time. We write the value functions exclusively as functions of debt, ignoring the potential variability of both productivity and bargaining power (that is, we now assume that z and η are constant).

The equity value of the firm can be written recursively as

$$\begin{aligned} J(b) \quad &= \quad \max_{b'} \left\{ z - g(b) - b + \frac{b'}{R} + \beta(1-\lambda)\, \mathbb{E} J'(b') \right\} \\ & \quad\quad \text{subject to} \\ & \quad\quad b' \leq \phi \mathbb{E} J'(b'), \end{aligned} \tag{46}$$

where $w = g(b)$ denotes the (to be determined) wage paid to the worker. As we will see, the wage will depend on the debt. Notice that we have also used a prime to denote the next period value of equity, since this also depends on the next period aggregate states, specifically, the unemployment rate and credit market conditions. To avoid cumbersome notation, we do not include the aggregate states as explicit arguments of the functions defined here. Instead we use the prime to distinguish current versus future functions.

The value of an employed worker is

$$W(b) = g(b) + \beta\, \mathbb{E}\Big[(1-\lambda)\; W'(b') + \lambda U'\Big], \tag{47}$$

which is defined once we know the wage $g(b)$. The function U' is the value of being unemployed and is defined recursively as

$$U = \overline{u} + \beta\, \mathbb{E}\Big[pW'(b') + (1-p)U'\Big],$$

where p is the probability that an unemployed worker finds a job and $\overline{u}$ is the flow utility for an unemployed worker. Although the value of an employed worker depends on the aggregate states and the individual debt b, the value of being unemployed depends only on the aggregate states, since all firms choose the same level of debt in equilibrium. Thus, if an unemployed worker finds a job in the next period, the value of being employed is $W'(b')$.

The determination of the current wage solves the same problem as in (38). We should take into account, however, that this solution also depends on b and on the function that determines future wages. Therefore, we write the solution of the bargaining problem as $w = \psi(g; b)$, where g is the function determining future wages. The equilibrium solution to the bargaining problem is the fixed point to the functional equation $g(b) = \psi(g; b)$.

Also in this case, the values for the firm and the worker satisfy

$$J(b) \;=\; (1-\eta)S(b), \tag{48}$$

$$W(b) - U \;=\; \eta S(b), \tag{49}$$

where the surplus is defined as $S(b) = J(b) + W(b) - U$. This can be written recursively as

$$S(b) = z - a - b + \frac{b'}{R} + (1-\lambda)\beta\, \mathbb{E}S'(b') - \eta\, \beta\, p\, \mathbb{E}S'(B'). \tag{50}$$

When a vacancy is filled, the newly created firm starts producing and pays wages in the next period. The only decision made in the current period is the debt b'. Therefore, the value of a vacancy just filled with a worker is

$$\begin{aligned} Q \;\; = \;\; & \max_{b'} \left\{ \frac{b'}{1+r} + \beta(1-\eta)ES'(b') \right\} \\ & \text{subject to} \\ & b' \leq \phi(1-\eta)\mathbb{E}S'(b'). \end{aligned} \tag{51}$$

Since the new firm becomes an incumbent starting in the next period, $S'(b')$ is the surplus of an incumbent firm defined in (50). Notice that in the choice of b', a new firm faces a problem similar to that of incumbent firms (see problem (46)). Even if the new firm has no initial debt and it does not pay wages, it will choose the same stock of debt b' as incumbent firms. We can then focus on a "representative" firm and in equilibrium $B = b$.

The value of posting a vacancy is equal to $V = qQ - \kappa$. Since in this version of the model there are no firm-specific productivity draws, $\kappa = \kappa_0 + \kappa_1$. As long as the value of a vacancy is positive, more vacancies will be posted. Free entry implies that $V = 0$ and in equilibrium we have

$$qQ = \kappa. \tag{52}$$

We now characterize the optimal choice of debt, that is, problem (46). Denoting by μ the Lagrange multiplier associated with the enforcement constraint, the first order condition is

$$\eta - R\Big[1 + (1-\eta)\phi\Big]\mu = 0. \tag{53}$$

In deriving this expression, we have used the property of the model for which the choice of b' does not depend on the existing debt b and, therefore, $\frac{\partial S(b)}{\partial b} = -1$. We have also used the equilibrium condition $\beta R(1-\lambda) = \beta(1+r) = 1$.

Looking at the first order condition, we can see that the enforcement constraint is binding (that is, $\mu > 0$) if $\eta \in (0,1)$. Thus, provided that workers have some bargaining power, the firm always chooses the maximum debt and the borrowing limit binds. In this way, bargaining introduces a mechanism through which the financial structure is determined (Modigliani and Miller [1958] does not apply). The reason is clear: by using outside finance, the firm is able to reduce the surplus that is bargained with the worker, increasing the possible rewards to equity.

To gather some intuition about the economic interpretation of the multiplier μ, it will be convenient to rearrange the first order condition as

$$\mu = \underbrace{\left(\frac{1}{1+(1-\eta)\phi}\right)}_{\substack{\text{Total change}\\ \text{in debt}}} \times \underbrace{\left(\frac{1}{R} - \frac{1-\eta}{R}\right)}_{\substack{\text{Marginal gain}\\ \text{from borrowing}}}.$$

The multiplier results from the product of two terms. The first term is the change in next period liabilities b' allowed by a marginal relaxation of the enforcement constraint, that is, $b' = \phi(1-\eta)\,\mathbb{E}S(b') + \bar{a}$, where $\bar{a} = 0$ is a constant. This is obtained by marginally changing $\bar{a}$. In fact, using the implicit function theorem, we obtain $\frac{\partial b'}{\partial \bar{a}} = \frac{1}{1+(1-\eta)\phi}$, which is the first term.

The second term is the actualized net gain from increasing the next period liabilities b' by one unit (marginal change). If the firm increases b' by one unit, it receives $1/R$ units of consumption today in the form of additional dividends. In the next period, the firm has to repay 1 unit. However, the effective cost for the firm is lower than 1, since the higher debt allows the firm to reduce the next period wage by η, that is, the part of the surplus going to the worker. Thus, the effective repayment incurred by the firm is $1-\eta$. This cost is discounted by $R = (1+r)/(1-\lambda)$ because the debt is repaid only if the match is not separated, which happens with probability $1-\lambda$. Thus, the multiplier μ is equal to the total change in debt (first term) multiplied by the gain from a marginal increase in borrowing (second term).

Using the property for which the enforcement constraint is binding, that is, $\phi \ EJ(b') = b'$, Appendix B shows that the wage can be written as

$$w = (1-\eta)\bar{u} + \eta(z - B) + \frac{\eta[p + (1-\lambda)\phi]\kappa}{q(1+\phi)}. \tag{54}$$

This equation makes clear that the initial debt B acts like a reduction in output in the determination of wages. Instead of getting a fraction η of output, the worker gets a fraction η of output "net" of debt. Thus, for a given bargaining power η, the larger the debt the lower the wage received by the worker. This motivates the firm to maximize the debt, as we have already seen from the first order condition.

Figure 3 plots the impulse responses to a credit shock, that is, a shock that raises ϕ and increases the credit available to firms. The credit expansion generates an increase in the capital income share and an increase in employment. Thus, changes in financial markets could alter the distribution of income and, with it, affect the incentives to create jobs. This is another example of how the distribution of income and macroeconomic performance are directly interconnected.

3.1.2 Financial accelerator and inequality

A well-established tradition in macroeconomics introduces financial market frictions in business cycle models. The key ingredients are based on two assumptions: market incompleteness and heterogeneity. Although not often emphasized, inequality plays a central role in these models. For example, in the seminal work of Bernanke and Gertler [1989] and Kiyotaki and Moore [1997], entrepreneurial net worth is central to the amplification of aggregate shocks. When more resources are in the hands of constrained producers (*i.e.*, these agents are richer), they can expand production and enhance macroeconomic activities. This can happen because they earn higher

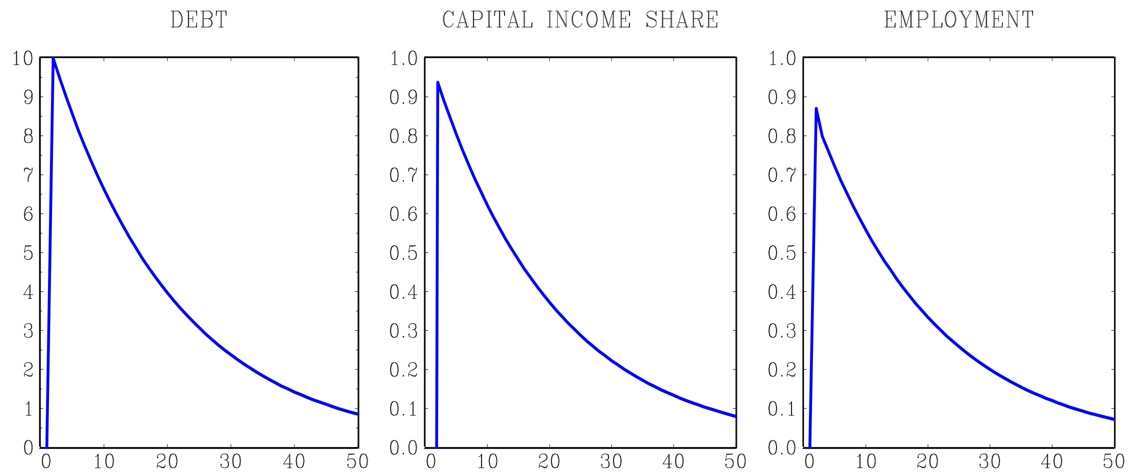


FIGURE 3
Impulse response to credit shock. The parameters are $\beta = 0.985$, $\alpha = 0.5$, $\mathbb{E}z = 1$, $\rho_z = 0.95$, $\sigma_z = 0.01$, $\overline{u} = 0.473$, $\kappa = 0.316$, $\lambda = 0.103$ and $A = 0.807$, $\bar{\phi} = 0.0022$, $\rho_\phi = 0.95$, $\sigma_\phi = 100$.

incomes or because their assets are worth more following asset price appreciations. Thus, these models posit a close connection between profit shares and the business cycle.

These models share some similarities with the matching models reviewed above: when a larger share of output goes to investors/entrepreneurs, the economy expands. At the same time, as the economy expands, a larger share of output or wealth (or both) is allocated to entrepreneurs. The mechanism of transmission, however, is different. In the matching model, the mechanism is the higher profitability of employment or investment. In the financial accelerator model, instead, it is the relaxation of the borrowing constraints. For a detailed review of the most common models used in the literature to explore the importance of financial frictions for macroeconomic fluctuations, see Quadrini [2011].[9]

3.2 Low frequency movements in inequality

We discuss here some theoretical ideas that have been proposed in the literature to explain some of the trends in the distribution of income that have occurred since the early eighties. In Section 3.2.1 we look at the reduction in labor share and in the following two sections we look at the increased inequality in wages and earnings. In Section 3.2.2 we examine the potential role of

[9] In addition to cyclical movements in the shares of income that go to labor and capital, there are cyclical movements in the shares of income earned by different groups of households. Castañeda et al. [1998] find that movements in unemployment rates by skill groups as well as movements in factor shares account for the bulk of the cyclical share of income earned by the various quintiles of households.

increased competition for human capital and in Section 3.2.3 the changes in the prices of skills due to skills-biased technical changes.

3.2.1 Labor Share's Reduction since the Early Eighties

In a recent paper, Karabarbounis and Neiman [2014] document that labor share has significantly declined since the early 1980s for a majority of countries and industries. They pose a CES production function with nonunitary elasticity of substitution between labor and capital and argue that the well-documented decline in the relative price of investment goods (see Gordon [1990], Krusell et al. [2000] and Cummins and Violante [2002]) induced firms to substitute away from labor and toward capital. The consequence was the reduction in the price of labor. They conclude that roughly half of the observed decline in the labor share can be attributed to this mechanism.

To see how this mechanism works, consider the following aggregate production function:

$$Yt = F(K_t, N_t) = \left[\alpha_k (A_{K_t} K_t)^{\frac{\sigma-1}{\sigma}} + (1-\alpha_k)(A_{N_t} N_t)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \tag{55}$$

where σ denotes the elasticity of substitution between capital and labor in production, α_k is a distribution parameter, and A_{K_t} and A_{N_t} denote, respectively, capital-augmenting and labor-augmenting technology processes. As σ approaches 1, this becomes a Cobb-Douglas production function. Under perfect competition, marginal productivities yield factor prices and one can easily obtain expressions for labor share that crucially depend on the elasticity of substitution σ.

Karabarbounis and Neiman [2014] estimate σ by using only trends in the relative price of investment and the labor share in a cross section of countries. They find a value of 1.25. With this value, the decline in the (observable) relative price of investment accounts for 60% of the observed reduction in labor share. The remaining reduction can be imputed to larger increases in capital-augmenting technology relative to labor augmenting technology, and to changes in non-competitive factors (the most important, perhaps, are the permanent changes in the bargaining power of workers relative to firms along the lines of the models discussed earlier). Additional explanations arise from changes in the sectoral composition of output toward industries with higher capital share (which does not really seem to be the case), in part associated with globalization or the increase in the share of output that is traded with other countries, especially developing countries.

3.2.2 Increased wage inequality: The role of competition for skills

There has been a big increase in earnings dispersion that is well documented in in the literature. A glimpse of it can be seen in Tables 3 and 4. We now explore a possible explanation based on increased competition for skills in the context of human capital accumulation. We have already seen in Section 2.2.1 that human capital accumulation could be an important mechanism through which income becomes heterogeneous and that the higher the incentive to invest in human capital, the greater the degree of income inequality. We now look at one of the mechanisms that could affect the incentive to invest in human capital: competition for skills.

