Monetary Policy and Top Inequality*

Jacob Sundram†

June 19, 2025

Abstract

How do monetary policy and inequality interact? Recently, there has been renewed interest in this question. However, models used to study this severely understate wealth inequality. I argue this is due to missing that wealthy households earn higher returns. I therefore construct a model calibrated to a new dataset of returns across US households. The model matches distributions of households and the pass-through of aggregates to households, which are sufficient statistics for the distributional effects of monetary policy. Standard models do not match this, so they miss how much expansionary monetary policy benefits the ultra-rich. With heterogeneous returns, the top 0.1% gain 11% of the income generated by monetary policy—an order of magnitude more than standard. This redistribution makes monetary policy less effective. Fiscal policy is more equal. Policymakers concerned about inequality should consider this.

Keywords: Monetary policy, business cycles, heterogeneous households
JEL Codes: D31, E21, E32, E52

*I am thankful for useful comments from Jeppe Druedahl, Raphaël Huleux, Moritz Lenel, Søren Hove Ravn, Karthik Sastry, Gianluca Violante, and seminar participants at the 2025 DAEiNA meeting, University of Copenhagen, and Princeton University. I am grateful for financial support from the Carlsberg Foundation (Grant CF20-0546). First version: April 2025.

†Department of Economics, University of Copenhagen, jacob.sundram@econ.ku.dk

1 Introduction

Who gains and who loses from monetary policy? And how does inequality affect the transmission of monetary policy? Recently, there has been a renewed interest in these questions. This has spawned a new class of macroeconomic models in the last decade: Heterogeneous Agent New Keynesian (HANK) models. These models add inequality to the workhorse New Keynesian model, letting the researcher study the interaction of inequality and monetary policy. However, these models severely understate wealth inequality, particularly at the top of the wealth distribution: In standard models, there are no households with more than around 30 million USD in wealth. This is despite these households accounting for almost 40% of all wealth in the US. This restricts our understanding of the interaction between monetary policy and inequality.

In this paper, I argue that standard models understate top wealth inequality due to missing heterogeneity in returns. In particular, it is a robust feature of the data that wealthy households earn higher returns on their wealth. For this reason, I add heterogeneous returns to a HANK model. To calibrate the heterogeneous returns, I construct a new panel dataset of returns across US households for 2001–2019 based on the Panel Survey of Income Dynamics (PSID).

I find that adding a realistic degree of heterogeneity in returns to a HANK model makes the model fit the top of the wealth distribution. I then study who gains and who loses from easy monetary policy. Using a sufficient statistics approach, I find that 4 moments shape the distribution of income in response to monetary policy: The wealth distribution, the income distribution, and the pass-through of aggregate earnings and returns to individual earnings and returns. Standard models provide a poor fit to all 4 moments. The model with heterogeneous returns in this paper fits all 4 moments. Using the model, I find that easy monetary policy disproportionately benefits the ultra-rich: The top 0.1% gain 11% of the income generated by monetary policy, which is an order of magnitude more than in standard models. Since the ultra-rich have low marginal propensities to consume, this implies that monetary policy is less effective than in standard models.

The key to these results is the heterogeneous returns. To calibrate these, I turn to two different datasets. First, I construct a new panel dataset of returns across US households for 2001-2019 based on the PSID. A key advantage of this dataset compared to the literature—including register-based datasets—is that households report their net investment into assets.[1] The dataset reveals significant heterogeneity in returns, with wealthy households earning higher returns. This is a well-established feature in the literature referred to as *scale dependence*. I also use the dataset to establish a novel empirical fact: The correlation between the average return and individual returns is increasing in the level of wealth. This means that in periods when returns are higher than normal, returns for wealthy individuals are even higher, while returns for poor individuals tend to stay the same. In addition to the new dataset of returns in the PSID panel, I also study return heterogeneity in the Survey of Consumer Finances (SCF). Here, I confirm the results from the PSID: Returns are heterogeneous and rich households earn higher returns.

I add these returns to the model in a parsimonious way: Returns follow a Markov chain. Thus, solving the model is only marginally slower than standard models and much simpler than bigger models like two-asset models. In addition to this, I model that the pass-through of aggregate to individual returns is heterogeneous and increasing in wealth, consistent with the empirical evidence.

I find that adding a realistic degree of heterogeneity in returns to a HANK model makes the model match empirical distributions of returns, wealth, and income. First, adding heterogeneous returns makes the model replicate the dispersion in returns across households in the US data. In addition to this, the model replicates the scale dependence in returns, i.e. the well-established fact that returns are increasing in the level of wealth. By definition, the standard HANK model cannot replicate this fact because all households earn the same return on their wealth. Scale dependence is not an assumption but occurs endogenously in the model with heterogeneous returns: Households who earn high returns *choose* to save more because they get a higher return on their wealth, creating a positive relationship between wealth and returns.

Since returns are heterogeneous, it opens a new question relative to the model with common returns: How is the pass-through of aggregate to individual returns? I show that the model with heterogeneous returns replicates the correlation of aggregate to individual returns in the data. In particular, the model replicates the fact that when aggregate returns increase, returns for rich households tend to increase more than for poor households. This is not possible in standard HANK models because returns are common.

In addition to matching heterogeneous returns, adding heterogeneous returns

1. Without this, one risks biasing the estimates of returns.

makes the model replicate the concentration of wealth at the very top of the wealth distribution in the US. This is a well-established feature that remains elusive to standard HANK models, which cannot replicate the concentration of wealth. In particular, these models typically find that the richest households hold only less than 30 million USD in wealth, while empirically the richest households hold around 4 orders of magnitude more wealth, exceeding 100 billion USD. This is because the empirical wealth distribution has a fat right tail, while the wealth distribution in standard models is bounded. The model with heterogeneous returns replicates the wealth concentration, exhibiting a fat right tail. This is despite the model featuring no permanent heterogeneity, no preference heterogeneity, and only a single asset.

How does introducing return heterogeneity increase the concentration of wealth at the top? This is explained by scale dependence: Households who earn high returns both save more and earn higher returns on their wealth, quickly accumulating significant wealth.

Matching the wealth concentration also makes the model replicate a high average marginal propensity to consume (MPC), consistent with empirical evidence. The MPC has been emphasized in the literature as a key moment—or even *the* key moment—in the study of business cycles and the propagation of shocks to aggregate demand.

The key insight is that the model with heterogeneous returns can replicate the MPC *while* simultaneously replicating a high level of aggregate wealth. This is something simple HANK models are unable to do. Instead, these models face an MPC-wealth trade-off (Kaplan and Violante 2022): Replicate a realistically high MPC at the cost of wealth an order of magnitude less than in the data, or replicate the wealth in the data at the cost of an MPC well outside the range of high-quality empirical estimates. In the model with heterogeneous returns, one does not have to choose: It is possible to replicate both a realistically high MPC and a realistically high level of wealth. This is the case because of heterogeneity in returns: Some households earn high returns and are very wealthy, increasing the average wealth. Other households earn low returns and are close to the borrowing constraint, increasing the average MPC.

Having established the microeconomic fit of the model, I now consider if this matters for the transmission of monetary policy. To do so, I compare the aggregate outcomes in the model with heterogeneous returns to a standard HANK model. I find that the effect of monetary policy on consumption is dampened by around 15% in the model with heterogeneous returns. To understand why, I decompose the response of consumption in both models. I find that the model with heterogeneous returns features a new re-distribution channel not present in the standard model, which drags down consumption.

To understand this re-distribution, I turn to the distributional effects of monetary policy. In particular, I ask the following question: For each $100 generated by monetary policy, how much goes to the top x%? To do this, I take a sufficient statistics approach. In particular, I derive sufficient statistics for determining which households gain from expansionary monetary policy.

I find that 4 moments shape the distribution of income in response to monetary policy. These are sufficient statistics in the sense that any model that matches these will yield the same distribution of income in response to monetary policy. The moments are: (1) The wealth distribution, (2) the income distribution, (3) the pass-through of aggregate returns to individual returns, and (4) the pass-through of aggregate earnings to individual earnings. Standard models provide a poor fit to all 4 moments. The model with heterogeneous returns in this paper fits all 4 moments.

The fact that the standard HANK model does not match the sufficient statistics means that it misses how much expansionary monetary policy benefits the ultra-rich. In particular, I find that the richest 0.1% gain 11% of the total increase in income when monetary policy is eased in the model with heterogeneous returns. In the standard HANK model, the top 0.1% gain less than 2% from monetary policy.

Why do the rich gain so much when returns are heterogeneous? This is due to the interaction between the concentration of wealth and capital income among the rich and the increase in capital income induced by expansionary monetary policy. Expansionary monetary policy increases capital income for two reasons: Directly through lower discount rates on firm profits and indirectly through higher profits, which are pro-cyclical in the model.

Due to the replication of the microdata, the increase in capital income almost entirely goes to the top of the wealth distribution in the model with heterogeneous returns. In the standard HANK model, both top wealth and capital income at the top are significantly understated. Thus, the capital income gains at the top are missing or severely understated.

This explains why monetary policy is less effective in the model with heterogeneous returns: Monetary policy re-distributes to rich households and rich households have lower marginal propensities to consume (MPCs), so consumption increases less.

If the gains of monetary policy are so unevenly distributed, should a policymaker use fiscal policy instead? In particular, consider a policymaker who wants to stabilize the business cycle in response to aggregate shocks. Which tool should the policymaker use? While both monetary and fiscal policy can stabilize *aggregate* outcomes, I find that they have distinctly different effects across the distribution. In particular, monetary policy achieves its stimulus mainly by benefiting the very top of the wealth distribution. In contrast to this, the gains from fiscal policy are much more evenly spread out. In particular, I find that the richest 0.1% gain 2.4% and 4.6% of the income increase from transfers and government consumption in the model with heterogeneous returns. This sharp distinction between the winners and losers from monetary and fiscal policy is only clear in the model with heterogeneous returns due to its replication of the microeconomic data.

Ultimately, this highlights the role of taking seriously the fit to microeconomic data in macroeconomic models. A policymaker who cares not only about aggregate stabilization but also who gains from different policies should consider these effects.

1.1 Related Literature

The main contribution of my paper compared to the literature is to study the interaction of monetary policy and inequality in a model with a realistic degree of inequality. My second contribution is to show that adding empirically realistic heterogeneous returns to a HANK model is exactly what lets the model match the inequality in the data.

In this way, my paper contributes to three strands of the literature. First, my paper contributes to the literature on HANK models and in particular to the literature on quantitative HANK models. Key contributions in this literature includes Kaplan et al. (2018), Ferra et al. (2020), Auclert et al. (2024a), Bayer et al. (2024). My key contribution here is to add heterogeneous returns to the model, which allows me to take the distributional effects of shocks more seriously due to replicating the microeconomic data.

In this literature, my paper is particularly close to the seminal contribution of Kaplan et al. (2018), who also study the effects of monetary policy in a HANK model which takes the microeconomic data seriously. Compared to their paper, I have a model with only a single asset instead of two assets. Instead, I have (a continuum of) heterogeneous returns. This addition does not introduce a new choice variable, so solving the household problem is not difficult. Despite this, I still replicate a realistically high MPC and a realistically high level of wealth. Furthermore, I also replicate the concentration of wealth at the top of the wealth distribution.

Second, my paper contributes to the literature on the distributional effects of policy shocks and in particular monetary policy. These papers can be split into empirical and theoretical. Key contributions are in particular Coibion et al. (2017), Holm et al. (2021), Andersen et al. (2023), and McKay and Wolf (2023a). My contribution to the theoretical literature is to replicate the empirical distributions, which is key to understanding the *distributional* effects of shocks. In particular, I replicate the concentration of wealth at the very top. My contribution compared to the empirical literature is to emphasize the importance of the very top of the wealth distribution, say the top 0.1%. Empirical papers often distribute households into coarse buckets of 20% or 10%, missing the very top. Furthermore, empirical papers often use censored data, completely missing the top of the wealth distribution. This highlights my approach of using a model to study the distributional effects, as this approach does not face these deficiencies.

Third, my paper contributes to the growing literature on heterogeneous returns and their implications in economic models. This literature includes Benhabib et al. (2011), Benhabib et al. (2015), Gabaix et al. (2016), Benhabib et al. (2017), Jones and Kim (2018), Xavier (2021), Guvenen et al. (2023), and Daminato and Pistaferri (2024). This literature generally emphasizes the importance of heterogeneous returns in shaping the distributions of wealth and income and their dynamics, as does my paper. The contribution of my paper is to embed this in a HANK model and to study the implications for business cycle dynamics.

1.2 Structure

I start my paper by constructing a dataset of heterogeneous returns in Section 2. I then add heterogeneous returns to a HANK model in Section 3. I calibrate my model to the data on heterogeneous returns and other moments in Section 4. In Section 5, I show how adding heterogeneous returns to the model makes it replicate the microeconomic data. I then show how this shapes the effects of macroeconomic shocks in Section 6. Finally, I conclude in Section 7.

2 Empirics

In this Section, I present two key aspects of the microeconomic data that standard HANK models miss: The concentration of wealth at the top and the heterogeneity in returns. The goal of doing so is to build a HANK model that replicates these aspects and then study the distributional aspects of monetary policy in this model.

2.1 Concentration of Wealth

I start by discussing the concentration of wealth at the top of the wealth distribution. It is a well-established fact that wealth is very concentrated among the very richest households. Table 1 demonstrates this, presenting top wealth shares in the data. The figure shows that the majority of wealth is held by just 5% of all households. Additionally, the top 1/1000th of households hold 14% of wealth.

Existing HANK models are unable to match this concentration of wealth. To see this, Table 1 reports top wealth shares in selected HANK models. I have selected these HANK models because of their prominence in the literature and their emphasis on matching the distribution of wealth. As such, the models in Table 1 provide the best case of matching the wealth concentration in standard HANK models, with most HANK models featuring *less* wealth concentration.

Table 1 shows that all four models understate the concentration of wealth at the top. In particular, Auclert et al. (2020) (ARS) and Bayer et al. (2024) (BBL) report only top 10% and 5% shares, understating these compared to the data. Their fit to the top 1% or 0.1% shares is then almost surely worse. McKay and Wolf (2023b) (MW) fit the top 5% share, but understate the top 1% share and do not report the top 0.1% share. Kaplan et al. (2018) (KMV) does the best, fitting the top 1% share. They are also the only ones to report the top 0.1% share, but they underestimate this by half—something they explicitly mention in their paper.

	Data	KMV	ARS	MW	BBL
Top 10% share	76%	82%	70%	82%	67%
Top 5% share	65%	69%	58%	66%	—
Top 1% share	37%	38%	—	27%	—
Top 0.1% share	14%	7%	—	—	—
Pareto tail index	1.52	—	—	—	—

Table 1: Wealth Concentration in the Data and HANK Models

Note: The table shows wealth shares and the Pareto tail index of wealth in the US data and selected models. The data is the 2019 SCF, except the Pareto tail index, which is from Vermeulen (2018). "KMV" is Kaplan et al. (2018), "ARS" is Auclert et al. (2020), "MW" is McKay and Wolf (2023b), and BBL is Bayer et al. (2024). The numbers for ARS are for the illiquid wealth distribution.

Furthermore, data shows that the wealth distribution is "fat-tailed". In particular, it has been shown that the right tail of the wealth distribution is well approximated by a Pareto distribution, which has a fat right tail. Mathematically, I note that if a variable X is Pareto distributed, it holds that

$$\log P(X > x) \sim -\alpha \log x, \tag{1}$$

where α is the Pareto tail index and $\log P(X > x)$ is the log counter-CDF (CCDF). The tail index then controls how "fat" the tail is, i.e. how concentrated wealth is at the top. In particular, a lower tail index, α, corresponds to *more* concentrated wealth.

Table 1 reports an estimate of the Pareto tail index in the US, which is 1.52. This implies a significant concentration of wealth at the top. For instance, this means that the variance of wealth is undefined. The models in Table 1 do not report their Pareto tail index. This is natural, as the models do not feature a Pareto tail and thus the Pareto tail index is not defined in the models.

2.2 Return Heterogeneity

Having studied the concentration of wealth at the top, I now turn to the degree of return heterogeneity. This is in contrast with standard HANK models, which feature a common return for all households. To study heterogeneous returns, I take two different approaches using two different datasets. First, I study return heterogeneity in the Survey of Consumer Finances (SCF) by comparing the distributions of capital income and wealth. Second, I directly construct a dataset of heterogeneous returns across US households using the Panel Survey of Income Dynamics (PSID). The two approaches complement each other. The advantage of the first approach is that it is straight-forward using readily available data and is less prone to measurement error. The advantage of the second approach is that it gets directly at the heterogeneous returns, which allows me to study additional aspects.

2.2.1 Capital Income in the SCF

I start by studying the degree of return heterogeneity in the cross-section using the SCF for 2019. The details of the data are provided in Appendix A.1. This approach requires neither measuring returns directly nor does it require a panel. The approach starts by computing shares of wealth and capital income. To do so, note that capital income is defined by

$$x_{i,t} = r^a_{i,t} a_{i,t-1}.$$

If returns are common across households, $r^a_{i,t} = r^a_t$, capital income can be written as $x_{i,t} = r^a_t a_{i,t-1}$. Consider now looking at the bottom p% of households in the wealth distribution. Denote these households by $i \in \mathcal{P}$. How large a share of total wealth and capital income is held by these households? In the case with common returns, the bottom shares of wealth and capital income are given by

$$S(a) = \frac{\sum_{i \in \mathcal{P}} a_i}{\sum a_i},$$
$$S(x) = \frac{\sum_{i \in \mathcal{P}} x_i}{\sum x_i} = \frac{r \sum_{i \in \mathcal{P}} a_i}{r \sum a_i} = S(a).$$

Intuitively, if returns are common, the share of wealth and capital income held by the bottom x% is the same. If returns are heterogeneous, they can be different. This allows me to test if returns are heterogeneous. Figure 1 plots this. In particular, the figure shows the share of wealth held by the bottom p% in the wealth distribution and their share of capital income. The figure clearly shows the shares of capital income as a function of the shares of wealth lying below the 45-degree line, i.e. $S(x) < S(a)$. This implies that returns are heterogeneous. Not only this, it implies that wealthier households earn higher *rates* of return on average.

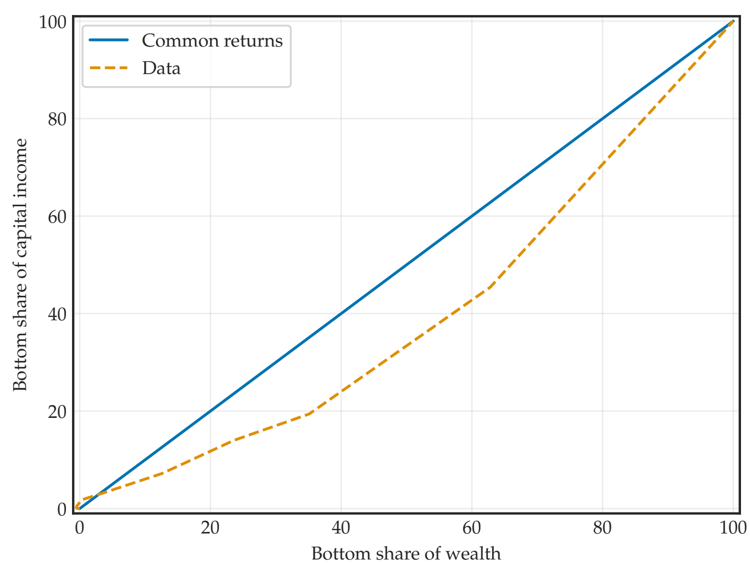

Figure 1: Shares of Wealth and Capital Income

Note: The figure shows the shares of wealth and capital income at different points in the distribution of households in the 2019 SCF. Households are sorted by their level of wealth. The x-axis shows the share of wealth held by the bottom x% of households, while the y-axis shows the share of capital income held by the same group.

