# Who should work how much?

Timo Boppart[*] Per Krusell[†] Jonna Olsson[‡]

September 1, 2024

A production efficiency perspective naturally leads to the prescription that more productive individuals should work more than less productive individuals. Yet, systematic differences in actual hours worked across high- and low-wage individuals are barely noticeable. We highlight that the insurance available to households is an important determinant behind this fact. Using a dynamic heterogeneous-agent model with insurance frictions, income effects calibrated to match aggregate hours across time and space, and financial frictions that deliver realistic wealth dispersion, we report stark effects of insurance: perfect insurance would raise aggregate labor productivity by 9.6 percent and decrease hours worked by 7.7 percent.

**JEL Classification:** E21, E24, J22.

[*]timo.boppart@iies.su.se
[†]per.krusell@iies.su.se
[‡]jonna.olsson@nhh.no

## 1. Introduction

In this paper we raise what we believe is a fundamental question for our economies: who should work how much? Assume that the purpose is to produce as much as possible, while taking the cost of working (the value of leisure) into account. Then if the cost of work is the same for all workers, it appears obvious that the people who are the most productive should work more than others. If, moreover, wages are a good measure of productivity, the difference in hours worked across people should be large—at least to the extent the costs of effort are not highly convex—since wage differences in the data are huge. We do not, however, observe large systematic differences in hours worked across different people in our available data sets. The (raw) correlation between wages and hours is, in fact, only very slightly above zero, if at all.

A straightforward interpretation of the weak correlation between wages and hours is that in our observed economies, wealth is very unevenly distributed: whereas high-wage earners have high incomes, they also choose to enjoy a significant amount of leisure, because they can afford it. Thus, a social welfare function placing a high weight on individuals with high labor productivity appears as a candidate for rationalizing what we observe. In this paper, we—roughly speaking—adhere to this straightforward explanation for what we observe, but without literally specifying a social welfare function. Instead, we argue that a heterogeneous-agent model with incomplete insurance that is rather standard and matches a range of basic micro- and macroeconomic facts can rationalize what we see. I.e., it roughly mimics an economy with a social welfare function that has high weights on high-wage individuals. This heterogeneous-agent model, building very directly on the Bewley (1986)/Imrohoroğlu (1989)/Huggett (1993)/Aiyagari (1994) literature by adding utility costs of working, has incomplete insurance at its core. A core contribution of our paper is to show that the degree of insurance in such a model is a key determinant of who works how much: the better is the insurance, the more will the high-wage workers work in relative terms. In particular, with perfect insurance, people work significantly less on average and output is significantly higher, all because work is shifted from low- to high-productivity workers.

We also argue that the connection between insurance and hours worked is consistent with cross-country and time series data, namely, if we take the point of view that higher-income economies are associated with more developed financial markets and social safety nets, therefore offering more effective insurance. In particular, Bick et al. (2018) run cross-household regressions relating hours worked to wages for a large number of countries, showing that the coefficients on wages are increasing with the level of development. As for time series evidence, Costa

(2000) runs the same kinds of regressions using U.S. household data and finds the coefficients to be increasing over time; we confirm her findings with more recent data and controlling for more individual characteristics. Thus, on a purely positive level, our paper proposes a theoretical mechanism behind these observations.

To implement our ideas, we first illustrate how insurance matters by considering a static model of labor supply with ex-ante random productivity draws. In such a model, no insurance can be associated simply with a decentralized equilibrium ex post where the only choice individuals have is that of choosing hours (and hence consumption) subject to a standard budget constraint. In the static model, we assume that all income is earned—there is no independent wealth—and in such a situation the relative strength of income and substitution effects will guide the extent to which high-productivity individuals work more or less. We take the view that the income effects are stronger, implying that high-productivity workers work less, not more. In fact, income effects are central to our analysis. Our modeling choice regarding this aspect does not reflect a desire to highlight a mechanism; instead, we hold a firm view that income effects are fundamental for understanding labor markets across time and space. Thus, we rather see our focus on insurance as an important implication (among many) of the presence of income effects. Armed with our simple model, we first consider ex-ante (full) insurance, which if all agents face the same probability distribution for wages can be thought of as equivalent to there being a social planner whose objective is "utilitarian": an equal-weighted sum of individual utilities. The case with full insurance implies a strong positive relation between productivity and hours worked: this is the production efficiency logic alluded to above. One can then change the set of planner weights so as to make hours wholly independent of productivity. We show that this must mean a higher Pareto (welfare) weight on high-productivity agents. The reason they do not work more than others is that the planner cares more about them: they are therefore assigned not to work so much.

A dynamic model with heterogeneous agents and incomplete insurance will, in its steady state, share some properties of the static model, but also be more realistic in a number of ways: insurance is somewhere in between and a more quantitatively convincing evaluation can be conducted.[1] We do not consider labor-market frictions in this paper, as our focus is lower-frequency movements in labor supply. The model is still formulated at an annual frequency, however, since it is of value to relate our results to those from the heterogeneous-agent literature. Our benchmark is therefore a straightforward extension of Aiyagari (1994) to include endogenous labor supply. We show that this model does deliver a roughly zero

[1]To our knowledge, the first paper to have pointed to the interplay between insurance, hours worked, and saving is Pijoan-Mas (2006).

hours-wage correlation, despite strong built-in income effects. A core assumption of this model class is that households can only insure through precautionary savings, and it is important in generating realistic outcomes in a number of ways. First, with complete insurance there would be no wealth or consumption mobility within the distribution. Relatedly, the model could not explain why wealth is so unevenly distributed; with incomplete insurance, the wealth distribution does spread out and attains a long-run shape that at least qualitatively looks like in the data (for its quantitative features, see below). Second, and most important from the present perspective, intertemporal substitution of work effort becomes an important feature of the framework: during times when wages are high, individuals work harder and save up the proceeds, enabling consumption to be smoothed over time. Thus, high wages will lead to high hours, causing a positive correlation between these variables. For reasonable calibrations of the wage process and of the size of the income effect in preferences, this mechanism roughly cancels with the strength of the income effect, which makes high-wage individuals work less.

However, the benchmark heterogeneous-agent model straight off the shelf does not readily produce a long-run wealth inequality that is as striking as what is observed in the data. In the literature, numerous model mechanisms have been proposed for resolving this challenge, including earnings risk with a superstar state, discount rate heterogeneity, and entrepreneurship (see, e.g., De Nardi and Fella (2017) for an overview). Building on recent empirical evidence on the existence of return heterogeneity (Bach et al., 2020; Fagereng et al., 2020; Daminato and Pistaferri, 2024) and insights about their quantitative importance for explaining wealth heterogeneity (Hubmer et al., 2021), our benchmark model generates the observed wealth inequality by introducing persistent return heterogeneity. In our baseline calibration, a financial intermediary, which offers households an opportunity to save but no opportunity to borrow, to some extent resembles a casino, because the interest rates it pays out to households contain idiosyncratic shocks. If a household is lucky and receives a high excess return in one period, it is also more likely to get a high excess return in the next period and it forms rational expectations over these idiosyncratic return realizations. The financial intermediary lends all the savings to the firm, receives the rental rate on capital (net of depreciation) in return, and makes zero profit.

To evaluate the benchmark model we compare outcomes to available micro data. First, since we stress the importance of income effects, we compare our results to a recent study by Golosov et al. (2024), who evaluate the strength of the income effect on labor supply using lottery winners in the U.S. We simulate their lottery winnings in our model and compare the outcomes of average labor earnings over the first five years after the win for different prize sizes. The results are striking:

our model replicates the size of the income effect as measured by Golosov et al. (2024) surprisingly well. Next, we turn to another endogenous outcome from the model, namely the dispersion of wealth and excess returns. The (persistent) return heterogeneity creates a wealth inequality that is in line with the data, with a resulting Gini coefficient of 0.85 and a wealth tail that is roughly Pareto shaped.[2] However, we can also evaluate the measured excess return across wealth deciles. We compare our model results to the evidence collected in Hubmer et al. (2021) and, again, the model performs very well. Both in the model and in the data the highest wealth deciles enjoy the highest average excess return, and the magnitudes of the differences across the wealth distribution are very similar. Despite preferences with strong income effects, the hours profile across the wealth distribution is relatively flat (with the exception for the very richest in the economy), a regularity in the data stressed by Yum (2018) and Ferraro and Valaitis (2024). Return heterogeneity is key for this result. Persistent heterogeneous returns flattens the hours-wealth relationship compared to a model with safe returns, due to the additional incentive to work and save for (some of) the wealthy agents.

In the baseline model with return heterogeneity the hours-productivity profile is relatively flat, and the estimated hours-wage elasticity is well within the range of numbers estimated from the data. The households with the lowest productivity on average work the most, and the hours-productivity distribution is U-shaped, with the households in the next to highest productivity state working the least. However, the hours dispersion across the productivity dimension is not very large: the highest-productive work only 14 percent less on average than the lowest-productive. Another way to evaluate the dispersion is the p90-to-p10 ratio in the hours distribution: in the model, the 90th percentile in the hours distribution works 37 percent more than the 10th percentile, while the corresponding figure in the data is 45 percent. Thus, our model delivers a reasonable dispersion of hours across the population. Looking at productivity by hours worked, we see an inverse U-shape relationship (in line with the empirical observations documented by Bick et al. (2022)), with households in the bottom or top 20 percent of the hours distribution having lower wages than those of the middle quintiles.

To address the importance of insurance frictions for hours choices, we then compare the outcomes from the baseline model to settings in which the financial intermediary is more or less sophisticated. First, we compare to a case in which the financial intermediary still does not offer any opportunity to borrow, but at least it can offer a safe and equal return on savings to all households. This improves the

[2]The Gini coefficient for wealth is 0.85 in the U.S.; see Kuhn, Ríos-Rull, et al. (2016) and their latest update to 2019 data, available at https://www.wiwi.uni-bonn.de/kuhn/paper/QR2021update.pdf.

hours-productivity distribution, in the sense that the observed hours-wage elasticity increases. We then compare to a case in which the financial intermediary is less well functioning than in the baseline: it does not even offer an opportunity to deposit or withdraw money, thus essentially removing the opportunity for self-insurance. The observed hours-wage elasticity in this model is substantially lower, i.e., in the economy with such a rudimentary financial sector the low-productive households work relatively more. The model results speak directly to the suggestive evidence in the data: a richer and more developed country could be assumed to have a more developed financial market with more opportunities for insurance, and the hours-wage correlation is higher (less negative) the more developed a country is, both across countries and within the U.S. over time. It turns out that the range of hours-wage elasticities that the model produces is well in line with the ranges observed in the data. Given those observations we conclude that the magnitude of the mechanism we discuss in this paper–how the level of insurance affects the hours-productivity distribution–is quantitatively both reasonable and important. In a final, more drastic experiment, we remove all insurance frictions.[3] As a result, consumption will be the same across all agents but, as predicted from the static model, the more productive individuals should work more. Now, however, we obtain quantitative results. It turns out that the hours-productivity distribution is strikingly different than for the benchmark model: the households with the highest productivity work 123 percent more than the agents with the lowest productivity. Moreover, the implied hours-productivity distribution improves aggregate outcomes. When households have access to full insurance, aggregate labor productivity is 9.6 percent higher than in our baseline model, while total hours is 7.7 percent lower. Thus, removing the lack-of-insurance friction improves the hours-productivity distribution, leading to higher aggregate labor productivity, lower total hours and higher consumption in a quantitatively meaningful way.

**Related literature.** In one way, this paper can be viewed as a labor market analogue to Mendoza et al. (2009) who examine the different level of financial market development across countries and its implications for capital markets. We focus on the labor markets, and assume no cross-country mobility.

From a modeling perspective this paper closely relates to the growing body of research using heterogeneous-agent models and an active labor choice. Early examples in this group include Aiyagari (1995), Krusell and Smith (1998), Pijoan-Mas (2006) (arguably the most closely related paper to ours), Domeij and Floden

[3]Here, we treat all agents as equal, thus corresponding to an equal-weighted utilitarian case, but the message remains the same also if one allows for full insurance within subgroups, e.g., by considering people in different education or occupational groups.

(2006), Floden (2006), Chang and Kim (2006, 2007) Marcet et al. (2007), Heathcote et al. (2014), and Heathcote et al. (2017). With the exception of the last paper, they all feature a non-degenerate wealth distribution. A key is that they all (with the exception of the first one) use standard "macro utility functions", in other words, they use a utility function where the income effect is relatively strong, which makes the wealth distribution important to consider.

**Roadmap.** In Section 2 we present empirical observations suggesting that with more developed financial markets, the productivity/hours correlation in the economy increases. In Section 3 we then describe a stylized model to highlight how the level of insurance affects hours worked, contrasting full insurance with no insurance in a static setting. Thereafter, in Section 4, we add realism to the model by introducing self-insurance via a financial intermediary in a full dynamic model. The calibration of the model is described in Section 5, while the results are presented in Section 6. In Section 7 we compare the outcomes depending on the sophistication of the financial intermediary. Finally, Section 8 concludes.

## 2. Hours-wage correlations in the data

Data from a cross-section of countries—see, e.g., Bick et al. (2018)—and from time series within countries—see, e.g., Boppart and Krusell (2020)—make clear that, at least with standard theories of labor supply, income effects must be strong: in less productive economies, people work significantly longer hours per capita and week, despite lower wages. That is, as we become more productive, though every hour of work pays better, we choose to work less: we buy more leisure. The quantitative magnitude of the effect may not seem huge—as productivity growth goes up by one percent, hours worked fall by perhaps 0.2 percent—but over time and between countries, it accumulates and generates large gaps. These insights are not novel on a qualitative level—they at least date back to famous remarks in Keynes (1930)—but the quantitative estimates of the size of the income effect are new.[4]

Given these insights, the same mechanism ought to be at work within a given population at any point in time. Here, on the surface at least, the data looks hard to square with strong income effects: wages vary greatly between individuals, and yet the cross-sectional correlation between hours and wages found in household data is near zero (see, e.g., Chang, Kim, Kwon, and Rogerson (2020)). However, even though the correlation is low, there are some systematic differences in the hours

[4]Additionally, recent studies on individual responses to exogenous wealth shocks point to strong income effects, see, e.g., Cesarini et al. (2017), Golosov et al. (2024), Giupponi (2019), and Artmann et al. (2023) using micro data, or Georgarakos et al. (2023) using a survey experiment.

distribution, both across countries and in the U.S. over time. As countries grow richer, the hours-productivity correlation increases. That is, in richer countries, highly productive individuals work relatively more. Likewise, in the U.S., the observed hours-wage correlation has risen over time.[5]

Thus, the hours-wage correlation is higher in countries that are richer/more developed. One feature of richness/development is that it appears to go hand in hand with financial development; see Greenwood and Jovanovic (1990) and Acemoglu and Zilibotti (1997) for early theory and King and Levine (1993) for early empirical work arguing for such a link.[6] Thus, regardless of whether financial development causes growth or the other way around, our documented correlation can motivate the theory we propose in the present research, at least to the extent financial development brings better insurance for households. Below, we also directly look at the relation between the hours-wage correlation and measured financial development. This section should be viewed as suggestive and a motivation for our theory work; we do not attempt to establish a causal link based on the data we present.

### 2.1. Evidence across countries

Bick et al. (2018) build a consistent set of microeconomic data on hours worked across a large number of countries, including both highly developed countries and developing countries.[7] For each country and gender they run the following regression:

$$\log h_i = \alpha + \beta_w \log w_i + \delta_i age_i + \delta_2 age_i^2 + \varepsilon_i. \tag{1}$$

Figure 1a shows the resulting $\beta_w$ coefficients by country against log GDP per capita for men.[8] Figure 1b shows the same information, but with ratio of domestic credit to GDP, a measure used by Rajan and Zingales (1996) as an indicator for the state of development of credit markets, on the x-axis. As the figures show, there is a clear positive relationship. Rich countries and countries with a high ratio of domestic credit to GDP have a higher (less negative) hours-wage elasticity.

The measurement of hours worked is challenging. In the Bick et al. (2018) data set, a substantial effort was made to ensure data consistency and comparability. Here, we only show results for the 49 countries the authors label as core countries

[5] This pattern is also observed by Alon et al. (2018) and Mantovani (2023). However, these two papers propose very different potential drivers than what we point to in this paper: Alon et al. (2018) explore the role of the family and joint household labor supply, while Mantovani (2023) focuses on hours-biased technical change.

[6] Mendoza et al. (2009) also build on the notion that richer countries have more developed financial markets; their application serves to account for real interest rates being higher in more developed countries.

[7] We are grateful to Bick, Fuchs-Schundeln, and Lagakos (2018) for sharing their data with us.

[8] We focus on men in the main text, due to large differences in female labor force participation across countries. The corresponding graphs for women can be found in the appendix, Figure A1.


**(a)** Hours-wage elasticities vs. log(GPD) per capita. The red dots indicate the historical observations from the US, taken from Costa (2000).

**(b)** Hours-wage elasticities vs. ratio of domestic credit to GDP (source IMF).

**Figure 1:** Country-specific elasticities of hours to wages for men (taken from Bick et al. (2018)) vs. different indicators of financial development.

in their study, for which international comparability of hours data is deemed high enough.[9,10]

For non-working individuals there is no information on wages. Thus, if a large part of the productivity-work adjustment is carried out on the extensive margin it would not be picked up in a hours-wage regression. However, using education level as a coarse proxy for productivity, there is again suggestive evidence that the same forces are present for the extensive margin choice as for the intensive margin choice: in more developed countries, high-productive individuals are relatively more likely to be working. Bick et al. (2018) look at the employment rates by education level and conclude that employment rates are essentially flat by education in the low-income countries for all gender-age groups. In contrast, in middle- and high-income countries, employment rates are increasing in education for all gender-age groups.

## 2.2. U.S. evidence over time

Costa (2000) runs the same regression (1) on U.S. data at three different points in time: 1895, 1973, and 1991. The $\beta_w$ coefficients from her study are also included in

[9] In particular, the authors require that the data from these core countries satisfy three basic criteria: (i) the surveys cover the entire calendar year (rather than, say, one month of the year); (ii) hours worked are measured as actual (rather than usual) hours in all jobs (not just the primary job), and in the week prior to the interview; and (iii) hours worked cover the production of goods or services counted in the National Income and Product Accounts (NIPA), which means that the hours measures cover unpaid work in agricultural or nonagricultural businesses, as well as wage employment, but do not cover home-produced services, such as child care.

[10] Because of measurement error in the hours data and the wage is constructed from information about earnings and hours, the hours-wage regression (1) suffers from what is known as "division bias": if an individual reports hours above those actually worked, their imputed hourly wage, computed as earnings divided by hours, mechanically becomes too low. Consequently, the estimated $\beta_w$ coefficient will be biased towards minus one (see, e.g., Borjas (1980)). Thus, one worry is that if the measurement error in hours is more severe in countries at a lower stage of development, the pattern we observe could be generated by measurement error rather economic factors. We discuss this issue more in the next section, where we use U.S. microeconomic data.