To study the importance of competition, we explore yet another version of the matching model described above by adding investment in human capital. We ignore here the possibility of firms to raise debt and limit the analysis to the version of the model without aggregate shocks. However, we now assume that output depends on the human capital of the worker, denoted by h. The production technology has a structure similar to the model presented in Section 2.2.1. The key feature of the model we look at now is that human capital investment requires both an output loss or pecuniary cost within the firm denoted by y and a utility cost for the worker that we assume quadratic, $\frac{\alpha y^2 h}{2}$. Given y, human capital evolves stochastically according to

$$h' = h(1 + y\varepsilon'),$$

where ε is an i.i.d. random variable. The gross growth rate of human capital is denoted by $g(y, \varepsilon') = 1 + y\varepsilon'$.

Since the outcome of the investment is stochastic, the model generates a complex distribution of human capital among workers. In the long run, the distribution will be degenerate since at the individual level, h follows a random walk. We assume that workers die with probability λ and that a match breaks down only when a worker dies. Thus, λ represents at the same time the death probability of a worker and the probability of separation of a match. In this way, the distribution becomes stationary and converges to a steady state.

There are contractual frictions that derive from the ability of the worker to control the investment y after bargaining over the wage. The worker unilaterally chooses an investment y that may be, and indeed is, different from the investment that maximizes the surplus of the match. This would be the investment that the worker would choose if he had been able to commit. Of course, when the firm bargains the wage, it anticipates the investment that the worker will choose in absence of commitment.

Let's define a few items. We can write the values of the investor and the worker in normalized form, that is, rescaled by human capital h. Then, the value for the investor can be written as

$$j = h - y - w + \beta(1-\lambda)\,\mathbb{E}\,g(y,\varepsilon')\,j', \tag{56}$$

where $j = J/h$ and w is the wage per unit of human capital. The total wage received by the worker is $w\,h$. The value for the worker is

$$\omega = w - \frac{\alpha y^2}{2} + \beta(1-\lambda)\,\mathbb{E}\,g(y,\varepsilon')\,\omega', \tag{57}$$

where $\omega = W/h$.

The value of being unemployed is

$$u = \overline{u} + \beta(1-\lambda)\Big[p\omega' + (1-p)u'\Big], \tag{58}$$

where $u = U/h$.

Even though in equilibrium employed workers do not lose their occupation, u is important because it affects the threat value in bargaining. In a steady state we have $v = v'$, $\omega = \omega'$, $u = u'$.

The optimal investment y chosen by the worker maximizes the worker's value, that is,

$$\max_y \left\{ w - \frac{\alpha y^2}{2} + \beta(1-\lambda)\,\mathbb{E}\,g(y,\varepsilon')\,\omega' \right\},$$

with the first order condition given by

$$y = \beta(1-\lambda)\omega'\,\mathbb{E}\,\varepsilon'. \tag{59}$$

The important part to remember is that bargaining happens before the worker chooses her investment, which means that the surpluses that enter the problem take the investment y as as determined by condition (59). From this condition we can see that y depends on ω' but not on the current value of ω, which implies that y is not affected by the outcome of the wage bargaining in the current period. Effectively, the current bargaining problem takes y as given and solves

$$\max_w \Big\{ j^{1-\eta}(\omega - u)^{\eta} \Big\}. \tag{60}$$

The first order condition implies that the parties split the net surplus, $s = j + \omega - u$, according to the bargaining weight η, that is,

$$\begin{aligned} j &= (1-\eta)s, \\ \omega &= \eta s + u. \end{aligned}$$

As a comparison, we can also characterize the optimal investment when the worker commits to a particular y chosen to maximize the surplus of the match. In this case, the bargaining problem maximizes the objective (60) over both w and y. The first order condition with respect to w does not change, whereas the first order condition with respect to y becomes

$$1 + \alpha y = \beta(1-\lambda)(j' + \omega')\,\mathbb{E}\,\varepsilon'. \tag{61}$$

Compared with the optimality condition when the investment is controlled by the worker, equation (59), we observe that the left-hand side and right-hand side terms in equation (61) are both bigger. Therefore, the optimal choice of y with commitment could be smaller or bigger. However, provided that α is sufficiently small, that is, the cost for the worker is not too large, the investment without commitment will be bigger.

General equilibrium and the impact of competition So far we have not worried about what happens outside the match, but there is a free entry condition that determines how many vacancies are posted. This is given by

$$q\,\beta\,j = \kappa, \tag{62}$$

where κ is the normalized cost of a vacancy.[10] One way of thinking about increased competition is that the entry cost κ is lower. We then have the following proposition.

Proposition 3.1. *The degree of competition κ affects the steady-state value of y only in the environment without worker commitment.*

This result has a simple intuition: A lower κ is associated with a higher probability that an unemployed worker finds an occupation. As a result, the value of being unemployed increases. Since this represents the threat value in bargaining, the worker can extract a higher wage w, which in turn increases the incentive to invest.

[10] For simplicity we are assuming that the cost of vacancies is proportional to the amount of human capital.

We can now show how an increase in competition (lower κ) affects inequality and aggregate outcomes simultaneously. In particular, we have that lower κ generates: (i) more risk taking and greater income inequality and (ii) higher aggregate income. The first effect can be seen from the first order condition (59). A lower entry cost increases the number of vacancies and, therefore, the value of finding another occupation if the worker quits. This allows the worker to bargain a higher wage, which in turn increases the employment value ω. We can then see from equation (59) that a higher value of ω is associated with a higher y. As we have seen in Section 2.2.1, a higher y implies greater inequality. The second property—the increase in aggregate income—is obvious since a higher y implies higher aggregate human capital. Thus, there is a trade-off between inequality and aggregate income.

Cooley et al. [2012] use a model with similar features but where the accumulation of human capital takes place in the financial sector. They show that greater competition for skills in the financial industry increased the incentive to invest in human capital and generated greater income inequality within and between sectors. This seems consistent with the recent increase in inequality, with income more concentrated at the very top of the distribution and in certain professions, namely, managerial occupations in the financial sector. This pattern is also observed in the United Kingdom, as documented by Bell and Van Reenen [2010].

The idea that competition may increase inequality may go against the common wisdom that wealth is very concentrated because those who control wealth are able to protect it by limiting competition. From this the call for increased enforcement of competition to reduce inequality. Of course, this does not mean that the theory described above is not valid. It depends on the particular environment we are studying: in certain sectors competition may lead to more inequality, in other sectors to lower inequality.

The degree of competition is just one way of affecting the equilibrium properties of aggregate income and inequality. Taxes are also important. In the context of this model, higher taxes discourage human capital investment (since the after-tax return from investing is lower), but this could be mitigated by the tax deductibility of the investment. Since the costs (curtailment of future earnings) and benefits (tax deductibility) occur at different stages of life when the individual has different incomes, the degree of progressivity becomes more important than the overall taxation. However, to the extent that taxes reduce investment, they also lower inequality (since lower investment reduces the volatility of individual incomes).

3.2.3 Skill-biased technical change

In addition to increased competition for skills who ends up rewarding those who are more skilled, a natural explanation for the increased earnings inequality is skill-biased technical change (Katz and Murphy [1992]). Although this term refers in general to changes in the distribution of earnings as a whole, it is often applied more specifically to the premium that college-educated people command compared with those without a college degree. This is motivated by the fact that the college wage premium, defined as the mean log wages of college graduates relative to high school graduates, has increased from 0.3 to 0.6 (see Goldin and Katz [2009]).

To illustrate how skill-biased technical change may have contributed to the increased earnings inequality, consider the following production function:

$$Y_t = F(A_{s,t}S_t, A_{u,t}U_t), \tag{63}$$

where S stands for the number of skilled workers (college education) and U for the number of uneducated workers (without college education). $A_{s,t}$ and $A_{u,t}$ are exogenous technical coefficients that could change over time. Under perfect competition in the labor market, wages are marginal productivities, that is

$$w_{s,t} = A_{s,t} \frac{\partial F(A_{s,t}S_t, A_{u,t}U_t)}{\partial S_t}, \tag{64}$$

$$w_{u,t} = A_{u,t} \frac{\partial F(A_{s,t}S_t, A_{u,t}U_t)}{\partial U_t}. \tag{65}$$

The skill wage premium is defined as

$$\frac{w_{s,t}}{w_{u,t}} = \frac{A_{s,t}\ \frac{\partial F_t}{\partial S}}{A_{u,t}\ \frac{\partial F_t}{\partial U}}. \tag{66}$$

Absent large changes in the relative quantities of skilled and unskilled workers, we can assume that $\frac{\partial F_{t+1}}{\partial S} / \frac{\partial F_{t+1}}{\partial U}$ is very close to $\frac{\partial F_t}{\partial S} / \frac{\partial F_t}{\partial U}$, and the same goes for the marginal productivity of the unskilled. Consequently, we have that the change in the skill wage premium is given by

$$\frac{\frac{w_{s,t+1}}{w_{u,t+1}}}{\frac{w_{s,t}}{w_{u,t}}} \sim \frac{\frac{A_{s,t+1}}{A_{u,t+1}}}{\frac{A_{s,t}}{A_{u,t}}} = \frac{\frac{A_{s,t+1}}{A_{s,t}}}{\frac{A_{u,t+1}}{A_{u,t}}}. \tag{67}$$

This implies that the increase in the wage premium is due to faster growth in the technology coefficient A_{s_t} relative to the growth of A_{u_t}. Hence, the commonly used term "skill-biased technical change". But is there something more tangible than just an exogenous and largely unobserved technological change, or can we track it down to something observable?

Krusell et al. [2000] argued that we can relate these changes to something observable. Gordon [1990] and later Cummins and Violante [2002] have documented that the price of equipment (which is the main part of capital) in terms of consumption goods has gone down dramatically during the period of the raising skill premium. At the same time, the quantity of equipment has gone up significantly relative to output. This is a measurable form of technical change. Combine this with the notion that equipment or capital and skilled labor are complements, whereas unskilled labor is a substitute, and we have an actual channel through which technical progress is skill biased. The formulation in Krusell et al. [2000] does not have factor-specific technical change because all the effects of skill-biased technical change are in the increased quantity of equipment and can be written as

$$Y_t = K_t^{\alpha} \left[\mu U_t^{\sigma} + (1-\mu) \left(\lambda E_t^{\rho} + (1-\lambda) S_t^{\rho} \right)^{\frac{\sigma}{\rho}} \right]^{\frac{1-\alpha}{\sigma}}, \tag{68}$$

where K_t stands for structures (buildings) that pay no role as they enter the production function in a Cobb-Douglas form, U_t and S_t are again unskilled and skilled labor, and E_t is equipment. Using observed measures of inputs, they estimated the elasticities of substitution ρ and σ and the share parameters α, λ, and μ, and found that unskilled labor is indeed a substitute for the aggregate of equipment and skilled labor, with both items being complementary to each other. They also found that this specification accounts very well for the observed wage premium under perfect competition for factor inputs.

Other forms of technical innovation indirectly generate skill-biased technical change. Suppose that technical change, regardless of its final effects on total productivity, is sometimes more dramatic than other change. The introduction of information technology could be one of these instances even if its impact on productivity is not as clear (Solow [1987]). Yet, the adaptation to this new technology may be easier for educated people. This is the approach taken by Greenwood and Yornkoglu [1997], Caselli [1999], and Galor and Moav [2000]. Alternatively, suppose that information technology reduces information and monitoring costs within firms, allowing for reorganizations with fewer vertical layers and with workers performing a wider range of tasks. This gives educated workers an advantage. See, for example, Milgrom and Roberts [1990] and Garicano and Rossi-Hansberg [2004]. Yet another form of skilled-biased technical change is an increase in competition for skills, as in the previous section, which could be the result of the technical change. In the context of the model studied earlier, the technical change can take the form of a lower vacancy cost κ. The lower κ increases the demand of skilled workers, which in turn increases the incentive to accumulate skills.

The technological innovations introduced in the 1970s seem to have affected the economy in other respects. Greenwood and Jovanovic [1999] and Hobijn and Jovanovic [2001] assume that new information technologies required a level of restructuring that incumbent firms could not face. As a result, their stock market value dropped. This is another form of redistribution in the sense that the owners of incumbent firms lost market value to the owners of new firms. Acemoglu [1998] has proposed a theory of the technical change itself being the result of a surge in college graduates.

The rise of superstars is another possible mechanism that increases the concentration of income. Rosen [1981] viewed the increase in earnings dispersion among people in some occupations as the result of an increase in their ability to reach more users of rare skills. Although this applies naturally to the case of artists and athletes, it also applies more generally to other types of skills. For example, Gabaix and Landier [2008] propose a theory of CEO pay where the value of managerial superstars is enhanced by the increase in the size of firms.

Skill-Biased Technical Change and Human Capital Accumulation How does human capital investment interact with skill-biased technical change? Heckman et al. [1998] provide an answer that relies on the difference between observed wages and the price of skills that is due to the unpaid on the job investment of the Ben-Porath [1967] type models. They find no especial role to capital in generating the increase in the skill wage premium. Instead, they find that the endogenous response of both more college attendance and the allocation of time to invest in further skills is sufficient to account for the patterns in the data. Guvenen and Kuruşçu [2010] also explore the interaction of skill-biased technical change and human capital accumulation emphasizing differences across people in the ability to acquire human capital. Guvenen and Kuruşçu [2010] argue that increase biased technical change induces immediately an increase in investment by talented individuals that first depresses the skill wage premium and then rises it and that is consistent with the observed bad performance of median wages and with the lack of increase in consumption inequality,

4 Inequality and financial markets

A large body of work links, theoretically, inequality and financial markets. The lack of complete markets helps to shape inequality through two channels. In Section 4.1 we study how the limited access to borrowing prevents poor households from undertaking valuable investments. This limited access keeps them and their descendants from climbing the social ladder. In Section 4.2 we study environments in which access to borrowing affects inequality, even when there are no household-specific investments.