Figure 1 also makes an additional point: Assuming common returns across households will understand the income of the rich as they earn higher returns. Thus, assuming common returns is also likely to understate wealth inequality. Gaillard et al. (2023) makes a very similar point theoretically, arguing that scale-dependent returns are necessary to match the tops of the distributions of wealth and income. Guvenen et al. (2023) finds a similar result. Thus, I conclude that matching the degree of return heterogeneity—and in particular that wealthy households earn higher returns—is crucial for matching the degree of wealth inequality, particularly at the top.

2.2.2 Heterogeneous Returns in the PSID

Having looked at the cross-sectional data in the SCF, I now take a different approach: Constructing data on heterogeneous returns directly using panel data. In particular, I use the PSID conducted from 1999 to 2019.[2] The panel is bi-annual and the unit of measurement is households. I present the data in detail in Appendix A.2.[3]

The main outcome is the return on wealth for household i at time t, which is

$$r_{i,t}^{a} = \frac{y_{i,t}}{a_{i,t-1}}, \tag{2}$$

where $a_{i,t-1}$ is wealth and $y_{i,t}$ is the income generated from this wealth, both realized and unrealized. Crucially, the return is measured *net of investment*. This is often not possible in other studies because net investment is unknown. For instance, Fagereng et al. (2020) use Norwegian register data, which does not include net investment. Without data on net investment, the measure of return is potentially biased. Consider for instance a household that buys a large amount of stocks in between measurements of wealth. The increase in wealth is then falsely counted as capital gains, biasing upward the measure of returns incorrectly. Fagereng et al. (2020) employ approximation methods to minimize the bias from this source. In the PSID, households are asked directly about net investments for many asset categories, so I avoid this potential source of bias.

Capital income, $y_{i,t}$, can be split into 8 sources: Trust fund and royalties, interest, dividends, primary housing, other housing, businesses, stocks, and other. I discuss how I measure these in Appendix A.2.3. The total wealth, $a_{i,t}$, can be split into 8

---

2. The resulting data for is every other year from 2000 to 2018, see Appendix A.2.1.

3. I am thankful to Stephen Snudden for making his replication files for Snudden (2021) publicly available. My return measurements take a starting point in his work but are different for numerous reasons, so any errors are purely mine.

asset categories: Primary housing, other housing, business, stocks, private annuities or IRAs, checking/savings accounts, vehicles, and other assets. I discuss these in Appendix A.2.4.

I annualize the returns from the bi-annual panel. Thus, the resulting measure of returns is *pre-tax annual returns*. For wealth, I normalize by the average level of wealth in the year. I report descriptive statistics of the dataset in Table 2. I show the time series of average returns in Appendix A.2.5.

	Mean	Std. dev.	P5	P50	P95
Return	3.5%	13.5%	−10.4%	0.0%	24.8%
Assets (USD)	317	710	1	154	1114
— Primary house	154	202	0	110	488
— Other house	25	179	0	0	125
— Business	34	365	0	0	25
— Stocks	25	190	0	0	100
— IRAs	37	155	0	0	200
— Savings	20	89	0	4	80
— Vehicles	16	20	0	10	50
— Other	10	98	0	0	25
Age (years)	44	10	29	43	61

Table 2: Descriptive Statistics for Heterogeneous Returns and Wealth in the PSID

Note: The values of wealth are in thousands of nominal USD. The age is in years.

Having constructed a dataset of heterogeneous returns, I start by plotting a histogram of these in Figure 2. As the figure shows, there is significant dispersion in returns. In particular, the standard deviation of returns is 14 pp. This is perfectly consistent with the literature, which finds standard deviations in the range of 7–31 pp. across settings, approaches, and datasets (Bach et al. 2020, Fagereng et al. 2020, Smith et al. 2022, and Snudden 2021). One might ask if the heterogeneity in returns is explained by the types of assets held by households. I show that this is not the case in Appendix A.2.6, implying that there is significant heterogeneity in returns within asset categories.

One thing is the cross-sectional dispersion in returns. Another is how returns

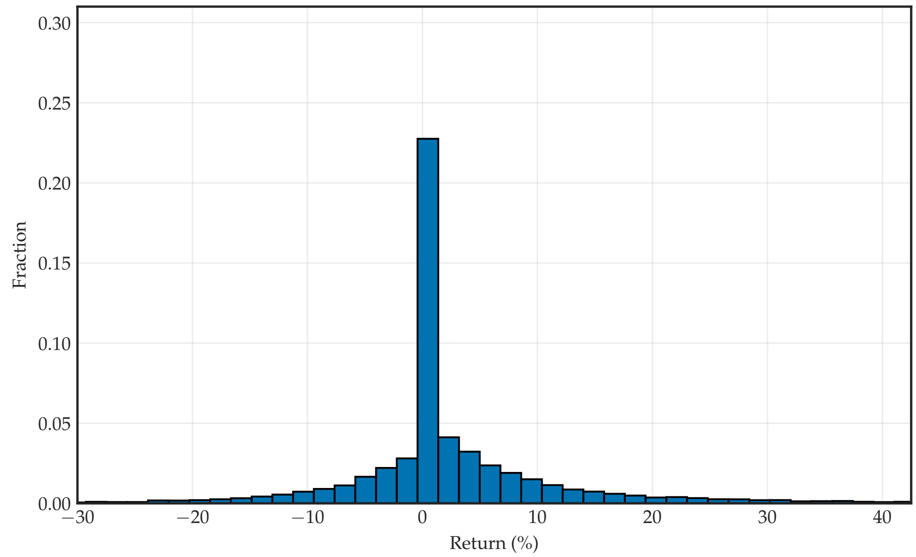

Figure 2: Distribution of Returns

Note: The figure shows a histogram of the returns of US households in the dataset constructed from the PSID.

change in response to aggregate shocks. To get at this, I estimate the following regression:

$$r^a_{i,t} = \alpha^{(q)} + \beta^{(q)}\bar{r}^a_t + \varepsilon^{(q)}_{i,t}, \tag{3}$$

where $\bar{r}^a_t$ is the average returns across households in year t. This is very similar to the idea of measuring the "β" of earnings, i.e. how much individual earnings change as aggregate earnings or GDP changes—see for instance Guvenen et al. (2017). However, to the best of my knowledge, I am the first to do such an exercise for returns.

Figure 3 plots the estimated β's by quintiles of the wealth distribution. The figure clearly shows that the pass-through of average to individual returns is stronger for richer households. In particular, the pass-through is non-existent for the poorest 20% households, while it is greater than 1-for-1 for the richest 40%.

3 A Model of Heterogeneous Returns

Having established that returns are heterogeneous using two different approaches with two different datasets, I now introduce heterogeneous returns in an otherwise standard HANK model. The goal of doing so is both to match the degree of heterogeneity in returns, but also the fit the concentration of wealth at the top.

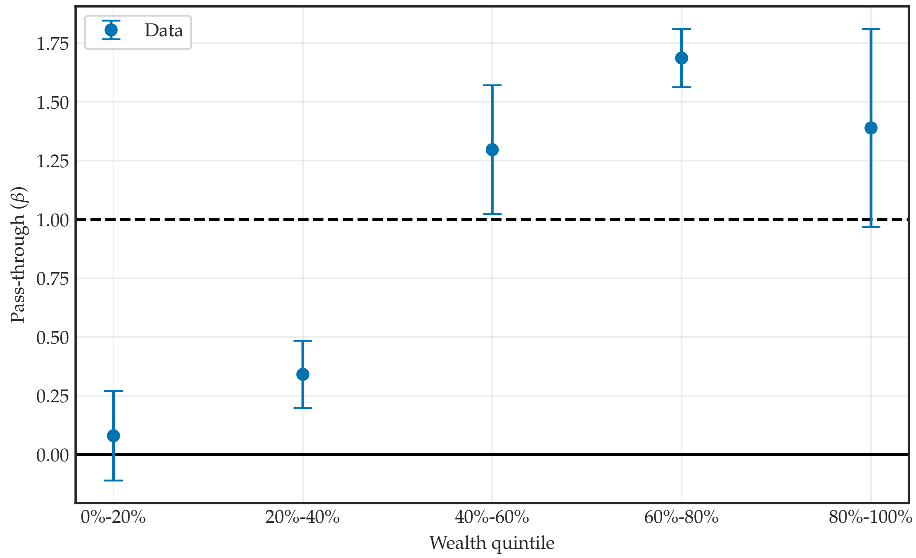

Figure 3: Pass-Through of Average to Individual Returns by Wealth Quintiles

Note: The figure shows estimates of β from eq. (3) by quintiles of wealth, $a_{i,t-1}$.

The model economy consists of households, firms, and a public sector. Households consume and save in an asset that pays back heterogeneous returns. Labor supply is set on behalf of households by a union subject to adjustment costs. Firms use labor to produce goods under monopolistic competition and flexible prices. The government issues bonds, raises taxes, pays transfers, and consumes. The central bank sets the real interest on bonds.

The key innovations compared to the literature have to do with incorporating heterogeneous returns in the New Keynesian framework. This amounts to changing both the asset demand and supply sides. Let me describe each component in more detail.

3.1 Households

Time is discrete and the horizon is infinite: $t = 0, 1, \ldots$ There is a continuum of households indexed by $i \in [0, 1]$. Each household chooses the sequence of consumption, $(c_{i,t})_{t=0}^{\infty}$, and wealth, $(a_{i,t})_{t=0}^{\infty}$, to maximize

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[u(c_{i,t}) - v(n_{i,t}) \right],$$

where u and v are the instantaneous utility of consumption and the disutility of labor supply, respectively. $\beta \in [0,1]$ is the common discount factor for all households. Labor supply is identical for all households and is not chosen directly by the households, $n_{i,t} = N_t$. Instead, it is chosen by the union as I describe later. Thus, the disutility of labor does not affect household behavior. I consider the standard CRRA utility function:

$$u(c) = \frac{c^{1-\sigma}}{1-\sigma},$$

where $\sigma > 0$ is the inverse elasticity of intertemporal substitution. Each household is subject to a budget constraint in every period:

$$c_{i,t} + a_{i,t} = (1 + r^a_{i,t})a_{i,t-1} + z_{i,t} + T_t - t_{i,t}. \tag{4}$$

Here, $z_{i,t}$ is the real pre-tax labor earnings of households. It depends on an idiosyncratic component and an aggregate component,

$$z_{i,t} = e^z_{i,t} Z_{ss} + e^z_{i,t}\beta^z_{i,t}(Z_t - Z_{ss}).$$

where $\int e^z_{i,t}\, di = 1$ and $\int e^z_{i,t}\beta^z_{i,t}\, di = 1$ such that Z_t is the average earnings. The idiosyncratic component of earnings follows a Markov chain:

$$e^z_{i,t} \sim \text{Markov}(\mathcal{S}_z, \mathcal{P}_z)$$

where $\mathcal{S}_z$ is the state space and $\mathcal{P}_z$ is the transition matrix. $\beta^z_{i,t}$ measures the elasticity of individual earnings to aggregate earnings, i.e. the "worker-β" in the style of Guvenen et al. (2017) c.f. Appendix B.1. It is calibrated to empirical evidence in Section 4. $\beta^z_{i,t} = 1$ nests the standard specification of $z_{i,t} = e^z_{i,t} Z_t$.[4]

The return on wealth, $r^a_{i,t}$, consists of an idiosyncratic component and an aggregate component,

$$r^a_{i,t} = r^a_{ss} + e^r_{i,t} + \beta^r_{i,t}(r^a_t - r^a_{ss}), \tag{5}$$

where $\int e^r_{i,t}\, di = 0$ and $\int \beta^r_{i,t}\, di = 1$ such that r^a_t is the average return. The idiosyncratic

4. In this case, the pass-through is common: $\partial \log z_{i,t} / \partial \log Z_t = 1$.

component follows a Markov chain

$$e^r_{i,t} \sim \text{Markov}(\mathcal{S}_r, \mathcal{P}_r)$$

where $\mathcal{S}_r$ is the state space and $\mathcal{P}_r$ is the transition matrix. Thus, the return on wealth is heterogeneous across households due to the randomness in $e^r_{i,t}$. Furthermore, returns may be persistent: If households earn a high return in one period, they tend to also earn a high return in the next period. $\beta^r_{i,t}$ controls the pass-through of aggregate to individual returns. I specify the functional form of this in the calibration. A relevant special case is equal pass-through, i.e. $\beta^r_{i,t} = 1$ for all i and t.

The last two components of the budget constraint reflect flows to and from the government: T_t is lump-sum transfers, while $t_{i,t}$ is the tax bill of the household. Households are taxed on both capital and labor income at rate τ:

$$t_{i,t} = \tau_t \left(r^a_{i,t} a_{i,t-1} + z_{i,t} \right).$$

Finally, households are subject to a borrowing constraint:

$$a_{i,t} \geq \underline{a}.$$

In this paper, I will focus on the case with heterogeneous returns, i.e. $e^r_{i,t} \neq 0$. Let me briefly mention a special case that I will compare the model to. This is the standard HANK model. This is nested when returns are common, $e^r_{i,t} = 0$ and $\beta^r_{i,t} = 0$, and the pass-through to earnings is $\beta^z_{i,t} = 1$ for everyone.

3.2 The Rest of the Model

Having presented the household side with heterogeneous returns, let me now present the rest of the model. For the most part, this is the standard New Keynesian model. The key difference is where the heterogeneous returns come from.

3.2.1 Firms

Firms produce output Y_t using only labor N_t with constant returns to scale technology, $Y_t = N_t$. They sell this output to households at price P_t and pay households a wage rate W_t for their labor, such that real labor income is $Z_t = w_t N_t$, where $w_t = W_t / P_t$ is the real wage rate.

Firms set the price, P_t, at a markup, $\mu \geq 1$, over marginal costs:

$$P_t = \mu W_t. \tag{6}$$

This means that prices are flexible. Instead, the nominal rigidity is in the form of sticky wages, as is standard in the HANK literature. I consider sticky prices in 6.3. All profits are paid period-by-period to households as dividends, which in real terms are

$$D_t = Y_t - Z_t,$$

Firms issue a unit mass of shares, which they sell at real price p_t.

3.2.2 Government

The government issues real bonds, B_t, which pay real interest rate r_t. It uses the bonds and taxes, $\mathcal{T}_t$, to finance government consumption, G_t, and lump-sum transfers to households, T_t, such that the real budget constraint is

$$B_t = (1 + r_{t-1})B_{t-1} + G_t + T_t - \mathcal{T}_t. \tag{7}$$

where tax receipts are $\mathcal{T}_t = \int t_{i,t}\, di$.

Shocks to government consumption or transfers are financed by both issuance of bonds and higher taxes in the short run. In the long run, fiscal policy is passive, in the sense that the tax rate, τ_t, adjusts to ensure that

$$B_t = B_{ss} + \phi_B(B_{t-1} - B_{ss}) + (G_t - G_{ss}) + (T_t - T_{ss}) \tag{8}$$

following Auclert et al. (2024a). The rule is chosen to ensure that bonds return to the initial steady state in the long run, i.e. $\lim_{t\to\infty} B_t = B_{ss}$, where "ss" indicates the steady state of the model.

3.2.3 Central Bank

The central bank sets the real interest rate directly:

$$r_t = r_{ss} + \varepsilon_t, \tag{9}$$

where ε_t is a monetary policy shock.[5] The Fisher equation defines the nominal interest rate,

$$i_t = (1 + r_t)(1 + \pi_{t+1}) - 1,$$

where $\pi_t = P_t / P_{t-1} - 1$ is inflation.

3.2.4 Union

In addition to sticky prices, there are also nominal rigidities in the model in the form of sticky wages. Specifically, a union sets nominal wages subject to Rotemberg adjustment costs. This yields the following non-linear New Keynesian wage Phillips curve (NKWPC) as in Auclert et al. (2024b):

$$\pi_t^W (1 + \pi_t^W) = \kappa^W \left(\frac{v'(N_t)}{u'(C_t)(1 - \tau_t) w_t} - 1 \right) + \frac{1}{1 + r_t} \pi_{t+1}^W (1 + \pi_{t+1}^W), \tag{10}$$

This NKWPC is written in such a way that heterogeneity does not matter directly as is common in the HANK literature.[6] $v(N_t) = \gamma N_t^{1+1/\phi}$ is the disutility of labor with $\phi > 0$ measuring the Frisch elasticity of labor supply and γ being a scalar parameter.

3.2.5 Asset Supply

Households do not have a portfolio choice but instead choose overall savings, $a_{i,t}$, with return $r_{i,t}^a$. The savings reflect two assets in the economy: Firm equity and government bonds. These two assets pay the same return along the perfect foresight transition path.[7] These returns are paid out to households each period according to the heterogeneous returns process in the household problem. This means that the total capital income coming from heterogeneous returns for all households equals the

5. This is equivalent to a Taylor rule, say $i_t = i_{ss} + \phi_\pi \pi_t + \varepsilon_t^{\text{Taylor}}$, and then backing out the monetary policy shock, $\varepsilon_t^{\text{Taylor}}$, to obtain the same path of the real interest rate.

6. This simplifies the comparison of different HANK models as only the household side is affected while the NKWPC is unchanged.

7. The only exception is on impact, $t = 0$. Here, the shock causes an unexpected revaluation of assets which means that returns can differ.

capital income in the economy:[8]

$$\underbrace{\int r^a_{i,t} a_{i,t-1}\, di}_{\text{HH's capital income}} = \underbrace{p_t + D_t + (1 + r_{t-1})B_{t-1} - A_{t-1}}_{\text{Capital income from bonds and equity}}. \tag{11}$$

Additionally, there is a no-arbitrage condition between government bonds and firm equity, such that the ex-ante expected returns are equalized:

$$\frac{p_{t+1} + D_{t+1}}{p_t} - 1 = r_t, \tag{12}$$

where r_t is the ex-ante short-term real interest rate.

It becomes convenient to define the *wealth-weighted* average return,

$$\tilde{r}^a_t = \frac{\int r^a_{i,t} a_{i,t-1}\, di}{A_{t-1}},$$

where $A_t = \int a_{i,t-1}\, di$. Note that the wealth-weighted average return generally differs from the average return since

$$\tilde{r}^a_t = r^a_t + \underbrace{\text{Cov}\left(r^a_{i,t}, \frac{a_{i,t-1}}{A_{t-1}}\right)}_{\text{Scale dependence}}.$$

The covariance measures the degree of scale dependence in returns: The degree to which wealthy households earn high returns on their wealth.[9] Intuitively, *if* wealthy households earn higher returns, i.e. $\text{Cov}\left(r^a_{i,t}, \frac{a_{i,t-1}}{A_{t-1}}\right) > 0$, the wealth-weighted average return is higher. In the case with common returns, the wealth-weighted average return and average return are identical, $\tilde{r}^a_t = r^a_t$, since $\text{Cov}\left(r^a_{i,t}, \frac{a_{i,t-1}}{A_{t-1}}\right) = 0$ as $r^a_{i,t} = r^a_t$.