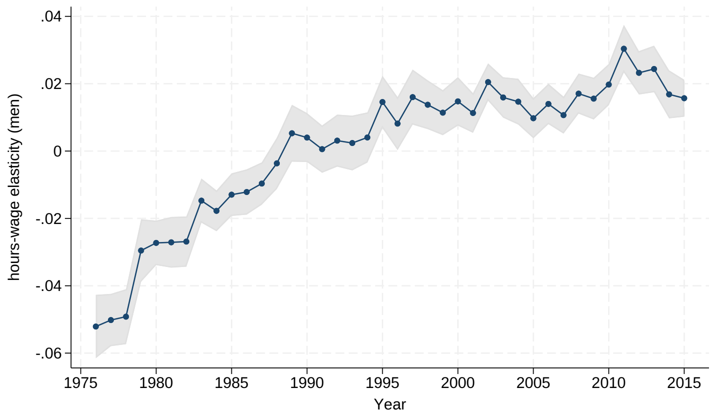



**Figure 2:** Hours-wage elasticity in the U.S. for men. Source: CPS ASEC.

Figure 1a (with red dots). As can be seen, the hours-wage elasticity was negative historically, but has become less negative over time, and is now positive. Thus, these data points add to the overall suggestive picture: the richer a country is, the higher is the hours-wage elasticity.

We then zoom in on the period 1976–2015, using the CPS March Annual Social and Economic supplement (ASEC).[11] We focus on men only due to the large changes in female labor force participation that have taken place over the same period, and exclude all individuals who are employed in the armed forces. For all analyses, the observations are weighted using ASECWT. The variables used are listed in the Appendix, Section A.

Figure 2 shows the hours-wage elasticity for men aged 25–54 by year, using the same type of regression as (1) but additionally controlling for education and race. As can be seen, the hours-wage elasticity has been increasing over time.[12]

However, an elasticity measure is not informative about the non-parametric relationship between hours and wages. To get a better picture of the distribution of hours across different wage levels Figure 3 shows average residualized hours for (residualized) wage ventiles.[13] The figure shows the average for three decades: the 1980s, the 1990s, and the 2000s. As the figure shows, the overall picture is that hours are reasonably flat over different wage levels, but there has been a slight tilt since the 1980s, which is consistent with the increase in the hours-wage elasticity.

A potential concern is that observations with very low (mis-measured) hours

[11] We start in 1976 since that is the first year for which we have data on "usual hours worked per week last year", which is our preferred variable for constructing yearly hours and yearly wages.

[12] As mentioned, when wages have to be constructed from data on hours, there is a "division bias" in the specification (1). In the Appendix, Section B.3, we show different specifications and robustness checks, and even though the level of elasticity differs across specifications, the conclusion that the elasticity has increased over time remains in all of them.

[13] Again, we control for age, age squared, education, and race.

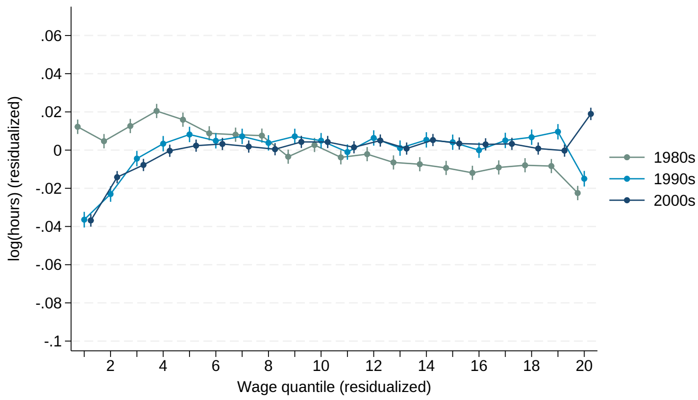



**Figure 3:** Hours-wage distribution (residualized values) in the U.S. for men. Source: CPS ASEC.

could skew the results, since those give rise to high wage figures as well. In Figure 2 and Figure 3 the bottom 1% of the hours observations each year are trimmed off. In the Appendix, Section B.4, we show the result from alternative cut-offs and demonstrate that this primarily affects the bottom and top ventiles, while the overall pattern of a tilt of the whole distribution is evident in all cases.

Again, the hours-wage elasticity and distribution only tells us something about those individuals for which we observe a wage, but nothing about the individuals who do not work. For this reason, we turn to the fraction working of the male population in the U.S. between 1976 and 2015.[14] Due to standard selection issues we cannot impute unobserved productivity based on time-varying characteristics. Instead, as a coarse proxy for potential productivity/wage we use three levels of education: less than high school, high school, and some college or more. Figure 4 shows the employment figures for the three educational groups between 1976 and 2015. A first observation is that the fraction employed has been falling in all groups over this time period. A second observation is that the fraction employed has fallen more in the less educated groups, as pointed out already by Juhn et al. (1991). In 1976, the fraction working among the less-than-high-school educated men was only 7 percentage points lower than the fraction working among those with at least some college education. In 2015, the gap had increased to 16 percentage points. One of the major contributing factors to the drop in employment in the lowest educated group is the increase in disability leave which sharply rose during the 1990s, but a large portion of the year-to-year variation in the fraction working is driven by differences in unemployment. In Figure A2 of our Appendix, we show the same type of graph but also include the unemployed. The same pattern remains. Thus,

[14]We use the variable EMPSTAT, employment status, and categorize those with answer 10 (at work) or 12 (has a job, not at work last week) as “working”, while the remainder are classified as “not working”.

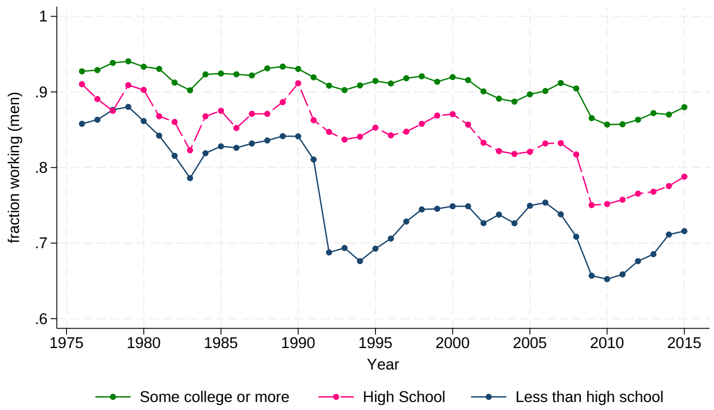



**Figure 4:** Fraction working in the male population aged 25–54, by education level. Source: CPS ASEC.

the extensive-margin analysis leads to the same conclusion as the intensive-margin analysis: the less educated (with lower wages) are relatively less likely to work over time.

An overview of the variables used for the analysis can be found in the Appendix, Section A.

## 3. Core static model for intuition

In this section we use the simplest possible model to illustrate the theoretical importance of insurance for the hours-productivity correlation and for aggregate outcomes such as aggregate labor productivity, total hours worked, and total consumption.

**Setup.** The economy consists of a unit mass of ex-ante homogeneous individuals. At time $t = 1$ they realize a shock to individual productivity $\omega_i$. After an individual observes its draw of $\omega_i$, they maximize utility by choosing how much to work and how much to consume, subject to a budget constraint.

Utility is a function of consumption $c$ and hours $h$: $u(c, h)$. Hours can be thought of either as clock-hours or effort-weighted hours, and can be chosen freely. To make the problem concrete, we use the MaCurdy (1981) utility function

$$u(c, h) = \frac{c^{1-\sigma} - 1}{1 - \sigma} - \psi \frac{h^{1+1/\theta}}{1 + 1/\theta}$$

for $\sigma \neq 1$ and

$$u(c, h) = \log c - \psi \frac{h^{1+1/\theta}}{1 + 1/\theta}$$

for $\sigma = 1$. Production in the economy is linear in (productivity/efficiency units of)

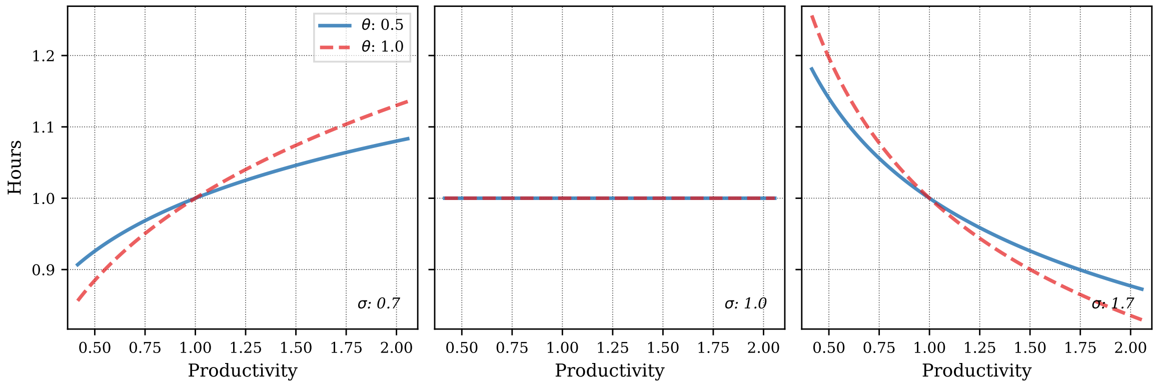



**Figure 5:** Hours as a function of productivity, $\sigma$, and $\theta$. Case 1: No insurance. ($\psi = 1$)

labor; total output is given by $Y = \int h_i \omega_i \, \mathrm{d}i$.

**Question.** How do the distribution of hours and other aggregate outcomes depend on the insurance mechanisms available at time $t = 0$?

**Economy 1: No insurance mechanism available.** In the first economy, there is no insurance mechanism available. Individuals at time $t = 1$ face the following (static) problem:

$$\max_{c,h} u(c, h) \quad \text{s.t.} \quad c = \omega_i h.$$

Solving for optimal hours and consumption for the individuals gives

$$h_i = \left( \frac{1}{\psi \omega_i^{\sigma - 1}} \right)^{\frac{\theta}{1+\sigma\theta}} \qquad \text{and} \qquad c_i = \left( \frac{1}{\psi} \right)^{\frac{\theta}{1+\sigma\theta}} \omega_i^{\frac{1+\theta}{1+\sigma\theta}}$$

for all $i$. Figure 5 shows the hours choice as a function of productivity for three values of $\sigma$ and two values of $\theta$. Note that with $\sigma = 1$ the utility function falls in the King, Plosser, and Rebelo (1988) class, with income and substitution effects exactly canceling out. With such preferences the hours choice becomes independent of the productivity/wage rate. This case is depicted in the middle panel of Figure 5.

We will focus on the $\sigma > 1$ case (the rightmost panel), since a utility function where the (long-run) income effect of higher wages slightly dominates the substitution effect better captures the observations described in Section 2, both across countries and within countries over time: as we become more productive, though every hour of work pays better, we choose to work less: we buy more leisure.

With a parameterization $\sigma > 1$, a key observation is that hours are falling in productivity: without any insurance, individuals with high productivity work the least.

**Economy 2: Perfect insurance.** In this second case, there is a perfect insurance market: in time $t = 0$ individuals can buy state-contingent bonds to insure themselves against the productivity risk.

The solution to the individual's problem coincides with the solution to the problem a utilitarian planner (assigning all individuals equal weights) faces:

$$\max_{\{c_i,h_i\}_{\forall i}} \int u(c_i, h_i)\,\mathrm{d}i \qquad \text{s.t.} \qquad \int c_i\,\mathrm{d}i = \int h_i\omega_i\,\mathrm{d}i.$$

The solution is characterized by the first-order conditions

$$c_i = c_j, \qquad \frac{h_i}{h_j} = \left(\frac{\omega_i}{\omega_j}\right)^{\theta}, \quad \text{and} \quad c_i^{-\sigma} = \psi\frac{h_i^{1/\theta}}{\omega_i}$$

and the optimal choices for hours and consumption are given by

$$h_i = \omega_i^{\theta}\left(\frac{1}{\psi^{\frac{1}{\sigma}}\left(\int \omega_j^{1+\theta}\,\mathrm{d}j\right)}\right)^{\frac{\sigma\theta}{1+\sigma\theta}} \qquad \text{and} \qquad c_i = \left(\frac{\int \omega_j^{1+\theta}\,\mathrm{d}j}{\psi^{\theta}}\right)^{1+\sigma\theta}.$$

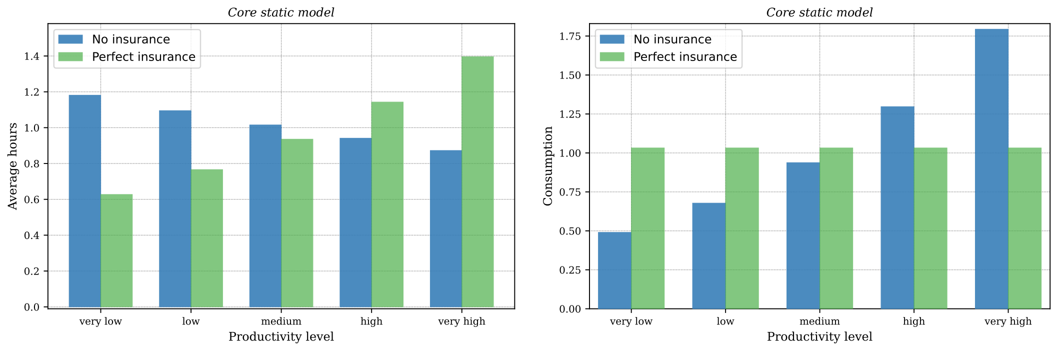
In contrast to the previous case, consumption is equalized across individuals (or, equivalently, across all potential states of an individual, since we assume ex-ante homogeneity) and hours are increasing in own (realized) productivity.

Now also the ex-post distribution of productivity among the individuals (or equivalently, the individual's productivity in other states of the world, and the ex-ante probabilities for those) matter for hours and consumption choices.

**Quantitative comparisons of outcomes.** For a more detailed comparison between the outcomes of the two economies we need to assign a productivity distribution for the population (equivalent to ex-ante probabilities). To do this, we develop a numerical example where we employ the stationary distribution used in the dynamic model developed and analyzed in Section 4 below. We assume that individuals can be of five different productivity levels, equidistant in logs, with medium productivity being the most common level, with symmetrically fewer individuals at the very low or the very high productivity states.[15] Furthermore, we parameterize the example with $\sigma = 1.7$, $\theta = 0.5$, and $\psi = 1.0$ (again we use the same values as in the full model described in Section 4). This numerical example is not a fully satisfactory summary of the much more ambitious dynamic model—e.g., it does not have non-earned income—but it does give us a first quantitative pass.

Figure 6 contrasts the resulting hours and consumption allocations from the two

[15] The distribution we use is shown in Figure A10 in the Appendix, Section C.




**(a)** Hours distribution

**(b)** Consumption distribution

**Figure 6:** Resulting hours and consumption allocations, comparing Economy 1 (no insurance) with Economy 2 (full insurance).

economies. What immediately stands out from subfigure 6a is the stark contrast in the hours choice along the productivity dimension. In the economy with full insurance hours are *increasing* in productivity, while in the model with no insurance hours are *decreasing* in productivity.

In addition, in the economy with full insurance we note that total hours worked is 6% lower while average consumption is 4% *higher* than in the economy with no insurance. The reason is the 11% higher average labor productivity arises due to improved correlation between hours and productivity. Full insurance increases aggregate productivity, and this gain is taken out partly as more consumption and partly as lower hours.

**Summary and remarks.** This highly stylized model shows the importance of insurance for the allocation of hours along the productivity dimension. With higher (in this case perfect) insurance, individuals work the more when they are more productive, which increases aggregate labor productivity, aggregate consumption, and aggregate leisure.

Of course the quantitative results depends on parameterization, one element of which is the Frisch elasticity $\theta$. The higher the Frisch elasticity, the higher are the labor productivity gains from insurance (since then it is less costly for a planner to redistribute work across individuals). In this example we use $\theta = 0.5$, a value guided by microeconometric estimates that do not factor in the extensive-margin choice nor human capital accumulation; though conservative, we use it as our baseline value. With $\theta = 1.0$ the gain from insurance for aggregate labor productivity increases to 21%. In Section C of the Appendix we show these results.

In the Appendix, we also report results from other preference types. In Section C, we consider non-separable preferences (we use Cobb-Douglas preferences for this example, thus falling into the King, Plosser, and Rebelo (1988) class). In this case, the difference in allocations between the no-insurance and full-insurance cases is even starker. With full insurance, the individuals with bad productivity draws do not work at all, while the individuals with high productivity draws work at their maximum. With no insurance, everyone works the same amount.[16] To show that the effect of insurance on the hours distribution requires an income effect, we also show results for the GHH type of preferences ((Greenwood, Hercowitz, and Huffman, 1988), preferences which are not consistent with balanced-growth facts absent exogenous shifters in the utility for leisure). Under GHH preferences, since they feature no income effect at all, the hours distributions of the no-insurance and full-insurance cases coincide.

**Ex-ante heterogeneity and equal hours worked.** So far we have assumed ex-ante homogeneity, for which the full insurance case is equivalent to a utilitarian planner with equal weights. We now consider a case with ex-ante heterogeneity and no risk, so that a planner can distinguish between individuals with different productivity outcomes and assign different Pareto weights to them. Thus, the planner problem now reads:

$$\max_{\{c_i,h_i\}\forall i} \int \lambda_i u(c_i,h_i)\,\mathrm{d}i \qquad \text{s.t.} \qquad \int c_i\,\mathrm{d}i = \int h_i\omega_i\,\mathrm{d}i.$$

where $\lambda_i$ is the Pareto weight for individual $i$ and $u(c_i,h_i)$ is the additively separable MaCurdy utility from before. Which Pareto weights would give rise to a planner solution where everyone works the same? Such an outcome would imply that the Pareto weights satisfy

$$\frac{\lambda_i}{\lambda_j} = \frac{\omega_i}{\omega_j}.$$

In other words, the resulting relative Pareto weights would be proportional to the productivity differences. The resulting weights for the numerical example we have used throughout this section are shown in Figure 7. The individuals with high productivity would need a high Pareto weight to be assigned the same hours as the low-productive. Note that this differences in Pareto weights would also imply that the high-productive individuals consume more than the low-productive.[17]

[16] Additive separability in consumption and leisure, our baseline assumption, is qualitatively distinct in that it does not feature aversion (or love) toward covariance between consumption and leisure, while of course allowing for aversion (or love) toward consumption risk and towards leisure risk. Absent empirical evidence or even introspective arguments, we consider additive separability to be reasonable.

[17] The intuition can easily be extended to a case with different "types" determining the probability of each outcomes, with some types having a higher probability to end up with, e.g., high productivity. If one imposes equal average hours worked by ex-post productivity outcome, individuals with higher ex-ante probability of high productivity would be assigned higher Pareto weights. The consumption level in such a planner problem would be determined by ex-ante Pareto weight,

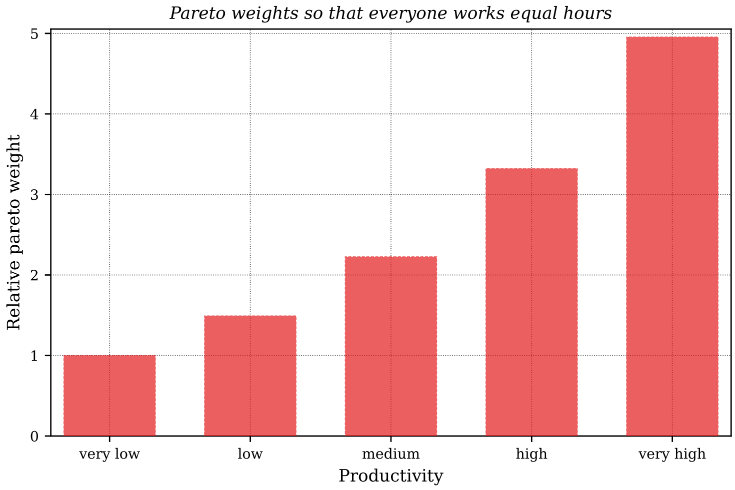



**Figure 7:** Relative Pareto weights that would give rise to a solution where everyone works equal amount of hours.

This case can be decentralized by assigning high-productivity individuals somewhat negative asset holdings, while low-productivity individuals have positive asset holdings so as to maintain a total sum of assets equal to zero. The high-productivity individuals will still be "richer" in total, given their higher labor productivity, and hence consume more. In a dynamic model, there will be capital, the returns on which on average will be consumed in steady state, and high-productivity individuals will have more assets on average. Two effects are then added. First, giving everybody an equal amount of non-earned income will, given our utility function, make substitution effects play out more relative to income effects and make high-productivity individuals work more in relative terms. Second, giving high-productivity individuals more wealth will make them work less in relative terms. These two effects work in opposite directions and, as we shall see, produce a hours-productivity pattern much like in the data.