In the environments studied in the first two sections, the borrowing limits are set exogenously. In Section 4.3 we start exploring endogenous theories of the borrowing limit by looking at environments in which the ability to borrow is limited by the incentive to default. In doing so, we follow the ideas suggested in Kehoe and Levine [1993]. In Section 4.4 we review recent papers in which the limits to borrow come from the legal ability to default on debts allowed by the U.S. bankruptcy code. In Section 4.5 we explore various extensions of these models. Finally, in Section 4.6 we briefly discuss the literature that links the long term performance of an aggregate economy with the ability of households to borrow.

4.1 Financial markets and investment possibilities

Agents with available funds are not necessarily those with the best opportunities to use the funds. It is then socially desirable that the funds are channeled from the former to the latter, which is the primary role played by financial markets. Financial market imperfections, however, limit the volume of funds that can be transferred, and as a result the allocation is inefficient.

Financial market imperfections can take different forms. In a simple overlapping generations model in which the only decision that agents make is how much to invest in the education of their children, the lack of borrowing possibilities implies that investments with a rate of return higher than the risk-free rate will not be undertaken. A similar mechanism operates when there are borrowing possibilities but investments are risky and there are no insurance possibilities. This implies that investing agents may be left with very little consumption if they are unlucky. As a result, risk-averse agents may choose not to undertake investments.

To illustrate the importance of financial market frictions, consider an environment in which agents can save but cannot borrow, similar to that of Aiyagari [1994] developed in Section 2.1.3.2. The difference now is that the amount of efficiency units of labor is not random but is the result of investment. Consequently, two different investment strategies are available: households can save in the financial asset a (which in this model is backed by real capital), or they can invest in their own human capital so that $s' = \phi(s, y)$, where y is the amount invested. The household's problem can be written as

$$\begin{aligned} v(s,a) &= \max_{c,y,a'} \; u(c) + \beta\, v(s',a';K), && (69)\\ & \qquad \text{subject to} && (70)\\ & c + a' + y = w\ s + a(1+r), && (71)\\ & s' = \phi(s,y), && (72)\\ & a' \geq 0. && (73) \end{aligned}$$

If constraint (73) is not binding, the first order conditions of this problem imply

$$w\ \phi_y(s,y) = 1 + r, \tag{74}$$

that is, the rate of return of the two types of investment is equalized. Moreover, imagine for simplicity that $\phi(s,y) = \phi(y)$; then in a steady state, all agents will have the same labor income. An interesting feature of this model is that convergence will arise immediately. That is, all households will have within a period the same labor income, since all agents will make the same investment in human capital. Differences in initial wealth perpetuate. For more general human capital production functions, we can get similar results, with the speed of convergence depending on the decreasing returns in s but not on y.

How would the analysis change when constraint (73) is binding? It all depends on the shape of function $\phi(s,y)$. Let's start with the case $\phi(s,y) = \phi(y)$. Assuming that the function $\phi(y)$ is strictly concave and $\phi_y(y)$ approaches infinity as y approaches zero, all households will make some investment and the first order condition is

$$u_c(c) = \beta\ \phi_y(y)\ u_c(c')w'. \tag{75}$$

This equation looks very similar to the Euler equation in the standard representative agent growth model, in which there is curvature in the production function. Consequently, no matter how poor they start out, all agents will slowly but steadily converge to a level of human capital that satisfies $\phi_y(y)w = \beta^{-1}$. So the economy converges to equal labor income, even if the wealth distribution can be very unequal. Policies that subsidize investment in human capital could speed up the equalization process but will not change the eventual convergence outcome.

A lot of concern remains about *poverty traps*, that is, situations in which households that start with insufficient initial resources never abandon their poverty status. For this situation to happen, some special assumptions are needed. In particular, $\phi(y)$ cannot be strictly concave. The typical assumption is to have a state of discontinuity such as a minimum expenditure that is needed to increase human capital. One example is the investment required for educational advancement. In this case, households compare the two options: whether to invest in education or not. If initial household wealth is very low, they may be unable to invest and still have positive consumption. But even households with slightly higher initial wealth may find that educational investment is feasible but not worth it, since it requires that initial consumption is way too low and the cost in utility terms too high. Clearly, in cases like these, government intervention can be fruitful because it is able to circumvent households inability to borrow. The government can tax richer households today and transfer resources to the poorer households, or it can borrow, transfer resources to the poor households, and tax them later after they have acquired more education. A policy that makes education compulsory even at the cost of severe current disutility will not be optimal because poor households could have chosen to do it themselves if this choice were preferable.

Another possibility in which the structure of financial markets matters is to have a stochastic return to the investment technology. Consider a version of equation (72) where higher investments in y yield a high expected value of s' but also a high variance. If the household had access to insurance markets, then it would happily undertake the investment, but if not, its risk aversion would prevent it from doing so. Again, in this case, certain government interventions that provide some form of insurance could be desirable.

4.2 Changes in the borrowing constraint

One way of assessing the role of financial constraints is to see what happens when they are relaxed. The Aiyagari [1994] model described in Section 2.1.3.1 assumes that financial markets are extremely underdeveloped: only one asset needs to be backed by physical capital, and there are no borrowing possibilities. What would happen if the financial constraints were to be relaxed, that is, what if we allowed for some noncontingent borrowing?

Table 7 shows the steady-state wealth distribution under various borrowing limits that go from a quarter of per-household yearly GDP to one year GDP. Figure 4 shows their associated Lorenz curves. We can see from the figure that, regardless of the calibration of the earnings process, inequality increases substantially with the relaxation of the borrowing constraint, in some cases to implausible levels (we cannot imagine an actual economy in which more than 60% of the population have negative financial assets). The Gini indices go up substantially, with one economy displaying a value above one, which is possible when we allow for negative values for the variable of interest. Looser borrowing constraints are associated with greater inequality because the poorest households want to borrow more. This result arises from the impatient nature of households in the general equilibrium. More specifically, households have a precautionary motive to save for the future when markets are incomplete, and on average they will never stop saving and will perpetually accumulate assets. But in a general equilibrium, the excess savings, which in aggregate takes the form of higher accumulation of capital, will drive down the marginal product of capital and, therefore, the return from savings. Consequently, in the steady-state equilibrium we have that $\beta^{-1} > 1 + r$, that is, households are more impatient than the return of their savings.[11] This result creates an incentive to anticipate consumption when the realization of earnings is low, which is made possible by the greater availability of credit (looser borrowing limit). This mechanism generates a greater concentration of wealth, as shown in Figure 4 and Table 7.

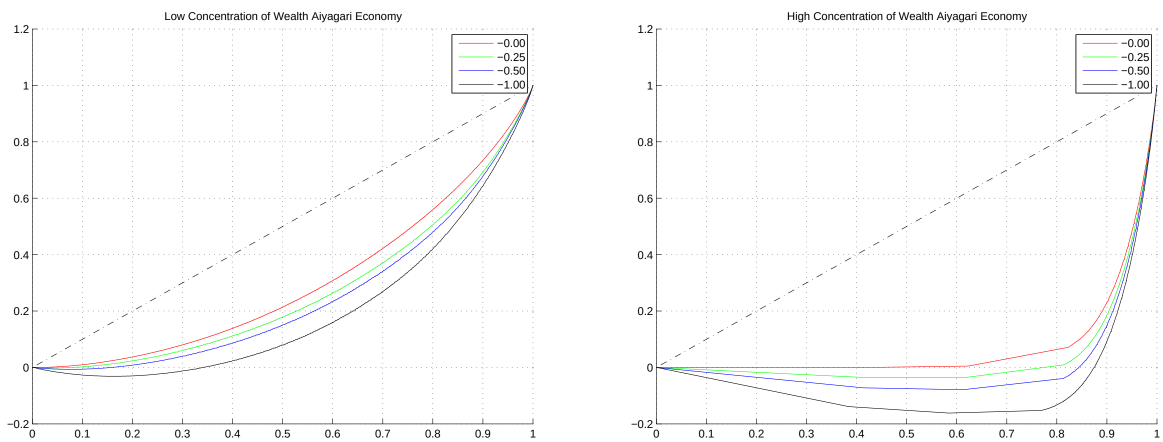
Another possibility is that improvements in the financial market allow agents not only to borrow more but also to buy insurance. In this case, agents could acquire assets or take liabilities with payments contingent on the realizations of idiosyncratic shocks. One consequence is that households will no longer save for precautionary motives, since they can completely insure their individual consumption. Furthermore, there would not be too much aggregate savings, as in the economy without insurance. So the economy would slowly reduce savings until the interest rate became equal to the rate of time preference. Individual consumption could differ across households, since consumption depends on the initial distribution of physical and human wealth, as in Chatterjee [1994] (see Section 2.1.1).

Another important question is, how much borrowing can be sustained? In a model without leisure choice, the maximum sustainable debt is the one that can be paid in all states of nature. The worst state of the world is the lowest possible value of s, which we refer to as $\underline{s}$. A household that receives the lowest realization of earnings forever has the capability to pay a maximum

[11] To better understand why β^{-1} must be bigger than $1 + r$ in a steady state general equilibrium, consider the following. Suppose that in a steady state $\beta^{-1} = 1 + r$. Given r, we can determine the stock of capital K from the equilibrium condition that equalizes the interest rate to the marginal product of capital. Lower interest rates must be associated with higher stocks of capital since the marginal product of capital is decreasing in K. Because agents face uninsurable risks, they save for precautionary reasons and, when $\beta^{-1} = 1 + r$, the average wealth accumulated by agents grows without bound (although individual wealth goes up and down stochastically, the average growth is positive). But in equilibrium the accumulated wealth ie equal to K. Therefore, if wealth increases, K also increases, reducing the marginal product of capital and, with it, the interest rate r. As the interest rate declines, households save less until the average growth rate of wealth for the aggregate economy is zero. We therefore conclude that in a steady state equilibrium $1 + r$ must be lower than β^{-1}.

FIGURE 4

Lorenz Curves for Various Economies and Borrowing Limits



(A) Low Concentration of Wealth Economy (PSID) Calibrated as in Aiyagari [1994]

(B) High Concentration of Wealth Economy (SCF) Calibrated as in Castañeda et al. [2003]

TABLE 7

Distribution of Wealth for Various Borrowing Limits

(in terms of per household yearly output)

Low Concentration of Wealth Economy (PSID)									
Borrowing Constraint	Quintiles 1st	2nd	3rd	4th	5th	Top 10%	Top 5%	Top 1%	Gini
0.00	3.75	10.14	16.97	25.06	44.08	26.73	15.51	3.99	0.41
-0.25	2.39	8.82	15.21	24.10	49.48	30.90	17.81	4.44	0.46
-0.50	0.82	7.77	14.72	24.56	52.13	32.52	18.62	4.69	0.50
-1.00	-2.98	5.37	13.60	26.11	57.90	35.88	20.73	5.32	0.60

High Concentration of Wealth Aiyagari Economy (SCF)									
Borrowing Constraint	Quintiles 1st	2nd	3rd	4th	5th	Top 10%	Top 5%	Top 1%	Gini
0.00	0.00	0.00	0.50	4.78	94.72	76.96	51.77	14.54	0.86
-0.25	-1.72	-1.72	-0.23	3.56	100.11	81.47	54.83	15.50	0.93
-0.50	-3.50	-3.50	-0.88	2.89	104.99	85.16	57.03	16.19	0.99
-1.0	-7.21	-7.10	-1.87	2.88	113.30	90.75	60.68	16.94	1.11

amount of interest $\underline{s}\, w$. Thus, the maximum sustainable debt is $\frac{\underline{s}\, w}{r}$, since the interest on this debt is exactly $\underline{s}$. Sometimes this is called the solvency constraint. Any debt larger than this value has a positive probability of not being paid.

4.3 Limits in the ability to borrow

So far, we have considered environments in which the access to credit markets is arbitrarily limited and there are no markets for contingencies. But why not? What limits the set of contracts that people can sign?

In a well-known and influential paper, Kehoe and Levine [1993] postulated that the ability of households to borrow is limited by their willingness to pay back when the alternative is to give up access to credit markets. In addition, a subset of the assets (physical or human or both) or endowments of the households can be seized, but not necessarily all of them. For example, future labor income may be outside of the reach of creditors.

Their approach does not preclude the existence of contingent markets. We would like to emphasize two features of this approach. The first feature is that it is the institutional environment that determines the set of contracts that are available. We think that this is an enormous advance compared with the literature that relies on exogenous borrowing limits. The second feature is that only the contracts that can be enforced ex post are available in the market. Therefore, once signed, there is complete compliance in the execution of these contracts. This feature is perhaps less appealing, since we see actual ex post reneging on formal contracts.

To show how this works, we could again modify slightly the Aiyagari economy. Let's define first the following object:

$$\overline{V}(s,a) = \max_{a' \geq 0} u\Big(a(1+r) + s\, w - a'\Big) + \beta \sum_{s'} \Gamma_{s,s'}\, \overline{V}(s',a'), \tag{76}$$

which is the household's value without having access to borrowing. Moreover, with some abuse of notation, let's define $\overline{V}(s) = \overline{V}(s,0)$ as the value attainable when the initial assets are zero.

Clearly, this limit depends on the value of the shock s. Now consider the following problem:

$$v(s, a) \quad = \quad \max_{c, y, a'} \;\; u(c) + \beta_i \; v(s', a'; K), \qquad \text{s.t.} \tag{77}$$

$$c + a' + y = w\, s + a(1 + r), \tag{78}$$

$$a' \geq \underline{a}(s), \tag{79}$$

where $\underline{a}(s)$ is such that

$$v\Big(s, \underline{a}(s)\Big) = \overline{V}(s). \tag{80}$$

In words, households can borrow up to the level in which they would be better off in an autarkic state, that is, in a state in which they start from zero assets and never borrow again. Notice that it is quite possible that in this situation high income people have more difficulties borrowing than low income people, and this is because $\overline{V}(s^H) > \overline{V}(s^\ell)$ when $s^H > s^\ell$. Notice also that $\underline{a}(s)$ is an endogenous variable. We do not know its value before solving for the equilibrium of this economy.