8. By capital income, I mean any income generated by holding assets. This could also be called asset income.

9. Other papers use "scale dependence" to mean that wealthier households earn higher returns *because* they are wealthier, see for instance Gaillard et al. (2023). I use the term scale dependence more generally to refer to a positive covariance between wealth and returns. This might occur directly *because* the returns process is increasing in wealth—as I consider in Appendix B.2—or it might occur endogenously, as it does in my baseline model.

3.2.6 Market Clearing

Final output is consumed by households and the government,

$$Y_t = C_t + G_t, \tag{13}$$

where private consumption is aggregated over households, $C_t = \int c_{i,t}\, di$. Asset market clearing is then

$$A_t = \int a_{i,t}\, di = p_t + B_t, \tag{14}$$

which follows from goods market clearing due to Walras' law c.f. Appendix B.3.

3.3 Equilibrium

In the following, I define a competitive equilibrium given a monetary policy or fiscal policy shock.

Definition 1 (Equilibrium)**.** *Given sequences for the shocks, $\{(\varepsilon_s)_{s=0}^{\infty}, (G_s)_{s=0}^{\infty}, (T_s)_{s=0}^{\infty}\}$, an initial household distribution over wealth, earnings, and returns, a competitive equilibrium is a path of household policies, distributions, prices, and quantities, such that (1) all households solve their dynamic programming problem, (2) firm behavior satisfies the equation in Section 3.2.1, (3) the central bank follows the monetary policy rule in eq.* (9)*, (4) the government satisfies the tax rule in eq.* (8) *and the budget constraint in eq.* (7)*, (5) the union sets labor supply according to the NKWPC in equation* (10)*, (6) capital income satisfies eq.* (11) *and enforces the no-arbitrage condition in eq.* (12)*, and (7) the goods market in eq.* (13) *clears.*

4 Calibration

I now present the calibration of the model. I first discuss the calibration of the new part: The returns process. To do so, I use the panel of heterogeneous returns discussed in Section 2. I discuss this in some detail. I then proceed with the calibration of the rest of the model, which is quite standard, so I give less detail.

4.1 The Returns Process

I start by discussing the calibration of the returns process, which is grounded by my data on household-level returns in the US. There are three key objects to be calibrated:

The average return, r^a_{ss}, the state space, $\mathcal{S}_r$, and the transition matrix, $\mathcal{P}_r$. Let me go through each.

The average return, r^a_{ss}, is straightforward to calibrate. I simply calibrate it to the average return in my data. Note here that the return in the data is computed conditional on $a_{i,t-1} > 0$. Otherwise, the numerator in the formula for returns is zero. Thus, I set r^a_{ss} such that the average return in the model conditional on positive wealth is the same as in the data.

Next, let me discuss the calibration of the idiosyncratic component of returns, $e^r_{i,t}$. This process is controlled by the state space, $\mathcal{S}_r$, and the transition matrix, $\mathcal{P}_r$. I start by constructing the equivalent of the idiosyncratic component of returns in the data. In particular, I care about the component of returns which is not permanent to the individual and which is not driven by aggregate movements in returns. For this reason, I estimate a regression of returns on an individual fixed effect, time fixed effect, and age dummies:

$$r^a_{i,t} = \gamma_i + \gamma_t + \beta D_{i,t} + \varepsilon_{i,t},$$

where $D_{i,t}$ is a vector of age dummies. The residual, $\varepsilon_{i,t}$, then measures the part of returns not explained by permanent heterogeneity—which is not present in the model—and aggregate returns—which comes from the average return component.

I divide the data on residual returns into 7 bins and compute the median return within each bin. This yields the following state space:

$$\mathcal{S}_r = \{-0.269, -0.116, -0.038, -0.001, 0.027, 0.11, 0.371\}$$

In addition, I get the stationary distribution of the process as the empirical fraction of observations in each bin:

$$\pi = \{0.027, 0.098, 0.226, 0.297, 0.226, 0.098, 0.027\}.$$

Next, I turn to the transition matrix. Denote the probability of staying in state i by p_i. I then make the following assumption: Households either stay in their current state, increase one state, or drop one state. Additionally, the probability of changing state is the same no matter if going up or down. Taking into account that transition probabilities have to sum to 1 in each state, this implies that there are 7 free parameters in the transition matrix. After matching the stationary distribution, π, this leaves one degree of freedom. I use this degree of freedom to match the number of billionaires in the US. I discuss the returns process in much more detail in Appendix B.4.

Finally, I choose the functional form of $\beta^r_{i,t}$ to match the pass-through of average to individual returns. In particular, I set[10]

$$\beta^r_{i,t} = \theta_0 + \frac{\theta_1}{1 + \exp\{-\theta_2 a_{i,t-1} + \theta_3\}}.$$

I estimate $\theta = (\theta_0, \theta_1, \theta_2, \theta_3)'$ by constructing β^g_{model} in the same way as in the data and then solving

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{G} \sum_{g=1}^{G} |\beta^g_{\text{data}} - \beta^g_{\text{model}}|.$$

I find that $\hat{\theta} = (-0.08, 1.98, 3.43, 2.54)'$ solves this, with the resulting fit shown in Figure 4.[11]

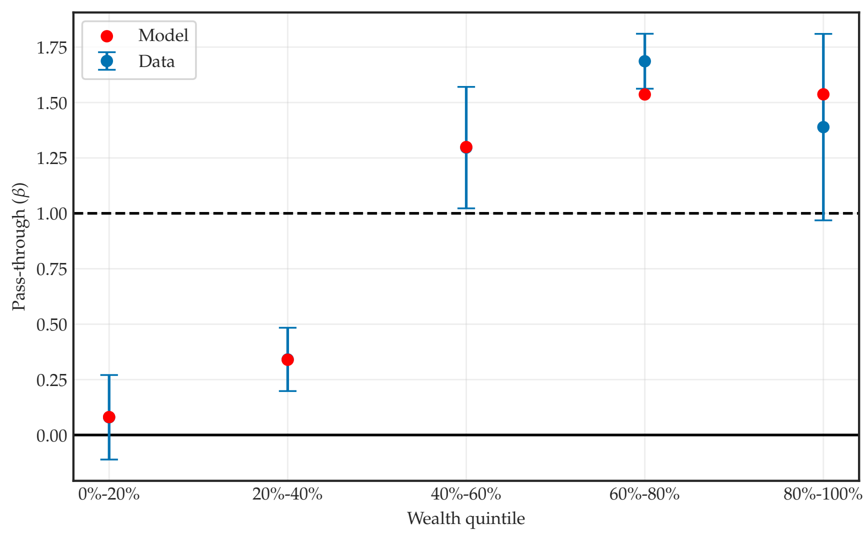

Figure 4: Pass-Through of Aggregate to Individual Returns, $\beta^r_{i,t}$

Note: The x-axis shows quintiles of the lagged asset distribution, while the y-axis shows $\beta^r_{i,t}$ for households at these points in the wealth distribution.

10. I normalize this such that $\mathbb{E}[\beta^r_{i,ss}] = 1$. Away from the steady state, it is possible that $\mathbb{E}[\beta^r_{i,t}]$ deviates slightly from 1 such that $\int_i r^a_{i,t}\, di \neq r^a_t$.

11. Whenever comparing the model to the returns data, I condition on positive wealth. If I did not, this could lead to significant bias. This is because the data on returns is also conditional on positive wealth—otherwise one cannot compute returns using eq. (2). This reflects that returns are well-defined in the model even when $a_{i,t-1} = 0$, while they are unobserved in the data in this case.

4.2 The Rest of the Model

I calibrate the model to match the US economy in 2019. The calibration is annual. For the household side, I mainly use the 2019 edition of the SCF.[12]

I start by presenting the internally calibrated parameters other than those in the returns process. These are parameters set to match moments in the data. I calibrate the discount factor to match the asset demand-to-GDP ratio of $\int a_{i,ss}/Y = 447\%\, di$, which yields $\beta = 0.933$. I let the earnings process be an AR(1) in logs, which is discretized using the Rouwenhorst method. I calibrate to an autocorrelation of log earnings to $\rho_z = 0.91$ as in Floden and Lindé (2001). I calibrate the standard deviation of log earnings to match the earnings share of the bottom 80%. This yields $\sigma_z = 0.561$, which is slightly higher than Guvenen et al. (2021).

	Description	Value	Target	Source
β	Discount factor	0.933	$\int a_{i,ss}\, di/Y = 4.47$	SCF 2019
r^a	Average return	1.1%	$\mathbb{E}[r^a_{i,t} \mid a_{i,t-1} > 0]$	PSID
T	Transfers	0.17	$T/\text{Income} = 14\%$	SCF 2019
B	Government bonds	1.05	$B/Y = 105\%$	US OMB
τ	Tax rate	0.39	$G/Y = 17.6\%$	US BEA
μ	Markup	1.30	$A/Y = 447\%$	SCF 2019
c	Returns persistence	0.026	788 billionaires	Forbes
σ_z	Log earnings std. dev.	0.561	Bot. 80% earnings = 34.9%	SCF 2019
ρ_z	Log earnings persistence	0.91	Floden and Lindé (2001)	
σ	CRRA	1	Kaplan et al. (2018)	
ϕ_B	Tax adjustment speed	0.9	Auclert et al. (2024a)	
κ^W	NKWPC slope	0.03	Auclert et al. (2024b)	

Table 3: Calibration

Note: $\varrho(x_t) = \frac{\text{Cov}(x_t, x_{t-1})}{\text{Var}(x_t)}$ is the autocorrelation of x. $\sigma(x_t) = \sqrt{\text{Var}(x_t)}$ i the standard deviation of x_t. Δ is the first difference operator, i.e. $\Delta x_t = x_t - x_{t-1}$. Total household income is given by $\text{Income}_t = Z_t + \int r^a_{i,t} a_{i,t-1}\, di + T_t$.

For transfers from the government to households, I match an average transfer income share of 14% from the SCF, yielding $T = 0.17$. For government bonds, I

12. The data follows Kuhn and Rios-Rull (2016) updated to the 2019 SCF.

match the debt-to-GDP ratio of 105% in 2019 by setting $B = 1.05$. I match the ratio of government consumption to GDP of 17.6%, which requires setting $\tau = 0.39$, slightly higher than the estimates in Barro and Redlick (2011). Finally, I calibrate the markup such that the supply of assets matches the demand for assets, yielding $\mu = 1.30$.

Next, I present the externally calibrated parameters, i.e. the parameters set to values from the literature. Here, I set a CRRA of $\sigma = 1$, i.e. I consider log utility, which is standard. I set the tax adjustment speed to $\phi_B = 0.9$ as in Auclert et al. (2024a). I set the slope of the Phillips curve, κ^W, in accordance with Auclert et al. (2024b).

Finally, let me discuss the calibration of $\beta^z_{i,t}$: The pass-through of aggregate earnings to individual earnings. For this purpose, I calibrate to the "worker β's" from Guvenen et al. (2017). In particular, I set $\beta^z_{i,t}$ as a function of $z_{i,t}$. The details are spelled out in Appendix B.1. The result is given in Figure 5. The figure shows a U-shaped relationship between earnings and the pass-through of earnings: The pass-through is the largest for the rich and the poor, while the pass-through is low for the middle of the earnings distribution.

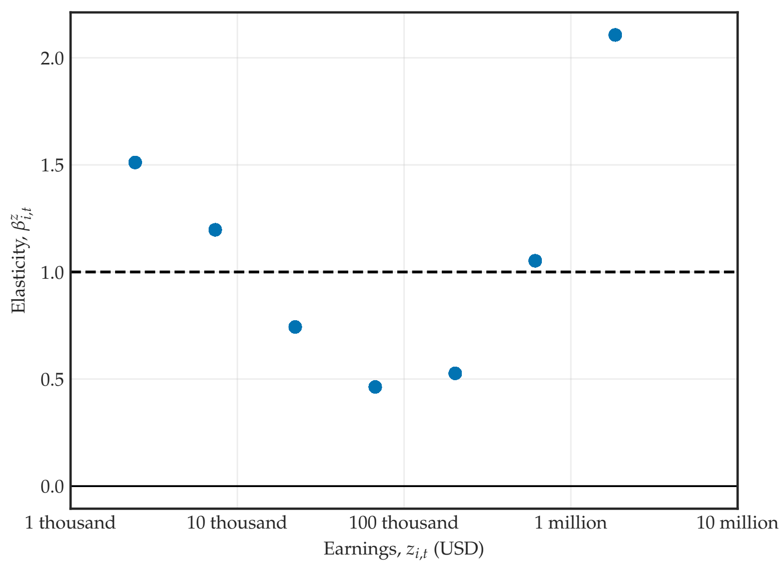

Figure 5: Pass-Through of Average to Individual Earnings by Earnings, $\beta^z_{i,t}$

Note: The figure shows the log pass-through of Z_t to $z_{i,t}$ by values of $z_{i,t}$.

In the rest of the paper, I will compare the model with heterogeneous returns to the standard HANK model. Table 3 refers to the model with heterogeneous returns, so let me discuss the calibration of the standard HANK model. All parameters are the same except the following. I set $e^r_{i,t} = 0$ and $\beta^r_{i,t} = 0$ to get common returns and $\beta^z_{i,t} = 1$ to get a common earnings pass-through. Finally, I consider permanent discount factor heterogeneity: Half of all households have a discount factor $\overline{\beta}$, while the other half have a discount factor $\underline{\beta}$. The first discount factor is set to match asset demand—as in the model with heterogeneous returns. The other discount factor is set to match the same average MPC as the model with heterogeneous returns. As I will argue shortly, this is an important moment to match, which the standard HANK model cannot do without permanent discount factor heterogeneity (or some other change).

4.3 Solution Method

I solve the households' problem using the endogenous gridpoint method (EGM) of Carroll (2006). I then use the methods from Auclert et al. (2021) to compute the Jacobians of the model. Lastly, I solve for the non-linear transition path under perfect foresight using a numerical equation solver.[13] This is equivalent to linearizing the model without perfect foresight (i.e. without aggregate risk) when shocks are small due to certainty equivalence. Note that while households have perfect foresight with respect to aggregate variables, they do not with respect to idiosyncratic variables, i.e. their labor income and returns.

Introducing heterogeneous returns changes the solution of the model compared to a standard HANK model in a few ways. First, it introduces a new state variable: The idiosyncratic return, $e^r_{i,t}$. Furthermore, heterogeneous returns stretch out the wealth distribution significantly. For this reason, it is important to have (i) a wealth grid with a very large maximum value, (ii) many grid points, and (iii) a highly non-linear grid. In particular, I (i) set the maximum wealth to 10^{14}, (ii) consider 501 grid points, and (iii) use the affine-exponential grid from Gouin-Bonenfant and Toda (2023).

Despite these differences, the solution of the model is not much slower than the standard HANK model. This is because the model does not introduce any new choice variables. Additionally, the grid for returns does not have to have very many grid points due to the efficiency of approximating a Markov chain using the log-Rouwenhorst method (Kopecky and Suen 2010 and Rouwenhorst 1995).[14] Thus, I can still use the fast EGM method of Carroll (2006) to solve the household problem with heterogeneous returns.

13. The code is written in Python and based on the GEModelTools package.

14. I use 7 grid points for both the labor income process and the returns process, i.e. $7^2 = 49$ grid points for both in total.

5 Microeconomic Fit

In this section, I show how the model with heterogeneous returns replicates several aspects of the microeconomic data that standard HANK models do not. To do so, I start by discussing heterogeneous returns and then turn to the wealth distribution since the former explains the latter.

5.1 Heterogeneous Returns

I now consider the heterogeneous returns in the model and their fit to the data. As discussed in Section 4, the returns process is calibrated to match the cross-sectional dispersion of returns. Thus, the model by assumption matches this aspect. In particular, Figure 6 shows the distribution of returns.

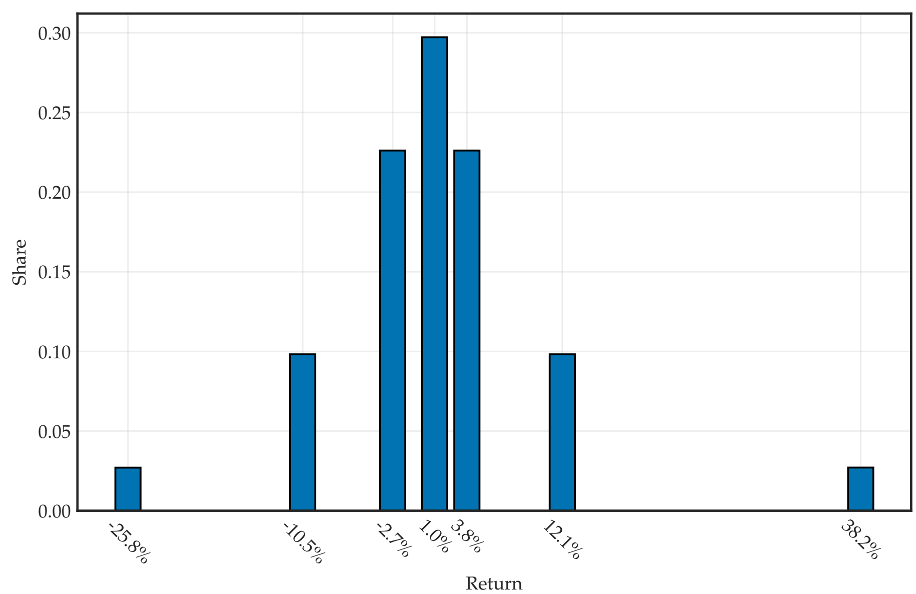

Figure 6: Histogram of the Returns in the Model

Note: The figure shows a histogram of the cross-sectional distribution of heterogeneous returns, $r^a_{i,t}$, in the model with heterogeneous returns.

A more interesting feature of the returns in the model is that the exhibit *scale dependence*. Scale dependence is the observation that wealthy households tend to earn higher returns. This is well-documented in the literature, see Fagereng et al. (2020), Bach et al. (2020), Xavier (2021), and Daminato and Pistaferri (2024). To show this, I split up the data sample and the model distribution by the wealth levels of households. For each group, I then compute the average return. Figure 7 plots this for quintiles of wealth. The figure shows that households with more wealth tend to earn higher returns, both in the model and in the data.