## 4. Full model

In the previous section we illustrated the gains from insurance in a stylized manner, comparing no insurance to full insurance. Reality is surely somewhere in between. Households do not have access to a full set of state-contingent bonds, but it seems reasonable to assume that they can at least self-insure via savings. To make the model more realistic we therefore add an opportunity to save. This makes the household problem dynamic and dependent on the amount and type of risk facing the household.

while the hours worked would be determined by the Pareto weight and the productivity outcome.

The model we employ is a version of the much-used setting dues to Bewley (1986)/Imrohoroğlu (1989)/Huggett (1993)/Aiyagari (1994). Time is discrete and every period is assumed to be one year. There is a continuum of infinitely lived, ex-ante identical households indexed by $i$. The households derive utility from consumption and disutility from supplying labor, which is endogenously chosen. Households can save via a financial intermediary, but face an exogenous borrowing constraint. Households face two types of idiosyncratic risks: shocks to productivity and shocks to returns on their savings. There is no possibility to insure against these risks except by savings and labor supply adjustments. For simplicity we abstract from growth and restrict attention to steady states.[18]

The household's saving is carried out through a financial intermediary, an element of our model that is new relative to the literature and that is motivated in part by microeconomic data on asset returns and in part by our aim to match the model's wealth distribution; we discuss the details of the calibration later, but let us already note here that a by-product of allowing this new element is that the model's implications generally align better with the data in several dimensions than do those generated by the standard framework. We thus consider financial intermediaries that can be more or less well-functioning. The intermediary offers households an opportunity to save but, as mentioned, no opportunity to borrow. It does offer households an interest rate on their savings but, depending on the "financial sophistication" of the intermediary, the interest rate may be *scrambled* to some extent on the household level. More precisely, different households obtain different ex-post returns on their savings, while on average giving the intermediary zero profits; the intermediary lends its funds to firms at a rate that is not scrambled. In our baseline calibration the financial intermediary is thus subject to scrambling, and the interest rate paid out to the households are subject to idiosyncratic shocks. The household is aware of the return randomness and thus forms rational expectations over their idiosyncratic interest rate realizations.

An overview of the model is given by Figure 8 and we now describe each actor in turn.

### 4.1. The household problem

Let $(a, \omega, r_x)$ be the household's (wholly idiosyncratic) state vector, where $a$ is beginning-of-period assets, $\omega$ is the productivity state, and $r_x$ is the current excess return draw. Each household solves the following recursive problem (suppressing

[18] For a treatment of the distribution of hours worked in the case of growth, see Boppart et al. (2023).

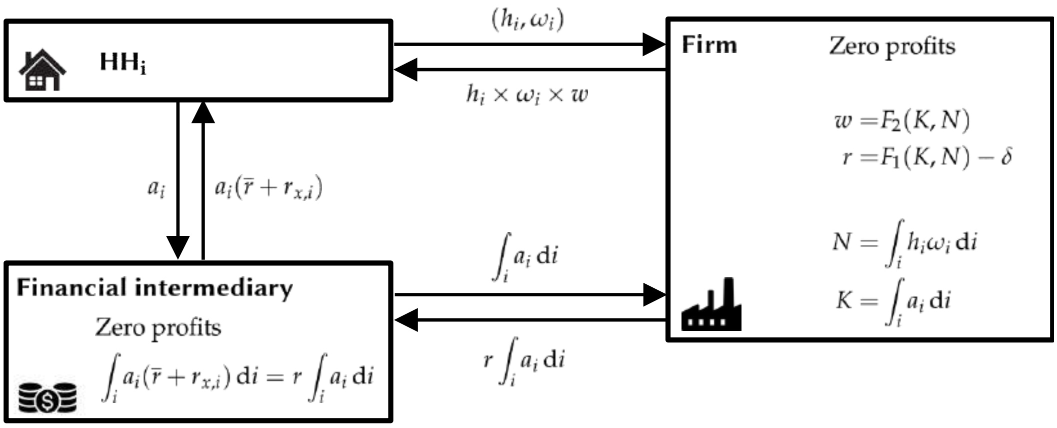



**Figure 8:** Schematic illustration of the baseline model.

index $i$ for readability):

$$V(a, \omega, r_x) = \max_{c,h,a'} \left\{ u(c,h) + \beta \mathbf{E} V(a', \omega', r_x') \right\}$$

$$\text{s.t.} \quad c + a' \leqslant w\omega h + a(1 + \bar{r} + r_x)$$
$$a' \geqslant 0.$$

The economy-wide wage rate per efficiency labor unit is denoted by $w$, while $\bar{r}$ denotes the aggregate return component. Households form expectations over idiosyncratic productivity $\omega'$ and excess returns $r_x'$.

Labor productivity follows (the discretized version of) an AR(1) process in logs with persistence $\rho_\omega$ and innovation $\varepsilon_\omega \sim N(0, \sigma_\varepsilon^2)$

$$\log(\omega') = \rho_\omega \log(\omega) + \varepsilon_\omega',$$

while the excess return each household receives on their savings follows (the discretized version of) a stochastic process given by

$$r_x' = \rho_x r_x + \varepsilon_x,$$

with persistence $\rho_x$ and innovation $\varepsilon_x \sim N(0, \sigma_x^2)$.

The instantaneous utility function $u(c,h)$ is of the MaCurdy (1981) type with constant relative risk aversion, separable in consumption and hours worked:

$$u(c,h) = \frac{c^{1-\sigma} - 1}{1-\sigma} - \psi \frac{h^{1+1/\theta}}{1+1/\theta}. \tag{2}$$

### 4.2. The financial intermediary

Each household $i$ can deposit its savings with the financial intermediary and earn interest rate $\overline{r} + r_{x,i}$, where $\overline{r}$ is the aggregate return component and $r_{x,i}$ is the risky idiosyncratic excess return. In each period the financial intermediary lends all the assets to the firms and obtain return $r$. The financial intermediary makes zero profits, thus, with $a_i$ denoting the individual household's asset holdings in the beginning of a particular period, we have

$$\int_i a_i(\overline{r} + r_{x,i})\,\mathrm{d}i = r\int_i a_i\,\mathrm{d}i.$$

Note that in the case where returns are serially uncorrelated ($\rho_x = 0$), $a_i$ and $r_{x,i}$ are uncorrelated by design and hence $r$ must equal the expectation of $\overline{r} + r_{x,i}$; in the case with serial correlation, $\overline{r}$ is nontrivially determined given $r$ (under positive serial correlation, $a_i$ and $r_{x,i}$ are positively correlated).

### 4.3. The production side

The production side of the model is completely standard. Competitive firms employ labor hired from households and capital hired from the financial intermediary to produce a homogeneous final good, which is used for both consumption and investment. The aggregate production function is assumed to be Cobb-Douglas:

$$F(K,N) = K^{\alpha}N^{1-\alpha},$$

with $K$ denoting the total amount of capital and $N$ denoting the total amount of efficiency units of labor: $N = \int_i \omega_i h_i\,\mathrm{d}i$. Capital is assumed to depreciate at the rate $\delta$.

### 4.4. Equilibrium

A household's state vector is given by $(a, \omega, r_x)$ where assets $a \in \mathcal{A} = [\underline{a}, \overline{a}]$, productivity $\omega \in \mathcal{W} = [\underline{\omega}, \overline{\omega}]$, and excess return $r_x \in \mathcal{R} = [\underline{r_x}, \overline{r_x}]$.[19]

A stationary competitive equilibrium is given by a set of prices $\{r, \overline{r}, w\}$, decision rules for consumption, savings and labor supply given by $C(a, \omega, r_x)$, $A(a, \omega, r_x)$, and $H(a, \omega, r_x)$, and a stationary measure $\Gamma(B, \omega, r_x)$ such that:

1. The decision rules solve the households' problem for all $(a, \omega, r_x)$.

[19] The domains are bounded due to the discretization of our processes.

2. Firms optimize, i.e, prices are given by

$$r = F_1(K,N) - \delta \qquad \text{and} \qquad w = F_2(K,N),$$

where

$$K = \int_{\mathcal{A}\times\mathcal{W}\times\mathcal{R}} A(a,\omega,r_x)\,\mathrm{d}\Gamma$$
$$N = \int_{\mathcal{A}\times\mathcal{W}\times\mathcal{R}} H(a,\omega,r_x)\omega\,\mathrm{d}\Gamma.$$

3. The financial intermediary makes zero profit:

$$\int_{\mathcal{A}\times\mathcal{W}\times\mathcal{R}} A(a,\omega,r_x)\sum_{r_x'}\pi_{r_x'|r_x}(\bar{r}+r_x')\,\mathrm{d}\Gamma = r\int_{\mathcal{A}\times\mathcal{W}\times\mathcal{R}} A(a,\omega,r_x)\,\mathrm{d}\Gamma$$

4. For all $(B,\omega,r_x)$,

$$\Gamma(B,\omega,r_x) = \sum_{\widehat{\omega}}\pi_{\omega|\widehat{\omega}}\sum_{\widehat{r}_x}\pi_{r_x|\widehat{r}_x}\int_{a:A(a,\omega,r_x)\in B}\Gamma(\mathrm{d}a,\widehat{\omega},\widehat{r}_x),$$

where the integrals over (subsets) of $\mathcal{A}$ are Lebesgue integrals, with measurable subsets denoted by $B$ (Borel sets on the positive real line). Intuitively, given the other elements of the equilibrium, the decision rules are determined by condition 1; the prices $(r,w)$ are determined by condition 2 (which indirectly ensures market clearing for capital and labor); and $\bar{r}$ is determined by condition 3; and the stationary distribution is determined by condition 4. The fixed point is non-trivial but straightforward to find numerically; there are two prices ($r$ and $\bar{r}$) to iterate on, which is one more than in the standard incomplete-markets model.

## 5. Baseline calibration

We calibrate the model by using standard parameters from the literature as far as possible. The model is yearly, so we set $\beta = 0.96$.

### 5.1. Preference parameters in the instantaneous utility function

We set the relative risk aversion $\sigma = 1.7$ and the Frisch elasticity $\theta = 0.5$, based on Heathcote et al. (2014). This parameterization allows for an income effect that is slightly stronger than the substitution effect. The MaCurdy (1981) type of utility function implies that if wages were to grow at gross rate $\gamma$, hours would fall at gross rate $\gamma^{\nu}$, with $\nu = (\sigma-1)/(\sigma+1/\theta)$, along a balanced growth path. Our choices of

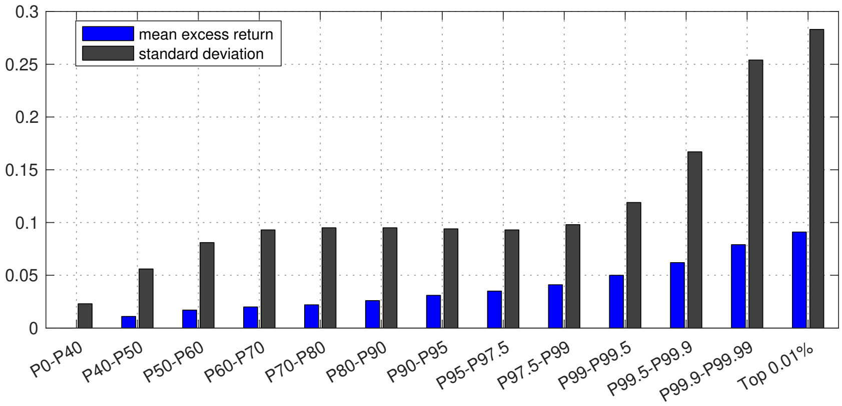



**Figure 9:** Schedule of excess returns, taken from Hubmer et al. (2021).

$\sigma$ and $\theta$ give a value of $\nu = -0.19$, meaning that if wages were to grow by 2% per annum, total hours would fall by about 0.4% each year. This number is well in line with aggregate time-series data from various countries (see Boppart and Krusell (2020)). The weight on the disutility of work $\psi$ is normalized to 1, as it only affects the level of hours worked.

### 5.2. Productivity process

The productivity process is calibrated following Aiyagari (1994). The values for persistence $\rho_\omega$ and innovation $\varepsilon_\omega$ of the productivity process corresponds to the values used in Aiyagari (1994) with the highest persistence ($\rho_\omega = 0.9$). Even so, the chosen persistence is still substantially lower than for instance what is used by Krueger et al. (2016), who estimate a persistence of earnings of 0.9695 (but also include a separate risk of unemployment). A higher persistence would result in a starker difference between less and more insurance; our choice of a persistence of 0.9 is therefore "conservative". The AR(1) process is discretized into a 5-state Markov chain with the Rouwenhorst method.

### 5.3. Excess return process

In the data, there is strong evidence of heterogeneity and persistence in the returns in the financial market (Bach et al., 2020; Fagereng et al., 2020; Daminato and Pistaferri, 2024). Figure 9 summarizes the excess return schedule computed by Hubmer et al. (2021) (who combine data from Bach et al. (2020), Piazzesi and Schneider (2016), Jordà et al. (2019), and Kartashova (2014)).

As can be seen, the standard deviation of excess returns for most of the population is around 10% (the top half percent of the wealth distribution has higher standard deviation, while the bottom 50% have lower). The process for the heterogeneous returns is therefore calibrated with a persistence $\rho_x = 0.8$ and conditional standard

| Parameter | Description | Value | Source |
|---|---|---|---|
| *Preference parameters* | | | |
| $\beta$ | Discount factor | 0.96 | Standard value |
| $\sigma$ | Relative risk aversion | 1.7 | Heathcote et al. (2014) |
| $\theta$ | Frisch elasticity | 0.5 | Heathcote et al. (2014) |
| $\psi$ | Disutility of work | 1.0 | Normalization |
| *Productivity process* | | | |
| $\rho_\omega$ | Persistence | 0.9 | Aiyagari (1994) |
| $\sigma_\omega$ | Cond. standard dev. | 0.1744 | Aiyagari (1994) |
| *Financial intermediary* | | | |
| $\rho_x$ | Persistence | 0.8 | Unconditional standard dev. |
| $\sigma_x$ | Cond. standard dev. | 6% | of 10%, wealth Gini $\approx$ 0.85 |
| *Technology* | | | |
| $\alpha$ | Capital share | 1/3 | Standard value |
| $\delta$ | Depreciation rate | 10% | Standard value |

**Table 1:** Parameter values used in baseline calibration.

deviation of returns $\sigma_x$ equal to 6% so that the unconditional standard deviation is 10% and the resulting Gini coefficient for wealth is 0.85, close to what is observed in the data in the U.S.[20] The AR(1) process is discretized into a 5-state Markov chain with the Rouwenhorst method, the resulting levels, ergodic distribution and transition matrix are shown in Table A2 in the Appendix. The excess return process is assumed to be independent of the productivity process.

An interesting observation from Figure 9 is that the mean excess return is increasing in wealth. The persistence of the return process naturally generates excess returns increasing in current wealth. We discuss this mechanism in some detail in Section 6.3 below.

### 5.4. Summary

Table 1 shows the baseline calibration of all parameters used in the model. We set the capital share $\alpha = 1/3$ and the depreciation rate at $\delta = 10\%$; these values are standard in the literature.

## 6. Results from the baseline model

Compared to the minimal static model described in Section 3, the dynamic setting with a possibility to save has three effects on the joint distribution of productivity and hours in the population.

[20] In the microeconomic data referred to above, returns are positively serially correlated on an individual level, thus motivating our assumptions. In Section F of the Appendix we report the results from a model with zero autocorrelation but equal unconditional standard deviation as in the baseline. The wealth Gini in such a model is substantially lower.

First, the possibility to save gives an opportunity to self-insure against the productivity risk. By saving, the households can intertemporally substitute labor supply across time periods and thereby take the opportunity to work more when they are highly productive.

Second, households can self-insure against the return risk. They can take the opportunity to work and save when they have a high expected return. This possibility generates more work among those with high current returns, as returns are positively serially correlated.

Third, as a consequence of the previous two points, households accumulate wealth. This wealth generates an income effect, which lowers the labor supply of a given household at any point in time. If households with high productivity are richer, the result is less work among the high-productive. We label this effect the wealth distortion of labor supply.[21]

How these three effects—the insurance against productivity shocks, the insurance against interest rate shocks, and the wealth distortion effect—interact in equilibrium is a quantitative question.

### 6.1. Work allocations in the baseline calibration

Figure 10 shows work allocations in the baseline calibration. As can be seen, the households with the lowest productivity are the ones working the most. Furthermore, hours worked display a slight U-shape along the productivity dimension: on average, households in the next-to-highest productivity state are working the least.

The hours-wage elasticity, as measured by running regression (1) in the model generated data, gives $\beta_w = -0.08$ which is well in line with the estimated value for men in countries in the upper half of the sample by GDP per capita (see Figure 1) or slightly lower than the estimates using U.S. time series (which started at $-0.06$ in 1975, see Figure 2).

Figure 11 shows the hours-productivity relationship but now with average productivity (i.e., the hourly wage) on the y-axis and how much the household is working on the x-axis. We note, quite interestingly, that the model delivers what resembles both a "part-time" and an "over-time" penalty (even though the causality is reversed in the model: the wage is fixed and the hours are chosen). The households in the second quintile of the hours distribution are the ones with the highest wages, while those working in the bottom or top 20 percent of the hours distribution have lower wages (for more details on the hours-productivity relationship in the data see Bick et al. (2022)).

[21]The term distortion is appropriate, as the steady state wealth distribution under complete insurance markets is indeterminate. The baseline model, in contrast, has a unique stationary distribution, which features positive covariance between productivity and wealth.