We have written problem (77) under the implicit assumption that assets can never be confiscated. If the legal system were such that assets could be taken away in absence of compliance, we could substitute equation (79) with

$$a' \geq \underline{\underline{a}}(s), \tag{81}$$

where $\underline{\underline{a}}(s)$ is such that $v(s, \underline{\underline{a}}(s)) = \overline{\overline{V}}(s)$ and $\overline{\overline{V}}(s) = u(s\, w) + \beta\, \sum_{s'} \Gamma_{s,s'}\, \overline{\overline{V}}(s')$. Essentially, the borrowing limit could be the amount that makes the agent indifferent between paying back the lender or defaulting and being forever unable to save or borrow.

In this model, all contracts are carried out, that is, loans and state-contingent contracts are always honored. In reality, however, many people file for bankruptcy. For example, in the twelve months between April 1, 2012, and March 31, 2013, 779,306 people filed for bankruptcy in the United States.[12] In some countries like the United States and Canada, debts are typically discharged after filing, whereas in other countries like Hungary, Romania, and Spain, there is no legal procedure to handle personal bankruptcy and people are always liable for previous debts. The rest of the countries lie somewhere in between these extremes.

One possible strategy to deal with the pervasiveness of bankruptcy is to model it as a contin-

[12] See United States courts, http://www.uscourts.gov/.

gency fully negotiated ex ante by the parties. This strategy is hard to justify, however, since filing for bankruptcy is a legal procedure that can be completed unilaterally by the debtor. Hence, it is a right that cannot be forfeited. We need, then, to have explicit models that explicitly incorporate bankruptcy filings. One approach, followed within the optimal contracting tradition, is to assume that there are information asymmetries and costly state verification, as in Townsend [1979]. The costly state verification model has been widely applied in macroeconomics, for example, in Bernanke and Gertler [1989], Bernanke et al. [1999], and Carlstrom and Fuerst [1995]. In these models, default arises in equilibrium even if agents sign fully optimal contracts. In the next section, we will describe other approaches that are more in line with the literature that excludes the applicability of fully optimal contracts.

4.4 Endogenous financial markets under actual bankruptcy laws

During the last few years, considerable work has been done to bring together models with imperfect insurance and models with a legal system that allows agents to file for bankruptcy in a way that is similar to that of Chapter 7 in the U.S. bankruptcy code (Chatterjee et al. [2007] and Livshits et al. [2007]). We now present a version of these models and describe how their implications for the income and wealth distribution change compared with the basic Aiyagari model. These studies take advantage of a feature of the legal system that lists people that have filed for bankruptcy in public records for a certain number of years. The literature interprets the implications of the listing as limiting accessibility to borrowing for the duration of the public record.

Consider the following household problem, yet another variant of the basic Aiyagari problem:

$$\begin{aligned} v(s,a,1) &= \max_{c,a'\geq 0} \; u\Big(w\;s + a(1+r) - a'\Big) + \\ &\qquad \beta \sum_{s'} \Gamma_{s,s'}\Big[(1-\delta)v(s',a',1) + \delta v(s',a',0)\Big], \end{aligned} \tag{82}$$

$$\begin{aligned} v(s,a,0) &= \max\Big\{u(sw) + \beta \sum_{s'} \Gamma_{s,s'}\; v(s',0,1), \\ &\qquad \max_{c,a'} \; u(c) + \beta \sum_{s'} \Gamma_{s,s'}\; v(s',a',0),\Big\} \end{aligned} \tag{83}$$

$$\text{s.t.} \qquad c + q(s,a')\; a' = w\; s + a. \tag{84}$$

The function $v(s,a,h)$ is the household's value function, with the last argument $h \in \{0,1\}$

denoting the household's credit history. When $h = 1$, the household's credit history is bad in the sense that the agent has defaulted in the near past and is prevented from having access to credit. Problem (82) depicts this case, showing that in the following period, its credit history may turn out to be good, $h = 0$, with parameter δ controlling the expected duration of market exclusion.

Problem (83) is of interest when the credit history is good, $h = 0$, and we have written it compactly, implicitly assuming that the household is in debt, $a < 0$. Here, the agent has two options: to file for bankruptcy or not. If the agent files, three things happen: household consumption equals current labor income, sw, credit history turns bad next period, $h' = 1$, and the household is prevented from saving. The latter property is a feature of the bankruptcy code, since the agent is not permitted to keep assets after filing for bankruptcy.[13] If the household does not file for bankruptcy, it can borrow or save as it wishes. Note, however, that we have written the budget constraint (84) differently from previous problems. The left-hand side, the uses of funds, has the asset position at the beginning of the following period multiplied by $q(s, a')$. This is the household-specific inverse of the interest rate. Lenders accurately forecast that the agent may file for bankruptcy and charge an extra premium so that in expected value they get the market return. The function $q(s, a')$ is an equilibrium object. If the household chooses to save, $a' \geq 0$, and the inverse of the interest rate is that of the safe asset: $q(s, a') = (1 + r)^{-1}$.

The optimal solution to the problem of a household with negative assets is to default for a range of its earnings. The set of earnings for which the household defaults increases with the stock of debt.

The solution to this problem has two interesting properties. First, since default is costly (the household will not be able to borrow for a while), the household would not default if the debt is very small. Second, in some circumstances, the household may be too poor to default, opting instead to borrow even more for a sufficiently low realization of earnings. Consequently, the equilibrium of this model requires that the inverse of the interest rate $q(s, a')$ is such that lenders break even in expected value, which in turn implies that interest rates are increasing in the amount borrowed and in the likelihood of bad earnings realizations.

This structure has proved useful in making sense simultaneously of the extent of unsecured borrowing in the United States as well as the frequency of bankruptcy filings, especially if the model is enhanced with a few bells and whistles such as expenditure shocks.

[13] In the United States, the agent can keep a maximum amount of assets, and this amount varies across states. Here, we have assumed zero retainable. In this discussion we are abstracting from are other subtleties of the bankruptcy code, such as the requirement that labor income is below the state median's income.

A weakness of this approach Why are households refused credit when they have a bad credit history? Nothing in the law requires this. In fact, the opposite is true: in the United States, a bankruptcy filing under Chapter 7, the relevant case, precludes additional filings over a period of years, making a recent filer appear to be a better creditor than somebody with a clean slate.

This question has three possibly answers, but none are completely satisfactory. First, a Nash equilibrium with a coordination problem can be constructed when lenders believe that agents with bad credit histories will not pay back and hence they will not lend, whereas prospective borrowers might as well choose to default, since they do not receive credit. Although this is indeed a Nash equilibrium, it is one that is always present in the event of lending, and there is no argument for why it happens only with a bad credit history. Another possibility is to construct a trigger strategy whereby lenders coordinate not to lend during a punishment period in the event of default. But like all triggers, this is not an equilibrium of the limit of finite economies. Hence, it is not a Markov equilibrium. Many economists are comfortable with trigger strategy equilibria, whereas while others are not. The last rationale for exclusion of those with bad credit is to postulate the existence of a regulator that prevents lenders from lending to those with bad credit, something that is not actually done by any of the banking regulators.

4.5 Credit Scoring

Chatterjee et al. [2008] and Chatterjee et al. [2004] propose a solution to the weakness of models based on exogenous exclusion after bankruptcy filings. These papers note that in the United States, there is pervasive use of credit scores, which are assessments of reliability made by independent companies. The authors then pose a model in which two types of people differ in some fundamental attribute associated with reliability that is not directly observable by outsiders—for instance, patience or even good driving habits. The credit score is then used as the market assessment of being a good type, meaning the type that is more likely to pay back debts or be reliable. In this context, both types of agents fall under the model of borrowing in which there are multiple types. The key here is that both types of agents—both the patient and the impatient agents—want to repay their debts to signal that they are patient, which allows them to have access to better borrowing terms. In this context, filing for bankruptcy increases the market-assessed likelihood that an agent is of the bad type, which translates to a severe worsening of loan terms, if not an outright exclusion of future credit. Moreover, because the market is assessing traits that are relevant not only for the repayment of credit but also for other things (e.g., cheap property insurance, access to rental property, personal relationships), timely repayment of debts carries a strong incentive that allows for the possibility of many contracts to be carried out, even if the law lacks the necessary teeth to enforce these contracts.

4.6 Financial development and long-run dynamics

We now look at the extent to which access to financial markets can help us to understand the long-run dynamics of the economy. In Section 4.6.1 we briefly focus on long-term growth, and in Section 4.6.2 we discuss how the evolution of financial markets can also help us understand the issue of global imbalances, that is, the emergence of large and persistent balance of payments deficits.

4.6.1 Long-run growth and financial development

The Schumpeterian view places entrepreneurship at the center stage of economic development. Due to financial constraints and the lack of insurance markets, however, entrepreneurial investment is sub-optimal. Essentially, when financial markets are not well developed, resources cannot be redistributed from those who control the resources but do not have the best uses of these resources to those who have the best investment opportunities but lack the funds. This efficiency problem is especially severe when the distribution of resources is particularly concentrated. We may then end up with a situation in which the poor become (relatively) poorer because they cannot take advantage of investment opportunities and the economy as a whole grows less. Examples of studies that emphasize the importance of inequality for growth in the presence of financial constraints are Galor and Zeira [1993], Banerjee and Newman [1993], and Aghion and Bolton [1997]. Because these studies were already reviewed by Bertola [2000] in a previous edition of the Handbook, we do not repeat their description in this chapter.

A more recent literature, however, also emphasizes that market incompleteness—that is, environments in which the trade of state-contingent claims is limited—could have both positive and negative effects on capital accumulation. In a world with only uninsurable and exogenous earning shocks as in Aiyagari [1994], market incompleteness generates more capital accumulation and, therefore, more growth. When risky income is endogenous, however, as in Angeletos [2007], market incompleteness may discourage investment. See also Meh and Quadrini [2006].

Another group of studies that investigate the relation between inequality and macroeconomic performance emphasize the importance of social conflict and expropriation. Greater inequality often associated with underdeveloped financial markets means that a larger group of individuals are at the bottom of the distribution and face poor economic conditions compared with the rest of the population. Faced with poor economic conditions and the feeling that the prospects for economic improvement are impaired by the excessive concentration of wealth, the resentment toward the rich starts to rise, which creates incentives to expropriate either by stealing or through revolutions. The risk of expropriation has two negative effects. First, it acts as an investment tax that discourages investment. Second, agents devote more resources to protect property rights instead of using the resources for productive and growth-enhancing activities. Benhabib and Rustichini [1996] develop a model that formalizes this idea. Although not explicitly considered in this paper, financial underdevelopment could contribute to this because it makes it more difficult for poor people to escape from poverty.

Another theory of inequality affecting growth is developed in Murphy et al. [1989]. This paper assumes that some technologies have increasing returns. These technologies become profitable only if the domestic market is sufficiently large, that is, enough demand exists for the goods produced with the new technologies. If wealth is highly concentrated, the domestic market remains small (since not enough consumers can afford these goods). As a result, growth-enhancing technologies will not be implemented. The paper does not explicitly explore the role of financial markets; however, to the extent that financial underdevelopment creates the conditions for greater concentration of wealth, the mechanism described in this paper becomes more relevant in economies in which the financial structure is relatively underdeveloped.

Kumhof and Rancière [2010] have proposed an explanation for the recent crisis based on the changes in income distribution pinpointing similarity with the Great Depression. The idea is that, due to an exogenous shock that affected the ability of the rich to grab earnings, income became more concentrated, and as a result, the poor started to borrow more, increasing the debt-to-income ratio in the economy. Eventually, the increase in borrowing triggered the crisis.

We conclude this section by citing the work of Greenwood and Jovanovic [1990]. Although this paper does not deal directly with inequality, it shows that improvements in financial markets (in this particular case, through the information gathered by financial intermediaries) have important effects on economic growth. As we have seen in previous chapters, market incompleteness also creates inequality. Therefore, once complemented with the previous analysis, this paper could also be relevant for understanding the link between financial market development, inequality, and growth. Also important is the work of Greenwood et al. [2010].

4.6.2 Global imbalances

We have not talked much about cross-country inequality because this topic is usually a concern for development-oriented economists. However, inequality may be shaped by the increase in trade that is properly known as globalization, which is due to the reduction in the trade barriers for both technological and policy reasons. We have already referred, if only obliquely, to a mechanism by which more trade across countries could affect inequality: opening to trade changes the relative price of skills and may be behind part of the recent increase in the wage-skill gap. But an increase in trade shapes inequality both within and between countries trough other, more subtle mechanisms. In this section we illustrate some potential mechanisms through which inequality is linked to globalization. Further analysis of the role of globalization for inequality is conducted in chapter 21.

The process of international globalization is commonly presented as taking the form of higher trade in goods and services (imports and exports) as a fraction of GDP. But there is another side to it. Several advanced countries, the United States in particular, have experienced over the last 30 years a persistent deficit in the balance of payments as a result of imports being higher than exports, with the consequent deterioration in their net foreign asset positions. On the other hand, oil-producing countries and several emerging countries, China in particular, have been accumulating positive net foreign asset positions. Global Imbalances is the term often use to refer to the situation in which some countries accumulate large negative net foreign asset positions while others accumulate positive net foreign asset positions. This situation has affected inequality, but to understand the impact on inequality we first need a theory of why imbalances could emerge in the wave of globalization.