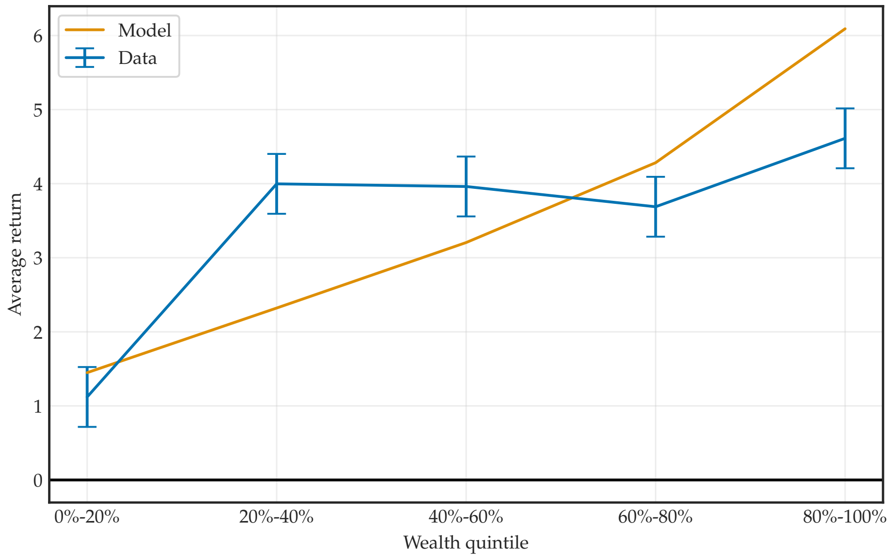

Figure 7: Average Return by Wealth (Scale Dependence)

Note: The figure shows the average return by quintiles of $a_{i,t-1}/a_{t,1}$, where a_t is the average wealth in year t.

Note that there is nothing in the specification of the returns process that says that wealthy households earn higher returns. Instead, the relationship between returns and wealth comes about as a result of household decisions: If households earn higher returns, they choose to save more, creating the relationship in Figure 7. I study this further in Appendix B.5, which plots the policy consumption functions for different rates of returns. The figure shows that households who earn higher returns have a lower MPC, saving more of income windfalls instead.

To study the congruence of return heterogeneity in the model and the data further, Figure 8 shows the equivalent of Figure 1 in the models. The figure shows that the bottom p% of households in the wealth distribution always hold the same shares of wealth and capital income in the standard HANK model, i.e. the curve is on the 45-degree line. This is due to the model having common returns. This clearly is at odds with the data, where for instance the bottom 95% holds 35% of all wealth but only 19% of capital income. In contrast, the model with heterogeneous returns fits the data well due to heterogeneity in returns and the scale dependence of returns.

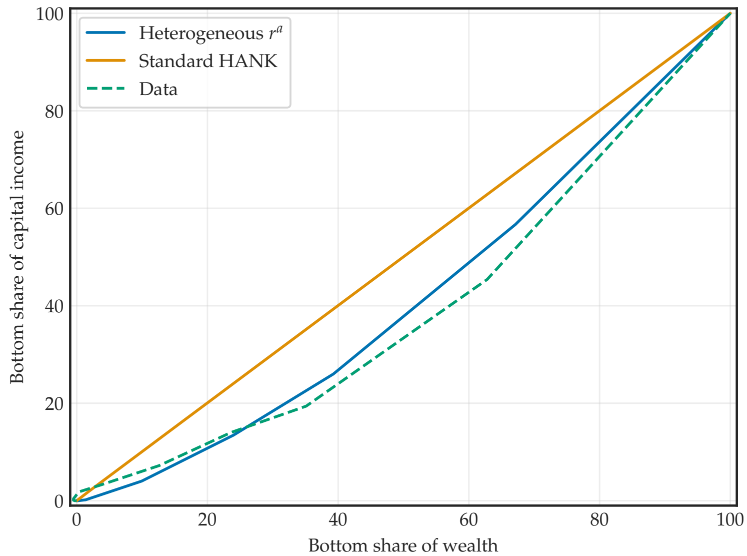

Figure 8: Shares of Wealth and Capital Income in the Models

Note: The figure shows the shares of wealth and capital income at different points in the distribution of households in the 2019 SCF and the two models. Households are sorted by their level of wealth. The x-axis shows the share of wealth held by the bottom x% of households, while the y-axis shows the share of capital income held by the same group.

5.2 Wealth

I now turn my attention to the wealth distribution and in particular the concentration in the upper right tail. I plot the right tail of the wealth distribution in Figure 9. Importantly, if wealth is Pareto-distributed, this plot should look like a line according to eq. (1), as I discussed in Section 2.1. This is not the case in the standard model, which significantly underestimates the wealth concentration at the top of the distribution. This is not due to the simplicity of the model: I also plot the larger 2-asset model of Kaplan et al. (2018), which also understates top wealth. On the other hand, the model with heterogeneous returns replicates the data almost perfectly, displaying a Pareto tail.

To elaborate on Figure 9, I provide key statistics on the wealth distribution in Table 4. This table highlights the key conclusion: The inclusion of heterogeneous returns allows the model to replicate the wealth distribution. Particularly the match to the top of the wealth distribution is a key innovation, as this is known to be difficult in standard heterogeneous agent models. This is despite the fact that only the number of billionaires in Table 4 is targeted in the calibration of the model. The improvement of the fit to the wealth distribution is not at the cost of matching the share of hand-to-mouth households, which is 33% in both models.

Finally, to summarize on the wealth distribution, I show the whole distribution in

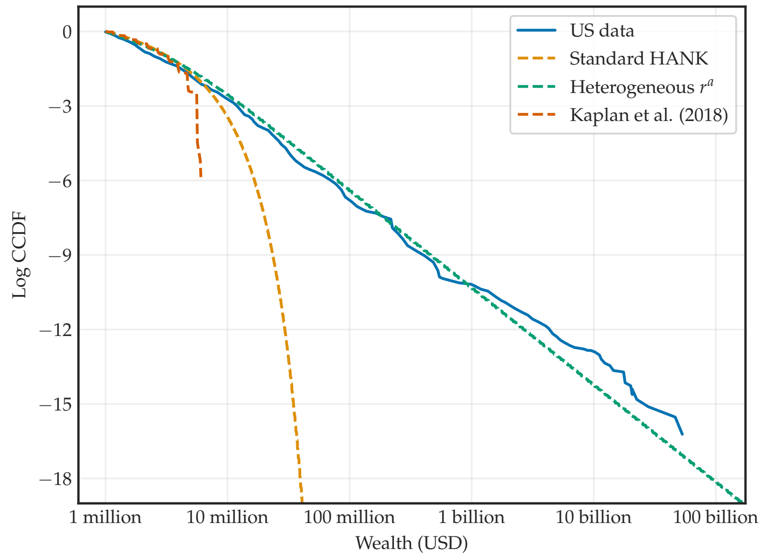

Figure 9: Wealth Concentration at the Top

Note: The figure shows the distribution of wealth in the two models. It does so by plotting the counter-CDF against the level of wealth in USD. Both axes are log-scale. The figure also shows the counter-CDF for the US based on data from Guvenen et al. (2023) and Vermeulen (2018).

Moment	Heterogeneous r^a	Standard HANK	Data
Top 20% share	90%	83%	87%
Top 10% share	76%	61%	76%
Top 1% share	33%	13%	37%
Top 0.1% share	13%	2%	14%
No. of billionaries	788	0	788
Top 0.0006% cutoff (bil. USD)	991	27	1000

Table 4: The Distribution of Wealth

Note: The data on the wealth distribution is the 2019 SCF. The data on the number of billionaires is from https://www.henleyglobal.com/publications/usa-wealth-report-2024. The values in the model are converted to USD by multiplying by GDP per household for the US in 2023.

Figure 10. In particular, Figure 10 shows the Lorenz curves of wealth in both models and the US data. The figure clearly shows how the model with heterogeneous returns significantly improves its fit to the wealth distribution compared to the standard HANK model, particularly at the top. Appendix B.6 shows Lorenz curves for the distributions of earnings and income. Additionally, Appendix B.7 compares the concentration of wealth, consumption, and capital income.

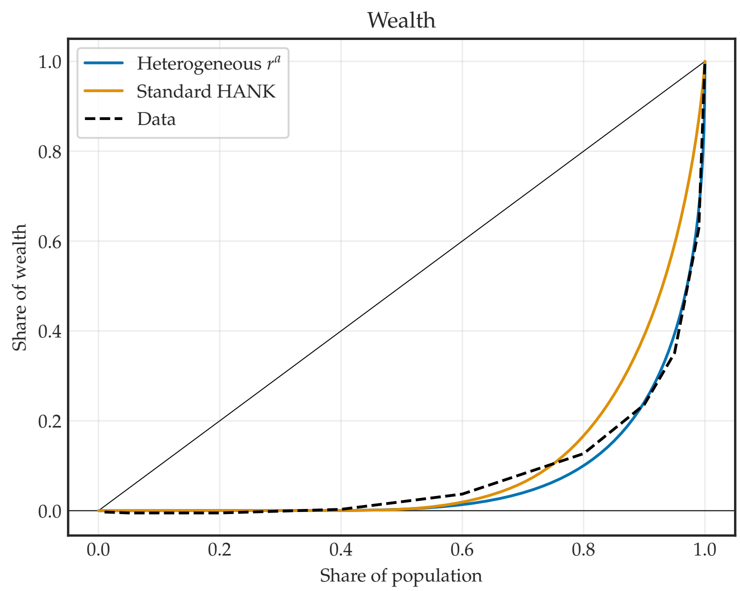

Figure 10: Lorenz Curves for Wealth in the Models and the Data

Note: The figure shows the Lorenz curves of wealth in both models and the 2019 SCF.

The fact that the model matches the wealth distribution is an attractive property in and of itself. But it is also an attractive property for another reason: It lets the model match a realistically high average marginal propensity to consume (MPC) *at the same* time it matches a realistically high average wealth. This is important because the literature on HANK models emphasizes that a key object to match is the MPC (Auclert et al. 2024b). However, it is a well-known issue that standard HANK models struggle to match simultaneously the MPC and a realistic level of wealth. This is the "MPC-wealth" tradeoff of Kaplan and Violante (2022). The intuition for this trade-off is straightforward: A high MPC is achieved by having a realistic amount of households close to or at the borrowing constraint. In contrast, a high level of wealth is achieved by having households away from the borrowing constraint.

The model with heterogeneous returns significantly reduces this tension. To see this, consider Figure 11. Figure 11 re-calibrates the model for different levels of the variation in returns.[15] As the figure shows, a larger variation in returns is associated with a higher MPC, keeping the aggregate level of wealth fixed.

The intuition follows from the considerations regarding the wealth distribution.

15. In particular, I multiply the return grid, $\mathcal{S}_r$ by some varying factor, and keep all other parameters fixed, except β and r, which are re-calibrated to ensure that the asset market still clears and the capital income of household still adds up to total capital income.

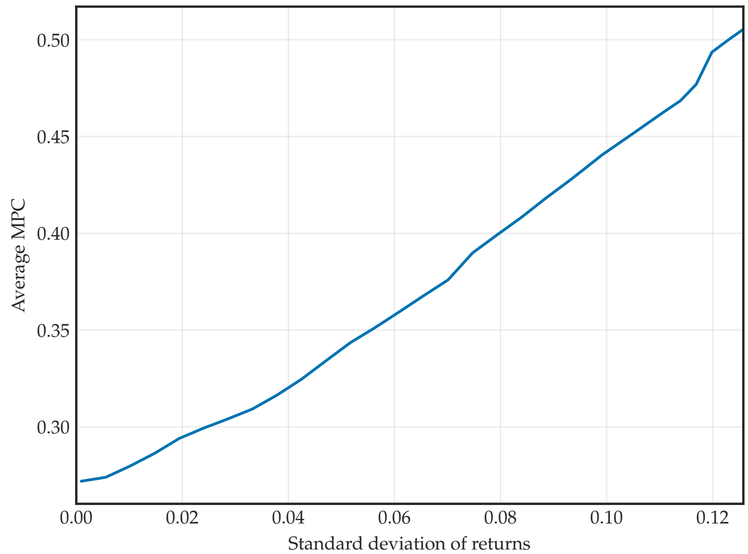

Figure 11: The MPC as Returns Are More Heterogeneous

Note: The figure shows the average MPC for the model with heterogeneous returns as the standard deviation of returns changes due to changes in $\text{Var}(r^a_{i,t})$.

With a higher variation in returns, the wealth distribution is more spread out. Thus, it is possible to simultaneously match a high MPC as many households are at the borrowing constraint, while matching a high level of wealth, as there are some extremely wealthy households contributing to a large aggregate wealth.

As such, the model with heterogeneous returns is an alternative—and arguably simpler—method of matching the level of wealth and the MPC compared to for instance the two-asset model.

6 Effects of Macroeconomic Policies

In this section, I compare the effects of macroeconomic policies when returns are heterogeneous with a focus on monetary policy. I first focus on the *aggregate* effects, i.e. the effects on aggregate outcomes such as output. Next, I focus on the *distributional* effects, i.e. the effect on individual outcomes such as income for different households.

6.1 Aggregate Effects of Macroeconomic Shocks

I start by considering the aggregate effects of monetary policy when returns are heterogeneous. To be specific, I consider the economy in the ergodic steady state. I then consider an unexpected 1 percentage point shock to the real interest with persistence 0.43, consistent with the estimated monetary policy shock in Auclert et al. (2020). I then report the resulting transition path back to the steady state. I will compare two models: The model with heterogeneous returns and the standard HANK model with common returns. The impulse response functions are given in Figure 12. Impulse response functions for additional variables are given in Appendix B.8.

Figure 12 shows that the effect of monetary policy is around 15% smaller in the model with heterogeneous returns compared to the model with common returns.

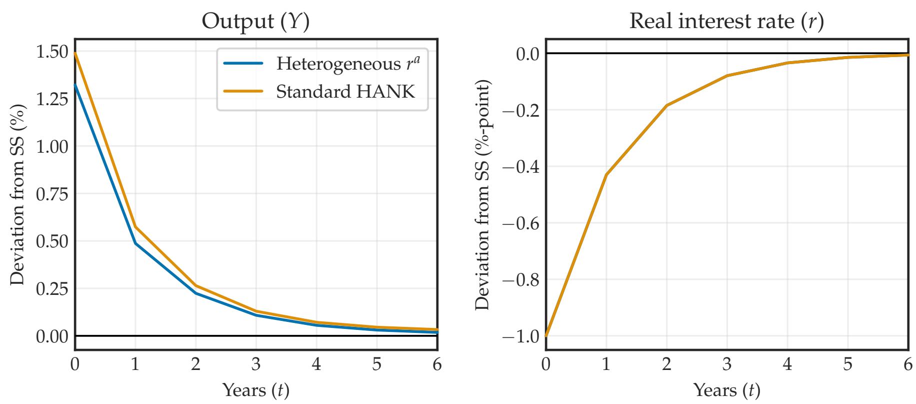

Figure 12: The Aggregate Effects of Monetary Policy

Note: The figure shows impulse response functions to a 1 percentage point fall in the real interest rate. The x-axis shows years after the shock.

To understand the differences in aggregate outcomes between the two models, I now decompose consumption into the channels through which monetary policy works. This decomposition is similar to the one in Kaplan et al. (2018). The decomposition is in Proposition 1.

Proposition 1. *Consider a monetary policy shock. The response of consumption is given by*

$$dC = \underbrace{\boldsymbol{M}^r d\boldsymbol{r}}_{\text{1. Direct}} + \underbrace{\boldsymbol{M}^Z d\boldsymbol{Z}}_{\text{2. Labor}} + \underbrace{\boldsymbol{M}^\tau d\boldsymbol{\tau}}_{\text{3. Taxes}} + \underbrace{\boldsymbol{M}^X dX_0}_{\text{4. Revaluation}} + \underbrace{\boldsymbol{M}^{Cov} dCov\left(r_i^a, \frac{a_{i,-1}}{A_{-1}}\right)}_{\text{5. Redistribution}},$$

where the $\boldsymbol{M}$*'s are defined in Appendix B.9.*

Proof. See Appendix B.9. □

I decompose the response of consumption, dC_0, into these channels in the two models in Table 5. Consider first the standard model with common returns. In this model, less than half of the response of consumption is driven by direct intertemporal substitution (channel 1). This channel reflects household pushing consumption forward in time and is the main channel operating in standard representative agent models. The remaining effect on consumption is through indirect channels: Higher labor income (channel 2), lower tax rates (channel 3), and higher capital income due to a revaluation of wealth (channel 4). Kaplan et al. (2018) makes the point that these indirect channels are strong drivers of consumption in HANK models, which is also the case here.

	Heterogeneous r^a	Standard HANK
1. Direct	1.37	0.71
2. Labor	0.12	0.15
3. Taxes	0.39	0.42
4. Revaluation	0.52	0.22
5. Redistribution	−1.08	0.00
Total	1.33	1.50

Table 5: Decomposition of Consumption in Response to Monetary Policy

Note: The table shows a decomposition of the response of consumption on impact to the monetary policy shock in both models.

Consider then the model with heterogeneous returns. In this model, one new channel is active: The redistribution channel (channel 5). The redistribution channel lowers consumption in the model with heterogeneous returns. This occurs because the capital gains disproportionately accrue to the wealthy as implied by $\beta^r_{i,t}$, consistent with the empirical evidence. As the wealthy have lower MPCs, this translates into a smaller increase in consumption as if returns increased 1-for-1 across the wealth distribution.

Why does capital income increase in response to a cut in interest rates? The important thing to understand is that monetary policy inflates asset prices. It does so for two reasons: Lower discount rates and higher dividends. To see this, note that eq. (12) implies that the price of firm equity is the discounted sum of future dividends:

$$p_t = \sum_{t=0}^{\infty} \frac{D_{t+1}}{1 + r_t}.$$

The direct effect of monetary policy is clear: Monetary policy lowers the interest rate, so decreases the discount rate. This increases the value of any (positive) stream of dividends. There is also an indirect effect through dividends: Higher economic activity implies higher dividends.[16] This follows from the fact that wages are more rigid than prices and is consistent with empirical evidence. This then makes the indirect effect of monetary policy on asset prices clear: Expansionary monetary policy increases output and hence dividends, boosting asset prices.

In Appendix B.10, I compare my models to the benchmark result from Werning (2015), who shows that monetary policy has the same effects on consumption with heterogeneous agents as in a representative agent model under certain assumptions. This result does not hold in my model with or without heterogeneous returns. However, when making certain assumptions, it does hold when comparing the model with common returns to a representative agent model. However—even under these assumptions—the Werning (2015) still does not hold in the model with heterogeneous returns, which has different effects of monetary policy compared to a model with common returns and a representative agent model, c.f. Appendix B.10

Having considered the effects of monetary policy, I now briefly turn my attention to fiscal policy. In particular, I consider two fiscal policies: Government consumption, G_t, and transfers, T_t. For both, I consider a shock of 1% of output with persistence 0.43. Figure 13 shows the output responses in the models with heterogeneous returns and the standard model. Appendix B.11 reports impulse response functions for additional variables. The figure shows that the effects of both fiscal policies on output are very similar. This is because both models are calibrated to the same MPC which is a key determinant of the effectiveness of fiscal policy. Additionally, the real interest rate is fixed, so the effects creating differences in response to monetary policy are mostly absent.

6.2 Distributional Effects of Macroeconomic Shocks

Having considered the aggregate effects of macroeconomic policies, I now turn my attention to the *distributional* effects.