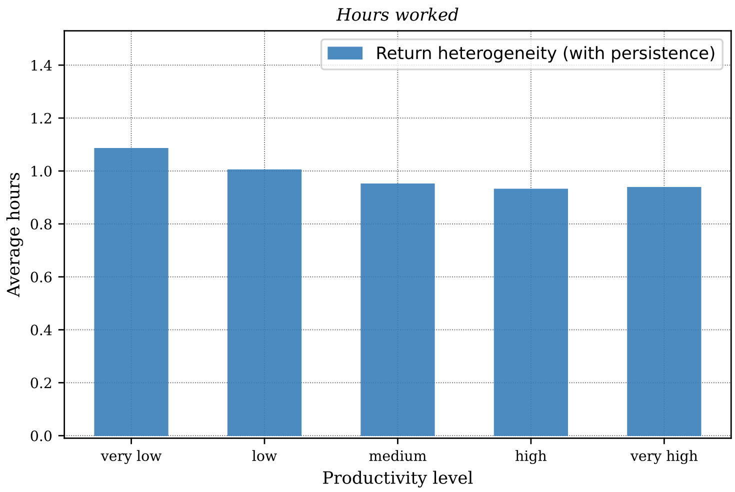



**Figure 10:** Average hours worked by productivity, baseline calibration.

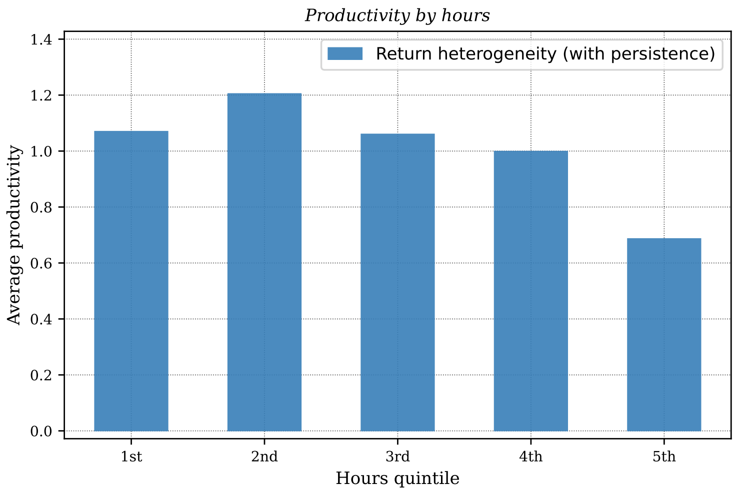



**Figure 11:** Average productivity (wage) by hours worked, baseline calibration.

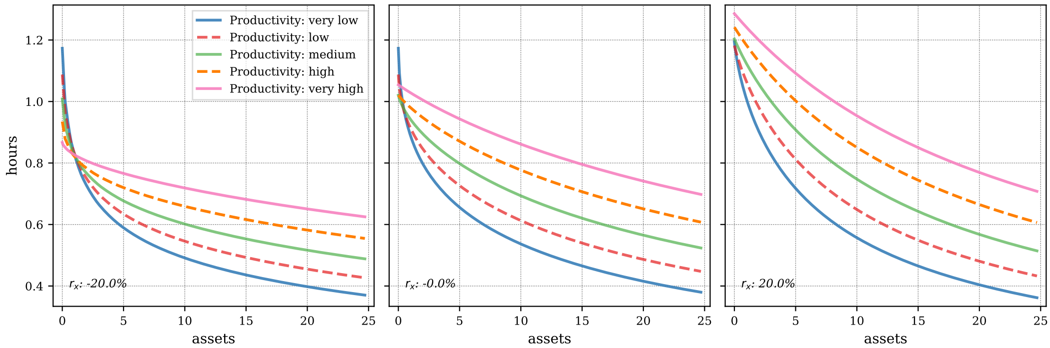



**Figure 12:** Policy functions for hours worked, baseline calibration. Each panel is a different current level of excess return: to the left the lowest excess return state, in the middle the middle state, to the right the highest excess return state. Each line represents a different productivity level. The x-axis indicates beginning-of-period assets (before interest rate income).

Another way to evaluate the dispersion in hours is the p90-to-p10 ratio of the hours distribution: in the model, the 90th percentile in the hours distribution works 37 percent more than the 10th percentile, while the corresponding figure in the data is 45 percent. Thus, our model delivers a reasonable dispersion of hours across the population.

To understand the hours allocation in the baseline calibration, Figure 12 is useful. The figure shows the policy functions for hours worked as a function of beginning-of-period assets, current productivity, and three different levels of current excess return (by panel). As can be seen, households with higher current return work more, conditional on current assets and productivity (seen by comparing across the three panels).[22] The reason is the persistence in the excess return: with a current high excess return there is a higher chance of a high current return also in the next period, and therefore it is beneficial to take the opportunity to work more and save more.

The same logic holds true for the different productivity states: in general, households with higher current productivity take the opportunity to inter-temporally shift their work load and work more when their productivity is high. The exception is households with very little wealth: for those, the income effect dominates the substitution effect, so that a high productivity leads to lower hours worked.

But why, then, do households with high productivity work less? The raw correlation of hours and productivity is weak, but noticeable in Figure 10. The reason is the wealth distribution. Households with high productivity are also on average richer. This is evidenced in Figure 13, which shows the average asset holdings (indicated by the size of the bubble) and average hours worked (indicated by the

[22] The exception is households in the lowest productivity state with some sizable wealth, for whom the high interest rate income has a direct wealth effect that pushes labor supply down.

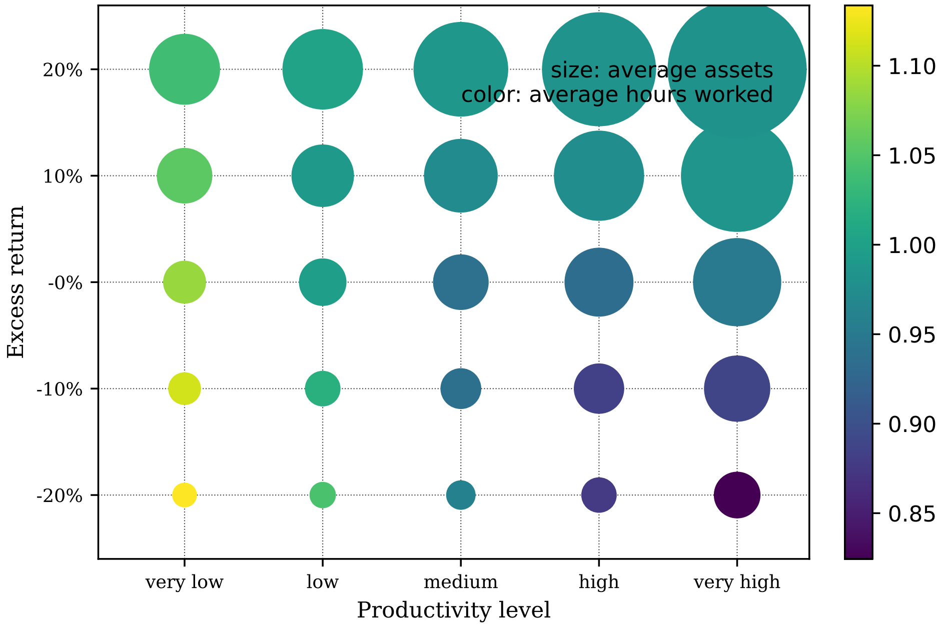



**Figure 13:** Average outcome for each productivity/excess return cell. Size of bubble indicates average asset holdings in the cell. Color is hours worked, with lighter color indicating more hours worked.

color) for each productivity/excess-return cell. As the figure shows, the households with the highest productivity and highest excess return state are on average the richest in the economy.

Figure 13 can be further used to understand the workings of the model. As a group, households with low productivity and low excess return work the highest hours (seen by the light yellow color of the bubble). The reason is that households in this group are on average very poor and therefore need to work.

The group that works the least are the households in the highest productivity state and lowest excess return state. There are two reasons for this: first, the long-run income effect dominates the substitution effect, so higher productivity leads to lower hours. Second, the group does *not* have any incentive to work for excess return reasons, since they have a low current excess return.

For households in the highest productivity state, hours are increasing in the excess return. Households with higher excess return have more wealth on average, but for these households the intertemporal substitution effect dominates: they are productive, and if they work they can take the opportunity to save more while their excess return is high.

For households in the lowest productivity state, hours are falling in excess return. On average, the households with higher excess returns are richer, have higher interest rate income, and the wealth effect pushes down labor supply. The incentive to work and save to take the opportunity to insure against the interest rate risk is not so appealing, since the productivity is low.

Along the productivity dimension, hours are falling in productivity for the households in the lowest excess return state. This is due to the income effect. However, for the households in the higher productivity states, hours actually display a non-monotonic pattern in productivity (even though it is difficult to see visually from Figure 13, the next to highest productivity state works the least for the higher excess return states). The income effect from asset holdings is counteracted by the willingness to intertemporally substitute hours due to a current high return.

In short, the interaction between the two risks—productivity risk and interest rate risk—is non-trivial but can be understood by close scrutiny. The insurance effect generates more hours worked among highly productive households and/or households with current high excess return: households intertemporally substitute their working hours and work more when they are highly productive and gain a lot from accumulating wealth. However, the wealth distortion effect is strong: the highly productive and the ones with currently high returns are also wealthier, and this pushes down their labor supply.

### 6.2. Comparing the wealth effects to those in Golosov et al. (2024)

The wealth effects are crucial for our results. Thus, to evaluate our model and its predictions for labor supply behavior it is useful to compare the results to empirical evidence on wealth effects.

Golosov et al. (2024) estimate how Americans respond to idiosyncratic and exogenous changes in household wealth by studying lottery winners. They find that Americans respond to an exogenous increase in household wealth by significantly reducing their employment and labor earnings. On average, an extra 100 dollars in wealth leads to a reduction of annual earnings by approximately 2.3 dollars per year in the five years after the lottery win.

We simulate the same type of unanticipated lottery wins in our model, and compare our results to the empirical evidence they provide.

**Size of win.** Golosov et al. (2024) report wealth effects by the size of the win, dividing the lottery winners into three groups: prizes between USD 30K and 300K, prizes between 300K and 1M, and prizes larger than 1M (see Figure B.6 in their paper).

Since we do not have the precise size distribution of wins, we simulate a number of different prize sizes. To facilitate the comparison, we evaluate the prize sizes at the edges of their size bins (30K, 300K, 1M) and also at 4M (which is the 99th percentile of the size of wins, see Table A.1 in their paper).

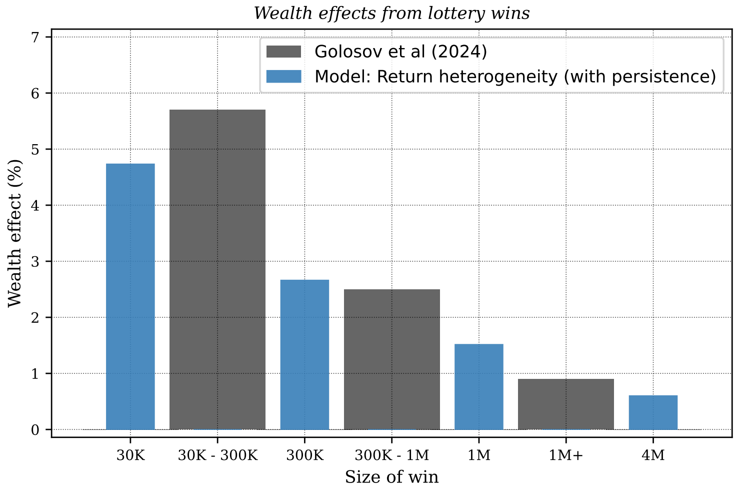



**Figure 14:** Wealth effects by prize size. The figure shows the effect of an extra dollar of wealth on total labor earnings by different prize sizes. To ease interpretability, we follow Golosov et al. (2024) and scale earnings responses by $100 (but report the decrease in labor earnings as a positive number). The numbers for Golosov et al. (2024) are taken from their Figure B.6. Note that the Golosov et al. (2024) numbers are reported for intervals of prize sizes, while the model numbers are reported for a specific prize size.

**Simulation.** For each prize size, we simulate 1 million individuals in our model economy, distributed according to the stationary distribution over wealth, productivity, and current excess return, over five years. During these five years each individual's productivity state and excess return follow the stochastic processes described in Sections 5.2 and 5.3.

We then give the same 1 million individuals a "lottery win" of a given size and re-simulate their behavior (using the same shock sequence as without the winning). We compute the average yearly labor earnings for the five years following the lottery win and compare the results to the case without the lottery win.

**Results.** Figure 14 compares the resulting wealth effects in the model to the estimates obtained by Golosov et al. (2024). As the figure shows, the model results and the empirical evidence are remarkably close: the larger the size of win, the smaller is the wealth effect, both in the model and according to the empirical findings. In addition, the magnitudes of the effects line up very well between model and data.

The reason for the stronger response for smaller prize sizes in the model is straightforward: it is due to the curvature of the policy function for labor. For all income levels and all excess return states, the policy functions for labor are convex and downward-sloping (see Figure 12), since the more current wealth an agent has, the income effect from additional wealth is muted.

Another observation by Golosov et al. (2024) is that richer individuals respond

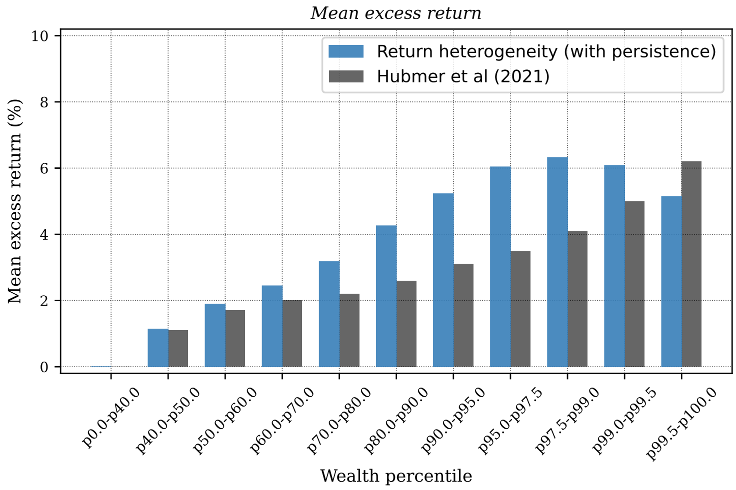



**Figure 15:** Excess return over wealth, normalized so that the asset-poorest group has 0. Agents with no assets are included in the mass of the poorest group, but their underlying excess return is not included in the return average of that group (only those with an observable return are included when calculating the average).

by decreasing their labor income *more* than do poorer individuals. Qualitatively, our model reproduces the stronger responses by the income-rich for all prize sizes except for the very smallest ones, even though the magnitude of the difference is somewhat smaller than indicated in the empirical evidence (for details, see the Appendix Section D.1).

In general, the comparisons made in this section between model results and empirical evidence should be interpreted with some caution, since our model does miss some elements, such as taxes and labor market frictions. However, it is both reassuring and interesting that the predictions from our model line up so well with the empirical evidence.

### 6.3. Excess return dispersion and wealth inequality

In the data, wealthier individuals have a higher excess return on average. In the model, excess return by wealth percentile is an endogenous outcome. Figure 15 shows the resulting mean excess return by wealth percentile and the corresponding data moments taken from Hubmer et al. (2021).

As the Figure shows, the results are well in line with the data. The excess return is increasing in wealth, and the magnitudes are of the right size, with an excess return for the top 0.5% of the wealth distribution of around 6%.

The model misses out on the highest return for the very richest. In the model the extremely rich have a slightly lower mean excess return than does the group around the 99th percentile, while in the data the relationship between excess return and wealth is monotone. The reason for the model result is that the very richest are

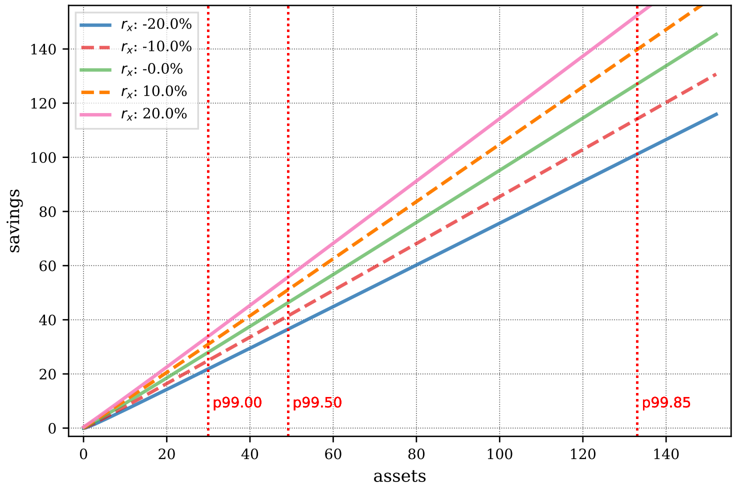



**Figure 16:** Policy functions for savings, baseline calibration, medium productivity. Each line represents a different excess return state. The x-axis indicates beginning-of-period assets (before interest rate income).

so rich, that these agents are still among the top 0.5%, even after a couple of low return draws. In the group of slightly less rich (e.g., in the 95% to 99% percentiles), a sequence of lower excess return draws quickly makes the individual drop out of the group and fall down the asset ladder.

Figure 16 shows the savings policy functions for different levels of current excess return (as shown by the different lines) for the medium productivity state.[23] As can be seen, if an agent is currently at the 99.85 percentile of the asset distribution, it can have 4 consecutive draws of the worst excess return state before it drops down below the 99.5 percentile. However, if the agent starts out at the 99.5 percentile, it falls below the 99 percentile after only two periods in the worst excess return state.

Note that the interpretation of the source of the higher excess return for the wealthy agents in the model is clear. It is not due to richer agents making better investment decisions. Rather, the causality goes the other way: the agents who were lucky enough to receive a high excess return in the financial casino have chosen to work more, save more, and have had a high return on their savings. Thereby, they have become richer.

The return heterogeneity mechanism gives rise to wealth inequality in line with the data. The Gini coefficient for wealth is around 0.85 in the U.S. The model delivers a wealth inequality of the right magnitude: the Gini coefficient is 0.85. Moreover, Figure 17 displays the tail behavior of the stationary wealth distribution. The logarithm of its counter-cumulative distribution function is close to becoming linear in the logarithm of assets as assets grow large. A linear relationship would

[23]For higher levels of savings, the current productivity state is not important for the savings decision. Figure A11 in the Appendix shows the same graph but for three different productivity states.

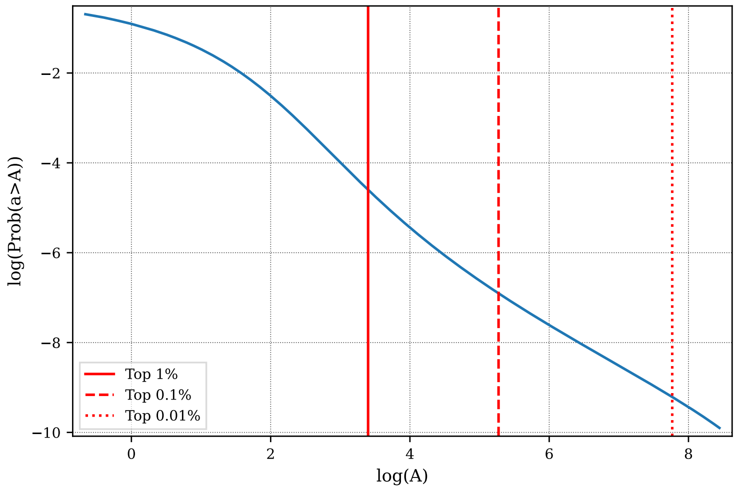
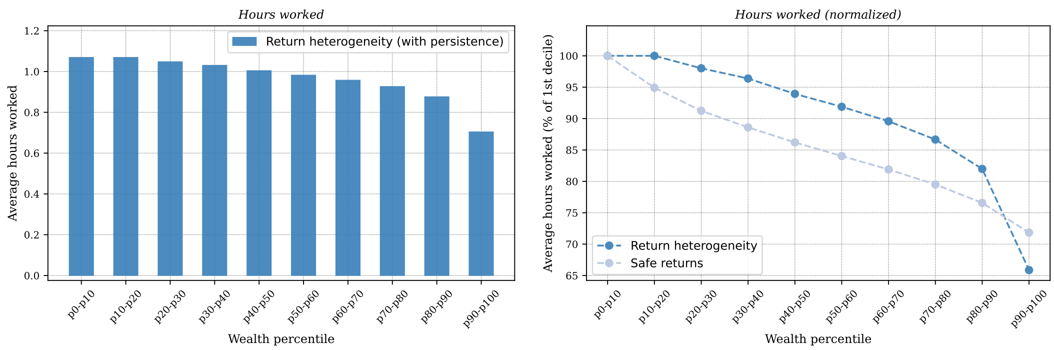



**Figure 17:** Log wealth distribution in the baseline calibration.




**(a)** Hours worked in the baseline model (with return heterogeneity).

**(b)** Hours worked, normalized as percentage compared to hours worked in the first decile.

**Figure 18:** Hours worked by wealth decile.

indicate that the right tail of the distribution follows a Pareto distribution. The reason for this tail behavior is the idiosyncratic return shocks, with the underlying intuition building on the results in Kesten (1973) and Beare and Toda (2022).

