Heterogeneous Risk Exposure and the Dynamics of Wealth Inequality

Riccardo A. Cioffi *

Princeton University

December 22, 2021

Job Market Paper†

[Click here for latest version]

Abstract

In this paper I argue that the dynamics of wealth inequality are largely driven by heterogeneous exposure to aggregate risk in asset returns. I propose a quantitative model of households' optimal portfolio choice that builds on evidence that housing is a necessary good. The model replicates households' portfolio heterogeneity along the wealth distribution: just like in the data, as households get wealthier they shift their portfolios away from safe assets, first towards housing, and then towards equity. Because households in different parts of the wealth distribution are exposed to different sources of aggregate risk, the model has strong implications for the evolution of inequality. In particular, temporary shocks in equity returns have large and persistent effects on top wealth shares. A key implication is that the observed rise in U.S. wealth inequality was mostly due to abnormal equity returns and it is therefore expected to revert back to lower levels.

JEL: D14, D31, E21, E44, E32, G11, G51

*Contact: rcioffi@princeton.edu

†I am incredibly thankful to my advisors Gianluca Violante, Richard Rogerson, Moritz Lenel, and Benjamin Moll for their continued guidance and support. I would also like to thank Andrea Ajello, Stefano Baratuche, Alberto Bisin, Andreas Fagereng, Martin Holm, Jonathan Payne, Elisa Rubbo, George Sorg-Langhans, Francisco Vazquez-Grande, Maximilian Vogler, and Christian Wolf for their thoughtful comments and discussions, as well as seminar participants at the Federal Reserve Board of Governors, the Federal Reserve Bank of St. Louis, Princeton University, and Statistics Norway.

1. Introduction

One of the key issues in macroeconomics and finance is understanding what explains the high degree of concentration in the distribution of wealth of modern economies. In particular, while most developed countries have extremely high degrees of wealth inequality and have experienced large increases in it since the second half of the twentieth century, we are still missing a theory that can explain both the high level and fast dynamics of inequality.

In this paper I develop a framework that can account for the observed dynamics of wealth inequality based on heterogeneous exposure to aggregate risk in asset returns. The connection between asset returns and wealth inequality is evidently composed of two distinct (albeit interrelated) parts: on the one hand, in the presence of systematic differences in portfolio composition along the wealth distribution, changes in returns induce changes in wealth inequality; on the other hand, by affecting their demand for assets, changes in households' wealth holdings have the potential to affect asset prices in return. Hereby I directly tackle the first half of such a connection: I build a model of households' optimal portfolio choice along the wealth distribution and show that, once paired with a realistic process for asset returns, it generates dynamics of wealth inequality consistent with the ones observed in the data. In particular, after feeding the realized sequence of asset returns into the model, I show that it can replicate all of the observed increase in U.S. top wealth shares since the 1980s.[1]

Obviously enough, one of the key inputs in any theory of the role of heterogeneous exposure to aggregate risk for wealth inequality is a realistic characterization of households' portfolios. In order to achieve such goal I therefore solve a heterogeneous-agent, partial equilibrium model of optimal portfolio choice with aggregate risk featuring a rich menu of assets which is meant to capture the main properties of households' portfolio shares as a function of wealth. Hence, by matching portfolio heterogeneity, the model is capable of replicating both the high level of wealth inequality – driven by differences in total returns to wealth – and the response of wealth inequality to movements in asset returns. I then use the model to run a series of counterfactual experiments that highlights the role of shocks to asset returns for the evolution of households' wealth.

The model has several implications for the dynamics of wealth inequality: First, I find that shocks to equity returns have large and persistent effects on wealth inequality. In particular, a one standard deviation increase in equity returns implies an increase in the top 10% wealth share of about 1 percentage point, an effect approximately 50% larger than an equally sized increase in housing returns. Second, I show that whether changes in returns are assumed to be permanent or temporary has extremely different implications for the evolution of wealth inequality: the long-run effect of a sequence of temporary shocks is in fact about eight times larger than that of a corresponding permanent change in returns. Third, by feeding the realized sequence of returns into the model, I

[1] In a companion paper, Cioffi (2021), I address the other half of the mechanism and look at how wealth inequality influences equilibrium asset prices and how this in turn changes our understanding of policy.

find that it is capable of replicating the observed increase in U.S. top wealth shares since the 1980s and that most of the increase was driven by high returns to equity in the late 1990s and early 2000s. Hence I use the model to conclude that, to the extent that changes in returns over the last 40 years were temporary, future inequality is expected to slowly revert back to its long-run average.