To do this, I consider a policymaker who wants to stabilize the economy in the face of economic shocks. One tool the policymaker could use is monetary policy. Other tools are fiscal policy, i.e. changes in government consumption or government transfers. Any of these tools can stabilize aggregate demand. However, they might

16. This can be seen clearly from the fact that dividends are $D_t = \frac{\mu-1}{\mu} Y_t$.

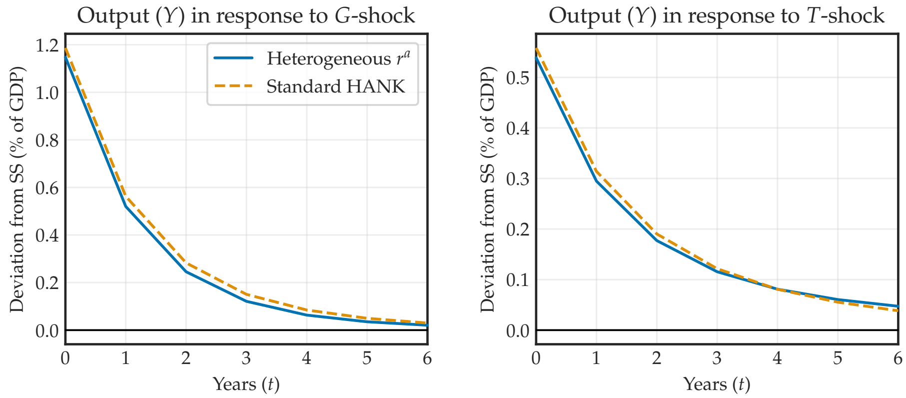

Figure 13: The Aggregate Effects of Fiscal Policy

Note: The figure shows impulse response functions of output to fiscal policy shocks of 1% of GDP with persistence 0.43. The x-axis shows years after the shock.

have different distributional effects, i.e. affect households differently. To study this, I ask the following question: For each $100 generated by a policy, how much goes to the top x%? This question helps clarify if policies favor the poor or the rich.

In Table 6 I ask exactly this question for the top 0.1%. I ask the question in both models for three different shocks: Monetary policy, transfers, and government consumption. The table shows a striking result: In the model with heterogeneous returns, 11% of all income generated by monetary policy goes to the top 0.1%. These numbers are much lower for transfers and government consumption, at 2.4% and 4.6%, respectively. Furthermore, the number is much larger than in the standard HANK model, where all policies modestly favor the top 0.1% with less than 2% of income going to the top 0.1%.

	Heterogeneous r^a	Standard HANK
Transfers	2.4%	0.5%
Government consumption	4.6%	1.1%
Monetary policy	10.8%	1.7%

Table 6: Shares of Income Going to the Top 0.1% After a Monetary Policy Shock

Note: The table shows the shares of income going to the top 0.1% on impact in response to the three different policies in the two models.

Next, I generalize Table 6 from the top 0.1% to any point in the wealth distribution. Figure 14 does so. The figure shows that transfers are by far the most equal policy: It benefits households across the income distribution fairly equally, with the bottom

x% getting almost x% of the increase in income. As an example, the bottom 50% get around 30% of the income generated by the policy. This is not surprising given that transfers are lump sum in the model, benefiting households equally. The fact that transfers are not completely equal is due to general equilibrium effects. The second most equal policy is government consumption. Here, the bottom 50% get almost 10% of the income generated. Consistent with Table 6, monetary policy is by far the least equal policy, with the bottom 50% getting essentially no income or even losing income.

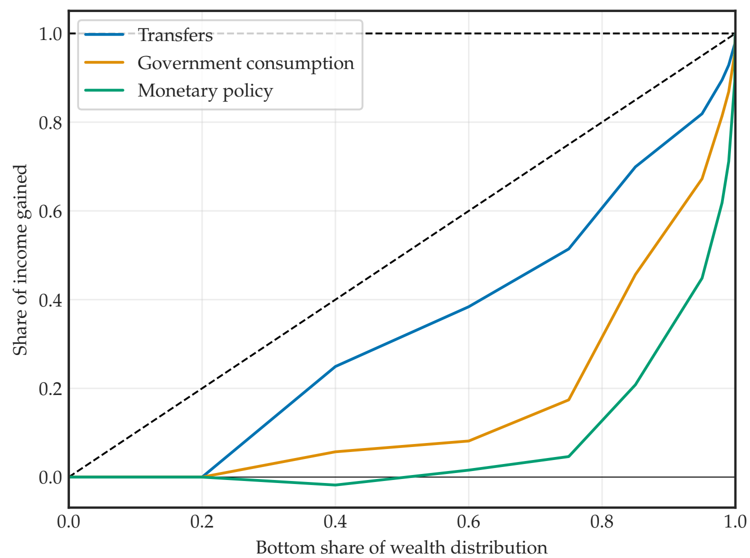

Figure 14: Shares of Income Going to Different Households

Note: The figure shows the composition of income generated by different policies on impact along the wealth distribution.

Why is this? The key thing to understand is that monetary policy implies a positive revaluation of wealth, $dr_0^a > 0$, as argued previously. How does this translate into income for different households? Note that at the top of the wealth distribution, income is essentially only capital income, $a_{i,ss}r_{i,0}^a$. This approximation certainly holds well for the top 0.1%. Top wealth is much larger in the model with heterogeneous returns than in the standard HANK model, so gains from higher asset prices are much higher at the top of the wealth distribution. This is why monetary policy benefits the ultra-rich much more in the model with heterogeneous returns than the standard HANK model, i.e. the third row in Table 6.

What explains the two first rows in Table 6, i.e. why do transfers and government consumption benefit the ultra-rich much less than monetary policy? This is quite simple: Even if the policies induce the same change in output, monetary policy inflates asset prices more due to lower discount rates, benefiting the ultra-rich more.

One way to see the importance of capital income is in Table 7. Table 7 splits the aggregate change of income into the three sources: Capital income, labor income, and government income. I define them as follows:

$$\begin{aligned} &\text{Labor income:} && z_{i,t}\ , \\ &\text{Capital income:} && x_{i,t}\ \equiv r^a_{i,t} a_{i,t-1}, \\ &\text{Gov. income:} && \omega_{i,t} \equiv T_t - t_{i,t}, \\ &\text{Total income:} && \psi_{i,t}\ \equiv x_{i,t} + z_{i,t} + \omega_{i,t}. \end{aligned}$$

The budget constraint is then simply $c_{i,t} + a_{i,t} = a_{i,t-1} + \psi_{i,t}$.

Table 7 shows each of the three income sources for each of the three policies. The table shows that transfers mostly generate government income because it *is* government income, while government consumption mostly generates labor income: This is because the government buys goods and firms have to pay households to produce these goods. Furthermore, monetary policy mainly generates capital income, as expected.

	Capital share	Labor share	Gov. share
Transfers	18%	24%	58%
Government consumption	35%	65%	0%
Monetary policy	85%	15%	0%

Table 7: Income Composition of Different Policies

Note: The table shows the composition of income generated by different policies on impact.

Figure 15 goes further and shows this decomposition into income sources along the wealth distribution. Consider first monetary policy. For monetary policy, the bottom 80% of the wealth distribution mainly benefit due to labor income and government income, which increase due to general equilibrium effects. At the top of the wealth distribution, the income gain is dominated by capital income, exactly as explained. For transfers and government consumption, the income sources are much more stable along the income distribution with everyone mainly gaining from the main sources generated by the shock as shown in Table 7.

One might ask if the distributional effects I find are consistent with the empirical literature outside the top of the wealth distribution. I do so in Appendix B.12. I find that the results are broadly consistent. However, due to the discussed data limitations,

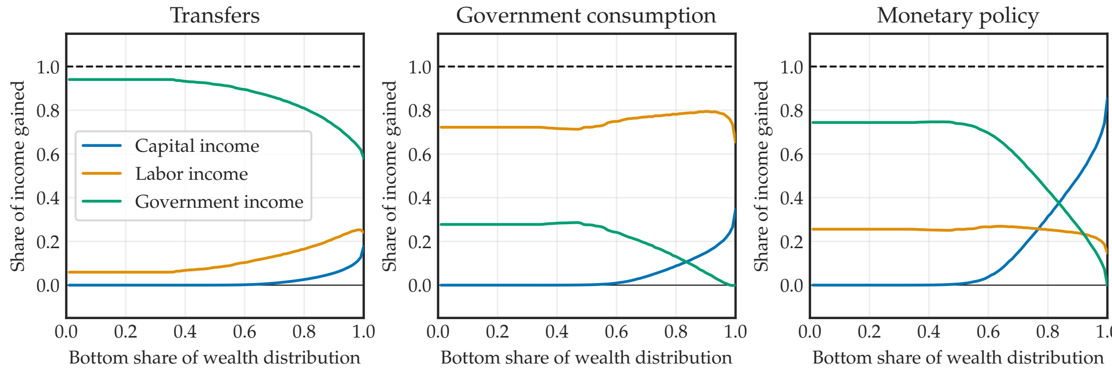

Figure 15: Income Composition of Policies for Different Households

Note: The figure shows the composition of income generated by different policies on impact along the wealth distribution ($a_{i,t}$).

they do not capture what happens at the very top of the wealth distribution.

One question is how robust the distributional effects in the model are. To answer this question, it is important to understand what shapes the distributional effects. For instance, does matching the wealth distribution matter? Does heterogeneity in returns matter? Proposition 2 answers these questions. It does so by a *sufficient statistics* approach: Using only households' budget constraints, what determines which households gain from different aggregate shocks?

Proposition 2. *Consider an aggregate shock. The share of income generated by this shock going to household i on impact is*

$$\frac{d\psi_i}{d\Psi} = \alpha_x \frac{a_{i,-1}}{A_{-1}} \beta_i^r + \alpha_z \frac{z_i}{Z} \beta_i^z + (1 - \alpha_x - \alpha_z) \frac{\omega_i}{\Omega} \beta_i^\omega, \tag{15}$$

neglecting time $t = 0$ subscripts when clear. $\alpha_x = dX/d\Psi$ and $\alpha_z = dZ/d\Psi$ are the capital and labor shares of the shock, while the β's are defined by

$$\beta_i^r \equiv \frac{dr_i^a}{d\bar{r}^a}, \quad \beta_i^z \equiv \frac{\frac{dz_i}{z_i}}{\frac{dZ}{Z}}, \quad \beta_i^\omega \equiv \frac{\frac{d\omega_i}{\omega_i}}{\frac{d\Omega}{\Omega}}.$$

Proof. See Appendix B.13. □

Proposition 2 holds for any model that satisfies the households' budget constraint. This means that it also holds for any dataset that can be cast in terms of such a budget constraint. With this result in hand, I now focus on the special case of monetary policy, which simplifies things further.

Corollary 1. *Consider a monetary policy shock. The share of income generated by this shock* *going to household i on impact is*

$$\frac{d\psi_i}{d\Psi} = \alpha_x \frac{a_{i,-1}}{A_{-1}} \beta_i^r + (1 - \alpha_x) \frac{z_i}{Z} \beta_i^z. \tag{16}$$

Proof. This follows from Proposition 2 when using the fact that monetary policy shocks do not affect the government's budget constraint in $t = 0$. □

This decomposition shows that one aggregate moment and four micro moments determine the distribution of income generated by monetary policy. The aggregate moment is the capital income share, which motivates focusing on this quantity. The four micro moments in the sufficient statistics decomposition are: The wealth distribution, $a_{i,-1}$, the labor income distribution, z_i, and the pass-through of aggregates to these, β_i^z and β_i^r. These four quantities are a key part of my calibration and the empirical fit of the model. Crucially, standard HANK models struggle to replicate these four quantities. They struggle particularly with the wealth distribution, $a_{i,-1}$, and β_i^r, which is the emphasis of this paper.

6.3 Robustness

Let me now consider the robustness of the results. To do so, I present a series of model changes and their implications for the model. In particular, I consider sticky prices, long debt, nominal debt, and different parameterizations.

6.3.1 Sticky Prices

I now consider what happens if the firm cannot set the price every period but instead has sticky prices. In particular, firms set the price, P_t, subject to quadratic adjustment costs, with discount factor $(1 + r_t)^{-1}$. This yields the following new Keynesian Phillips curve (NKPC) for inflation, $\pi_t = P_t / P_{t-1} - 1$, which replaces eq. (6):

$$\log(1 + \pi_t) = \kappa^P \left(w_t - \frac{1}{\mu} \right) + \frac{1}{1 + r_t} \frac{Y_{t+1}}{Y_t} \log(1 + \pi_{t+1}).$$

Here, $\kappa^P \geq 0$ is the slope of the Phillips curve. Flexible prices are obtained as $\kappa^P \to \infty$. I set $\kappa^P = 0.23$ as in Auclert et al. (2024b). Crucially, this implies $\kappa^W < \kappa^P$ such that wages are more sticky than prices (Broer et al. 2020).

6.3.2 Long Debt

I now consider what happens if government bonds have a duration longer than 1 year. In particular, the government issues a quantity of real bonds, B_t, with real price, q_t. The budget constraint in eq. (7) is then modified to

$$q_t B_t = (1 + \delta q_t) B_{t-1} + G_t + T_t - \mathcal{T}_t. \tag{17}$$

The bonds are long, paying a unit coupon each period. They are exponentially decaying with decay rate $\delta \in [0, 1]$, cf. Auclert and Rognlie (2018) and Auclert et al. (2020). I consider the case of long bonds because the literature emphasizes the importance of this for the transmission of monetary policy. In this case, the definition of capital income in eq. (11) is modified to

$$\underbrace{\int r_{i,t}^a a_{i,t-1}\, di}_{\text{HH's capital income}} = \underbrace{p_t + D_t + (1 + \delta q_t) B_{t-1} - A_{t-1}}_{\text{Capital income from bonds and equity}} \tag{18}$$

and the no-arbitrage condition in eq. (12) is simply

$$r_t = \frac{p_{t+1} + D_{t+1}}{p_t}.$$

Finally, asset market clearing in eq. (14) reads

$$A_t = p_t + q_t B_t.$$

For the calibration, I set $\delta = 0.8$ as in Auclert et al. (2020) to match the average US debt maturity of 5 years.

6.3.3 Nominal Debt

I now consider what happens if government bonds are nominal. In this case, the government's budget constraint in eq. (7) is modified to

$$B_t = \frac{1 + i_{t-1}}{1 + \pi_t} B_{t-1} + G_t + T_t - \mathcal{T}_t. \tag{19}$$

The definition of capital income in eq. (11) reads

$$\underbrace{\int r^a_{i,t} a_{i,t-1}\, di}_{\text{HH's capital income}} = \underbrace{p_t + D_t + \frac{1+i_{t-1}}{1+\pi_t} B_{t-1} - A_{t-1}}_{\text{Capital income from bonds and equity}} \tag{20}$$

Model	Output increase	Top 0.1% income share
Baseline	1.32%	10.8%
Long debt	1.36%	11.1%
Nominal debt	1.33%	10.8%
Sticky prices	1.31%	10.8%
More flexible wages	1.32%	10.8%
More bonds	1.37%	9.1%

Table 8: Robustness

Note: The table shows the robustness of the results to various changes to the model.

6.3.4 Parametrization

In addition to the model changes, I also consider two different parametrizations of the model. First, I consider what happens when wages are more flexible: $\kappa^W = 0.1$ instead of $\kappa^W = 0.03$. Second, I consider the case of more of liquidity being in the form of government bonds: $B/A = 50\%$ instead of the baseline of $B/A = 23\%$. This implies a different markup, μ, which is important c.f. Appendix B.14.

6.3.5 Results

The results of all model changes are shown in Table 8. As the table shows, the results are very robust to all changes. In particular, the increase in output remains in a narrow band. The same is the case for the top 0.1% income share of monetary policy, with one slight exception: When more liquidity is in the form of bonds, the top 0.1% income share drops slightly. This is because there is less capital income from holding firm equity. Despite the implausibly large value of government bonds, however, the top 0.1% income share remains large.

7 Conclusion

I study the distributional effects of monetary policy when returns are heterogeneous by adding them to an otherwise standard HANK model. I do so by constructing a dataset of heterogeneous returns across US households. I find that the model replicates empirical distributions of returns, wealth, and income. Crucially, the model matches the concentration of wealth at the top and the pass-through of aggregate to individual returns.

Using a sufficient statistics approach, I show that exactly these moments of the microeconomic data shape the distributional effects of monetary policy. Because the standard HANK model matches these moments, it understates how much monetary policy disproportionately benefits the ultra-rich. Because the model with heterogeneous returns replicates these moments, it is well suited for understanding this.

I find that income gains from expansionary monetary policy disproportionately benefit the ultra-rich: The top 0.1% take 11% of the income increase, more than 100 times their population share and an order of magnitude more than in standard HANK models. This is because monetary policy mainly increases capital income, which mostly goes to the wealthy. Existing empirical studies miss how unevenly distributed the gains from monetary policy are because of censored data, a limited sample, and/or distributing households into coarse buckets hiding these effects. Theoretical models also miss these effects, as they do not replicate the concentration of wealth in the data.

My paper thus constitutes new evidence on the distributional effects of monetary policy. These findings potentially have important policy implications. They suggest that policymakers aiming to stimulate aggregate demand through monetary policy should be aware of the disproportionate benefits to the wealthiest households. In contrast, fiscal policy tools such as transfers and government consumption distribute gains more equitably. Policymakers who care about distributional effects should consider this when designing stabilization policies.

References

Andersen, Asger Lau, Niels Johannesen, Mia Jørgensen, and José-Luis Peydró. 2023. "Monetary Policy and Inequality." *The Journal of Finance* 78, no. 5 (October): 2945–2989.

Auclert, Adrien, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub. 2021. "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models." *Econometrica* 89 (5): 2375–2408.

Auclert, Adrien, and Matthew Rognlie. 2018. *Inequality and Aggregate Demand.* Technical report w24280. Cambridge, MA: National Bureau of Economic Research, February.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2020. *Micro Jumps, Macro Humps: Monetary Policy and Business Cycles in an Estimated HANK Model.* Technical report w26647. Cambridge, MA: National Bureau of Economic Research, January.

———. 2024a. "Fiscal and Monetary Policy with Heterogeneous Agents" (September): NBER working paper 32991.

———. 2024b. "The Intertemporal Keynesian Cross." *Journal of Political Economy* (August): 732531.

Bach, Laurent, Laurent E. Calvet, and Paolo Sodini. 2020. "Rich Pickings? Risk, Return, and Skill in Household Wealth." *American Economic Review* 110, no. 9 (September): 2703–2747.

Barro, Robert J., and Charles J. Redlick. 2011. "Macroeconomic Effects From Government Purchases and Taxes *." *The Quarterly Journal of Economics* 126, no. 1 (February): 51–102.

Bayer, Christian, Benjamin Born, and Ralph Luetticke. 2024. "Shocks, Frictions, and Inequality in US Business Cycles." *American Economic Review* 114, no. 5 (May): 1211–1247.

Benhabib, Jess, Alberto Bisin, and Mi Luo. 2017. "Earnings Inequality and Other Determinants of Wealth Inequality." *American Economic Review* 107, no. 5 (May): 593–597.

Benhabib, Jess, Alberto Bisin, and Shenghao Zhu. 2011. "The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents." *Econometrica* 79 (1): 123–157.

———. 2015. "The wealth distribution in Bewley economies with capital income risk." *Journal of Economic Theory* 159 (September): 489–515.

Broer, Tobias, Niels-Jakob Harbo Hansen, Per Krusell, and Erik Öberg. 2020. "The New Keynesian Transmission Mechanism: A Heterogeneous-Agent Perspective." *The Review of Economic Studies* 87, no. 1 (January): 77–101.