### 6.4. Hours worked by wealth

Figure 18a shows average hours worked by wealth decile. Despite strong (long-run) income effects in the utility function, average hours by wealth decile is only slowly declining in wealth (with the exception of the top 10% of the wealth distribution who works substantially less than the other groups). One contributing reason for the relatively slow decline in hours along the wealth gradient is the persistent heterogeneous returns. A currently high excess return draw generates an additional incentive to work, and excess return is increasing in wealth (as shown in Figure 15).

To put the hours worked by wealth decile results in context, Figure 18b shows the same average hours by wealth but normalized by hours worked in the first decile and compares the results from the baseline model (with return heterogeneity) to the results from a "standard" model with safe returns. As the figure shows, the existence of heterogeneous returns makes the hours distribution fall less sharply along the wealth gradient.[24]

Thus, the existence of return heterogeneity improves how well the model can replicate the hours distribution along the wealth gradient as well, since in the data hours worked is relatively flat in this dimension (Yum, 2018; Ferraro and Valaitis, 2024).

## 7. The role of insurance for who works how much

Given that our model fits the data surprisingly well along several dimensions and, importantly, replicates key empirical facts on labor supply behavior we can use it as a laboratory for quantitative evaluation of the importance of access to insurance for the distribution of hours worked in the economy. That is, we now provide the dynamic counterpart of the comparisons obtained for the static model we used for illustration. It is however natural in our dynamic model to consider a richer set of financial institutions. We thus describe four levels of financial intermediary sophistication in turn.

### 7.1. No capital markets

In this model setup, we assume that the financial sector is extremely rudimentary: all individuals own an equal amount of capital (that sums up to the total capital holdings in the economy) that can be considered an exogenous endowment. The only role of the financial intermediary is to aggregate the capital holdings, lend them to the firm, and channel back the rental rate to the households, in this case "unscrambled" for simplicity. Importantly, the financial intermediary does not offer the households any opportunity to increase or decrease their account holdings. Therefore we even call this model the *"No capital markets"* economy. The households in each period solve a static problem, with a small non-labor income.

In any time period $t$ the household problem is given by

$$\max_{c,h} u(c,h) \quad \text{s.t.} \quad c = hw\omega + rk \tag{3}$$

where $w$ is the equilibrium wage rate and $rk$ is the capital income. Compared to the

[24]The graph corresponding to Figure 18a for the case of safe returns is shown in the appendix, Figure A18. Note also that the asset distribution is of course very different in the two models. A graph of the asset tail for the model with safe returns can be found in Section E.2 together with additional graphs of policy functions for hours and the resulting joint distribution of assets and productivity.

static model without capital presented in Section 3 there is one new equilibrium quantity: $K/N$. The first-order condition for the household reads

$$(\omega_i w h_i + rk)^{-\sigma} = \frac{\psi}{w\omega_i} h_i^{1/\theta}, \tag{4}$$

which implicitly gives $h$ as a function of the capital-to-labor ratio $K/N$.[25] The size of the capital stock is assumed to be such that the rental rate on capital net of depreciation in equilibrium is $1/\beta - 1$. The additional equilibrium conditions compared to the static model in Section 3 are from firm maximization: $r = F_1(K, N) - \delta$ and $w = F_2(K, N)$.

### 7.2. Return heterogeneity

This is the baseline calibration described in Section 5 in which the financial intermediary offers an opportunity to save, but cannot give any promise about the interest rate. Results from this model are described in detail in Section 6. Compared to the *"No capital markets"* economy, individuals have the opportunity to self-insure via savings, and can therefore intertemporally take advantage of periods of high productivity and work more. However, the fact that there is an individual persistent excess return risk adds a distortion to the labor choice, and so does the wealth accumulation.

### 7.3. Safe returns

In this version of the model, the financial intermediary offers an opportunity to save, and is sophisticated enough to be able to promise a safe return. Thus, the return on savings is risk-free and equal for all households. This model setup is equal to the baseline, but setting the excess return shocks to zero: $\varepsilon_x = 0$. This model highlights the importance of the additional labor distortion created by the persistent heterogeneous returns, since in this economy with safe returns, the heterogeneous returns channel is switched off.

### 7.4. Perfect insurance

In this economy the financial intermediary is fully sophisticated and offers the households a full set of state-contingent bonds, the returns on which are also not scrambled. The solution coincides with the planner allocation with equal Pareto

[25] The capital-to-labor ratio pins down both the interest rate $r = F_1 - \delta$ and the wage rate $w = F_2$.


**Figure 19:** Average hours worked by productivity. Comparing four models with different assumptions about the level of financial intermediary sophistication.

weights for all households. The intra-temporal problem is given by:

$$\max_{\{c_i,h_i\}\forall i} \int_i u(c_i, h_i)\, \mathrm{d}i$$

$$\text{s.t.} \quad \int_i c_i\, \mathrm{d}i = F(K, N) - \delta K \quad \text{with } N = \int_i \omega_i h_i\, \mathrm{d}i$$

and $K$ denoting the aggregate capital stock.[26] The characterization of equilibrium is almost unchanged compared to the problem presented in Section 3:

$$c_i = c_j, \qquad \frac{h_i}{h_j} = \left(\frac{\omega_i}{\omega_j}\right)^{\theta}, \qquad c_i^{-\sigma} = \psi \frac{h_i^{1/\theta}}{\omega_i F_2(K, N)}. \tag{5}$$

As in Section 3, consumption is equalized across households, while hours are increasing in own productivity.

### 7.5. Comparing outcomes across scenarios

We now compare the outcomes from the four different models, representing four different levels of insurance. We first focus on the resulting distribution of hours in the economy.

**Distribution of hours.** Figure 19 shows the resulting average hours over the different productivity states for the four models with different levels of insurance. As anticipated, in the setting with *"No capital markets"* the hours-productivity

[26]The intertemporal maximization problem gives that in steady state $1 = \beta\,(F_1(K, N) + 1 - \delta)$ as usual.

distribution is the "worst", with the agents with the lowest productivity working the most on average.[27]

In the *"Safe return"* scenario, i.e., when the return risk is shut down, the households with the lowest productivity are still working the most on average, but by a smaller margin. Average hours worked by productivity follows a U-shape, with the least hours worked by the medium productive. The scenario *"Return heterogeneity"* falls in between the previous two scenarios.

The difference between the *"Safe returns"* and the *"Return heterogeneity"* scenarios shows the impact of the return risk on the hours allocation. The heterogeneous returns have two effects: first, they create an additional distortion to the labor supply, since households want to work when their return is high (which not necessarily coincides with when their productivity is high). Even though the productivity risk and the return risk are orthogonal, the interaction is non-trivial since the labor supply decision also depends on current wealth of the household. Second, the heterogeneous returns create additional wealth distortion, since households save to insure themselves against this risk as well.

The differences between the three previously mentioned scenarios pale when compared to the *"Perfect insurance"* scenario. In this economy, the agents with the highest productivity work the most, by a large margin, and this drives up the aggregate productivity in the economy.

The difference in average hours over the different productivity (i.e., wage) states give rise to different measured hours-wage elasticities in the model. We run regression (1) on the different model results. Figure 20 shows the resulting estimates from the three models with different levels of frictions in the financial market, and compare the estimated range of elasticities to the ranges found in different empirical studies.[28]

The range of estimates the model produces for the different levels of insurance we consider in this exercise is approximately of the right magnitudes compared to the data. Moreover, the movements between a (rather extreme) scenario with *"No capital markets"* to a scenario with *"Safe returns"* (and no further distortions to the labor supply decision than the income risk) span a difference in hours-wage elasticity that is somewhat larger than the movements in the U.S. since the 1970s, but somewhat smaller than the range of cross-country estimates from Bick et al. (2018). Given those observations we conclude that the magnitudes of the mechanism we

[27] Section E.1 in the Appendix describes the intuition in more detail for the hours choice in a static model with non-labor income, and contrasts a model with $\sigma = 1.7$ (as in our calibration) with a model with $\sigma = 1.0$ (which gives KPR preferences where the long-run income effect and the substitution effect cancel).

[28] We do not display the results from the *"Full insurance"* scenario to focus the y-axis on the interesting part. The hours-wage elasticity in the *"Full insurance"* scenario is 0.5, trivially given by the Frisch elasticity.

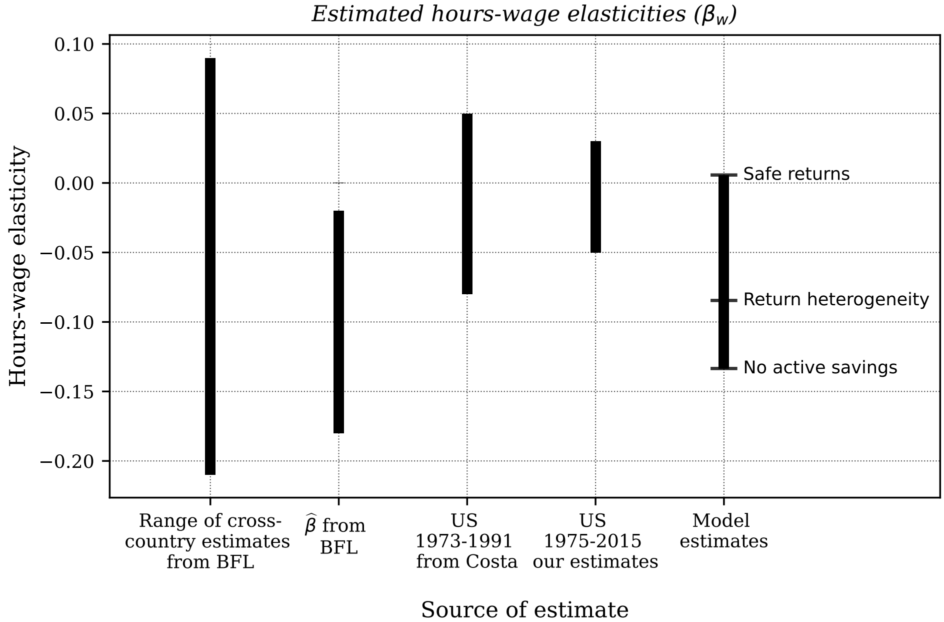



**Figure 20:** Estimated ranges of hours-wage elasticities. BFL refers to Bick et al. (2018). The range of cross-country estimates gives the range between the lowest and highest number they estimate in their core country sample for men. "$\widehat{\beta}$ from BFL" refers to the predicted value from the cross-country regression illustrated in Figure 1, evaluated at the lowest and highest GDP per capita in the core country sample. "Costa" refers to the values from Costa (2000) for men.

discuss in this paper–how the level of insurance affects the hours-productivity distribution–is quantitatively both reasonable and important.

**Aggregate labor productivity.** The different work allocations across productivity has implications for the aggregate outcomes in the economy. Figure 21 shows a key aggregate indicator of the hours/productivity allocation in the four different scenarios: the resulting aggregate labor productivity in the economy. Productivity in the model is normalized so that average productivity equals one if everyone in the economy works the same amount. An average labor productivity below one therefore indicates that low-productive households work relatively more, and vice versa.

As Figure 21 shows, the average labor productivity in the *"No capital markets"* scenario is below one, due to relatively more work being carried out by the households with low productivity. As noted above, the households with the lowest productivity work the most on average in this setting.

In the *"Return heterogeneity"* scenario (the baseline scenario) agents have some opportunity to self-insure, but the insurance market is not very well functioning. The resulting average labor productivity is only slightly higher than in the no-capital-markets economy. In the *"Safe return"* scenario, the average labor productivity

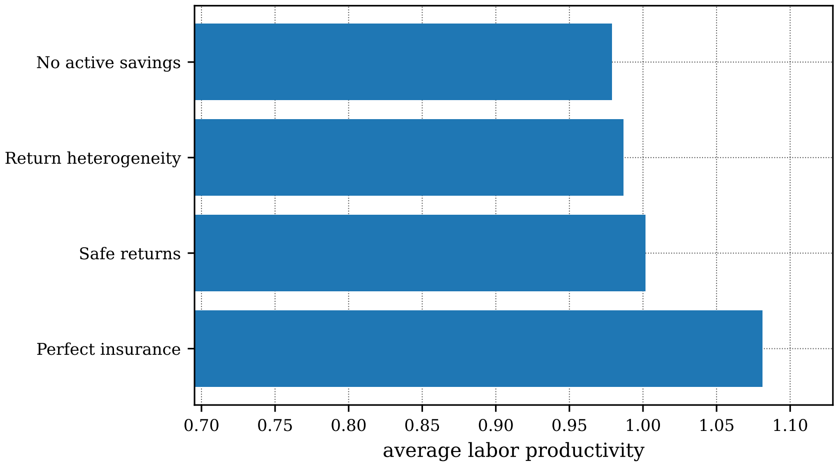



**Figure 21:** Average labor productivity from different model setups. Productivity is normalized so that average productivity equals one if everyone works the same amount.

| **Scenario** | **Aggregate labor productivity** | **Hours-wage elasticity** | **Total hours worked** |
|---|---|---|---|
| No capital markets | 0.975 | −0.13 | 97.6% |
| Return heterogeneity | 0.987 | −0.08 | 100.0% |
| Safe returns | 1.002 | +0.006 | 97.1% |
| Perfect insurance | 1.081 | +0.5 | 92.3% |

**Table 2:** Comparison of key aggregate outcomes across different levels of access to insurance. Aggregate labor productivity is normalized so that average productivity equals one if everyone works the same amount. Total hours worked is normalized so that the amount worked in the baseline model with heterogeneous returns is 100%.

increases compared to the baseline calibration with return risk.

Not surprisingly, the *"Perfect insurance"* scenario display the highest average labor productivity: 9.6% higher than in the *"Return heterogeneity"* scenario.

**Total hours worked.** The increase in aggregate labor productivity is partly taken out as consumption, partly as lower number of hours worked. In the *"Perfect insurance"* case, the total number of hours is the lowest: 7.7% lower than in the baseline with return heterogeneity.

Table 2 summarizes the resulting aggregate labor productivity, hours-wage elasticities, and total hours worked in the different economies. As can be seen, the better is access to insurance, the higher (less negative) is the hours-wage elasticity, and the higher is aggregate labor productivity.

**More on the wealth distortion channel.** To shed more light on the wealth distortion channel we conduct the following experiment: we use the resulting wealth distribution from the full model, but turn off the self-insurance possibility. In other

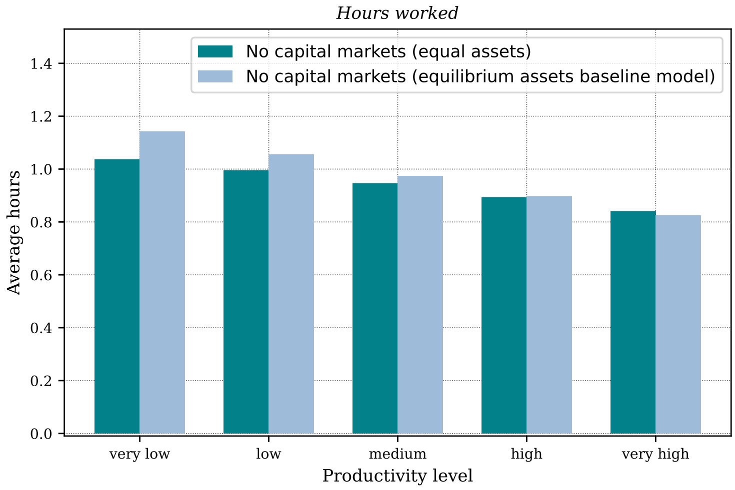



**Figure 22:** Average hours worked by productivity. Comparing the *"No capital markets"* model where everyone has the same amount of assets with a static model where the agents are given the equilibrium assets from the baseline model with return heterogeneity.

words, the households cannot increase or decrease their savings, but receive an interest rate income from their asset holdings, so that the resulting problem for the household is static with non-labor income depending on their wealth. We let all households receive the economy-wide interest rate times their assets.

The resulting hours allocation is shown in Figure 22 and is contrasted with the hours allocation in the *"No capital markets"* model, in which all agents are given an equal amount of assets.

Not surprisingly, the hours-productivity distribution is worse in the scenario with equilibrium assets. The reason is the positive association between productivity and wealth in the baseline steady state. Thus, the difference between the two sets of bars in Figure 22 is a direct measure of the wealth distortion in the baseline model.

## 8. Concluding remarks

In this paper, we advance the argument that the degree of financial sophistication in insurance markets matters not only for smoothing consumption across states of nature but also for the distribution of hours worked in the population. We show theoretically that better insurance increases the association between individual productivity and hours, and we argue that richer economies do display a higher wage-hours correlation. Using a rather standard heterogeneous-agent model, where we make an effort to match key facts on wealth dispersion and portfolio returns, we first show that the hours-wage pattern across households matches that in the data quite well; we also cross-validate our setting with microeconomic studies where income effects on labor supply are credibly identified. We then compute how much the sophistication of insurance intermediaries matters for the hours-wage correlation across households. We report results that also appear quantitatively reasonable given our empirical (cross-country as well as over time) estimates. We finally point out that the full insurance case is very different, qualitatively as well as quantitatively, from the benchmark model: the high-productive individuals now work a lot more in relative terms, total output is significantly higher, while total labor input is significantly lower.

It needs to be pointed out that, with these results, we still cannot answer the question in the title of our paper: who should work how much? A normative analysis would require pinpointing where the frictions originate and what, if anything, a benevolent government could do to overcome these frictions. Mirrlees (1971) pioneered one way to think about incomplete insurance and a dynamic contracting literature, considering multi-period extensions and generalization of Mirrlees's approach, later evolved (for early contributions, see, e.g., Green (1987) and Atkeson and Lucas Jr (1992); for a survey, see Kocherlakota (2010)). Parts of this literature touches on hours worked but the focus is on consumption insurance. In the present paper, in contrast, the focus is entirely on hours worked and its relation to productivity: aggregate productivity is endogenous since the weighting of hours and productivity on the individual level is endogenous and, as we show, influenced by the availability of insurance markets. Our paper is still motivated by the normative question, but the aim here is really more modest. Given that we find that insurance market matters quite a bit for production efficiency (in the sense of aggregate productivity), and therefore welfare. The next step is clearly to try to isolate how policy could potentially be used to diminish this welfare loss.[29]

We should finally stress that our paper is rather stylized: in making our main point as clearly as possible we have abstracted from a number of potentially important features. One is existing government policies—especially tax and transfer schemes—and these can potentially change our results quantitatively, especially given that we take income effects to be strong. Relatedly, there may be systematic differences in policies across different levels of development that also help explain the patterns we see in the data. Another relevant feature is labor-market frictions, which arguably are another important determinant of the hours-productivity correlation across individuals, as well as a life-cycle component of labor supply choices, including an endogenous extensive-margin consideration. We hope to incorporate at least some of these features in future work.