Mendoza et al. [2007] provide one such theory. They claim that sustained deficits cannot be explained solely with traditional trade forces (different factor prices, technological advantages, or lower transportation costs). We also need to understand the differential saving behavior of countries which in equilibrium lead to different rates of returns on savings (insofar as international financial markets are somewhat segmented). This is possible even if countries have identical preferences and production technologies, but agents in each country differ in the extent to which they are capable of insuring their individual risks. This can be illustrated with the now familiar Aiyagari economy.

Suppose that we compare two economies, both slightly modified versions of the Aiyagari environment described above, that differ only in the process for earnings, one being more volatile than the other. It is important to point out that the assumption that countries differ in the volatility of earnings is a short cut to captures other, more micro-founded differences. For example, in Mendoza et al. [2007] countries do not differ in the underlying process for earnings but in the sophistication of financial markets. Agents (consumers and firms) in countries with more advanced financial markets have a better opportunity to insure their idiosyncratic risk. Since in terms of savings the implication of higher insurance is similar to lower variability of earnings, here we illustrate the mechanism by assuming lower earning volatility. In some applications the higher ability to insure could derive from government policies (for example, the provision of public funded health insurance). In some cases, the differences could come from more uncertainty about the underlying process for earnings. For example, a country that is experiencing a process of transformation (like China during the last three decades) is also possibly characterized by greater uncertainty at the individual level. Independently of the actual sources (greater ability to insure or greater underlying uncertainty), it should be clear that the example provided here is just a short cut to illustrate something more fundamental such as differences in the characteristics of the financial system.[14]

To this end, we use two different processes for earnings. The first process is what we used above for the version of the model that we called PSID economy or low variability economy. The second process is what we used in the high variability economy, a version of the SCF economy with a slightly less extreme good state. Besides the process for earnings, the two economies are alike in all other dimensions.

The first two columns of Table 8 display the steady states of these two economies under autarky. The first column, the low variability economy, has a capital-to-output ratio of 3.34, implying an annual interest rate of 4.02%. Because in this economy wealth can only take the form of capital, total wealth is also 3.34 times output, and this is what households choose to hold to accommodate the shocks to earnings given the 4.02% interest rate. The second column of Table 8 refers to the economy with higher income variability also in the autarky regime. Households choose to hold more wealth (3.88 times output) to bear the high risk. Two things to note are that the interest rate is now much lower, 1.27% and that output is slightly higher due to the higher capital.

The determination of the equilibrium is depicted in panel a) of Figure 5. This figure plots the aggregate (steady-state) supply of savings as an increasing, concave function of the interest rate.[15]

[14]See Mendoza et al. [2007] for more details on how differences in the financial system can lead to lower ability to insure.

[15]Aggregate savings converge to infinity as the interest rate approaches the rate of time preference from below, because agents need an infinite amount of precautionary savings to attain a nonstochastic consumption profile.

TABLE 8
Two Economies before and after Being Able to Borrow from Each Other

Economy	Before: Autarky		After: Mobility	
	Low Var	High Var	Low Var	High Var
Capital to Output ratio	3.34	3.88	3.67	3.67
Interest Rate (%)	4.02	1.27	2.24	2.24
Wealth to Output ratio	3.34	3.88	0.39	6.95
Gini Index of Wealth	0.41	0.59	0.50	0.39
Coeff. of Var Wealth	0.76	1.09	0.88	0.96
1st quantile	3.40	0.00	0.40	0.00
2nd quantile	10.21	0.69	6.06	3.35
3rd quantile	17.11	15.50	17.58	17.45
4th quantile	25.16	30.21	25.94	21.00
5th quantile	44.15	53.60	50.01	58.20
top 10% (cumulative)	26.73	33.28	30.37	29.63
top 5% (cumulative)	15.51	19.72	18.09	17.21
top 1%	3.99	5.19	4.25	4.32

The demand for savings is downward sloping because of the diminishing marginal productivity of capital. Country 1 has a lower volatility of individual earnings and hence a lower supply of savings for each interest rate. As a result, the equilibrium in autarky implies a higher interest rate and lower total capital.

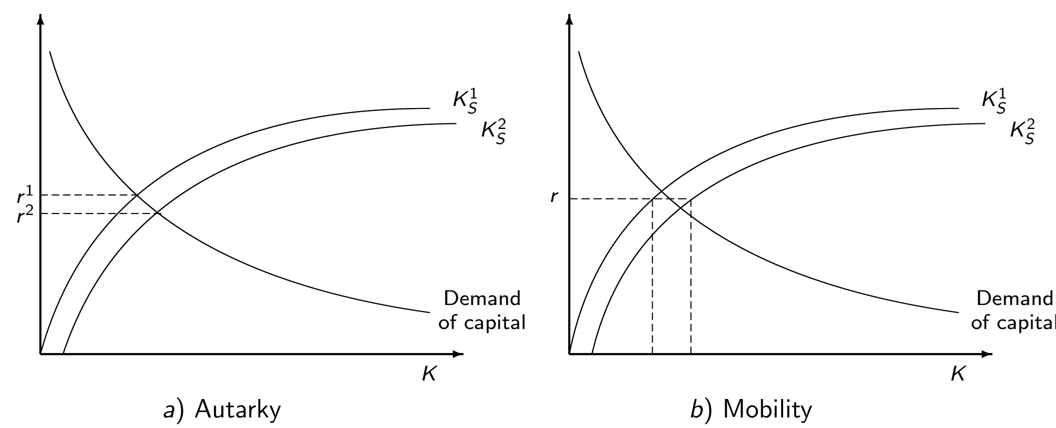


FIGURE 5
Steady-state equilibria with heterogeneous earning risks.

Imagine now that households in these two economies can start owning capital in the other country, that is the countries become financially integrated. After a brief period of transition that depends on the ease with which physical capital can flow or be reallocated, the interest rate in both countries will be equalized. This implies that the low variability economy will experience a reduction in the interest rate and the high variability economy will experience an increase in the interest rate. Then, in the country in which the interest rate decreases (low variability economy), savings will fall, while in the country in which the interest rate increases (high variability economy), savings will rise. The result is that households in the high variability economy end up owning part of the capital installed in the low variability economy. In this way, global imbalances may emerge as the low variability economy dis-saves. Effectively, the low variability economy consumes and invests more than it produces, with the difference covered by imports in excess of exports (trade deficit).

This process takes a long time until the aggregate savings of households in each country no longer change. This new steady state is depicted in panel b) of Figure 5. The world interest rate is somewhere between the pre-liberalization interest rates in the two countries. Compared with autarky, the interest rate and the supply of savings fall in country 1 and rise in country 2, and hence the country with lower volatility of earnings ends up with a negative foreign asset position. Moreover, the capital stock rises relative to its autarky level in country 1 and falls in country 2. Thus, financial globalization leads capital to flow from economies with more risk to those with lower risk.

For analytical simplicity, we have modeled this process as the outcome of countries that differ in their earnings risk. However, as emphasized above, this this just a short cut to capture other types of differences across countries that ultimately lead to difference exposure to risk. It could very well be the case that the underlying risk is identical across countries but that the lower risk in one country is just the result of more developed financial markets, which allow for higher insurability of risk. Formally, in the first country there are more markets for state-contingent claims. This is the approach taken in Mendoza et al. [2007], and the end result is similar to the case of differential processes for earnings: the country with a higher ability to insure saves less and has a higher interest rate than the country with less developed financial markets. When the two countries integrate, it is the more financially developed country that accumulates negative foreign assets while the less financially developed country accumulates positive foreign assets.

Therefore, financial market differences can affect the distribution of wealth across countries: in the long run, countries that are more financially sophisticated become poorer relatively to countries that are less financially sophisticated (compared to the pre-liberalization era). This, however, does not mean that liberalization is welfare reducing for developed countries and welfare improving for less developed countries. In Mendoza et al. [2007] we found, somewhat surprisingly, that liberalization was welfare improving for developed countries but slightly welfare reducing for less developed countries (based on equally weighted welfare function). In our example displayed in Table 8, the international redistribution of wealth is quite large, with the low variability country ending up with barely 5% of total wealth. Yet, it started with almost half. The large international redistribution of wealth follows from the assumption that there are large differences in risk between the two countries. In reality, especially among integrated countries, the differences in risk may not be that big. Also, when a country accumulates too many foreign liabilities, there could be an incentive to default on these liabilities. This imposes a limit on the redistribution of wealth that can be generated across countries through this mechanism. Nevertheless, this example suggests that differences in savings could generate significant inequality in wealth across countries.

Cross-country financial market heterogeneity also plays a central role in Caballero et al. [2008] for explaining global imbalances. The mechanism proposed in this paper does not rely on risk but on the availability of saving instruments. The idea is that in certain countries, savers have difficulty storing their savings in high return assets. The implications for global imbalances, however, are similar to Mendoza et al. [2007]. The two mechanisms are complementary ways of thinking about how the characteristics of financial systems can shape the distribution of wealth across countries in a globalized world. Interestingly, these contributions illustrate another mechanism through which financial globalization redistributes wealth. When productive inputs are not perfectly reproducible (as in the case of land), liberalization also leads to the equalization in the prices of these assets. Since under autarky these assets were cheaper in financially developed countries, these countries experience capital gains while countries with less developed financial markets experience capital losses.

The process of international redistribution of wealth also has consequences for the internal wealth distribution within each country. We see how wealth concentration increases in the low variability country as measured by either the Gini index of the coefficient of variation of wealth or even by the shares held by the richest households (see Table 8). The opposite process happens in the less financially developed country, where the wealth distribution becomes more equal after international financial integration. Perhaps this process has contributed, at least in part, to the increased wealth concentration in the United States that we documented earlier.

5 The political economy channel

We have already seen in the previous sections some channels through which the distribution of income and wealth is interconnected with the aggregate performance of the economy. In this section, we discuss one particular channel through which inequality affects economic activities, that is, through the political and institutional system. Since many policies have redistributive consequences, the degree of inequality plays a central role in the choice of policies because societies with more unequal distributions of resources might demand greater redistribution. Since redistributive policies are often distortionary, the result is that more unequal societies tend to experience lower income or growth (or both).

Many contributions emphasize this mechanism, starting with Meltzer and Richard [1981]. Examples are Persson and Tabellini [1994], Alesina and Rodrik [1994], Krusell and Ríos-Rull [1996], and Krusell et al. [1997]. Many of these contributions, however, ignore individual uncertainty which, in a dynamic environment could play an important role in affecting the demand for redistribution as well as the distortions associated with redistributive policies. The goal of this section is to present a simple framework that illustrates the central idea of the early literature. It shows how the consideration of idiosyncratic uncertainty enriches the analysis and makes the relation between inequality and redistribution more complex than in these early studies.

5.1 A simple two-period model

Suppose that there is a continuum of agents who are alive for two periods. Agents value consumption, c_t, but dislike working, h_t, according to the utility function

$$u\left(c_t - \frac{h_t^2}{2}\right).$$

There are two sources of income: endowment, η_t, and labor, h_t. Individual endowments evolve according to

$$\ln(\eta_{t+1}) = \rho \cdot \ln(\eta_t) + \epsilon_{t+1},$$

where $\epsilon_{t+1} \sim N\Big(0\,,\,(1-\rho^2)\cdot\sigma^2\Big)$. This implies that the economy-wide distribution of log-endowments is normal with mean zero and variance σ^2, that is, $\ln(\eta_t) \sim N(0\;,\;\sigma^2)$. By changing ρ we change the persistence of endowments but we keep the economy-wide distribution (inequality) constant. This parameter determines the degree of mobility: higher values of ρ imply lower mobility.

Before continuing, it will be helpful to derive some of the key moments of the cross-sectional distribution of endowments. Since endowments are log-normally distributed, that is, $\eta \sim LN(0, \sigma^2)$, the mean and the median are, respectively,

$$\text{Mean}(\eta_t) = e^{\frac{\sigma^2}{2}}, \qquad \text{Median}(\eta_t) = 1.$$

These are unconditional moments. Also convenient is to derive the expected next period endowment for an individual with current endowment η_t^i. The conditional mean is

$$\mathbb{E}[\eta_{t+1}^i | \eta_t^i] = e^{\rho \ln(\eta_t^i) + \frac{(1-\rho^2)\sigma^2}{2}}.$$

This conditional expectation will play an important role in the analysis of the model. Here we observe that, if $\rho = 0$, the expected next period endowment is the same for all agents. For an agent with median endowment, that is, $e_t^m = 1$, the conditional expectation becomes

$$\mathbb{E}[\eta_{t+1} | \eta_t^m] = e^{\frac{(1-\rho^2)\sigma^2}{2}}.$$

We can then compute the ratio of next period economy-wide average endowment over the next period endowment expected by an agent whose current endowment is the median value. This is equal to

$$\frac{\text{Mean}(\eta_{t+1})}{\mathbb{E}[\eta_{t+1} | \eta_t^m]} = e^{\frac{\rho^2\sigma^2}{2}}.$$

This expression makes clear that the difference between the average endowment and the endowment expected by the agent with the median endowment in the current period depends on the persistent parameter ρ. The difference becomes zero if there is no persistence, that is, $\rho = 0$, and it is maximal when $\rho = 1$. Although the parameter ρ affects the ratio between the average endowment and the expected endowment by the median agent, ex post inequality does not depend on ρ. In fact, we have that

$$\frac{\text{Mean}(\eta_{t+1})}{\eta_{t+1}^m} = e^{\frac{\sigma^2}{2}}.$$

We will use these moments below, after completing the description of the model.

The government taxes incomes, from endowment and labor, at rate τ_t and redistributes the revenues as lump-sum transfers. The budget constraint for the government is

$$T_t = \tau_t \int_i (\eta_t^i + h_t^i) di,$$

where i is the index for an individual agent.