In general, differences in wealth accumulation can come from at least two sources: heterogeneity in rates of return, and heterogeneity in savings. Benchmark models of wealth inequality (e.g., Aiyagari 1994) assume that all households face a constant rate of return to total wealth. However, an old result by Kesten (1973) – later revisited in the context of wealth accumulation by Benhabib and Bisin (2018) – tells us that, in models without any heterogeneity in rates of return, tail inequality in wealth will exactly mirror tail inequality in income, a feature that is simply not true in the data.[2] Then, to generate wealth inequality above and beyond income inequality, a mechanism that is often used is to introduce idiosyncratic risk in rates of return. In fact, by generating *ex-post* positive correlation between wealth and returns, idiosyncratic return risk acts as an amplifying force for wealth inequality which allows the distribution of wealth to decouple from that of earnings. Nonetheless, Gabaix et al. (2016) show that, compared to the data, a simple model of stochastic rates of return does not generate fast-enough movements in wealth inequality. To generate wealth dynamics that are closer to the observed ones, they then propose two mechanisms that make the rate of return explicitly depend on either individual characteristics (type dependence), or on wealth (scale dependence). Also in light of the recent empirical work suggesting that these are indeed features of the data (e.g. Bach, Calvet, and Sodini 2020; Fagereng et al. 2020, among others), the new generation of wealth-inequality models has taken exactly this approach and usually features some form of either type or scale dependence (e.g. Hubmer, Krusell, and Smith 2021; Xavier 2020). For the purposes of our discussion, notice however that, in all models of idiosyncratic returns just mentioned, movements in inequality over time can only be generated by changes in the economy's fundamentals (such as, for instance, changes in the dispersion of returns).

More often than not, these models do not take a stand on what is the underlying source of the heterogeneity in rates of return to total wealth. However, for most households, such heterogeneity is actually the result of the interaction of portfolio choice and differences in returns across assets. It should in fact be apparent that even in the absence of idiosyncratic return risk, as long as different agents hold different portfolios, rates of return can still be very different along the wealth distribution. Consider for instance an (empirically realistic) economy in which everyone has access to the same assets and in which households, as they get richer, gradually shift their portfolios away from low-risk/low-return assets (e.g. cash and bonds) to high-risk/high-return ones (e.g. stocks). On the one hand, this generates an *ex-ante* positive correlation between wealth and returns to total wealth which – just as in the idiosyncratic risk example – acts as an amplification mechanism for the level of wealth inequality. On the other hand, because households hold the same assets, returns to wealth also move together over time which – differently from an economy with only idiosyncratic risk –

[2] Both across countries and over time wealth has been found to be always much more concentrated than income.

generates fluctuations in wealth inequality. For example in the economy just described, with wealthier households having a larger exposure to the stock market than poorer ones, wealth inequality will increase during stock-market booms and decrease during crashes, which is exactly what Kuhn, Schularick, and Steins (2020) show to be the case for the post-war U.S. economy.[3]

In this paper I specifically highlight this mechanism: in section 2, I build a partial equilibrium model of wealth inequality that carefully accounts for the role of portfolio choice and asset returns. Importantly, I show that the dual role of housing as a risky investment and a necessary good is crucial to generate the right schedule of portfolio shares: the fact that, in the data, the expenditure share of housing is declining in total consumption indicates that housing is a necessary good and, consequently, that preferences are non-homothetic. Such non-homotheticity in the utility aggregator of housing services and non-housing consumption causes both the elasticity of intertemporal substitution (EIS) and relative risk aversion (RRA) to change with wealth. In particular, RRA decreases with wealth and the EIS increases with it. Hence, by effectively introducing decreasing relative risk aversion, the model generates an optimal share of equity that is increasing in wealth.[4]

In section 3 I then show that the model is in fact capable of replicating the observed heterogeneity in households' portfolios along the wealth distribution: poor households mostly hold low-risk/low-return bonds, the middle class is heavily invested in housing, and the top of the wealth distribution holds most of the equity in the economy. Such portfolio composition generates exactly the sort of scale dependence in returns to total wealth that is necessary to generate a high level of wealth inequality. On top of that, the fact that portfolio shares are different along the wealth distribution, also implies wealth inequality will respond differently to different changes in asset returns.