[29] In this framework, lump-sum transfers from high-wage to low-wage individuals, in the spirit of Mankiw and Weinzierl (2010), would clearly be welfare-improving from an ex ante perspective, but they might endogenously change the nature of insurance markets once the roots of incompleteness is specified and taken into account.

## References


Acemoglu, D. and F. Zilibotti (1997). Was Prometheus unbound by chance? Risk, diversification, and growth. *Journal of Political Economy 105*(4), 709–751.

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

Aiyagari, S. R. (1995). Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting. *Journal of Political Economy 103*(6), 1158–1175.

Alon, T., M. Doepke, and S. E. Coskun (2018). Trends in work and leisure: It's a family affair. In *2018 Meeting Papers*, Number 552. Society for Economic Dynamics.

Artmann, E., N. Fuchs-Schündeln, and G. Giupponi (2023). Forward-looking labor supply responses to changes in pension wealth: Evidence from Germany. Technical report.

Atkeson, A. and R. E. Lucas Jr (1992). On efficient distribution with private information. *The Review of Economic Studies 59*(3), 427–453.

Bach, L., L. E. Calvet, and P. Sodini (2020). Rich pickings? risk, return, and skill in household wealth. *American Economic Review 110*(9), 2703–2747.

Beare, B. K. and A. A. Toda (2022). Determination of pareto exponents in economic models driven by markov multiplicative processes. *Econometrica 90*(4), 1811–1833.

Bewley, T. (1986). Stationary monetary equilibrium with a continuum of independently fluctuating consumers. *Contributions to mathematical economics in honor of Gerard Debreu 79*.

Bick, A., A. Blandin, and R. Rogerson (2022). Hours and wages. *The Quarterly Journal of Economics 137*(3), 1901–1962.

Bick, A., N. Fuchs-Schundeln, and D. Lagakos (2018). How do hours worked vary with income? Cross-country evidence and implications. *American Economic Review 108*(1), 170–99.

Boppart, T. and P. Krusell (2020). Labor supply in the past, present, and future: a balanced-growth perspective. *Journal of Political Economy 128*(1), 118–157.

Boppart, T., P. Krusell, and J. Olsson (2023). Labor supply when productivity keeps growing. *Review of Economic Dynamics 50*, 61–87.

Borjas, G. J. (1980). The relationship between wages and weekly hours of work: The role of division bias. *The Journal of Human Resources 15*(3), 409–423.

Cesarini, D., E. Lindqvist, M. J. Notowidigdo, and R. Ostling (2017). The effect of wealth on individual and household labor supply: Evidence from Swedish lotteries. *American Economic Review 107*(12), 3917–46.

Chang, Y. and S.-B. Kim (2006). From individual to aggregate labor supply: A quantitative analysis based on a heterogeneous agent macroeconomy. *International Economic Review 47*(1), 1–27.

Chang, Y. and S.-B. Kim (2007). Heterogeneity and aggregation: Implications for labor-market fluctuations. *American Economic Review 97*(5), 1939–1956.

Chang, Y., S.-B. Kim, K. Kwon, and R. Rogerson (2020). Cross-sectional and aggregate labor supply. *European Economic Review 126*, 103457.

Costa, D. L. (2000). The wage and the length of the work day: From the 1890s to 1991. *Journal of Labor Economics 18*(1), 156–181.

Daminato, C. and L. Pistaferri (2024, May). Returns heterogeneity and consumption inequality over the life cycle. Working Paper 32490, National Bureau of Economic Research.

De Nardi, M. and G. Fella (2017). Saving and wealth inequality. *Review of Economic Dynamics 26*, 280–300.

Domeij, D. and M. Floden (2006). The labor-supply elasticity and borrowing constraints: Why estimates are biased. *Review of Economic Dynamics 9*(2), 242–262.

Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri (2020). Heterogeneity and persistence in returns to wealth. *Econometrica 88*(1), 115–170.

Ferraro, D. and V. Valaitis (2024). Consumption quality and employment across the wealth distribution. *Review of Economic Studies*, rdae052.

Floden, M. (2006). Labour supply and saving under uncertainty. *The Economic Journal 116*(513), 721–737.

Georgarakos, D., T. Jappelli, G. Kenny, and L. Pistaferri (2023). Labor supply response to windfall gains. *Available at SSRN 4792886*.

Giupponi, G. (2019). When income effects are large: Labor supply responses and the value of welfare transfers. Technical report.

Golosov, M., M. Graber, M. Mogstad, and D. Novgorodsky (2024). How americans respond to idiosyncratic and exogenous changes in household wealth and unearned income. *The Quarterly Journal of Economics 139*(2), 1321–1395.

Green, E. J. (1987). Lending and the smoothing of uninsurable income. Volume 1 of *Contractual arrangements for intertemporal trade*, pp. 3–25. University of Minnesota Press.

Greenwood, J., Z. Hercowitz, and G. W. Huffman (1988). Investment, capacity utilization, and the real business cycle. *The American Economic Review*, 402–417.

Greenwood, J. and B. Jovanovic (1990). Financial development, growth, and the distribution of income. *Journal of Political Economy 98*(5, Part 1), 1076–1107.

Heathcote, J., K. Storesletten, and G. L. Violante (2014). Consumption and labor supply with partial insurance: An analytical framework. *American Economic Review 104*(7), 2075–2126.

Heathcote, J., K. Storesletten, and G. L. Violante (2017). Optimal tax progressivity: An analytical framework. *The Quarterly Journal of Economics 132*(4), 1693–1754.

Hubmer, J., P. Krusell, and A. A. Smith Jr (2021). Sources of US wealth inequality: Past, present, and future. *NBER Macroeconomics Annual 35*(1), 391–455.

Huggett, M. (1993). The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control 17*(5-6), 953–969.

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

Jordà, Ò., K. Knoll, D. Kuvshinov, M. Schularick, and A. M. Taylor (2019). The rate of return on everything, 1870–2015. *The Quarterly Journal of Economics 134*(3), 1225–1298.

Juhn, C., K. M. Murphy, R. H. Topel, J. L. Yellen, and M. N. Baily (1991). Why has the natural rate of unemployment increased over time? *Brookings Papers on Economic Activity 1991*(2), 75–142.

Kartashova, K. (2014). Private equity premium puzzle revisited. *American Economic Review 104*(10), 3297–3334.

Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica 131*, 207–248.

Keynes, J. M. (1930). Treatise on money.

King, R. G. and R. Levine (1993). Finance and growth: Schumpeter might be right. *The Quarterly Journal of Economics 108*(3), 717–737.

King, R. G., C. I. Plosser, and S. T. Rebelo (1988). Production, growth and business cycles. *Journal of Monetary Economics 21*(2), 195–232.

Kocherlakota, N. R. (2010). *The new dynamic public finance*. Princeton University Press.

Krueger, D., K. Mitman, and F. Perri (2016). Chapter 11 - Macroeconomics and household heterogeneity. Volume 2 of *Handbook of Macroeconomics*, pp. 843–921. Elsevier.

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

Kuhn, M., J.-V. Ríos-Rull, et al. (2016). 2013 update on the US earnings, income, and wealth distributional facts: A view from macroeconomics. *Federal Reserve Bank of Minneapolis Quarterly Review 37*(1), 2–73.

MaCurdy, T. E. (1981). An empirical model of labor supply in a life-cycle setting. *Journal of Political Economy 89*(6), 1059–1085.

Mankiw, N. G. and M. Weinzierl (2010, February). The optimal taxation of height: A case study of utilitarian income redistribution. *American Economic Journal: Economic Policy 2*(1), 155–76.

Mantovani, C. (2023). Hours-biased technological change. Technical report, Working Paper.

Marcet, A., F. Obiols-Homs, and P. Weil (2007). Incomplete markets, labor supply and capital accumulation. *Journal of Monetary Economics 54*(8), 2621–2635.

Mendoza, E. G., V. Quadrini, and J.-V. Rios-Rull (2009). Financial integration, financial development, and global imbalances. *Journal of Political Economy 117*(3), 371–416.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies 38*(2), 175–208.

Piazzesi, M. and M. Schneider (2016). Chapter 19 - Housing and macroeconomics. Volume 2 of *Handbook of Macroeconomics*, pp. 1547–1640. Elsevier.

Pijoan-Mas, J. (2006). Precautionary savings or working longer hours? *Review of Economic Dynamics 9*(2), 326–352.

Rajan, R. and L. Zingales (1996). Financial dependence and growth.

Yum, M. (2018). On the distribution of wealth and employment. *Review of Economic Dynamics 30*, 86–105.

# Online appendix

## A. Data description

We list the central variables from the CPS ASEC used in our empirical analysis in Table A1. Some of them are computed from several variables, a process we describe below.

| Variable | Description | CPS ASEC variable(s) |
|---|---|---|
| Hours worked | Total hours worked previous calendar year. | WKSWORK1, UHRSWORKLY |
| Unemployment | Number of weeks that the respondent looked for work or was on layoff during the previous calendar year. | WKSUNEM1 |
| Labor force status | Indicator whether the respondent participated in the labor force during the previous week. | LABFORCE |
| Labor earnings | Total pre-tax wage and salary income for the previous calendar year. | INCWAGE |
| Age | Respondent's age | AGE |
| Race | Respondent's race, recoded to *black* indicator variable | RACE |
| Sex | Respondent's sex | SEX |
| Education | Educational attainment, recoded to three groups | EDUC |
| Weight | Respondent-level sampling weights | ASECWT |

**Table A1:** Variables from CPS ASEC used in the analysis

**Hours worked** We calculate the hours worked previous year by combining information about how many weeks the respondent worked, and how many hours per week the respondent normally worked. For weeks worked we use the variable WKSWRK1 which gives the answer to the question how many weeks the respondent worked the preceding year (for profit, pay or as an as an unpaid family worker). The variable UHRSWORKLY gives information about the number of hours per week that respondents usually worked if they worked during the previous calendar year. We use this variable (as opposed to, e.g., hours worked last week) to get a consistent measure of the hours worked *the preceding year*.

**Unemployment** WKSUNEM1 gives the number of weeks that the respondent looked for work or was on layoff during the preceding calendar year. If this variable is non-zero we discard the respondent, since then the choice of hours and weeks was likely not driven by labor supply considerations, but rather lack of labor demand.

**Labor force participation** LABFORCE indicates whether the respondent participated in the labor force during the preceding week. Even though the question is about the preceding week, we discard the respondent if he/she reports to not be in the labor force, since this might be an indicator of having dropped out of the labor force already at some point the previous year.

**Labor earnings** INCWAGE indicates the respondent's total pre-tax wage and salary income for the previous calendar year. We CPI adjust the values using 2010 as base year.

**Hourly wage** We construct the hourly wage by dividing "Labor earnings" by our preferred measure of "Hours worked".

**Education** EDUC indicates respondents' educational attainment, as measured by the highest year of school or degree completed. We recode this into three groups: *less than high-school*, *high-school diploma or equivalent*, and *some college and above*.

**Alternative 2 hours worked** As an alternative for hours worked we also use the variable AHRSWORKT, actual hours worked last week.

**Alternative 3 hours worked** As an alternative for hours worked we also use a complementary measure of the total hours worked in a year, using the variables AHRSWORKT (actual hours worked last week) and WKSWORK2 (weeks worked last year, intervalled). For the latter variable, we assign the midpoint of the interval.

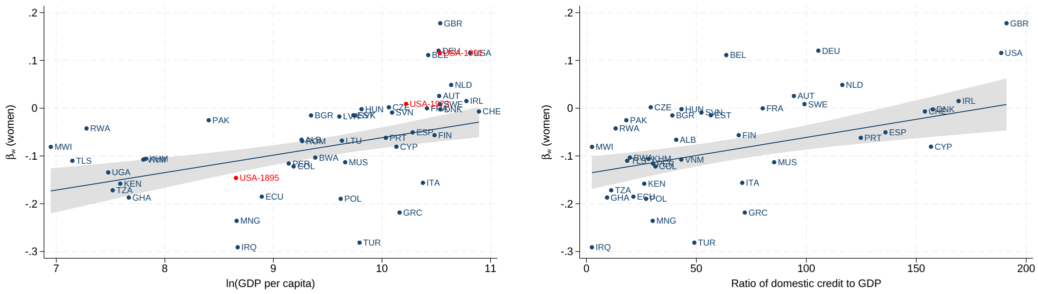
## B. Additional empirical results

### B.1. Hours-wage correlations for women

Figure A1 shows the hours-wage correlation for women against different indicators of financial development across countries. The corresponding figure for men is Figure 1 in the main text.

### B.2. Extensive margin analysis of the U.S.

Figure A2 shows the same information as Figure 4 in the main text, except that the enumerator includes both individuals who work and the unemployed (EMPSTAT $\in \{10, 12, 20, 21, 22\}$).




**(a)** Hours-wage elasticities vs. log(GPD) per capita. The red dots indicate the historical observations from the US, taken from Costa (2000).

**(b)** Hours-wage elasticities vs. ratio of domestic credit to GDP (source IMF).

**Figure A1:** Country-specific elasticities of hours to wages for **women** (taken from Bick et al. (2018)) vs. different indicators of financial development.

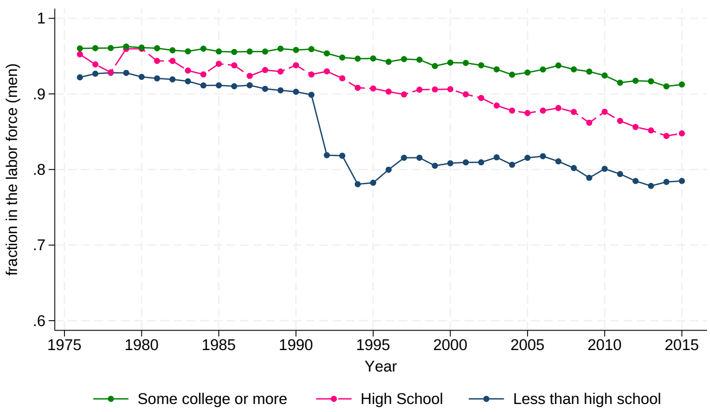



**Figure A2:** Fraction of the male population aged 25–54 in the labor force, by education level (men). Source: CPS ASEC.

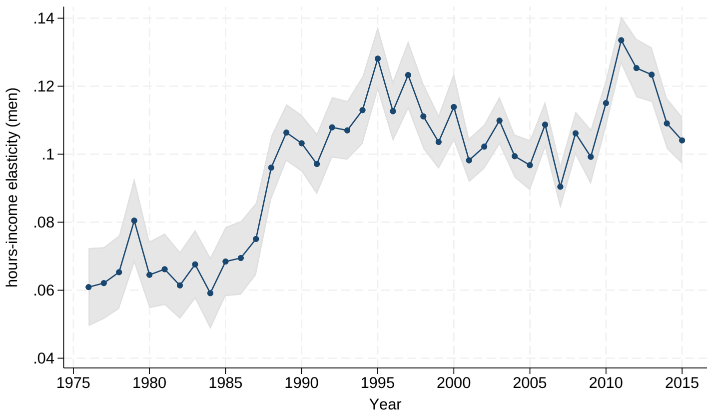



**Figure A3:** Hours-earnings relationship. The graph shows the estimate of $\beta_w/(1+\beta_w)$ from (A1). Source: CPS ASEC.

### B.3. Robustness checks of the hours-wage elasticity

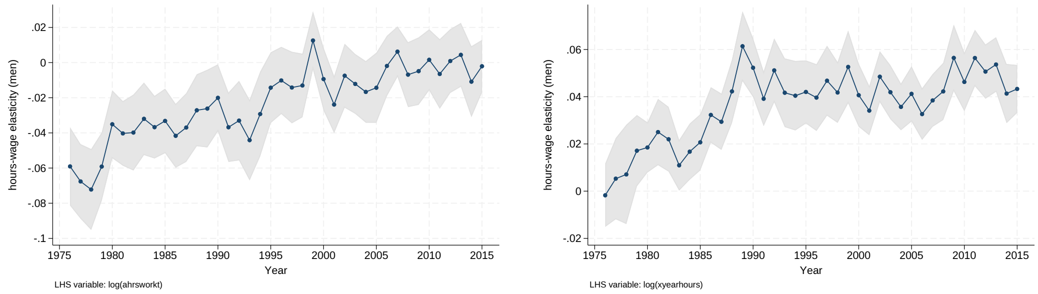
With measurement errors in the hours data and the wage constructed from information about earnings and hours, the hours-wage regression (1) suffers from what Borjas (1980) calls "division bias". If an individual reports too high hours, his/her imputed hourly wage, computed as earnings divided by hours, is automatically too low. As a consequence, the estimated $\beta_w$ coefficient will be biased towards minus one.

To avoid the division bias, one could replace the wage in (1) with earnings/hours explicitly and then solve for $\log h$, i.e., run the following regression instead:

$$\log h_i = \frac{\alpha}{1+\beta_w} + \frac{\beta_w}{1+\beta_w} \log e_i + \frac{\delta}{1+\beta_w} \mathbf{Z}_i + \frac{\varepsilon_i}{1+\beta_w} \quad \text{(A1)}$$

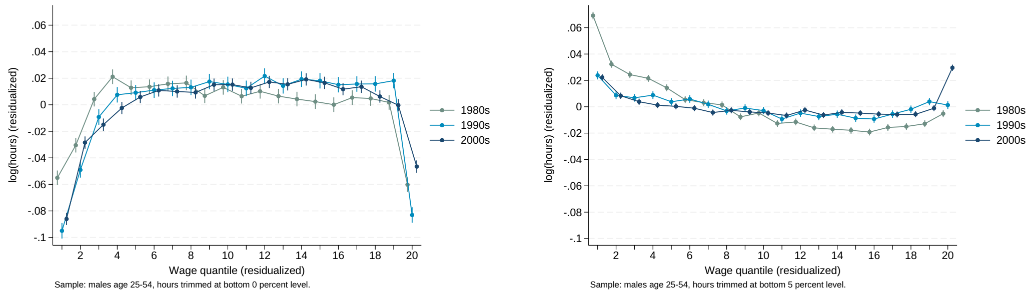
with $e$ denoting earnings and $\mathbf{Z}$ being the vector of control variables included in the regression. However, if (1) is the true behavioral relation, the earnings on the right-hand side of (A1) is endogenous, and it can be shown that if one applies ordinary least squares, the estimate of the wage elasticity $\beta_w$ will be upward biased (see Borjas (1980)). Nevertheless, we run (A1) and the results are shown in Figure A3. The time series starts at $\beta_w/(1+\beta_w) = 0.06$, in other words corresponding to $\beta_w = 0.064$, as anticipated higher than what is estimated with equation (1), but the trend over time is the same: upward-sloping.