Agents do not save and solve a static optimization problem. Given the tax rate and the transfer, an individual agent i maximizes the period utility by choosing the labor supply h_t^i, subject to the following budget constraint:

$$c_t^i = (\eta_t^i + h_t^i)(1 - \tau_t) + T_t.$$

Taking first order conditions with respect to h_t for an individual worker with endowment η_t^i, we get the supply of labor $h_t = 1 - \tau_t$. Substituting in the utility function and using the equation that defines the government transfers, we get the indirect utility for period t:

$$U^i(\tau_t) = u\left(\tau_t \int_\eta \eta dF_t + \tau_t(1-\tau_t) + \eta_t^i(1-\tau_t) + \frac{(1-\tau_t)^2}{2}\right).$$

Now suppose that agents vote for the next period tax rate τ_{t+1}. The tax rate preferred by an agent with current endowment η_t^i maximizes the expected next period indirect utility, that is,

$$\max_{\tau_{t+1}} \mathbb{E}_t \left[u\left(\tau_{t+1} \int_\eta \eta dF_{t+1} + \tau_{t+1}(1-\tau_{t+1}) + \eta_{t+1}^i(1-\tau_{t+1}) + \frac{(1-\tau_{t+1})^2}{2}\right) \Bigg| \eta_t^i \right],$$

where we have denoted by $F(\eta)$ the distribution of endowments. Since the log-endowments are normally distributed, $F(\eta)$ is a log-normal distribution.

Notice that the voter forms expectations about the future endowment conditional on the current endowment. Of course, the higher is the persistence, the higher the dependence of the expected value from the current value.

Taking the first order condition, we derive

$$\tau_{t+1}^i = \int_\eta \eta \, dF_{t+1} - \mathbb{E}[\eta_{t+1}^i | \eta_t^i] - \frac{\text{Cov}(dU_{t+1}^i, \eta_{t+1}^i | \eta_t^i)}{\mathbb{E}[dU_{t+1}^i | \eta_t^i]}, \tag{85}$$

where dU_{t+1}^i denotes the derivative of the indirect utility for agent i with respect to the next period tax rate. Notice that this term also depends on the tax rate. The above condition implicitly determines the tax rate.

The first term on the right-hand-side of (85) is the mean value of the economy-wide endowment, which is equal to $e^{\frac{\sigma^2}{2}}$. This term is the same for all agents. The second term is the expected endowment of agent i given the current endowment. This term is increasing in η_t^i, unless $\rho \leq 0$, which is excluded by assumption. Therefore, ignoring the third term, the preferred tax rate decreases with the current endowment.

The third term captures the role of risk aversion. Because the utility function is strictly concave and its derivative is strictly decreasing, $dU_{t+1}^i(.)$ decreases with the realization of next period endowment η_{t+1}^i, implying that the covariance term is negative. Therefore, preferences for taxes increase with the concavity of the utility function. This is the effect of risk aversion.

The case of risk neutrality. Because the third term in the first order condition (85) is itself a function of τ_{t+1}^i, it is difficult to derive an analytical expression for the tax rate. Therefore, we first specialize to the case with risk-neutral agents so that $\text{Cov}(dU_{t+1}^i, \eta_{t+1}^i | \eta_t^i) = 0$ and the preferred tax rate reduces to the first two terms in equation (85). We can then establish that the preferred tax rate is monotonically decreasing in the current endowment η_t^i and the equilibrium tax rate is the one preferred by the agent with the median endowment. Using the fact that endowments are log-normally distributed and the log-endowment of the median voter is zero, the conditional expectation of the median voter for the next period endowment is $\mathbb{E}[\eta_{t+1} | \eta_t^m] = e^{\frac{(1-\rho^2)\sigma^2}{2}}$, whereas the economy-wide average is $\int_\eta \eta dF_{t+1} = e^{\frac{\sigma^2}{2}}$. Substituting in the preferred tax rate we obtain the equilibrium tax rate

$$\tau_{t+1}(e_t^m) = e^{\frac{\sigma^2}{2}} - e^{\frac{(1-\rho^2)\sigma^2}{2}}. \tag{86}$$

The first two terms capture the standard politico-economy theory: since the average endowment, $e^{\frac{\sigma^2}{2}}$, is bigger than the median endowment, $e^{\frac{(1-\rho^2)\sigma^2}{2}}$, there is demand for redistribution. If we increase inequality by raising σ, the demand for redistribution increases. Since the optimal effort chosen by all agents is $h = 1 - \tau$, higher taxes discourage effort with negative effects on aggregate production. In some of the models proposed in the literature, taxes distort the accumulation of capital instead of effort, but the idea is similar.

The mechanism described above links inequality to redistribution and macroeconomic activity and captures the key features of the model studied in Meltzer and Richard [1981]. In addition to this mechanism, the model presented here emphasizes the role of mobility captured by the parameter ρ. If we reduce ρ so that the economy experiences higher mobility, the cross-sectional inequality does not change. In fact, the ratio of average endowment and median endowment remains $e^{\frac{\sigma^2}{2}}$. However, the tax rate preferred by the median voter declines, as we can see from equation (86). Even if the median voter has low endowment in the current period, what matters for next period taxes is the future endowment. If mobility is high, the median voter does not expect to keep the low endowment in the future. Thus, it is not optimal to choose high tax rates. In the limiting case with $\rho = 0$, the expected future endowment for all agents will be the average endowment and, in expected terms, the future benefit of redistribution is zero for all agents.

The importance of mobility for political preferences has received less attention than cross-sectional inequality. But the simple model presented here shows that mobility is also an important factor in the determination of political preferences. More importantly, if inequality and mobility are not independent, either across countries or across times, by focusing only on inequality we may reach inaccurate conclusions. Suppose, for example, that an increase in cross-sectional inequality, σ, is associated with a decrease in ρ, that is, with an increase in mobility. Therefore, we have two contrasting effects: the increase in σ leads to higher taxes, whereas the decrease in ρ leads to lower taxes.

This example may help to explain why, in certain episodes of increasing inequality, such as in the United States before the recent crisis, we do not see a significant increase in demand for redistribution. Perhaps the reason is that voters perceive higher mobility as coincident with greater inequality. Then, thanks to the perceived mobility, voters do not demand higher taxes and the economy continues to perform well even if income becomes more concentrated. However, if the performance of the economy changes and voters start to perceive lower mobility, they will start demanding more redistribution, which will further deteriorate the performance of the economy. This idea has been developed in Quadrini [1999] in a model that features two equilibria. The first equilibrium is characterized by high growth, high inequality, and low redistribution. The second equilibrium is characterized by low growth, low inequality, and high redistribution.[16] The idea that the prospect of upward mobility reduces the demand for redistribution has also been studied in Benabou and Ok [2001].

[16]The increase in inequality in the United States is not a recent phenomenon. However, voters and politicians started to focus more on this issue after the recent crisis. During the good times in which financial markets were expanding, low income households had access to credit, allowing them to own houses. For many this appeared as a new opportunity (mobility). However, with the crisis and the credit market and the freeze, these opportunities dried up and many households lost faith in the possibility of improving their current position (mobility). Not surprisingly, they turned to the government for help and asked for more populist policies.

The case of risk aversion. We now assume that the utility function is concave and takes the following form:

$$u\left(c_t - \frac{h_t^2}{2}\right) = \frac{\left(c_t - \frac{h_t^2}{2}\right)^{1-\nu}}{1-\nu},$$

where the parameter ν captures the curvature of the utility function.

Figure 6 plots the preferred tax rate as a function of current endowment η_t^i for different values of ν. As can be seen from the figure, the preferred tax rate is monotonically decreasing in current endowment, and therefore, the median voter theorem also applies in the case of risk-averse agents. Furthermore, we see that, for each endowment level η_t^i, the preferred tax rate increases with risk aversion.

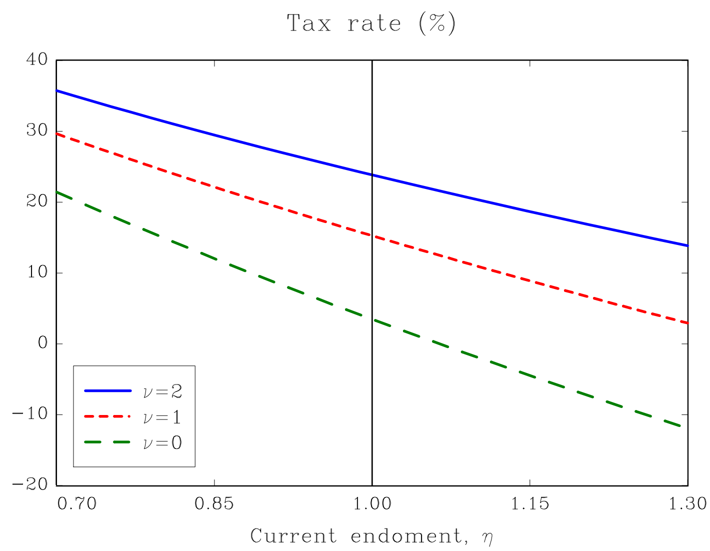


FIGURE 6
Preferred tax rates for different degrees of risk aversion. The inequality and mobility parameters are $\sigma = 0.5$ and $\rho = 0.5$. The vertical line denotes the median endowment.

Figure 7 plots the preferred tax rate as a function of current endowment for different degrees of mobility and risk aversion. The first panel is for the case of risk neutrality. In this case, we see that lower mobility (ρ changes from 0.5 to 0.9) increases the equilibrium tax rate, that is, the tax rate preferred by the median voter, which in the figure is identified by the vertical line. These are the properties we have shown analytically in the previous subsection. However, when agents are risk averse, lower mobility reduces the equilibrium tax rate. The reason is that, conditional on the current endowment, lower mobility means that agents face lower risk. In fact, with $\rho = 1$, next period endowment is equal to current endowment. Thus, there is less demand for insurance.

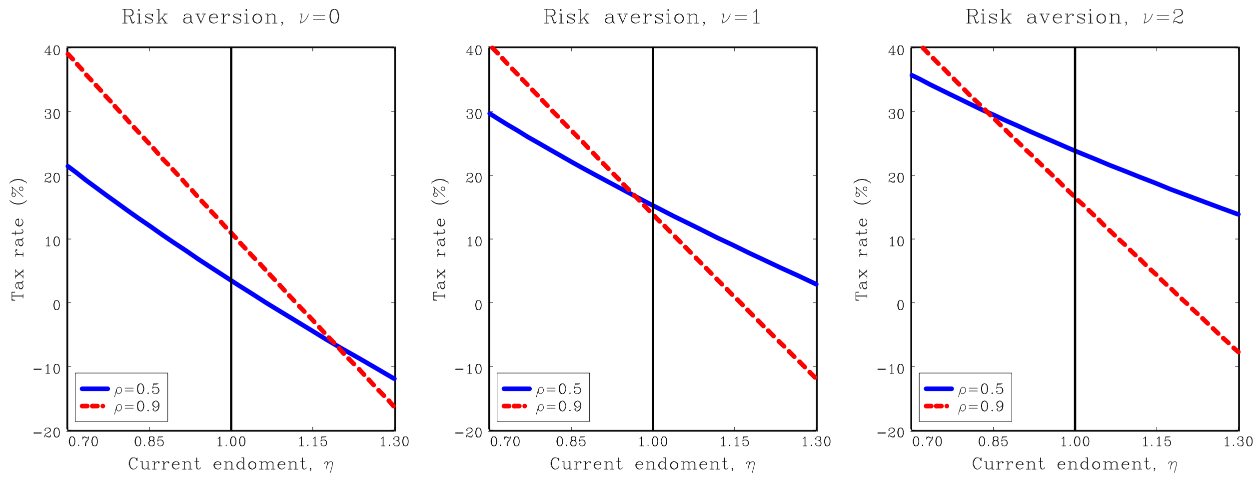


FIGURE 7
Preferred tax rates for different degrees of mobility, $\rho \in \{0.5, 0.9\}$, and risk aversion, $\nu \in \{0, 1, 2\}$. The inequality parameter is $\sigma = 0.5$. The vertical line denotes the median endowment.

This example shows that mobility affects equilibrium policies through two mechanisms. The first mechanism works through the impact of mobility on the redistributive gains from next period taxes. When mobility is low, the expected redistributive gains are high. These gains vanish if mobility is perfect, that is, $\rho = 0$. The second mechanism works through the impact of mobility on individual risk. Given the current endowment, higher mobility (lower ρ) increases the conditional volatility of next period endowment and the median agent faces higher risk. Thus, greater redistribution is preferred if preferences are concave. This second mechanism is irrelevant when agents are risk neutral but becomes important when agents are risk averse. For a sufficiently high degree of risk aversion, the second mechanism dominates and the equilibrium tax rate declines with lower mobility.

Corbae et al. [2009] and Bachmann and Bai [2013] are two papers that study infinite horizon political economy models with income taxes and uninsurable idiosyncratic risks. Thus, these two papers are potentially capable of capturing the mechanisms described in this section.

5.2 More on the political economy channel

Some theories formulate channels through which redistributive taxes have a beneficial effect on the macroeconomy in the presence of financial constraints. For example, in the Schumpeterian view where entrepreneurship is central to economic growth, financial constraints and the absence of insurance markets make entrepreneurial investment suboptimal. Under these conditions, redistribution may provide extra resources to constrained entrepreneurs and could facilitate more investments in growth-enhancing activities. At the same time, a redistributive system provides an implicit mechanism for consumption smoothing (a person pays high taxes when he or she earns high profits but receives payments in case of losses). Therefore, it provides insurance. Thus, if entrepreneurs are risk averse, redistribution could encourage investment.