Carroll, Christopher D. 2006. "The method of endogenous gridpoints for solving dynamic stochastic optimization problems." *Economics Letters* 91, no. 3 (June): 312–320.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia. 2017. "Innocent Bystanders? Monetary policy and inequality." *Journal of Monetary Economics* 88 (June): 70–89.

Daminato, Claudio, and Luigi Pistaferri. 2024. "Returns Heterogeneity and Consumption Inequality Over the Life Cycle" (May): NBER working paper 32490.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. 2020. "Heterogeneity and Persistence in Returns to Wealth." *Econometrica* 88 (1): 115–170.

Ferra, Sergio de, Kurt Mitman, and Federica Romei. 2020. "Household heterogeneity and the transmission of foreign shocks." *Journal of International Economics* 124 (May): 103303.

Floden, Martin, and Jesper Lindé. 2001. "Idiosyncratic Risk in the United States and Sweden: Is There a Role for Government Insurance?" *Review of Economic Dynamics* 4, no. 2 (April): 406–437.

Gabaix, Xavier, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2016. "The Dynamics of Inequality." *Econometrica* 84 (6): 2071–2111.

Gaillard, Alexandre, Philipp Wangner, Christian Hellwig, and Nicolas Werquin. 2023. "Consumption, Wealth, and Income Inequality: A Tale of Tails." *SSRN Electronic Journal.*

Gouin-Bonenfant, Émilien, and Alexis Akira Toda. 2023. "Pareto extrapolation: An analytical framework for studying tail inequality." *Quantitative Economics* 14 (1): 201–233.

Guvenen, Fatih, Gueorgui Kambourov, Burhan Kuruscu, Sergio Ocampo, and Daphne Chen. 2023. "Use It or Lose It: Efficiency and Redistributional Effects of Wealth Taxation." *The Quarterly Journal of Economics* 138, no. 2 (April): 835–894.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song. 2021. "What Do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics?" *Econometrica* 89 (5): 2303–2339.

Guvenen, Fatih, Sam Schulhofer-Wohl, Jae Song, and Motohiro Yogo. 2017. "Worker Betas: Five Facts about Systematic Earnings Risk." *American Economic Review* 107, no. 5 (May): 398–403.

Holm, Martin Blomhoff, Pascal Paul, and Andreas Tischbirek. 2021. "The Transmission of Monetary Policy under the Microscope." *Journal of Political Economy* 129, no. 10 (October): 2861–2904.

Jones, Charles I., and Jihee Kim. 2018. "A Schumpeterian Model of Top Income Inequality." *Journal of Political Economy* 126, no. 5 (October): 1785–1826.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante. 2018. "Monetary Policy According to HANK." *American Economic Review* 108, no. 3 (March): 697–743.

Kaplan, Greg, and Giovanni Violante. 2022. *The Marginal Propensity to Consume in Heterogeneous Agent Models.* Technical report. Cambridge, MA: National Bureau of Economic Research, May.

Kopecky, Karen A., and Richard M.H. Suen. 2010. "Finite state Markov-chain approximations to highly persistent processes." *Review of Economic Dynamics* 13, no. 3 (July): 701–714.

Kuhn, Moritz, and Jose-Victor Rios-Rull. 2016. "2013 Update on the U.S. Earnings, Income, and Wealth Distributional Facts: A View from Macroeconomic Modelers."

McKay, Alisdair, and Christian K. Wolf. 2023a. "Monetary Policy and Inequality." *Journal of Economic Perspectives* 37, no. 1 (February): 121–144.

———. 2023b. *Optimal Policy Rules in HANK.* Technical report.

Rouwenhorst, K. Geert. 1995. "10 Asset Pricing Implications of Equilibrium Business Cycle Models." In *Frontiers of Business Cycle Research,* edited by Thomas F. Cooley, 294–330. Princeton University Press, December.

Smith, Matthew, Owen Zidar, and Eric Zwick. 2022. "Top Wealth in America: New Estimates Under Heterogeneous Returns." *The Quarterly Journal of Economics* 138, no. 1 (December): 515–573.

Snudden, Stephen. 2021. "Leverage and Rate of Return Heterogeneity among U.S. Households." *Working paper.*

Vermeulen, Philip. 2018. "How Fat is the Top Tail of the Wealth Distribution?" *Review of Income and Wealth* 64, no. 2 (June): 357–387.

Werning, Iván. 2015. "Incomplete Markets and Aggregate Demand" (August): NBER working paper 21448.

Xavier, Inês. 2021. "Wealth Inequality in the US: the Role of Heterogeneous Returns." *SSRN Electronic Journal.*

Appendix

A Empirical Appendix

A.1 Survey of Consumer Finances

I use data on the 2019 SCF. The data is from Kuhn and Rios-Rull (2016), which is updated to 2019 online. They provide data on earnings, income, and wealth directly. They also provide data on transfers as a share of income, which I convert to USD using the income data. I do the same for "other" income.

In the model, income is separated into earnings, capital income, and transfers. There are two issues to overcome in order to make the data consistent with the model: (1) What to consider capital income and (2) how to attribute "other" income. For (1), I compute capital income as the part of income not due to earnings, transfers, or other income. Regarding (2), I distribute other income to the three remaining components (earnings, transfers, and capital income). The result is a variable for income identical to the one in the data provided by Kuhn and Rios-Rull (2016), but which is made up of the three components in the model: (1) Earnings, (2) capital income, and (3) transfers.

A.2 Panel Survey of Income Dynamics

In this section, I discuss how I use data from the PSID to construct a dataset of heterogeneous returns across US households.

A.2.1 Data Structure

In addition to the SCF, I use panel data from the PSID. The panel is bi-annual, starting in 1999 and ending in 2019. Crucially, the unit of time for flow variables is still years. Thus, the panel can be thought of as annual with missing data.

The variables I use can be considered as being part of two categories: Wealth *stocks* and income *flows*. For a survey conducted in year t, respondents are asked about their *stock* of wealth at that point in time, i.e. year t, and about income flows during the previous year, i.e. year $t-1$, see Figure A.1. For this reason, both the numerator and denominator in eq. (2) are observed, though not in the same survey. Thus, I can construct returns for year $t-1$ for each survey year t, but not t itself. Since surveys are conducted in 1999, 2001, ..., 2019, I can construct returns for 2000, 2002, ..., 2018.[17]

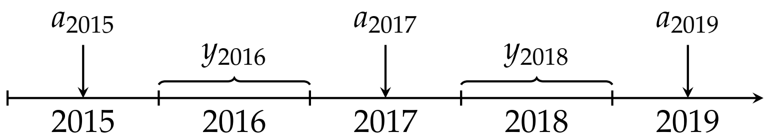

Figure A.1: Example of Survey Timeline

A.2.2 Sample Restrictions

I consider a sample where the head of the household is working full time. Thus, I exclude students and retirees. I also restrict the age of the head to be 25–65. I drop the SEO sample consisting of over-sampled low-income households. I restrict attention to households that have the same head and with no change in the family composition. I winsorize the sample by capping returns at the 0.5th and 99.5th percentiles.

A.2.3 Capital Income

I now discuss how I compute the capital income from the 8 sources. First, I discuss income from trust funds and royalties, interest income, and dividend income. Respondents are asked directly about these 3 in the survey, so I simply use the responses. I discuss the remaining 5 in turn.

- *Primary Housing.* Income from primary housing can be split into two parts. The first part is rental income, which is reported directly in the survey, but is not attributed to primary or other housing. I attribute the rental income to primary housing if the household is a homeowner and does not own other real estate. Otherwise, I attribute it to other housing.

  The second part is capital gains. Capital gains are computed as

  $$\text{capital gains} = \frac{\text{price change} - \text{improvements}}{2},$$

  The price change is the change in the price of the house, which depends on if the house was sold or not. If the household was sold, the selling price is used. If the price was not used, the self-assessed value is used. Both are net of an 8% commission. Improvements are reported directly in the survey but are not

17. I lose the first year due to the lagged wealth in eq. (2).

attributed to primary or other housing. I attribute it to other housing if the respondents have primary housing and other housing otherwise.

- *Other Housing.* Income from other housing can be split into the same two parts as primary housing. The first part is rental income, which is reported directly in the survey, but is not attributed to primary or other housing. I attribute it as presented when discussing primary housing.

  The second part is capital gains. Capital gains are computed as

  $$\text{capital gains} = \frac{\text{price change} - \text{improvements} - \text{net investment}}{2}.$$

  The price change is simply the change in the price, while net investments are the difference between the price of real estate sold and the price of real estate bought. Improvements are as discussed with primary housing.

- *Businesses.* Business income is split into two types: Realized and unrealized capital gains. The realized part is the sum of the income associated with owning the business—reported directly in the survey—and the income associated with owning the farm.[18]

- *Stocks.* Income from stocks is given by

  $$y_{i,t}^{\text{stocks}} = \frac{\Delta a_{i,t}^{\text{stocks}} - f_{i,t}}{2},$$

  where $a_{i,t}^{\text{stocks}}$ is the value of the stocks and $f_{i,t}$ is the net investment into stocks.

- *Other.* In the survey, the respondent is asked for the total capital income of other members of the family unit. I include this as *other* capital income.[19]

A.2.4 Wealth

The value of wealth can be split into two types in the survey: Assets for which net investment *is* reported, and assets for which net investment is *not* reported. This

18. Farm income is not split into capital and labor, so I do this. If farm income is negative, I attribute all of it to capital income. If it is positive, I attribute half to capital income.

19. Before 2005, respondents are only asked about the *total* income of other members of the family unit, not how much of it is labor income. I attribute this income as labor income before 2005, except if it is negative, in which case I attribute it as capital income.

distinction is important because it matters for how returns are computed. In particular, it matters for the measurement of the numerator in eq. (2). When net investments are reported, I simply use the lagged value of wealth in the numerator. When net investments are not reported, I use an average of the lagged value of wealth and the contemporaneous value of wealth. Let me start by discussing the assets where net investment is reported.

- *Primary housing.* Respondents are asked about the value of their house. I report the value net of 8% commission.
- *Other housing.* Respondents are asked about the value of their real estate, i.e. how much it would sell for.
- *Stocks.* Respondents are asked about the value of their stocks if they paid off everything owed on them.
- *Businesses.* Respondents are asked about the value of their farm/business, i.e. how much it would sell for.

Let me then discuss the assets where net investment is *not* reported.

- *Private annuities or IRAs.* Respondents are asked about the value of their annuities/IRAs.
- *Checking/savings accounts.* Respondents are asked about the value of their checking/savings accounts.
- *Vehicles.* Respondents are asked about the value of their vehicles if they paid off everything owed on them.
- *Other assets.* Respondents are asked about the value of other assets if they paid off everything owed on them.

A.2.5 Time Series of Average Returns

Figure A.2 shows the time series of average returns estimated in the PSID for each survey. In addition to the average return, I also report the average return for the bottom 20% and the top 20%.

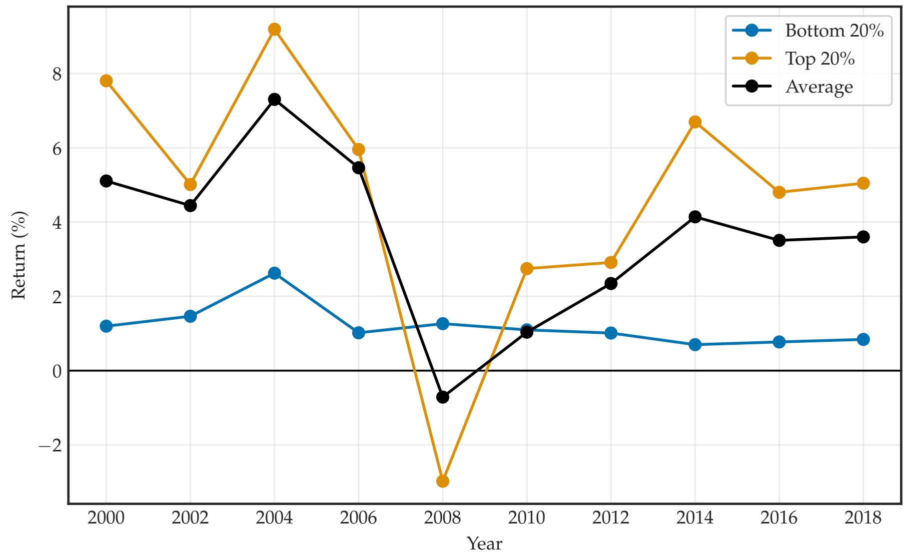

Figure A.2: Time Series of Average Returns

Note: The figure shows a time series of average returns based on the PSID data.

A.2.6 Explaining Returns With Portfolio Shares

Consider splitting household wealth into N assets indexed by $j = 1, \ldots, N$ such that total capital income and wealth are the sum over these assets:

$$y_{i,t} = \sum_{j=1}^{N} y_{i,t}^{j} \quad \text{and} \quad a_{i,t} = \sum_{j=1}^{N} a_{i,t}^{j}.$$

Consider the case where households only earn different returns because they hold different assets, but each asset gives the same return, i.e. $y_{i,t}^{j} = r_t^j a_{i,t-1}^{j}$. In this case, the total return for a household is just a weighted average of individual returns,

$$r_{i,t} = \frac{y_{i,t}}{a_{i,t-1}} = \sum_{j=1}^{N} \frac{a_{i,t-1}^{j}}{a_{i,t-1}} \frac{y_{i,t}^{j}}{a_{i,t-1}^{j}} = \sum_{j=1}^{N} \omega_{i,t-1}^{j} r_t^j \tag{21}$$

where the weights are $\omega_{i,t}^{j} = a_{i,t}^{j}/a_{i,t}$. An immediate implication is that a regression of individual returns on individual weights should yield a R^2—coefficient of determination—of exactly 1 *within each year*.

Figure A.3 reports the R^2 of such regressions.[20] The R^2 is in the range of 0.01–0.06,

20. I include a constant in the regression even though eq. (21) suggests that this should be 0. This should only bias upwards the R^2.

far away from 1. Even significant measurement error cannot explain this, suggesting that household returns are idiosyncratic also within asset categories and years. This is consistent with Fagereng et al. (2020), who find that returns are heterogeneous also within narrow asset classes.

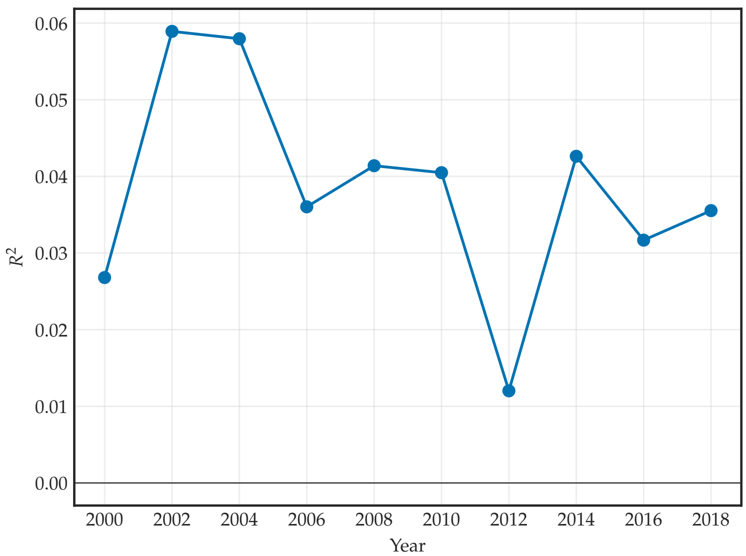

Figure A.3: R^2 Of Regression of Returns on Portfolio Shares by Year

Note: The figure shows the coefficient of determination, i.e. R^2, of a regression of household-level returns on their portfolio shares by year.

B Model Appendix

B.1 The Elasticity of Earnings

I start by showing that the elasticity of $z_{i,t}$ with respect to Z_t is $\beta^z_{i,t}$. To see this, note the definition of an elasticity:

$$\frac{\partial z_{i,t}}{\partial Z_t}\frac{Z_t}{z_{i,t}} = e^z_{i,t}\beta^z_{i,t}\frac{Z_t}{z_{i,t}}. \tag{22}$$

Evaluating this in the steady state yields that this elasticity is $\beta^z_{i,t}$.

Let me now discuss how to ground $\beta^z_{i,t}$ empirically. Here, I turn to Guvenen et al. (2017). They estimate a regression that recovers something very similar to the elasticity, $\beta^z_{i,t}$. The only difference is that their elasticity is of earnings with respect to GDP and not aggregate earnings. However, these two elasticities are the same in the model. To see this, note that

$$Z_t = w_t N_t = \frac{W_t}{P_t} Y_t = \frac{1}{\mu} Y_t.$$

Thus, earnings are given by

$$z_{i,t} = \frac{e^z_{i,t}}{\mu} Y_{ss} + \frac{e^z_{i,t}}{\mu}\beta^z_{i,t}(Y_t - Y_{ss}).$$

Repeating the calculations in eq. (22) for Y_t instead of Z_t then yields that the elasticity of $z_{i,t}$ with respect to Y_t is $\beta^z_{i,t}$. Thus, I use the estimates of $\beta^z_{i,t}$ from Guvenen et al. (2017). In particular, I use their estimates of $\beta^z_{i,t}$ as a function of earnings percentiles for males. To do so, I look at percentiles of the 7 grid points for $e^z_{i,t}$ in the model and interpolate between the estimates in Guvenen et al. (2017). Finally, I divide all $\beta^z_{i,t}$ by the same value such that $\int_i e^z_{i,t}\beta^z_{i,t}\, di = 1$, which ensures that $\int_i z_{i,t}\, di = Z_t$ for all t.

B.2 Robustness to Returns Process

I now consider 3 alternative specifications for the returns process. The first specification is that $e^r_{i,t}$ follows a mean-zero AR(1), discretized by the Rouwenhorst method. I set the standard deviation and autocorrelation to match the baseline specification. This specification has three problems. First, it features a much too low wealth concentration, c.f. Table A.1. Second, it features far too much negative income compared to the data, where there is essentially no negative income whatsoever. This negative income occurs because some very wealthy households get unlucky and earn negative returns. The baseline process almost entirely avoids this, because households can only transition up or down one state. Third, it features too low scale dependence, measured as the difference between returns of the top 20% and bottom 20%, c.f. Table A.1.

The second specification is that there are 2 permanent returns, i.e. $e^r_{i,t} = e^r_i \in \{e^r_{\text{low}}, e^r_{\text{high}}\}$. The values are pinned down by (i) having a mean of zero, and (ii) matching the standard deviation of the baseline specification. This specification has two main problems. First, it features too little wealth concentration as measured by the top 1% wealth share, c.f. Table A.1. Second—and in contrast with the AR(1)—if features too *much* scale dependence.

The third specification is a schedule for returns. In particular, I set the idiosyncratic returns process as a function of wealth: $e^r_{i,t} = f(a_{i,t-1})$. I choose this schedule to match the schedule in the baseline model. By construction—and in contrast with the other two specifications—it has a realistic degree of scale dependence. However, this comes at the cost of too low a standard deviation of returns.

In conclusion, the baseline specification provides the best fit to the data. However, let me emphasize that combining the different specifications would probably provide the best fit and most realistic specification at the cost of parsimony and without improving the fit much compared to the baseline specification.