As a robustness we also perform a "cross-division" as suggested by Borjas (1980). We run the same regression as (1) but replace the left-hand-side with an alternative hours variable that does not have the spurious correlation with the wage estimate. I.e., we run the following regression:




**(a)** Hours, Alternative 2. **(b)** Hours, Alternative 3.

**Figure A4:** Estimates of hours-wage elasticity using other alternative measures of hours.




**(a)** Trimming bottom 0% of yearly hours obs. **(b)** Trimming bottom 5% of yearly hours obs.

**Figure A5:** Hours-wage distribution (residualized values) in the U.S. for men. Source: CPS ASEC.

$$\log h_i^{alt} = \alpha + \beta_w \log\left(e_i/h_i\right) + \delta Z_i + \varepsilon_i \tag{A2}$$

As a first alternative for $h^{alt}$ we use actual hours worked last week, and as a second alternative we create a measure for yearly hours by using actual hours last week times another variable for weeks worked last year than what was used for the wage calculation. The variables are described in Table A1. On the right hand side, $h_i$ is the preferred measure of hours worked just as before.

Figure A4 shows the resulting $\beta_w$ coefficients. Reassuringly, both alternative measures also display an upward time trend in the estimated coefficient.

## B.4. Robustness check trimming of low-hours observations

A concern is that some of the patterns in Figure 3 could be due to people misreporting their hours. Especially observations with very low hours could result in low hours and high imputed wage. Figure A5 shows the hours-wage distribution in the U.S. for men aged 25–54 for two alternative choices of trimming low hours.

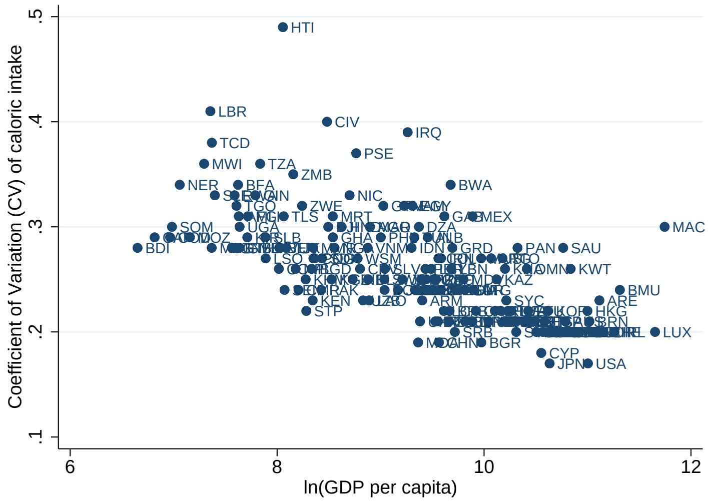
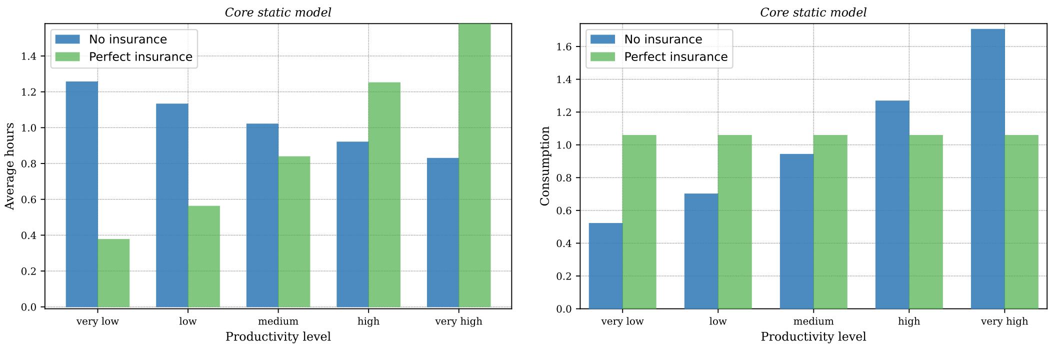



**Figure A6:** Inequality in food consumption vs. GDP per capita. Inequality in food consumption is measured by the coefficient of variation in per capita caloric consumption. Source: FAO; World Bank.

## B.5. Consumption inequality across countries

Assuming that richer countries have access to better insurance, another prediction is that the consumption inequality should be lower in rich countries. Consumption is difficult to measure, especially if one wants to include countries from the whole development spectrum. Figure A6 shows one take: the coefficient of variation (CV) of caloric intake for a number of countries, as estimated by FAO (World Bank).

As can be seen, the pattern is clear: the richer the country, the less inequality in caloric intake. Food is only part of the consumption basket, and there is severe issues with measurement for this data (some numbers are even imputed), but the overall picture is in line with the prediction: there is less consumption inequality in richer countries.




**(a)** Hours distribution

**(b)** Consumption distribution

**Figure A7:** Resulting hours and consumption allocations, comparing Ecoomy 1 (no insurance) with Economy 2 (full insurance). Frisch elasticity $\theta = 1.0$.

## C. Alternative calibrations of core static model

In this section we present results from the core static model with alternative calibrations.

### C.1. A higher Frisch elasticity

Figure A7 shows the hours allocation and the consumption allocation from a model identical to the one in Section 3 except the Frisch elasticity, which is here set to $\theta = 1$.

Compared to the model in Section 3 (with $\theta = 0.5$), the higher Frisch elasticity gives larger differences between the Economy 1 (with no insurance) and Economy 2 (with full insurance). Average labor productivity is here 21% higher with perfect insurance. Hours worked are 12% lower and aggregate consumption is 7% higher. With a higher Frisch elasticity $\theta$, it is less costly for a planner to redistribute work.

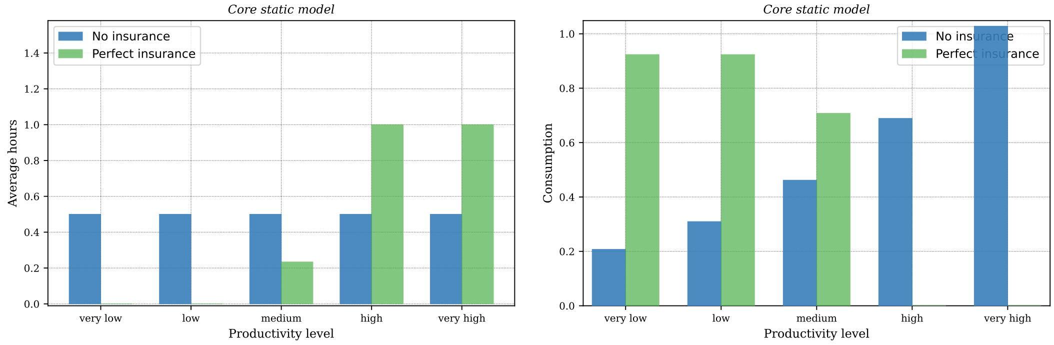
### C.2. Non-separable preferences

Figure A8 shows the hours allocation and the consumption allocation from a model identical to the one in Section 3 except that the utility function used is here of the Cobb-Douglas type:

$$u(c, h) = c^{\alpha}(1-h)^{1-\alpha}$$

and we restrict the hours choice to $h \in [0, 1]$. In the case of no insurance, households would work an equal amount, regardless of their resulting productivity level. This is a direct consequence of the fact that Cobb-Douglas preferences are part of the KPR class (King, Plosser, and Rebelo (1988)) where the income and substitution effects exactly cancel.

The hours distribution in the case of full insurance is qualitatively similar but




**(a)** Hours distribution **(b)** Consumption distribution

**Figure A8:** Resulting hours and consumption allocations, comparing Economy 1 (no insurance) with Economy 2 (full insurance). Utility function of the Cobb-Douglas type, $\alpha = 1/3$.

quantitatively very different to the model presented in section 3. As in the previous case, the planner lets the high-productive work the most, but the difference is now even larger than before. The planner let the households with the lowest productivity not work at all, while the ones with high productivity work maximum hours (capped at 1 in this case).

This translates into large consumption differences as well for the planner (full insurance) solution. A planner would let the high-productive households who work a lot not consume at all, while the low-productive households both enjoy leisure and consume a lot.

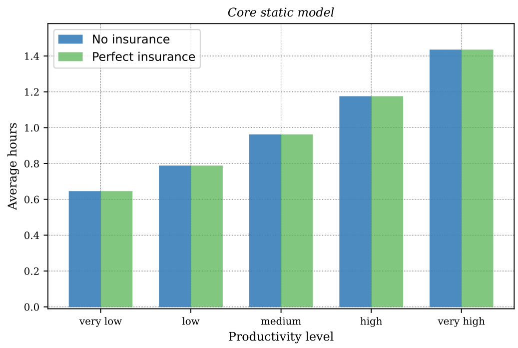
## C.3. GHH preferences

For completeness, we also include an example where the preferences are of the GHH type (Greenwood et al., 1988) (even though such preferences are at odds with the empirical evidence and do not allow for balanced growth):

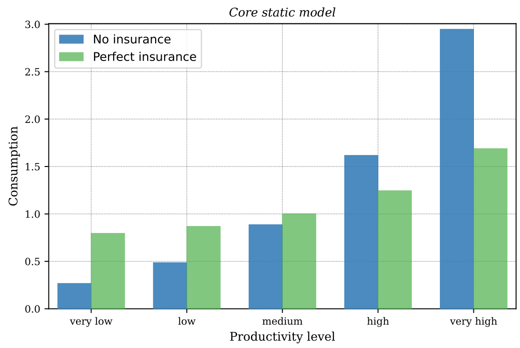
$$u(c,h) = \frac{1}{1-\gamma}\left(c - \psi\frac{h^{1+1/\theta}}{1+1/\theta}\right)^{1-\gamma}.$$

With these preferences, there is no income effect at all.

The hours allocation is thus identical in the two economies with no insurance or full insurance. However, the consumption allocation differs between the no-insurance and full-insurance case. A planner wants to equalize the sum of consumption and disutility from working across households and is therefore smoothing consumption, not perfectly, but more than in the no-insurance case. The productivity levels are normalized so that average productivity is one if all households work the same amount.




**(a)** Hours distribution




**(b)** Consumption distribution

**Figure A9:** Resulting hours and consumption allocations, comparing Economy 1 (no insurance) with Economy 2 (full insurance). Utility function of the GHH type.

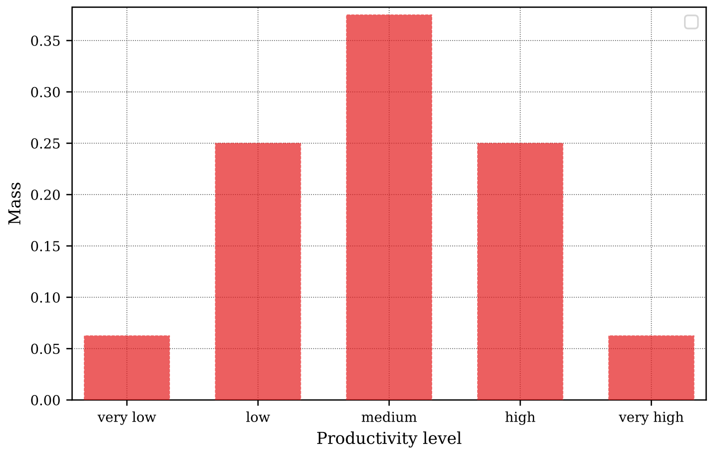



**Figure A10:** Productivity distribution.

## C.4. Distribution used

Figure A10 shows the distribution across productivity states used for the minimal working example model.

The productivity states are equidistant in logs, and the productivity levels and the mass at each are given by the ergodic distribution from the discretization of the AR(1) productivity process in the full model described in Section 4.

| Return state | $r_x^1$ | $r_x^2$ | $r_x^3$ | $r_x^4$ | $r_x^5$ |
|---|---|---|---|---|---|
| Excess return | -20.0% | -10.0% | -0.0% | 10.0% | 20.0% |
| Ergodic mass | 0.0625 | 0.2500 | 0.3750 | 0.2500 | 0.0625 |
| *Transition matrix* | | | | | |
| $r_x^1$ | 0.66 | 0.29 | 0.05 | 0.00 | 0.00 |
| $r_x^2$ | 0.07 | 0.68 | 0.22 | 0.02 | 0.00 |
| $r_x^3$ | 0.01 | 0.15 | 0.69 | 0.15 | 0.01 |
| $r_x^4$ | 0.00 | 0.02 | 0.22 | 0.68 | 0.07 |
| $r_x^5$ | 0.00 | 0.00 | 0.05 | 0.29 | 0.66 |

**Table A2:** Resulting levels, ergodic distribution and transition matrix for the excess return process.

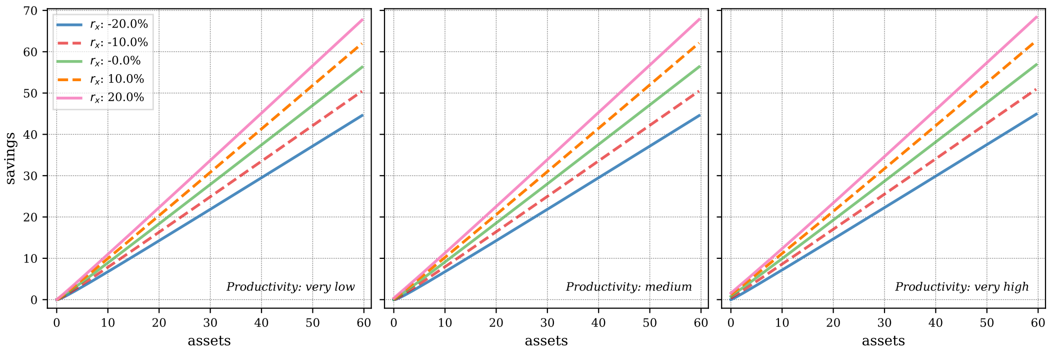



**Figure A11:** Policy functions for savings, baseline calibration. Each panel is a different productivity state: to the left the lowest productivity, in the middle the middle state, to the right the highest productivity. Each line represents a different excess return state. The x-axis indicates beginning-of-period assets (before interest rate income).

# D. Additional results for baseline calibration

Table A2 shows the resulting excess return levels, the ergodic distribution and the transition matrix for the persistent heterogeneous returns in the baseline calibration.

Figure A11 shows the policy functions for savings for lowest productivity state (left panel), the middle productivity state (middle panel) and the highest productivity state (right panel). The lines indicate different levels of current excess return. As the different panels show, current productivity does not impact the savings decision to any larger extent once the agent is rich enough.

Figure A12 shows the mass of households within each productivity/excess return cell, as indicated by the size of the bubble. The color shows hours worked, with lighter color indicating more hours worked.

## D.1. More details, comparing the wealth effects to those in Golosov et al. (2024)

Another observation by Golosov et al. (2024) is that richer individuals respond by decreasing their labor income *more* than do poorer individuals. Figure A13 shows

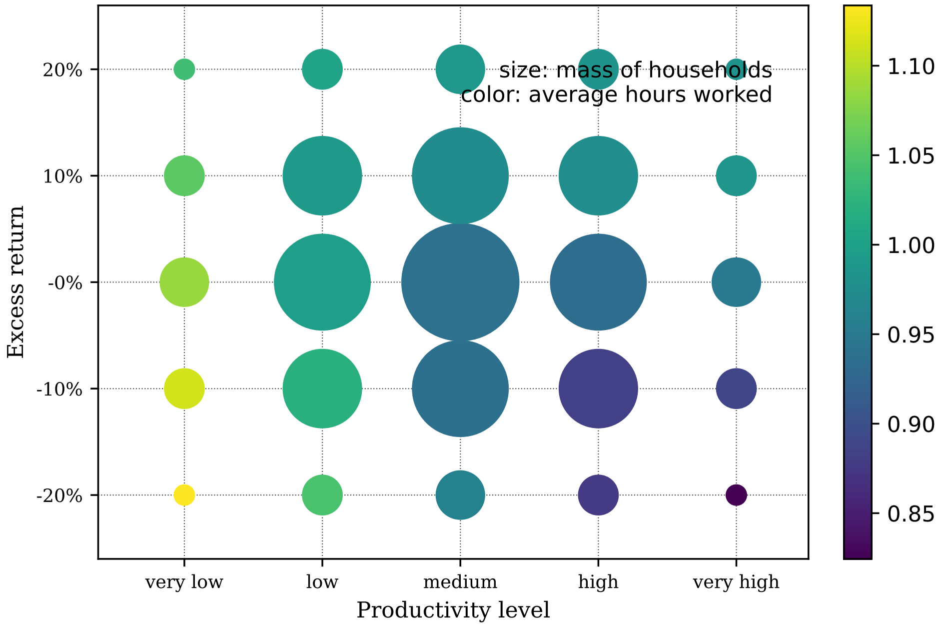



**Figure A12:** Average outcome for each productivity/excess return cell. Size of bubble indicate mass of households within that cell. Color is hours worked, with lighter color indicating more hours worked.

the ratio of the response by the fourth quartile (in pre-win income) to the response by the first quartile. The model results are again reported by size of win, while the Golosov et al. (2024) estimate is for the whole sample (the paper does not report these results by size of win). A ratio above 1 means that the income-rich lottery winners respond more than the income-poor winners. Qualitatively, our model reproduces the stronger responses by the income-rich for all prize sizes except the smallest ones (as indicated by the ratio being larger than one for prize sizes 100K or larger), even though the magnitude of the difference is somewhat smaller than indicated in the empirical evidence. For the very smallest wins, our model estimates give the reverse: richer individuals respond more.

The earnings response from a lottery win depends on on the wage rate and on how much hours worked change. In the model, for all prize sizes, the income-poor individuals adjust their hours more, while the income-rich have a higher wage rate on average. Thus, the total effect on earnings from a wealth shock depends on the relative strength of the the hours vs. wage effects. The response to a small win is stronger in terms of hours. Thus, for small wins, the difference in the change in hours by the income-poor and the income-rich is more pronounced, and thereby the hours effect dominates. We therefore see a stronger response by the income-poor for the small prize sizes. For the larger prize sizes, the differences in wage levels matter more: the hours response is still slightly stronger for the income-poor, but the difference is not pronounced enough to outweigh the fact that the income-rich

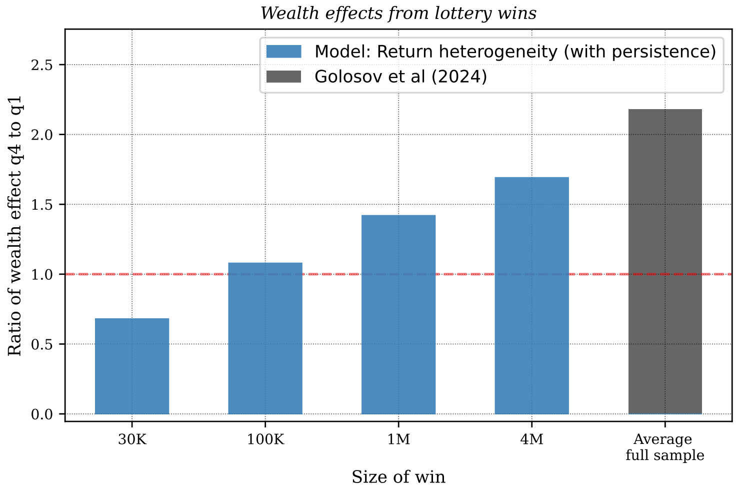



**Figure A13:** Ratio of the wealth effect estimated for the income-richest quartile to that for the income-poorest quartile, by prize size. The number for Golosov et al. (2024) is taken from their Table 3.1, row "Winner Wage Earnings" (i.e., = 3.06/1.40). Note that the Golosov et al. (2024) numbers are reported for the whole distribution of sizes of wins. A ratio above 1 indicates that rich individuals respond more strongly.

have higher wages and thereby their earnings change more for each hour of reduced working time.