A similar mechanism applies to the investment in education or human capital. If education is important for economic growth, and parents choose suboptimal levels of education because of financial constraints, then government transfers may allow for greater investment and growth. A more direct mechanism could work through the financing of public education, as in Glomm and Ravikumar [1992].

Political economy forces are also important for the choice of government borrowing. Azzimonti et al. [2013] propose a theory of public debt where greater income inequality could increase the incentive of the government to borrow more if the higher inequality is associated with greater individual risk. This is because higher risk increases the demand for safe assets, which are undersupplied when markets are incomplete. If financial markets are integrated, the increase in inequality (risk) in a few countries could induce a worldwide increase in public debt. In this way, the paper proposes one of the possible mechanisms for explaining the rising public debt observed in most of the industrialized countries since the early 1980s.

We close this section by mentioning that, although a large branch of the political economy literature has been developed on the assumption that voters are self-motivated and agree on how their views of the world, so that their assessment of a policy is based on how much they benefit, some authors have proposed alternative frameworks. Especially interesting is Piketty [1995]. This study develops a model in which agents prefer different policies not because they are selfish but because they have different beliefs. All voters care about social welfare, but some believe that luck is more important in generating income, whereas others believe that effort is more important. These beliefs evolve over time based on personal experience, but they never converge. Thus, at any point in time, preferences are heterogeneous. Although not explicitly explored in the original article, it is possible to introduce factors that could change the distribution of beliefs and with them the properties of the macro-economy. This could be an interesting direction for future research.

6 Conclusion

In this chapter we have discussed a variety of topics that lie in the somewhat fuzzy intersection of income distribution and macroeconomics. The choice of topics and approaches has surely been idiosyncratic, reflecting our tastes, interests, and expertise, and we have left out many topics from behavioral and non-optimizing models to issues in development, to the analysis of the impact of the rise of inequality on the U.S. economy. We have also touched other aspects only superficially such as the role of globalization on the economy. In addition, we have only looked marginally to the implications of income inequality for consumption inequality or even for inequality in the duration of life[17] that is in the end what really matters to determine the welfare costs of inequality.

We are very aware that a very different chapter covering the same could be written by other authors (in fact next chapter includes an example of this by providing some very different ideas of macro modelling of the wealth distribution). Yet we hope that this chapter has provided an idea of how macroeconomics is incorporating explicitly the analysis of inequality to improve our understanding of the dynamics of the aggregate economy, and also of how the discipline that macroeconomics brings to the table –that all pieces have to be mutually consistent, and that dynamics is at the core of economics— shapes the way we think about income and wealth inequality.

[17] Pijoan-Mas and Ríos-Rull [2014] argue that the welfare cost of inequality in life spans dwarfs that of inequality in consumption.

Appendix

A Derivation of the inequality index

In each period, there are different cohorts of workers who have been employed for j periods. They also differ in terms of initial human capital h_0^k at birth. Because workers die with probability λ, the fraction of workers in the j cohort (composed of workers who survive for j periods) is equal to $\sum_k x_0^k(1-\lambda)^j$. Denote by h_j^k the human capital of a worker born with initial human capital h_0^k of age $j+1$. Since human capital grows at the gross rate $g(y,\varepsilon)$, we have that $h_j^k = h_0^k \Pi_{t=1}^j g(y,\varepsilon_t)$. Of course, this differs across workers of the same cohort because the growth rate is stochastic. The average human capital is then computed as

$$\bar{h} = \sum_k x_0^k \sum_{j=0}^{\infty} (1-\lambda)^j \mathbb{E} h_j^k, \tag{87}$$

where $\mathbb{E}$ averages the human capital of all agents in the j-k cohort. Because growth rates are serially independent, we have that $\mathbb{E}h_j^k = h_0^k \mathbb{E} g(y,\varepsilon)^j$. Substituting in the above expression and solving we get

$$\bar{h} = \frac{\bar{h}_0}{1-(1-\lambda)\mathbb{E}g(y,\varepsilon)},$$

where $\bar{h}_0 = \sum_k x_0^k h_0^k$ is the aggregate human capital of newborn agents.

We now turn to the variance which is calculated as

$$\text{Var}(h) = \sum_k x_0^k \sum_{j=0}^{\infty} (1-\lambda)^j \mathbb{E}(h_j^k - \bar{h})^2.$$

This can be rewritten as

$$\text{Var}(h) = \sum_k x_0^k \sum_{j=0}^{\infty} (1-\lambda)^j \left[\mathbb{E}(h_j^k)^2 - \bar{h}^2\right].$$

This can be further rewritten as

$$\text{Var}(h) = \sum_k x_0^k (h_0^k)^2 \sum_{j=0}^{\infty} (1-\lambda)^j \mathbb{E}\left(\frac{h_j^k}{h_0^k}\right)^2 - \bar{h}^2.$$

The term h_j^k/h_0^k is independent of the initial human capital h_0^k. Taking into account the serial independence of the growth rates, we have that $\mathbb{E}(h_j^k/h_0^k)^2 = [\mathbb{E}g(y,\varepsilon)^2]^j$. Substituting and solving, we have

$$\mathrm{Var}(h) = \left(\sum_k x_0^k (h_0^k)^2\right)\left(\frac{1}{1-(1-\lambda)\mathbb{E}g(y,\varepsilon)^2}\right) - \bar{h}^2.$$

To compute the inequality index, we simply divide the variance by $\bar{h}^2$, where $\bar{h}$ is given by (87). This returns the inequality index (27).

B Wage equation with endogenous debt

Consider the value of a filled vacancy defined in (51). Using the binding enforcement $B' = \phi(1-\eta)\mathbb{E}S'(B')$ to eliminate B', this value becomes

$$Q = (1+\phi)\beta(1-\eta)\mathbb{E}S'(B').$$

Notice that at this stage we are imposing $b = B$ and $b' = B'$, which hold in a symmetric equilibrium.

Next we use the free entry condition $V = qQ - \kappa = 0$. Eliminating Q using the above expression and solving for the expected value of the surplus, we obtain

$$\mathbb{E}S'(B') = \frac{\kappa}{q(1+\phi)\beta(1-\eta)}. \tag{88}$$

Substituting into the definition of the surplus—equation (50)—and taking into account that $b' = \phi(1-\eta)\mathbb{E}S'(B')$, we get

$$S(B) = z - \bar{u} - B + \frac{[1-\lambda-p\eta+\phi(1-\lambda)(1-\eta)]\kappa}{q(1+\phi)(1-\eta)}. \tag{89}$$

Now consider the net value for a worker,

$$W(B) - U = w - \bar{u} + \eta(1-\lambda-p)\beta\mathbb{E}S'(B').$$

Substituting $W(B) - U = \eta S(B)$ in the left-hand side and eliminating $\mathbb{E}S'(B')$ in the right-hand side using equation (88), we obtain

$$\eta S(B) = w - \bar{u} + \frac{\eta(1-\lambda-p)\kappa}{q(1+\phi)(1-\eta)}. \tag{90}$$

Finally, combining (89) and (90) and solving for the wage, we get

$$w = (1-\eta)\bar{u} + \eta(z-b) + \frac{\eta[p+(1-\lambda)\phi]\kappa}{q(1+\phi)},$$

which is the expression reported in (54).

C References

J. M. Abowd and D. Card. Intertemporal labor supply and long term employment contracts. *American Economic Review*, 77(1):50–68, 1987.

D. Acemoglu. Why do new technologies complement skills? Directed technical change and wage inequality. *Quarterly Journal of Economics*, 113(4):1055–1089, 1998.

P. Aghion and P. Bolton. A theory of trickle-down growth and development. *Review of Economic Studies*, 64(2):151,172, 1997.

M. A. Aguiar and M. Bils. Has consumption inequality mirrored income inequality? Working Paper 16807, National Bureau of Economic Research, 2011.

S. R. Aiyagari. Uninsured idiosyncratic risk and aggregate saving. *Quarterly Journal of Economics*, 109(3):659–684, 1994.

A. Alesina and D. Rodrik. Distributive politics and economic growth. *Quarterly Journal of Economics*, 109(2):465–490, 1994.

F. Alvaredo. A note on the relationship between top income shares and the gini coefficient. *Economics Letters*, 110(3):274–277, 2011. URL `http://EconPapers.repec.org/RePEc:eee:ecolet:v:110:y:2011:i:3:p:274-277`.

G.-M. Angeletos. Uninsured idiosyncratic investment risk and aggregate saving. *Review of Economic Dynamics*, 10(1):1–30, 2007.

A. B. Atkinson, T. Piketty, and E. Saez. Top incomes in the long run of history. *Journal of Economic Literature*, 49(1):3–71, 2011.

O. Attanasio, E. Battistin, and H. Ichimura. What really happened to consumption inequality in the us? NBER Working Papers 10338, National Bureau of Economic Research, Inc, March 2004.

A. J. Auerbach and L. J. Kotlikoff. *Dynamic Fiscal Policy*. Cambridge University Press, Cambridge, 1987.

M. Azzimonti, E. de Francisco, and V. Quadrini. Financial globalization, inequality, and the rising public debt. *American Economic Review*, forthcoming, 2013.

R. Bachmann and J. H. Bai. Politico-economic inequality and the comovement of government purchases. *Review of Economic Dynamics*, 16(4):565–580, 2013.

A. Banerjee and A. Newman. Occupational choice and the process of development. *Journal of Political Economy*, 101(2):274–298, 1993.

B. Bell and J. Van Reenen. Bankers' pay and extreme wage inequality in the UK. Economic Performance special papers CEPSP21, Centre for Economic Performance, London School of Economics and Political Science, 2010.

Y. Ben-Porath. The production of human capital and the life cycle of earnings. *Journal of Political Economy*, 75(4):352–365, 1967.

R. Benabou and E. A. Ok. Social mobility and the demand for redistribution: The poum hypothesis. *Quarterly Journal of Economics*, 116(2):447–487, 2001.

J. Benhabib and A. Rustichini. Social conflict and growth. *Journal of Economic Growth*, 1(1): 125–142, 1996.

J. Benhabib, A. Bisin, and S. Zhu. The distribution of wealth and fiscal policy in economies with finitely lived agents. *Econometrica*, 79(1):123–157, 2011.

B. Bernanke and M. Gertler. Agency costs, net worth, and business fluctuations. *American Economic Review*, 79(1):14–31, 1989.

B. S. Bernanke, M. Gertler, and S. Gilchrist. The financial accelerator in a quantitative business cycle framework. In J. B. Taylor and M. Woodford, editors, *Handbook of Macroeconomics*, volume 1 of *Handbook of Macroeconomics*, chapter 21, pages 1341–1393. Elsevier, 1999.

G. Bertola. Macroeconomics of distribution and growth. In A. Atkinson and F. Bourguignon, editors, *Handbook of Income Distribution*, volume 1 of *Handbook of Income Distribution*, chapter 9, pages 477–540. Elsevier, 2000.

T. Bewley. The permanent income hypothesis: A theoretical formulation. *Journal of Economic Theory*, 16(2):252–292, 1977.

O. J. Blanchard. Debt, deficits, and finite horizons. *Journal of Political Economy*, 93(2):223–247, 1985.

S. Budría, J. Díaz-Gimenez, V. Quadrini, and J.-V. Ríos-Rull. New facts on the U.S. distribution of earnings, income and wealth. Unpublished manuscript, 2001.

F. J. Buera. A dynamic model of entrepreneurship with borrowing constraints: Theory and evidence. *Annals of Finance*, 5(3):443–464, 2009.

R. J. Caballero, E. Farhi, and P.-O. Gourinchas. An equilibrium model of "global imbalances" and low interest rates. *American Economic Review*, 98(1):358–393, March 2008.

M. Cagetti and M. De Nardi. Entrepreneurship, frictions, and wealth. *Journal of Political Economy*, 114(5):835–869, 2006.

M. Cagetti and M. De Nardi. Estate taxation, entrepreneurship, and wealth. *American Economic Review*, 99(1):85–111, 2009.

C. T. Carlstrom and T. S. Fuerst. Interest rate rules vs. money growth rules: A welfare comparison in a cash-in-advance economy. *Journal of Monetary Economics*, 36(2):247–267, November 1995.

C. Carrillo-Tudela and L. Visschers. Unemployment and endogenous reallocation over the business cycle. IZA Discussion Paper 7124, Institute for the Study of Labor (IZA), 2013.

C. D. Carroll. Buffer-stock saving and the life cycle/permanent income hypothesis. *Quarterly Journal of Economics*, 112(1):1–55, 1997.

F. Caselli. Technological revolutions. *American Economic Review*, 89(1):78–102, March 1999. URL `http://ideas.repec.org/a/aea/aecrev/v89y1999i1p78-102.html`.

F. Caselli and J. Ventura. A representative consumer theory of distribution. *American Economic Review*, 90(4):909–926, 2000.

A. Castañeda, J. Díaz-Giménez, and J.-V. Ríos-Rull. Exploring the income distribution business cycle dynamics. *Journal of Monetary Economics*, 42(1):93–130, August 1998.

A. Castañeda, J. Díaz-Giménez, and J.-V. Ríos-Rull. Accounting for the U.S. earnings and wealth inequality. *Journal of Political Economy*, 111(4):818–857, 2003.

S. Chatterjee. Transitional dynamics and the distribution of wealth in a neoclassical growth model. *Journal of Public Economics*, 54(1):97–119, 1994.

S. Chatterjee, D. Corbae, and J.-V. Rios-Rull. A competitive theory of credit scoring. Mimeo, University of Pennsylvania, CAERP, 2004.

S. Chatterjee, D. Corbae, M. Nakajima, and J.-V. Ríos-Rull. A quantitative theory of unsecured consumer credit with risk of default. *Econometrica*, 75(6):1525–1589, 2007.

S. Chatterjee, D. Corbae, and J.-V. Rios-Rull. A finite-life private-information theory of unsecured consumer debt. *Journal of Economic Theory*, 142(1):149–177, 2008.