	Baseline	AR(1)	Permanent	Schedule
Average return	1.1%	1.1%	1.1%	2.6%
Std. dev. of returns	9.3%	9.3%	9.3%	1.6%
Top 1% wealth share	32.7%	15.4%	11.6%	14.9%
Scale dependence	4.6%	−0.4%	9.4%	4.0%
1st income percentile	0.18	−0.97	0.19	0.19

Table A.1: Alternative Returns Processes

Note: The table shows descriptive statistics from models with alternative returns processes. “Scale dependence” refers to the difference in returns of the top 20% and bottom 20% in the wealth distribution.

B.3 Walras' Law

In this appendix, I show that the goods market clearing condition in eq. (13) implies the asset market clearing in (14). To start, note that aggregating households' budget constraints in eq. (4) yields

$$C_t + A_t = (1 + \tilde{r}^a_t)A_{t-1} + Z_t + T_t - \mathcal{T}_t.$$

Inserting the government's budget constraint in eq. (7) gives

$$C_t + A_t = (1 + \tilde{r}^a_t)A_{t-1} + Z_t + B_t - (1 + r_{t-1})B_{t-1} - G_t.$$

Using eq. (11) implies that

$$C_t + A_t = p_t + D_t + Z_t + B_t - G_t.$$

Using $D_t = Y_t - Z_t$ yields

$$C_t + A_t = p_t + Y_t + B_t - G_t.$$

Goods market clearing in eq. (13) then implies that

$$A_t = p_t + B_t,$$

which is exactly the asset market clearing in eq. (14).

B.4 Returns Process

The idiosyncratic part of returns, $e^r_{i,t}$, follows a discrete Markov chain with 7 grid points. These are set as the median within 7 bins, yielding the grid points for $e^r_{i,t}$. The ergodic distribution over these grid points, π, is taken as the empirical distribution.

What remains to be specified is the transition matrix. I parametrize the transition matrix as follows. For each of the 7 states, there is some probability of staying in that state. Additionally, there is some probability of going up one state and the same probability of going down that state. This means that the transition matrix can be written as

$$\Pi = \begin{pmatrix} p_1 & 1-p_1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1-p_2}{2} & p_2 & \frac{1-p_2}{2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1-p_3}{2} & p_3 & \frac{1-p_3}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1-p_4}{2} & p_4 & \frac{1-p_4}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1-p_5}{2} & p_5 & \frac{1-p_5}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1-p_6}{2} & p_6 & \frac{1-p_6}{2} \\ 0 & 0 & 0 & 0 & 0 & 1-p_7 & p_7 \end{pmatrix}$$

There is the following relationship between the transition matrix and the stationary distribution:

$$p_i = \begin{cases} 1 - \frac{c}{\pi_i} & \text{for} \quad i = 1,7 \\ 1 - \frac{2c}{\pi_i} & \text{for} \quad i = 2,3,4,5,6 \end{cases},$$

for some parameter c satisfying

$$0 < c \leq \min\left(\pi_1, \frac{\pi_2}{2}, \ldots, \frac{\pi_6}{2}, \pi_7\right).$$

Thus, a choice of c and taking π from the data pins down the full transition matrix, P. c is closely related to the persistence of the process: As $c \to 0$, states are permanent. For higher values of c, the probability of changing state is higher.

B.5 Household Consumption Policy Functions

In this Appendix, I show how the policy functions for households depend on their return. To be specific, the policy function for consumption in the ergodic steady state can be written as[21]

$$c_{i,t} = c(a_{i,t-1}, e^z_{i,t}, e^r_{i,t}).$$

Instead of plotting directly the policy functions, I plot a more easily interpretable object: The MPC. I define the MPC as the marginal increase in consumption from a

21. The subscript t refers to variables for individual households changing even in the ergodic steady state. Only aggregate variables are fixed at their steady state values. This explains also why there is no subscript t on $c(\cdot)$.

marginal increase in cash-on-hand:

$$\text{mpc}_{i,t}(a_{i,t-1}, e^z_{i,t}, e^r_{i,t}) = \frac{\partial c(a_{i,t-1}, e^z_{i,t}, e^r_{i,t})}{\partial a_{i,t-1}} \frac{1}{1 + (1 - \tau_t) r^a_{i,t}}.$$

The factor $(1 + (1 - \tau_t) r^a_{i,t})^{-1}$ simply adjusts for the factor that a unit increase in wealth increases cash-on-hand not by a unit, but by $1 + (1 - \tau_t) r^a_{i,t}$. This ensures that the MPC is between 0 and 1.

I plot the MPC policy function in Figure A.4. In particular, I fix a value of $e^z_{i,t}$ and then plot consumption as a function of wealth for the 7 different values of $e^r_{i,t}$ and the corresponding returns. The figure shows that households with higher returns have a much lower MPC, instead saving more of income windfalls. This is what creates scale dependence.

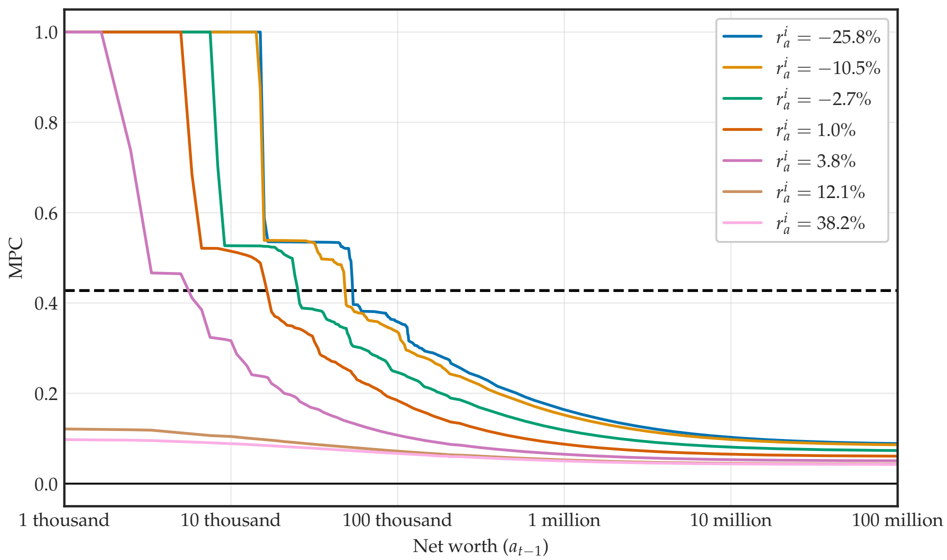

Figure A.4: MPC Policy Functions

Note: The figure shows the MPC policy function in the ergodic steady state for the model with heterogeneous returns. In particular, I fix a value of $e^z_{i,t}$ and plot consumption as a function of wealth for different values of returns.

B.6 Income and Earnings Distributions

Figure A.5 shows the Lorenz curve for earnings in the model and the data. Figure A.6 shows the Lorenz curve for earnings in the model and the data.

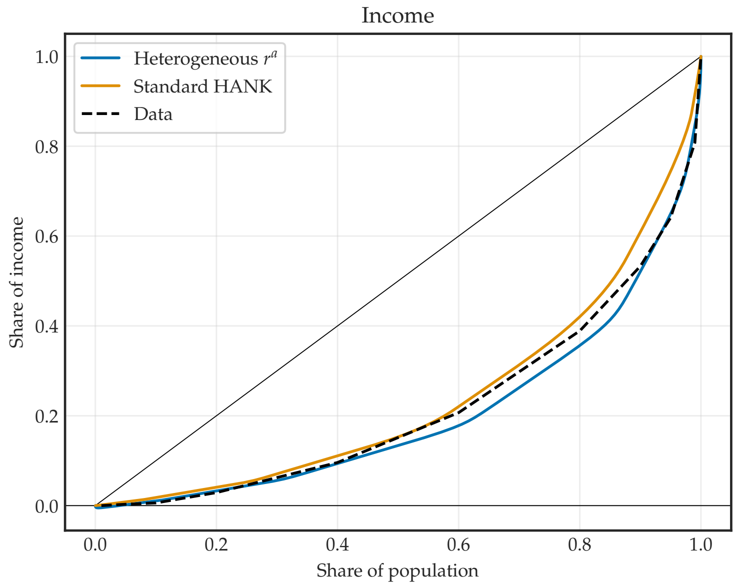

Figure A.5: Lorenz Curves for Income in the Models and the Data

Note: The figure shows the Lorenz curves of income in both models and the 2019 SCF.

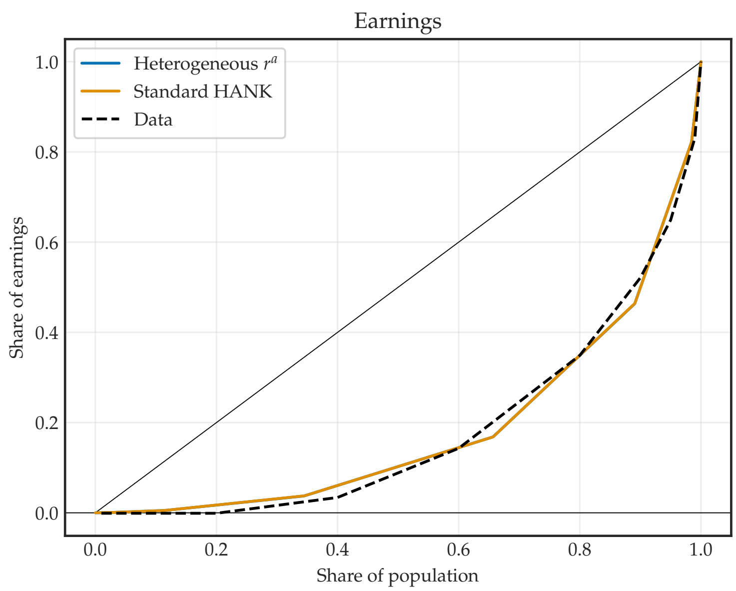

Figure A.6: Lorenz Curves for Earnings in the Models and the Data

Note: The figure shows the Lorenz curves of earnings in both models and the 2019 SCF.

B.7 The Concentration of Household Variables

One feature of the distributions of household variables in the data is that they follow an ordering of concentration. In particular:

$$g(c_{i,t}) < g(a_{i,t}) < g(x_{i,t}), \tag{23}$$

where $g(\cdot)$ denotes a measure of concentration (inequality) like the Gini index or the Pareto tail index. Gaillard et al. (2023) studies this in a standard heterogeneous agents model with common returns and finds that the tail index of all 3 variables is the same, in contrast with the data.[22]

How does my model do regarding the ranking in (23)? To study this, I report top shares of the 3 variables in Table A.2. I find that my model satisfies the ranking: Consumption is the most equal, capital income is the least equal, and wealth is somewhere in between.

Variable	Top 5%	Top 1%	Top 0.1%	Top 0.01%
Capital income	87%	55%	23%	9%
Wealth	60%	33%	13%	5%
Consumption	26%	11%	4%	1%

Table A.2: The Concentration of Household Variables

Note: The table shows top shares of selected household variables in the model with heterogeneous returns.

B.8 Additional IRFs to Monetary Policy

B.9 Proof of Proposition 1

The sequence space consumption function can be written as

$$C_t = C_t(\{Z_s\}_{s=0}^{\infty}, \{r_s^a\}_{s=0}^{\infty}, \{\tau_s\}_{s=0}^{\infty}, \{T_s\}_{s=0}^{\infty}).$$

22. I do not consider earnings here, as it has a simple 7-point discrete distribution implied by the Markov chain.

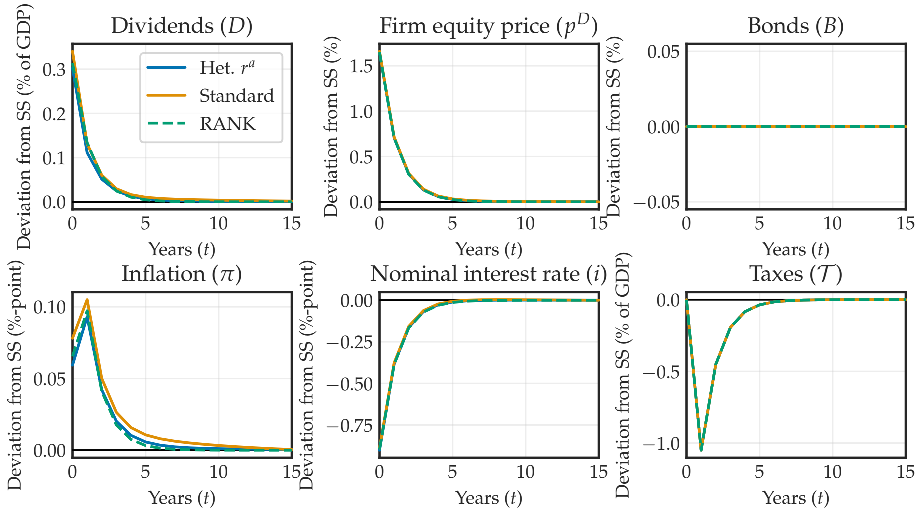

Figure A.7: Additional Aggregate Effects of Monetary Policy

Note: See Figure 12.

Linearizing this in the sequence space gives

$$d\boldsymbol{C} = \frac{\partial \boldsymbol{C}}{\partial \boldsymbol{Z}} d\boldsymbol{Z} + \frac{\partial \boldsymbol{C}}{\partial \boldsymbol{r}^{\boldsymbol{a}}} d\boldsymbol{r}^{\boldsymbol{a}} + \frac{\partial \boldsymbol{C}}{\partial \boldsymbol{\tau}} d\boldsymbol{\tau},$$

ignoring changes in lump-sum transfers, T_t, as I am considering a monetary policy shock. The sequence Jacobians of this function are then the partial derivative of the consumption function at time t with respect to one of the inputs at time s. These are presented by the matrices $\frac{\partial \boldsymbol{C}}{\partial \boldsymbol{Z}}$, $\frac{\partial \boldsymbol{C}}{\partial \boldsymbol{r}^{\boldsymbol{a}}}$, and $\frac{\partial \boldsymbol{C}}{\partial \boldsymbol{\tau}}$. Figure A.8 shows selected columns of the sequence space Jacobians for Z_s (the i-MPC matrix) and r^a_s for the model with heterogeneous returns and the standard HANK model.

To proceed with the proof, note that

$$d\tilde{\boldsymbol{r}}^{\boldsymbol{a}} = \boldsymbol{L} d\boldsymbol{r} + \iota d r^a_0,$$

where $\boldsymbol{L}$ is the lag operator and $\iota = (1, 0, 0, \dots)'$. Note also that

$$d\tilde{r}^a_t = d r^a_t + d\mathrm{Cov}\left(r^a_{i,t}, \frac{a_{i,t-1}}{A_{t-1}}\right),$$

Combining these two expressions yields

$$d\boldsymbol{r}^{\boldsymbol{a}} = \boldsymbol{L} d\boldsymbol{r} + \iota d\tilde{r}^a_0 - d\mathrm{Cov}\left(r^a_i, \frac{a_{i,-1}}{A_{-1}}\right),$$

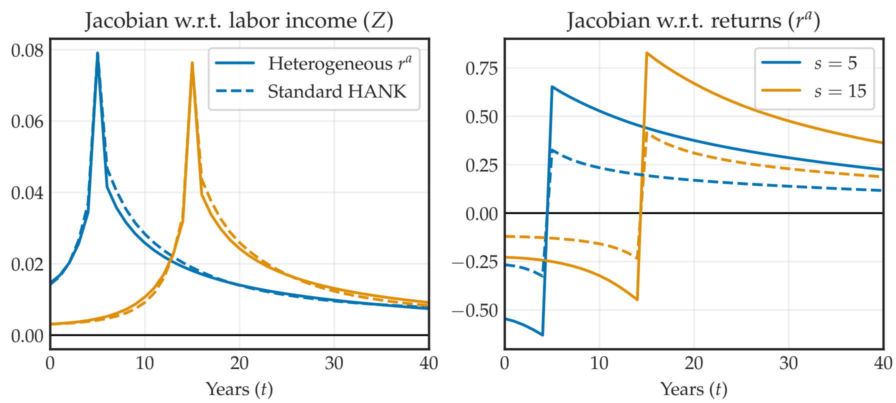

Figure A.8: Sequence Space Jacobians of Household Problem

Note: The figure shows two columns ($s = 5$ and $s = 15$) of the sequence space Jacobians of the household problem for real labor income and returns, i.e. $\partial C_t / \partial Z_s$ and $\partial C_t / \partial r_t^a$. It does so for both the model with heterogeneous returns and the standard HANK model.

Inserting this into the consumption function yields

$$d\boldsymbol{C} = \frac{\partial \boldsymbol{C}}{\partial \boldsymbol{r^a}} \boldsymbol{L} d\boldsymbol{r} + \frac{\partial \boldsymbol{C}}{\partial \boldsymbol{Z}} d\boldsymbol{Z} + \frac{\partial \boldsymbol{C}}{\partial \boldsymbol{\tau}} d\boldsymbol{\tau} + \frac{1}{A_{ss}} \frac{\partial \boldsymbol{C}}{\partial \boldsymbol{r^a}} \iota dX_0 - \frac{\partial \boldsymbol{C}}{\partial \boldsymbol{r^a}} d\text{Cov}\left(r_{i,t}^a, \frac{a_{i,t-1}}{A_{t-1}} \right).$$

Re-writing slightly and defining Jacobians, I arrive at the equation:

$$d\boldsymbol{C} = \underbrace{\boldsymbol{M}^r d\boldsymbol{r}}_{\text{1. Direct}} + \underbrace{\boldsymbol{M}^Z d\boldsymbol{Z}}_{\text{2. Labor}} + \underbrace{\boldsymbol{M}^\tau d\boldsymbol{\tau}}_{\text{3. Taxes}} + \underbrace{\boldsymbol{M}^X dX_0}_{\text{4. Revaluation}} + \underbrace{\boldsymbol{M}^{\text{Cov}} d\text{Cov}\left(r_i^a, \frac{a_{i,-1}}{A_{-1}} \right)}_{\text{5. Redistribution}},$$

where

$$\boldsymbol{M}^r \equiv \frac{\partial \boldsymbol{C}}{\partial \boldsymbol{r^a}} \boldsymbol{L}$$
$$\boldsymbol{M}^Z \equiv \frac{\partial \boldsymbol{C}}{\partial \boldsymbol{Z}}$$
$$\boldsymbol{M}^\tau \equiv \frac{\partial \boldsymbol{C}}{\partial \boldsymbol{\tau}}$$
$$\boldsymbol{M}^X \equiv \frac{1}{A_{ss}} \frac{\partial \boldsymbol{C}}{\partial \boldsymbol{r^a}} \iota$$
$$\boldsymbol{M}^{\text{cov}} \equiv -\frac{\partial \boldsymbol{C}}{\partial \boldsymbol{r^a}}.$$

B.10 Relation to Werning (2015)

Werning (2015) shows that the response of aggregate consumption to real interest rate changes are identical in incomplete markets and complete markets models under certain assumptions. The models in this paper do *not* satisfy this assumption—even with common returns.