In section 4, I in fact show that the model generates significant variation in wealth inequality over time with a standard deviation for the Gini coefficient of 0.05. The main reason why the model generates such large dispersion in wealth inequality is that movements in returns – in particular in the equity market – have large and persistent effects on top wealth shares: a one standard deviation increase in equity returns implies an increase in the top 10% wealth share of about 1 percentage point, an effect about 50% larger than an equally-sized housing shock. With a half-life of approximately 40 years, the effect of movements in equity returns are extremely persistent and significantly contribute to generate large swings in wealth inequality. Importantly, I also show that whether we assume changes in returns are temporary or permanent has extremely different implications for wealth in-

[3]In the above discussion, I have been agnostic about where the changes in asset returns are originating from. This is consistent with the approach I take in the rest of the paper, which treats returns as completely exogenous objects. In truth, however, returns are general equilibrium objects that might in turn depend – among other things – on the distribution of wealth, therefore closing the general equilibrium loop between households' portfolio choices, wealth holdings, and asset returns. Analyzing such general equilibrium feedback is beyond the scope of this paper, and is instead tackled in Cioffi (2021).

[4]The basic intuition is that, if preferences are non-homothetic, the share of luxury goods in the consumption basket is increasing in wealth. Because households care less about fluctuations in consumption when they consume a higher share of luxury goods, RRA declines with wealth and EIS increases with wealth; see Browning and Crossley (2000) for an early formal analysis of this argument.

equality: in particular, a continuous sequence of unexpected shocks giving rise to an additional 1% excess return to equity in each period implies an ever rising top 10% wealth share with a total increase of about 5 percentage points; on the other hand, if the exact same sequence of returns were to arise from a single, permanent shock, the top 10% wealth share would actually decrease in the short run and only increase by about 0.6 percentage points in the long run. Because the model generates fluctuations in wealth inequality even in the absence of any change in fundamentals (as opposed to models of idiosyncratic return risk), I then ask how much of the observed increase in U.S. wealth inequality since the 1980s can be explained by fluctuations in realized returns alone: I feed the observed sequence of returns into the model and show that the evolution of top wealth shares closely tracks the data. Then, by running a series of counterfactual experiments I also show that: First, large positive changes in inequality are a perfectly reasonable outcome of a model in which the wealth distribution moves over time if initial inequality is below average. The model-implied probability of observing the data (conditional on initial inequality being less than in 1986) is as high as 14.3%. Second, the rise in U.S. top wealth shares was mostly the result of abnormal equity returns in the late 90s and early 2000s. Third, to the extent that such abnormal returns were temporary, inequality is expected to slowly revert back to its long-run average.

One of this paper's main conclusions is therefore that, given return fluctuations have such important implications for the evolution of wealth, if we want to have a better understanding of wealth-inequality dynamics we should also have a good model of price determination. In this paper I take a very reduced-form approach to the evolution of asset returns, which are simply assumed to follow an exogenous process calibrated to fit some data moments. I do so for two reasons: First, by keeping the return process fixed I can directly focus on the role of shocks to returns without having to isolate it from the endogenous response of prices to changes in the wealth distribution. Second, solving for prices in equilibrium would also require me to take a stand on how to generate a high equity premium. While the model does include features that are likely to help in that respect, this would involve a whole different set of challenges that are just beyond the scope of this paper (I do tackle some of these issues in companion work, Cioffi 2021).

Compared to the rest of the literature, in this paper I therefore reach a very different conclusion about the increase in U.S. wealth inequality observed in the last 40 years; namely that the observed history might just have been the result of chance, rather than of structural changes. Such a sharp rise in inequality is in fact perfectly compatible with an unchanging economy in which, among the many possible realizations of asset returns, the observed one happened to be especially favorable to the portfolios of the rich.