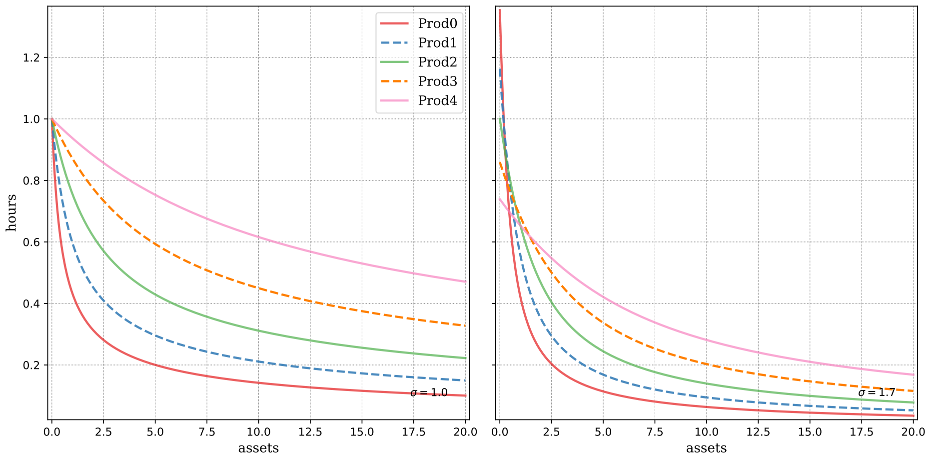



**Figure A14:** Policy functions for hours worked for a static problem, as a function of assets (non-labor income) on the x-axis and productivity level. Plots for two different values of $\sigma$: to the left, $\sigma = 1.0$, to the right, $\sigma = 1.7$.

# E. Additional results for different financial intermediary sophistication

## E.1. No capital markets

Figure A14 shows policy functions for hours in a static model as a function of non-labor income (on the x-axis) and different productivity levels, for two different values of $\sigma$.

The case of $\sigma = 1.0$ gives standard KPR preferences where the income effect and the substitution effect exactly cancel, and therefore the hours choice is the same for all productivity levels, given zero non-labor income. However, with positive assets the income effect from an increase in productivity (wage) is muted, the substitution effect dominates, and higher productivity therefore leads to more hours worked.

The case of $\sigma = 1.7$ is what is used in our baseline calibration. With this value of $\sigma$, the income effect slightly dominates and therefore, with zero non-labor income, higher productivity (wage) leads to lower hours worked. The income effect from a wage increase dominates for lower wealth levels. However, when the non-labor income becomes large, the income effect becomes muted and eventually the substitution effect dominates and higher productivity (wage) means more hours worked.

In the model described in Section 7.1 the non-labor income is relatively small, and therefore the income effect dominates the substitution effect, and the low-productive work more than the high-productive in equilibrium.

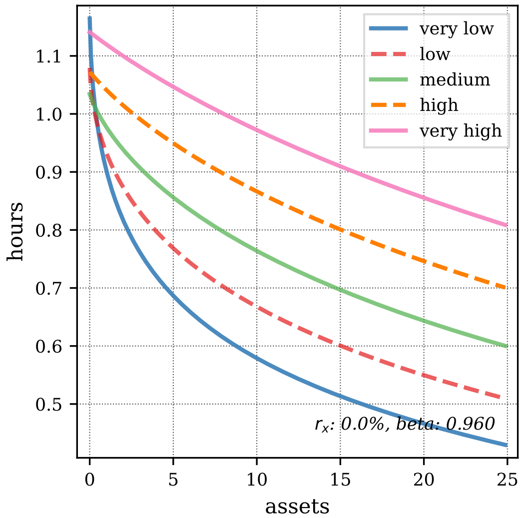



**Figure A15:** Policy functions for hours worked, model with safe returns.

## E.2. Safe returns

Figure A15 shows the policy functions for hours worked in the model with safe returns. All households have the same excess return in this model, therefore the hours choice depends only on the current asset holdings and current productivity. As the figure shows, for very low asset holdings the income effect dominates the substitution effect, and low-productive individuals work the most. For higher asset levels, the substitution effect dominates and the high-productive work the most. The intuition for the shape of the policy functions can be found in the static model policy functions shown in Figure A14, since the productivity process is relatively persistent.

To understand the hours distribution in the model, we need to know the joint distribution over assets and productivity in equilibrium. Figure Figure A16 shows the mass of households (as indicated by the size of the bubbles) in the asset/productivity space. As can be seen, households with higher productivity are on average richer.

Figure Figure A17 shows the resulting wealth distribution. As can be seen, there is no trace of any Pareto tail in this model.

Figure A18 shows average hours by wealth decile for the model with safe returns.

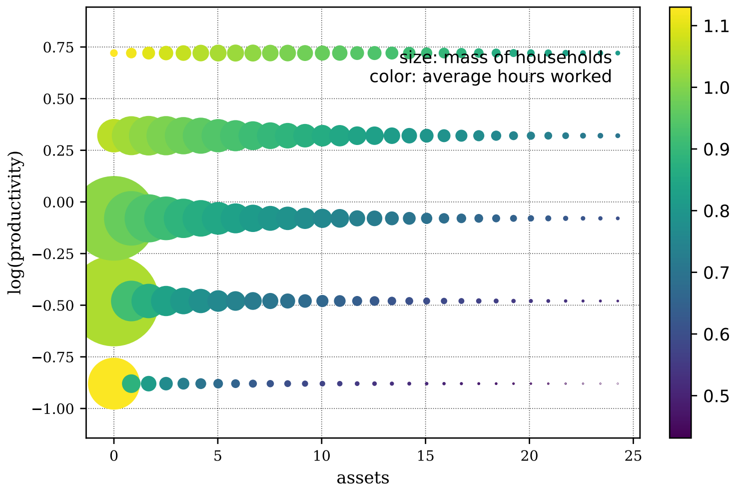



**Figure A16:** Distribution of households, model with safe returns.

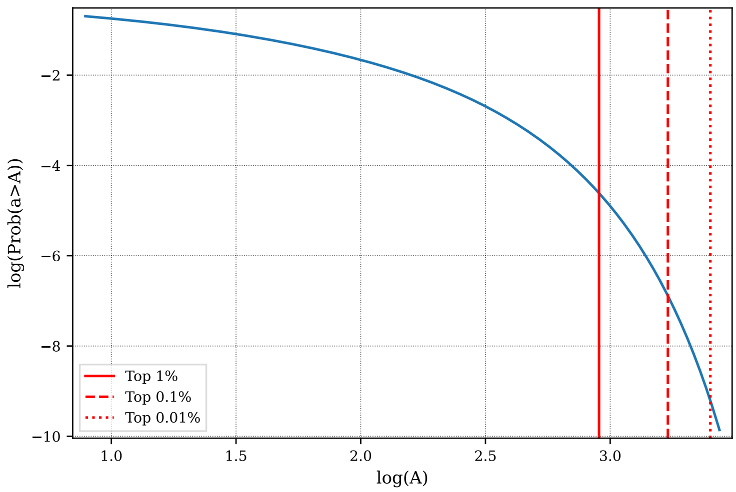



**Figure A17:** Log wealth distribution, model with safe returns.

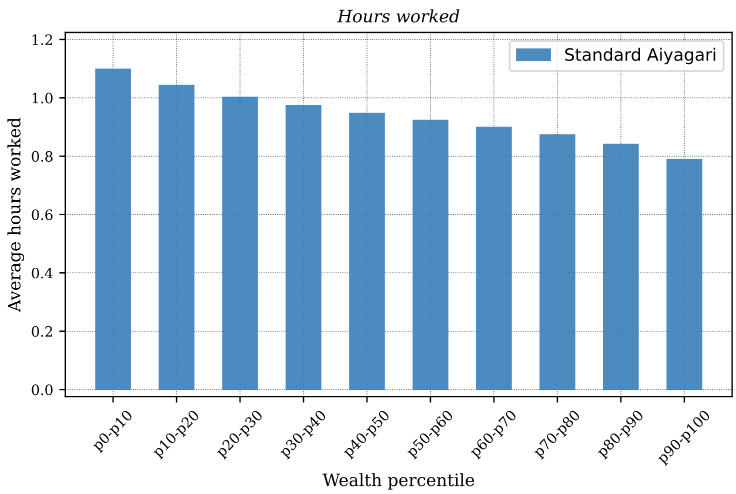



**Figure A18:** Hours worked by wealth decile, model with safe returns.

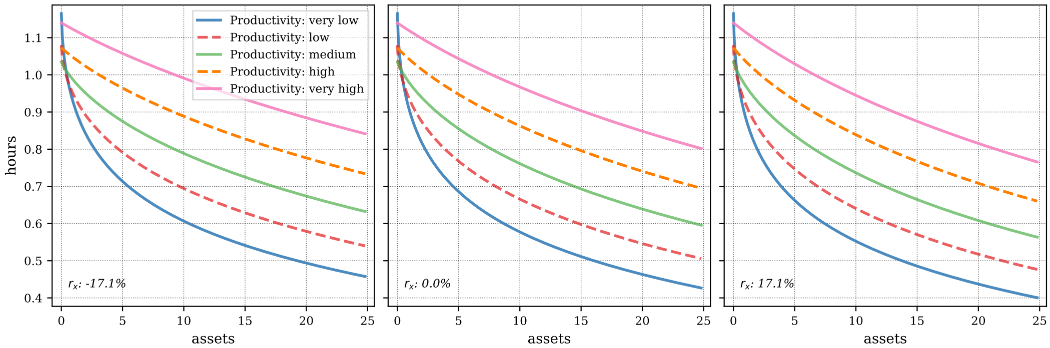



**Figure A19:** Policy functions for hours worked, model with iid excess return shocks.

## F. A model with with iid return shocks

To evaluate the importance of the persistence in the excess return shocks we also solve a model in which the return heterogeneity displays the same unconditional standard deviation (10%) but no persistence.

Figure A19 shows the policy functions for hours as a function of beginning-of-period assets (x-axis) and current productivity (the different lines) for the lowest, the middle, and the highest excess return draw. As the figure shows, agents with a higher excess return draw work slightly less (most visible for the relatively rich agents). The reason is that the graph shows assets in the beginning of the period: thus, with higher excess return the agent will have more assets at hand after they receive their interest rate and is about to make their hours choice.

Figure A20 shows the hours choice for each productivity/excess return cell (as indicated by the color). Here, as opposed to in the baseline model (as described in Figure 13 in the main text) the relationships are monotone. Hours are decreasing in excess return due to the immediate interest rate payments giving rise to a wealth effect. Hours are decreasing in productivity due to the wealth effect. There is no desire to intertemporally take advantage of high returns (since they are iid), and therefore the additional distortion that we have in the baseline model is not present here.

Figure A21 shows the resulting hours worked by wealth. As can be seen, hours are more clearly downward sloping than in the baseline model, precisely because we do not have the additional excess return distortion of the hours choices. In the baseline model, wealth and current excess return is positively correlated, and agents with a good current excess return draw want to intertemporally substitute their work effort while they are in that state.

Figure A22 finally shows the resulting average hours worked by productivity level, and contrast it to the outcome with perfect insurance (which of course coincides with the perfect insurance case in the main paper, since the fundamentals of

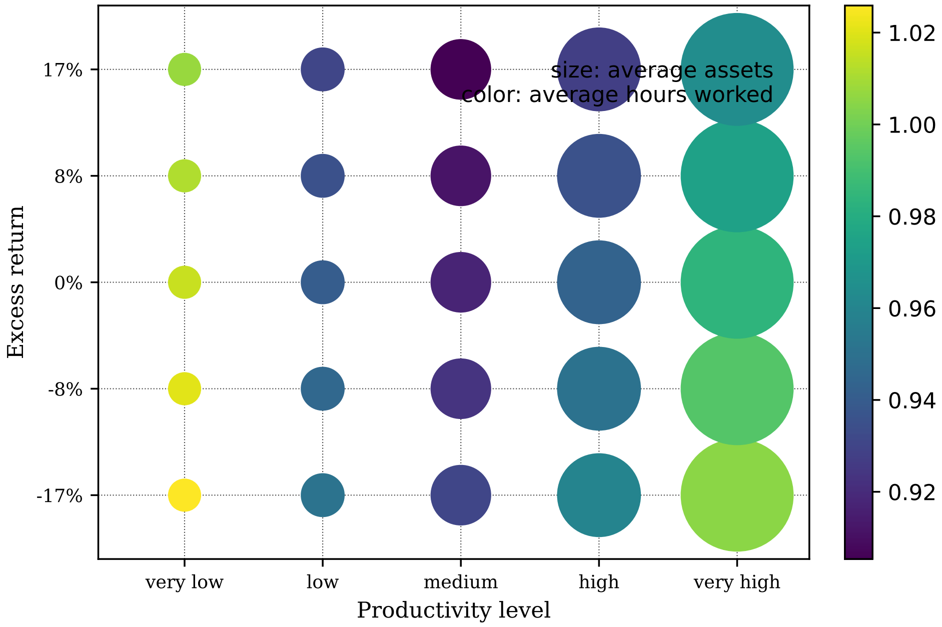



**Figure A20:** Average outcome for each productivity/excess return cell. Size of bubble indicates average asset holdings in the cell. Color is hours worked, with lighter color indicating more hours worked. Model with iid excess return shocks.

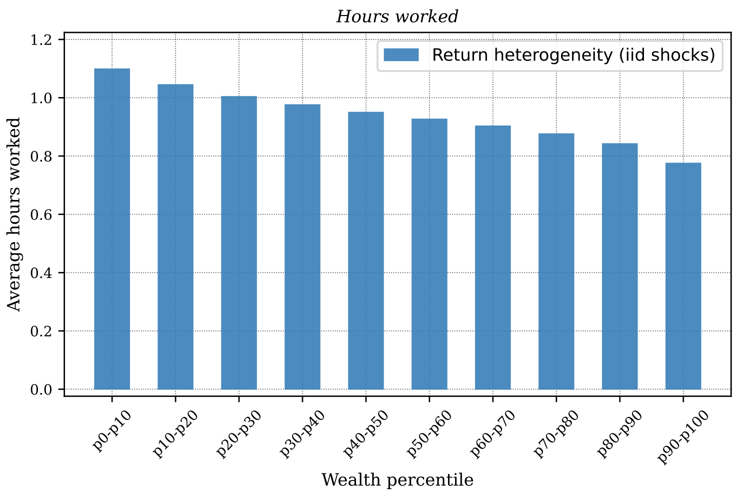



**Figure A21:** Average hours by wealth, model with iid excess return shocks.


**Figure A22:** Average hours by productivity level, model with iid excess return shocks.

the economy is kept constant).

Not surprisingly, the wealth Gini in this model is substantially lower than in the model when the excess return shocks are persistent. With excess return shocks being iid, the wealth Gini is 0.59.

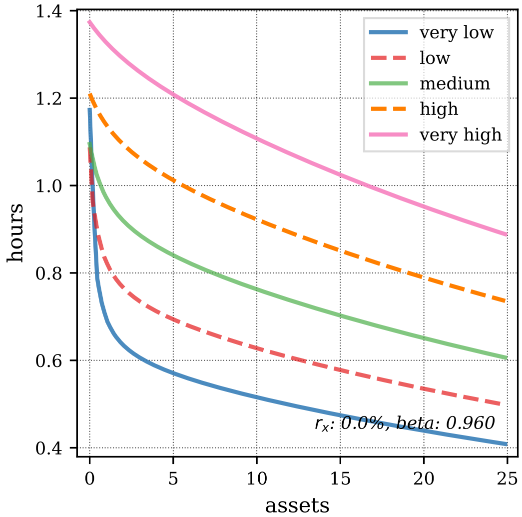



**Figure A23:** Policy functions for hours worked, model with iid productivity shocks.

## G. A model with iid productivity shocks

A hypothetical and counterfactual scenario (at odds with the evidence from the data) is one in which the productivity shocks are iid. This scenario is included to highlight the importance of the intertemporal insurance in a less risky environment.

With an iid productivity process it is not as costly to insure against adverse (temporary) bad productivity shocks by self-insurance via savings. Since the productivity process is iid, there is no correlation between assets and productivity, and therefore the wealth distortion is low. The intertemporal substitution is consequently high: households can take the opportunity to work a lot when their productivity is high, and work less when they have lower productivity.

Figure A23 shows the policy functions for hours worked in the model with iid productivity shocks. All households have the same excess return in this model, therefore the hours choice depends on the current asset holdings and current productivity. As the figure shows, the intratemporal substitution force is strong, and even for low asset levels, the high-productive work the most.

Figure A24 shows the resulting joint distribution over assets and productivity, as can be seen there is no correlation between assets and productivity. Figure A25 shows the resulting average hours by productivity, and contrast the results to the outcome from a model with perfect insurance. The high-productive households work the most in this model, and the aggregate labor productivity comes close to a model with full insurance.

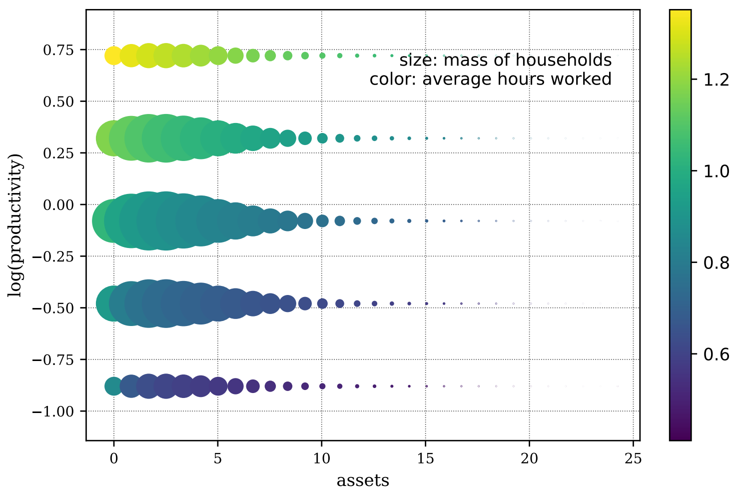



**Figure A24:** Distribution of households, model with iid productivity shocks.

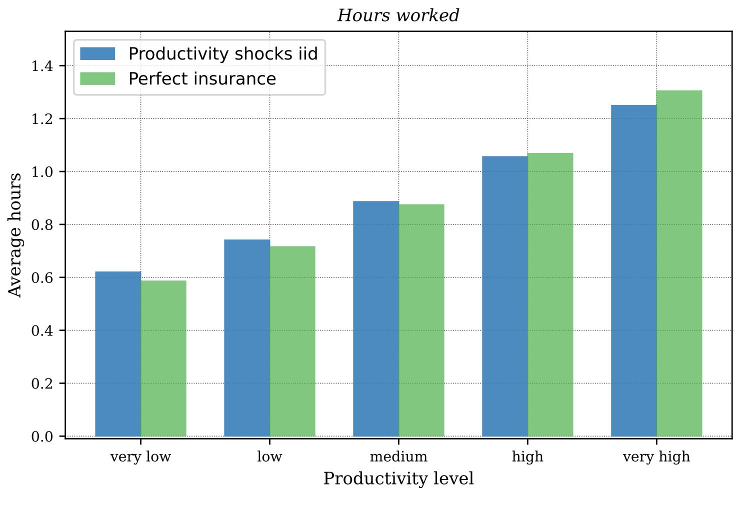



**Figure A25:** Average hours by productivity level, model with iid productivity shocks.