I now consider simplifying my model such that it fits the setup in Werning (2015). In particular, I start with the fully calibrated model with heterogeneous returns. I then remove the government: $\tau_{ss} = B_{ss} = T_{ss} = \delta = 0$, which also implies $G_{ss} = 0$. I also remove the heterogeneous pass-through: $\beta^r_{i,t} = \beta^z_{i,t} = 1$. With this parametrization, I re-calibrate β and r_{ss}. This gives me a simplified version of the model with heterogeneous returns. I also consider a version of this simplified model without heterogeneous returns by setting $e^r_{i,t} = 0$ and re-calibrating β. Additionally, I consider a representative agent (RANK) version of the model, which is obtained when also $e^z_{i,t} = 1$ and $\underline{a} \to -\infty$.

Figure A.9 shows the response of consumption in all 3 models. Note first that the response of consumption is completely identical in the standard HANK and RANK models. This is exactly Werning (2015). Note additionally that consumption in the heterogeneous returns model is *almost* the same, but not quite. The difference is small enough to not be economically meaningful but large enough to not be an artifact of the numerical solution.

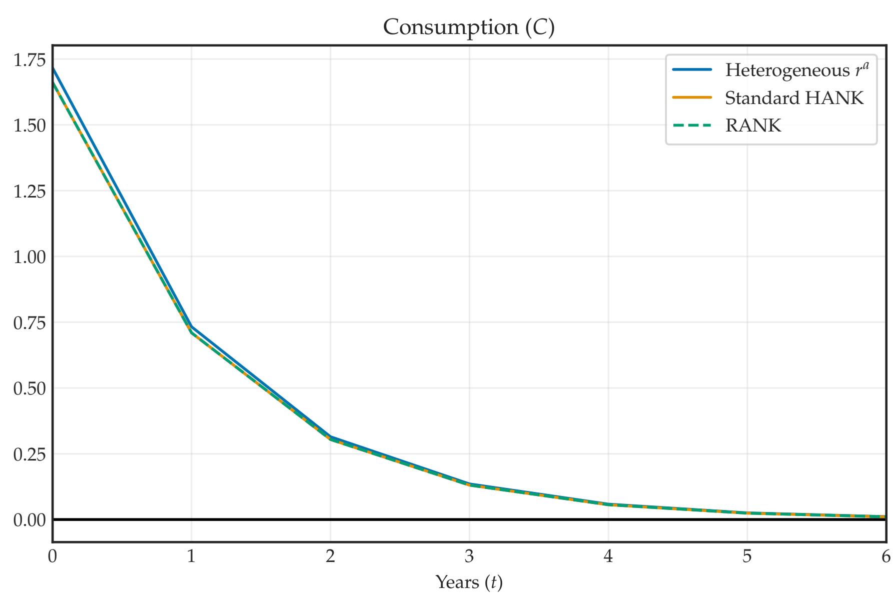

Figure A.9: IRFs to a Monetary Policy Shock in Simplified Models

Note: The figure shows impulse response functions to a 1 percentage point drop in the real interest rate in simplified versions of 3 models. The x-axis shows years after the shock.

To see the math, let $\tilde{c}_{i,t}$ denote the consumption of household i at time t in the absence of the monetary policy shock, i.e. when aggregate variables are in steady state. Werning (2015) shows that

$$\frac{c_{i,t}}{\tilde{c}_{i,t}} = \frac{C_t^{\mathrm{RA}}}{C_{ss}^{\mathrm{RA}}},$$

where C_t^{RA} is consumption in the RANK model faced with the same monetary policy shock. This does not hold with heterogeneous returns. Instead, it holds that

$$\frac{c_{i,t}}{\tilde{c}_{i,t}} = s_{i,t}\frac{C_t^{\mathrm{RA}}}{C_{ss}^{\mathrm{RA}}},$$

where $s_{i,t}$ is simply defined as the scaling factor such that this holds. Aggregate consumption is then

$$C_t = \int c_{i,t}\, di = \int s_{i,t}\frac{C_t^{\mathrm{RA}}}{C_{ss}^{\mathrm{RA}}}\tilde{c}_{i,t}\, di = C_t^{\mathrm{RA}} \int s_{i,t}\tilde{c}_{i,t} = C_t^{\mathrm{RA}}\left[S_t + \mathrm{Cov}(s_{i,t}, \tilde{c}_{i,t})\right],$$

where $S_t = \int s_{i,t}\, di$. Under which assumptions is $C_t = C_t^{\mathrm{RA}}$? The assumptions in Werning (2015) imply that $s_{i,t} = 1$ for all i and t. However, while $s_{i,t} = 1$ is *sufficient*, it is not *necessary*. For instance, one obtains $C_t = C_t^{\mathrm{RA}}$ despite $s_{i,t} \neq 1$ if both (i) $S_t = 1$, and (ii) $\mathrm{Cov}(s_{i,t}, \tilde{c}_{i,t}) = 0$. This is (approximately) the case in the model with heterogeneous returns. Intuitively, the consumption of households *is* affected differently by the monetary policy shock, but the way it is affected is (almost) independent of initial consumption, so all the changes wash out in the aggregate.

Figure A.10 illustrates this. The first row of the figure shows the change of consumption across households for both the standard HANK model and the model with heterogeneous returns. Each dot corresponds to one of the discretized states in the model. In the standard HANK model, each household's consumption rises in the same proportion, i.e. $s_{i,t} = 1$ as in Werning (2015). In the model with heterogeneous returns, there is a difference in how much consumption increases, with consumption increases ranging from around 1.2% to 2.4%. This clearly shows that the Werning (2015) result does *not* apply in the model with heterogeneous returns. However, the figure also shows that the changes in consumption are (almost) uncorrelated with the initial level of consumption. The second row makes this point, dividing the distribution of consumption into 20 quantiles of 5%. In this case, consumption increases by the same proportion for households at different points in the consumption distribution. In other words, $\mathrm{Cov}(s_{i,t}, \tilde{c}_{i,t}) \approx 0$.

Table A.3 shows the decomposition of consumption in response to monetary

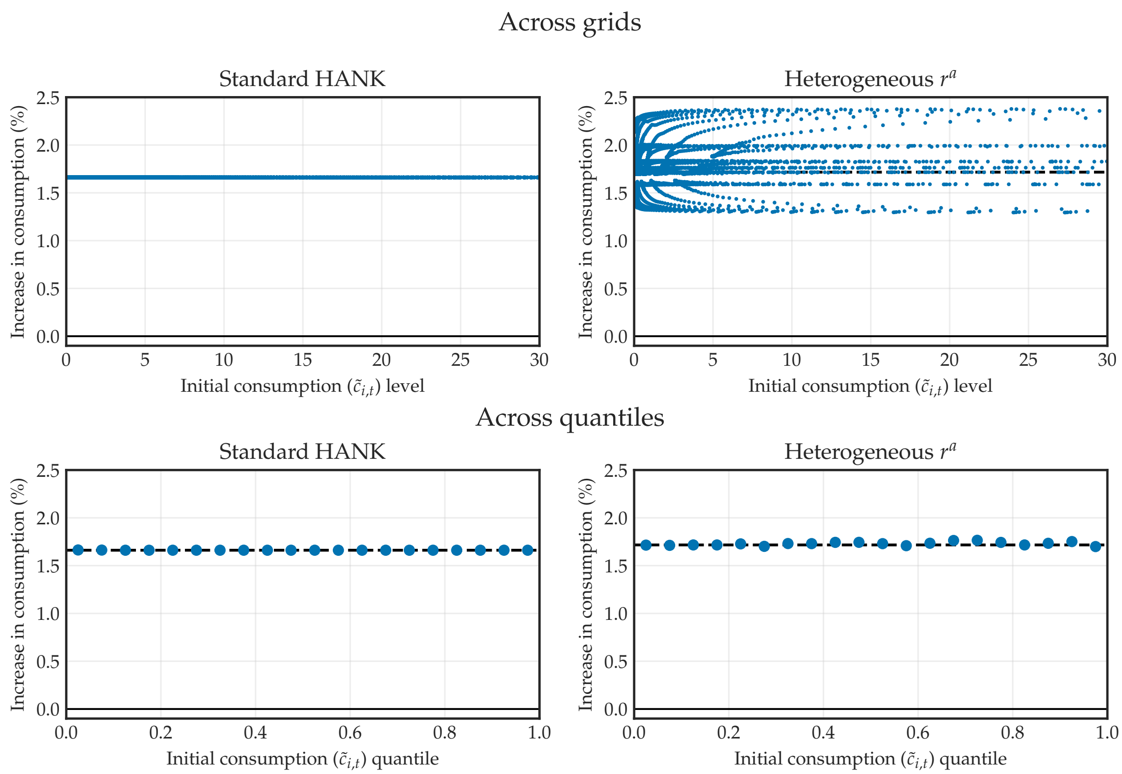

Figure A.10: Heterogeneity in Consumption Response

Note: The figure shows the percent change in consumption to a monetary policy shock, $100(\frac{c_{i,t}}{\tilde{c}_{i,t}} - 1)$, at different points in the initial distribution of consumption in two models. The first row shows the response across all grid points. The second row shows the average response within 20 groups based on the initial consumption level.

policy like in Table 5 in the models consistent with Werning (2015). The table shows that the transmission of monetary policy is very similar in this version of the model, consistent with Figures A.9 and A.10. Additionally, the channels are very similar, which is in contrast with Table 5. In particular, the re-distribution is zero in both models. This is because $\beta^r_{i,t} = 1$ in this version.

B.11 Additional IRFs to Fiscal Policy

B.12 Comparison to the Empirical Literature

In this Appendix, I compare the distributional effects of monetary policy in my model to the empirical literature. First, I compare to McKay and Wolf (2023a). They estimate the effects of expansionary monetary policy on consumption across the wealth distribution. They find that the effect on consumption is increasing in the level of wealth beyond the first quintile. The consumption increase at the top is largely due to stocks. I do a similar exercise in Figure A.12. The figure shows that the change

	Heterogeneous r^a	Standard HANK
1. Direct	0.60	0.57
2. Labor	0.29	0.21
3. Taxes	0.00	0.00
4. Revaluation	0.81	0.86
5. Redistribution	-0.00	0.00
Total	1.69	1.64

Table A.3: Decomposition of Consumption in Response to Monetary Policy: Werning (2015) Case

Note: The table shows a decomposition of the response of consumption on impact to the monetary policy shock in both models adjusted to be consistent with Werning (2015).

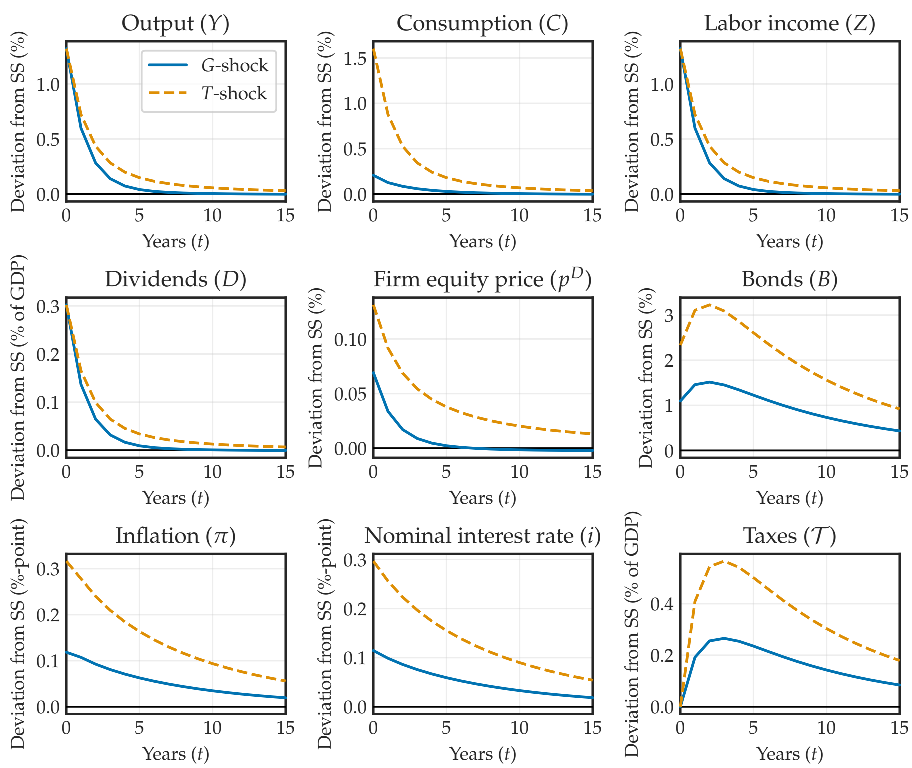

Figure A.11: Additional Aggregate Effects of Fiscal Policy

Note: The figure shows impulse response functions of output to fiscal policy shocks with persistence 0.43. The size of the shock is chosen to cause the same change in output as with the monetary policy shock in Figure 12—this is simply a scaling of the size to make comparison across shocks easier. The x-axis shows years after the shock.

in consumption is U-shaped, with the poorest and richest gaining the most. This is largely consistent with McKay and Wolf (2023a), though the magnitudes are slightly larger in my model.

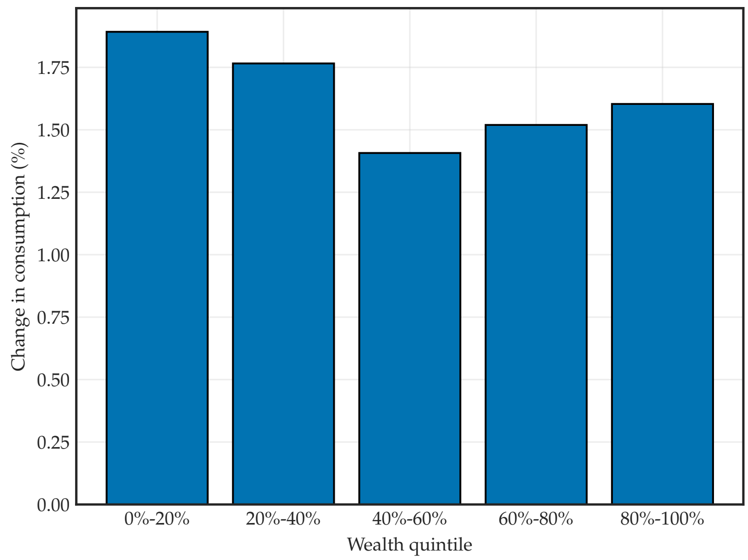

Figure A.12: Consumption Effect Across the Wealth Distribution

Note: The figure shows the percent change in consumption on impact across quintiles of the wealth distribution.

Next, I compare to Andersen et al. (2023), who studies the effect of expansionary monetary policy on households in Denmark. Their headline result is a clear income gradient of monetary policy: Households with low income (poor households) lose income and households with high income (rich households) households gain income *relative to the median income change*. The two-year effects range from around a -2% change in income for the poorest and around 4% for the richest, with the median effect being 0% by definition. I do a similar exercise in Figure A.13. The figure shows a clear income gradient as in Andersen et al. (2023).

B.13 Proof of Proposition 2

The change in total income for household i at time $t = 0$ is given by

$$d\psi_{i,0} = dx_{i,0} + dz_{i,0} + d\omega_{i,0}.$$

The change in income going to this household is then

$$\frac{d\psi_{i,0}}{d\Psi_0} = \frac{dx_{i,0} + dz_{i,0} + d\omega_{i,0}}{d\Psi_0} = \frac{dx_{i,0}}{d\Psi_0} + \frac{dz_{i,0}}{d\Psi_0} + \frac{d\omega_{i,0}}{d\Psi_0}.$$

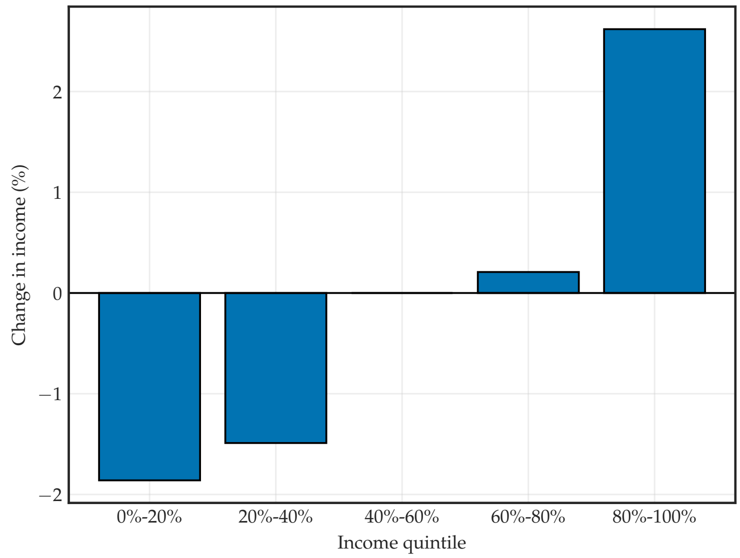

Figure A.13: Income Effect Across the Wealth Distribution

Note: The figure shows the percent change in income on impact across quintiles of the income distribution. The figure is conditioned on positive income.

Furthermore, assuming $dX_0 \neq 0$, $dZ_0 \neq 0$, and $d\Omega_0 \neq 0$:

$$\frac{d\psi_{i,0}}{d\Psi_0} = \frac{dX_0}{d\Psi_0}\frac{dx_{i,0}}{dX_0} + \frac{dZ_0}{d\Psi_0}\frac{dz_{i,0}}{dZ_0} + \frac{d\Omega_0}{d\Psi_0}\frac{d\omega_{i,0}}{d\Omega_0}.$$

Write now $x_{i,0} = r^a_{i,0}a_{i,-1}$, such that $dx_{i,0} = a_{i,-1}dr^a_{i,0}$. Thus,

$$\frac{d\psi_{i,0}}{d\Psi_0} = \frac{dX_0}{d\Psi_0}\frac{a_{i,-1}}{A_{t-1}}\frac{dr^a_{i,0}}{d\tilde{r}^a_{i,0}} + \frac{dZ_0}{d\Psi_0}\frac{z_{i,0}}{Z_0}\frac{\frac{dz_{i,0}}{z_{i,0}}}{\frac{dZ_0}{Z_0}} + \frac{d\Omega_0}{d\Psi_0}\frac{\omega_{i,0}}{\Omega_0}\frac{\frac{d\omega_{i,0}}{\omega_{i,0}}}{\frac{d\Omega_0}{\Omega_0}}.$$

Following the definitions of the α's and β's and neglecting the $t = 0$ subscript yields the result.

B.14 Capital Income in the Model

The gains of the rich in the model with heterogeneous returns largely go through capital income. But what drives capital income in the model? And is the magnitude of the capital income gains empirically realistic? In this Appendix, I try to answer this question. The key to answering this question is Proposition 3, which decomposes the effects of aggregate shocks on capital income.

Proposition 3. *The effect of an aggregate shock on capital income can be decomposed into two components: The firm equity price and dividends. This decomposition can be written as* *follows:*

$$dx_0 = dp_0 + dD_0. \tag{24}$$

These components are given by:

$$D_0 = \frac{\mu - 1}{\mu} Y_0,$$
$$p_0 = \frac{1}{1 + r_0} \frac{\mu - 1}{\mu} Y_1 + \frac{1}{(1 + r_0)(1 + r_1)} \frac{\mu - 1}{\mu} Y_2 + \ldots$$