T. Cooley, R. Marimon, and V. Quadrini. Risky investments with limited commitment. NBER Working Paper No. 19594, 2012.

D. Corbae, P. D'Erasmo, and B. Kuruscu. Politico-economic consequences of rising wage inequality. *Journal of Monetary Economics*, 56(1):43–61, 2009.

J. S. Costain and M. Reiter. Business cycles, unemployment insurance, and the calibration of matching models. *Journal of Economic Dynamics and Control*, 32(4):1120–1155, 2008.

J. G. Cummins and G. L. Violante. Investment-specific technical change in the united states (1947-2000): Measurement and macroeconomic consequences. *Review of Economic Dynamics*, 5(2): 243–284, 2002.

M. De Nardi. Wealth inequality and intergenerational links. *Review of Economic Studies*, 71(3): 743–768, 2004.

J. DeBacker, B. Heim, V. Panousi, and I. Vidangos. Rising inequality: transitory or permanent? new evidence from a panel of U.S. household income 1987-2006. Finance and Economics Discussion Series 2011-60, Board of Governors of the Federal Reserve System (U.S.), 2011.

A. Díaz, J. Pijoan-Mas, and J.-V. Ríos-Rull. Precautionary savings and wealth distribution under habit formation preferences. *Journal of Monetary Economics*, 50(6):1257–1291, September 2003.

J. Díaz-Giménez and J. Pijoan-Mas. Flat tax reforms: Investment expensing and progressivity. Centre for Economic Policy Research, Discussion Paper No. 8238, 2011.

J. Díaz-Giménez, A. Glover, and J.-V. Ríos-Rull. Facts on the distributions of earnings, income, and wealth in the united states: 2007 update. *Federal Reserve Bank of Minneapolis Quarterly Review*, 34(1):2–31, February 2011.

J. Duras. Amplification of shocks in a model with labor and goods market search. Manuscript, University of Minnesota, 2013.

J. Fernandez-Villaverde and D. Krueger. Life-cycle consumption, debt constraints and durable goods. Forthcoming in *Macroeconomic Dynamics*, 2010.

X. Gabaix and A. Landier. Why has CEO pay increased so much? *Quarterly Journal of Economics*, 123(1):49–100, 02 2008.

O. Galor and O. Moav. Ability-biased technological transition, wage inequality, and economic growth. *Quarterly Journal of Economics*, 115(2):469–497, 2000.

O. Galor and J. Zeira. Income distribution and macroeconomics. *Review of Economic Studies*, 60(1):6035–52, 1993.

L. Garicano and E. Rossi-Hansberg. Inequality and the organization of knowledge. *American Economic Review*, 94(2):197–202, May 2004.

M. Gertler and A. Trigari. Unemployment fluctuations with staggered nash wage bargaining. *Journal of Political Economy*, 117(1):38–86, 2009.

G. Glomm and B. Ravikumar. Public versus private investment in human capital: Endogenous growth and income inequality. *Journal of Political Economy*, 100(4):818–834, 1992.

A. Glover, J. Heathcote, D. Krueger, and J.-V. Ríos-Rull. Inter-generational redistribution in the Great Recession. NBER Working Paper No. 16924, 2011.

C. Goldin and L. F. Katz. *The Race between Education and Technology*. Cambridge, MA: Belknap Press, 2009.

R. J. Gordon. *The Measurement of Durable Goods Prices*. National Bureau of Economic Research Monograph Series, Chicago: University of Chicago Press, 1990.

P.-O. Gourinchas and J. A. Parker. Consumption over the life cycle. *Econometrica*, 70(1):47–89, 2002.

J. Greenwood and B. Jovanovic. Financial development, growth, and the distribution of income. *Journal of Political Economy*, 98(5):1076–1107, 1990.

J. Greenwood and B. Jovanovic. The information-technology revolution and the stock market. *American Economic Review, Papers and Proceedings*, 89(2):116–122, 1999.

J. Greenwood and M. Yornkoglu. 1974. *Carnegie-Rochester Conference Series on Public Policy*, 46(2):49–95, April 1997.

J. Greenwood, J. M. Sanchez, and C. Wang. Financing development: The role of information costs. *American Economic Review*, 100(4):1875–1891, September 2010.

F. Guvenen. An empirical investigation of labor income processes. *Review of Economic Dynamics*, 12(1):58–79, 2009.

F. Guvenen and B. Kuruşçu. A quantitative analysis of the evolution of the u.s. wage distribution, 1970–2000. *In D.Acemoglu, K. Rogoff, and M. Woodford, editors, NBER Macroeconomics Annual*, 24(1):227–276, 2010.

F. Guvenen and B. Kuruşçu. Understanding the evolution of the u.s. wage distribution: A theoretical analysis. *Journal of the European Economic Association*, 10(3):489–517, 2012.

F. Guvenen, B. Kuruşçu, and S. Ozkan. Taxation of human capital and wage inequality: A cross-country analysis. Federal Reserve Bank of Minneapolis, Research Department Staff Report 438, November 2009.

F. Guvenen, S. Ozkan, and J. Song. The nature of countercyclical income risk. NBER Working Papers 18035, National Bureau of Economic Research, May 2012.

M. Hagedorn and I. Manovskii. The cyclical behavior of equilibrium unemployment and vacancies revisited. *American Economic Review*, 98(4):1692–1706, 2008.

G. D. Hansen and S. Imrohoroglu. Consumption over the life cycle: The role of annuities. *Review of Economic Dynamics*, 11(3):566–583, July 2008.

J. Heaton and D. J. Lucas. Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy*, 104(3):443–487, 1996.

J. Heckman, L. Lochner, and C. Taber. Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics*, 1(1):1–58, Jan 1998.

B. Hobijn and B. Jovanovic. The information-technology revolution and the stock market: Evidence. *American Economic Review*, 91(5):1203–1220, December 2001.

A. Hornstein, P. Krusell, and G. Violante. Frictional wage dispersion in search models: A quantitative assessment. *American Economic Review*, 101(7):2873–2898, 2011.

M. Huggett. The risk-free rate in heterogeneous-agent, incomplete-insurance economies. *Journal of Economic Dynamics and Control*, 17(5):953–969, 1993.

M. Huggett. Wealth distribution in life-cycle economies. *Journal of Monetary Economics*, 38(3): 469–494, December 1996.

M. Huggett, G. Ventura, and A. Yaron. Sources of lifetime inequality. *American Economic Review*, 101(7):2923–2954, December 2011.

A. İmrohoroğlu. Cost of business cycles with indivisibilities and liquidity constraints. *Journal of Political Economy*, 97(6):1364–1383, 1989.

G. Kambourov and I. Manovskii. Occupational specificity of human capital. *International Economic Review*, 50(1):63–115, 2009a.

G. Kambourov and I. Manovskii. Occupational mobility and wage inequality. *Review of Economic Studies*, 76(2):731–759, 2009b.

L. Karabarbounis and B. Neiman. The global decline of the labor share. Forthcoming at *The Quarterly Journal of Economics*, 2014.

L. F. Katz and K. M. Murphy. Changes in relative wages, 1963–1987: Supply and demand factors. *Quarterly Journal of Economics*, 107(1):35–78, 1992.

T. Kehoe and D. Levine. Debt-constrained asset markets. *Review of Economic Studies*, 60(4): 865–888, 1993.

N. Kiyotaki and J. Moore. Credit cycles. *Journal of Political Economy*, 105(2):211–248, 1997.

D. Krueger and F. Kubler. Pareto-improving social security reform when financial markets are incomplete!? *American Economic Review*, 96(3):737–755, 2006.

D. Krueger and F. Perri. Does income inequality lead to consumption inequality? Evidence and theory. *Review of Economic Studies*, 73(1):163–193, 2006.

D. Krueger, F. Perri, L. Pistaferri, and G. L. Violante. Cross-sectional facts for macroeconomists. *Review of Economic Dynamics*, 13(1):1–14, January 2010.

P. Krusell and J.-V. Ríos-Rull. Vested interests in a positive theory of stagnation and growth. *Review of Economic Studies*, 63(215):301–331, 1996.

P. Krusell and A. A. Smith Jr. Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896, 1998.

P. Krusell, V. Quadrini, and J.-V. Ríos-Rull. Politico-economic equilibrium and economic growth. *Journal of Economic Dynamics and Control*, 21(1):243–272, January 1997.

P. Krusell, L. E. Ohanian, J.-V. Rios-Rull, and G. L. Violante. Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica*, 68(5):1029–1054, September 2000.

M. Kuhn. Trends in income and wealth inequality. University of Bonn working paper, May, 2014.

M. Kumhof and R. Rancière. Inequality, leverage and crises. *IMF Working Paper*, pages 1–37, 2010.

I. Livshits, J. MacGee, and M. Tertilt. Consumer bankruptcy: A fresh start. *American Economic Review*, 97(1):402–418, 2007.

R. E. Manuelli and A. Seshadri. Human capital and the wealth of nations. Unpublished manuscript, University of Wisconsin-Madison, 2010.

C. A. Meh and V. Quadrini. Endogenous market incompleteness with investment risks. *Journal of Economic Dynamics and Control*, 30(11):2143–2165, 2006.

A. H. Meltzer and S. F. Richard. A rational theory of the size of government. *Journal of Political Economy*, 89(5):914–27, 1981.

E. G. Mendoza, V. Quadrini, and J.-V. Rios-Rull. Financial integration, financial deepness and global imbalances. NBER Working Papers 12909, National Bureau of Economic Research, 2007.

P. Milgrom and J. Roberts. The economics of modern manufacturing: Technology, strategy, and organization. *American Economic Review*, 80(3):511–528, June 1990. URL `http://ideas.repec.org/a/aea/aecrev/v80y1990i3p511-28.html`.

F. Modigliani and M. H. Miller. The cost of capital, corporate finance and the theory of investment. *American Economic Review*, 48(3):261–279, 1958.

T. Monacelli, V. Quadrini, and A. Trigari. Financial markets and unemployment. NBER Working Paper No. 17389, 2011.

K. M. Murphy, A. Shleifer, and R. Vishny. Income distribution, market size, and industrialization. *The Quarterly Journal of Economics*, 104(3):537–564, 1989.

T. Papageorgiou. Learning your comparative advantages. 2009 Meeting Papers 1150, Society for Economic Dynamics, 2009.

T. Persson and G. Tabellini. Is inequality harmful for growth? *American Economic Review*, 84 (3):600–621, 1994.

J. Pijoan-Mas and J.-V. Ríos-Rull. The welfare cost of inequality in life expectancies. Mimeo, University of Minnesota, 2014.

T. Piketty. Social mobility and redistributive politics. *Quarterly Journal of Economics*, 110(3): 551–84, 1995.

T. Piketty. *Capital in the Twenty-First Century*. Harvard University Press, 2014. ISBN 9780674369559. URL https://encrypted.google.com/books?id=J222AgAAQBAJ.

T. Piketty and E. Saez. Income inequality in the united states, 1913-1998. *Quarterly Journal of Economics*, 118(1):1–39, 2003. Longer updated version published in A.B. Atkinson and T. Piketty eds., Oxford University Press, 2007.

C. A. Pissarides. Search, wage bargains and cycles. *Review of Economic Studies*, 54(3):473–83, July 1987. URL http://ideas.repec.org/a/bla/restud/v54y1987i3p473-83.html.

C. A. Pissarides. *Equilibrium Unemployment Theory*. Oxford: Basil Blackwell, 1990.

V. Quadrini. Growth, learning and redistributive policies. *Journal of Public Economics*, 74(2): 263–297, November 1999.

V. Quadrini. Entrepreneurship, saving, and social mobility. *Review of Economic Dynamics*, 3(1): 1–40, 2000.

V. Quadrini. Financial frictions in macroeconomic fluctations. *Economic Quarterly*, 97(3Q): 209–254, 2011.

J. Ríos-Rull. Life-cycle economies and aggregate fluctuations. *Review of Economic Studies*, 63 (3):465–489, 1996.

J.-V. Ríos-Rull. On the quantitative importance of market completeness. *Journal of Monetary Economics*, 34(3):463–496, 1994.

J.-V. Ríos-Rull and R. Santaeulàlia-Llopis. Redistributive shocks and productivity shocks. *Journal of Monetary Economics*, 57(8):931–948, 2010.

S. Rosen. The economics of superstars. *American Economic Review*, 71(5):845–858, 1981.

E. Saez and G. Zucman. The distribution of us wealth, capital income and returns since 1913. Unpublished Manuscript, U.C. Berkeley, July, 2014.

R. Shimer. The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review*, 95(1):25–49, 2005.

R. Solow. We'd better watch out. *New York Times Book Review*, page 36, July 1987.

J. E. Stiglitz. Distribution of income and wealth among individuals. *Econometrica: Journal of the Econometric Society*, pages 382–397, 1969.

K. Storesletten, C. I. Telmer, and A. Yaron. How important are idiosyncratic shocks? evidence from labor supply. *American Economic Review, Papers and Proceedings*, 91(2):413–17, 2001.

Y. Terajima. Education and self-employment: Changes in earnings and wealth inequality. Bank of Canada Working Paper, No. 2006-40., 2006.

R. M. Townsend. Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, 21(2):265–293, October 1979.

R. K. von Weizsäcker. *A Theory of Earnings Distribution*. Cambridge University Press, 1993.

D. Wiczer. Long-term unemployment: Attached and mismatched? Unpublished manuscript, University of Minnesota, 2013.

M. E. Yaari. Uncertain lifetime, life insurance and the theory of the consumer. *Review of Economic Studies*, 32(2):137–150, 1965.

S. Yamaguchi. Tasks and heterogeneous human capital. *Journal of Labor Economics*, 30(1):1 – 53, 2012.