Relation to Literature — This paper is connected to several strands of literature: First, it directly relates to models of optimal portfolio choice in the presence of non-homothetic preferences and housing. Most of the literature on portfolio choice focuses on how portfolio shares vary over the life-cycle. One important reason for why this is the case is that in the classic Merton

(1971) model of portfolio choice with CRRA preferences, the optimal equity share is independent of wealth (while age, often mainly through its effect on human capital, directly affects households' portfolio decisions). Instead, I hereby emphasize the direct role of wealth. In my model, non-homothetic preferences in fact imply that wealth directly affects portfolio shares through its effect on total consumption expenditure and therefore on expenditure shares. In this sense, my work is more closely related to papers of non-homothetic preferences exhibiting decreasing RRA such as Meeuwis (2020) and Wachter and Yogo (2010). Using a representative sample of U.S. retirement investors Meeuwis (2020) finds substantial evidence in favor of decreasing RRA and a significant degree of non-homotheticity in risk preferences. This is consistent with results in Wachter and Yogo (2010) who are able to replicate the fact that the equity share is increasing in wealth by incorporating non-homothetic utility over necessary and luxury goods. Relative to Wachter and Yogo (2010), my model also features a form of non-homotheticity that generates decreasing RRA. One notable difference is that, while their model does not feature housing (neither as a consumption good nor as an investment asset), I assume the non-homotheticity is directly driven by housing. This is motivated by the novel observation that, in the data, the expenditure share of housing is declining in total consumption – implying housing is in fact a necessary good; a point that, to the best of my knowledge, has not been made before.

Due to the effect of non-homothetic preferences on RRA and EIS my paper is also loosely related to papers of preference heterogeneity such as Gomes and Michaelides (2008); Gomez (2019); Guvenen (2009) and Vestman (2019). These papers often feature two different types of investors: a high risk-aversion, wealth-poor agent and a low risk-aversion wealthy agent. In this paper, on the other hand, because I assume agents are identical *ex-ante*, preference heterogeneity is completely endogenous. My model can therefore also be interpreted as a microfoundation for such form of preference heterogeneity based on the necessary-good nature of housing.

Aside from being a consumption good, housing is obviously also an investment asset, which also ties my model to the large literature on portfolio choice in the presence of housing. An exhaustive list of such literature would obviously require a paper on its own (which already exists in Piazzesi and Schneider 2016), rather it is here worth mentioning that in this paper I do not strive to achieve a comprehensive model of housing, but rather of the interaction between households' portfolio shares and wealth inequality. Because of this I take some shortcuts in the characterization of the housing problem and the model here is therefore more closely related to early papers such as Cocco (2005); Flavin and Nakagawa (2008); Flavin and Yamashita (2002); Piazzesi, Schneider, and Tuzel (2007).

Second, my paper connects to the recent literature on return heterogeneity as a key driver of wealth concentration. Relative to already existing work (see the empirical contributions in Bach, Calvet, and Sodini 2020; Fagereng et al. 2020; and the model-based analyses in Gabaix et al. 2016; Hubmer, Krusell, and Smith 2021; Xavier 2020), I specifically focus on the role of households' *optimal* portfolio choice as a driver of heterogeneity in returns. As quickly mentioned before, Gabaix et al. (2016) provides the first theoretical analysis of the role of return heterogeneity for the dynamics of wealth inequality: although their main argument focuses on inequality in incomes, they also show that a simple model of stochastic returns is not capable of matching the fast dynamics of wealth inequality. They then show that allowing for scale dependence in rates of return can overcome this shortcoming of models of idiosyncratic return risk. While my paper does feature a scale-dependent component to returns (wealthier household investing in higher-return assets), I show that the presence of aggregate risk alone also allows us to match the dynamics of wealth inequality by making the whole wealth distribution stochastic. Xavier (2020) instead solves a model of type dependence and shows that return heterogeneity is necessary to match the wealth shares of the top 10%. While I focus more closely on the dynamics of wealth inequality, my paper also corroborates her result in a model of scale dependence in which rates of return are directly tied to households' optimal portfolio choices.

Finally, my paper also speaks to the literature that links the evolution of the wealth distribution to movements in asset returns. My findings are in fact broadly consistent with empirical results relating changes in wealth inequality to the succession of booms and busts in equity and housing markets (Kuhn, Schularick, and Steins 2020; Martinez-Toledano 2020). More closely related, however, are papers that try to explain the observed increase in U.S. wealth inequality since the 1980s via the effect of changes in asset returns: first, from a methodological perspective, I extend the quantitative analyses in Favilukis (2013) and Hubmer, Krusell, and Smith (2021) who find that, even though movements in asset returns do generate variation in wealth inequality, they are not in and of themselves sufficient to explain such a large increase as observed in the data; second, my paper also complements recent findings by Gomez and Gouin-Bonenfant (2020) and Greenwald et al. (2021) who argue that the decline in real rates has been a key driver for the increase in top wealth shares.

In a similar model without housing Favilukis (2013) shows that, in order to explain the increase in wealth inequality one needs to account for the combined role of increased wage inequality, loosening borrowing constraints, and decreasing participation costs. However, his model ignores returns to private equity which were a substantial source of excess returns at the top of the wealth distribution. In fact Kartashova (2014) shows that, particularly after 2002, returns to private equity outpaced returns to public equity. By allowing for the role of higher private equity returns in that period, I therefore show that return movements alone can explain the increase in top wealth shares.

Hubmer, Krusell, and Smith (2021) also attempts to explain the observed variation of wealth inequality by feeding the realized sequence of asset returns into a model of portfolio heterogeneity. They find that while changes in asset returns are key to explain the U-shape of top wealth shares, they "have also dampened the increase in wealth concentration on net, in particular explaining the initial dip" (Hubmer, Krusell, and Smith 2021, p. 430). My paper differs from theirs in two key aspects: First, households' portfolio choice in their model is completely *exogenous*; that is, whenever households make a consumption-saving choice, they take the change in average rates of return as given and cannot change it by investing in different assets. Second, they assume all movements are driven by changes in expected returns rather than by realized returns. In this paper I show that temporary changes in returns have a much larger effect on inequality than permanent ones, which helps explain the difference in results.

The distinction between temporary and permanent changes in returns also differentiates my paper from Gomez and Gouin-Bonenfant (2020) and Greenwald et al. (2021): Gomez and Gouin-Bonenfant in fact claim that, by making it cheaper for entrepreneurs to raise capital, lower interest rates have supported the growth of new fortunes thereby increasing wealth inequality; Greenwald et al. on the other hand argue that a decline in interest rates causes higher (financial) wealth inequality due to wealthier households needing to save more to finance the same consumption stream. Relative to both, who start from a decline in expected returns, I instead emphasize that a large increase in wealth inequality can also be consistent with an economy in which all movements in asset returns are entirely temporary.

2. Model

In this section I present the model setup, which will later be used to quantify the effect of aggregate risk in asset returns on both the level of wealth inequality and its dynamics. The model is set in continuous-time and is a partial-equilibrium version of the standard heterogeneous-agent model in macroeconomics in which households face idiosyncratic income risk and smooth consumption over time by saving in a menu of assets (e.g. Achdou et al. 2017). However, especially when compared to the rest of the literature on wealth inequality, the model directly innovates in the richness of households' portfolio choice.

The model purpose is therefore to propose a parsimonious yet realistic characterization of households' optimal portfolio choice along the wealth distribution. In particular, we would like the model to replicate the fundamental observation that households at the bottom of the wealth distribution mostly hold safe assets, that the middle class is heavily invested in housing, and that equity is mainly held by the rich. This is achieved by a combination of features: First, to generate an increasing equity share along the wealth distribution, the model features both participation costs in the equity market and non-homotheticity in preferences. Second, to capture the main features of the housing market, I allow for both an illiquidity component and transaction costs. Finally, due to the nature of preferences, households at the bottom of the distribution have a strong preference for housing consumption; in order to avoid their housing share to be counterfactually high it is therefore important to allow for a rental housing market.

2.1. Main environment

Demographics — There is a continuum of households facing a constant death rate ζ. Population is constant and normalized to 1 so every time an agent dies she is immediately replaced by a newborn agent. Agents are born according to distribution Ψ, to be specified later.[5]

Preferences — Households have recursive inter-temporal preferences as in Duffie and Epstein (1992):

$$v_t = \mathbb{E}\left[\int_t^{\infty} f(u_s, v_s) ds\right] \tag{1}$$

where $f(u, v)$ is an aggregator of intra-temporal utility u and continuation utility v and takes the form:

$$f(u, v) = \frac{\rho + \zeta}{1 - \psi^{-1}}(1-\gamma)v \left\{ \left[\frac{u}{((1-\gamma)v)^{\frac{1}{1-\gamma}}}\right]^{1-\psi^{-1}} - 1 \right\} \tag{2}$$

where ρ is the rate of time preference, and γ and ψ govern risk aversion and elasticity of intertemporal substitution, respectively.

Intra-temporal utility u is an aggregator of housing services n and non-housing consumption c. The aggregator is of the addi-log form as in Wachter and Yogo (2010):

$$u(c, n) = \left((1-\omega)\frac{1-\varepsilon_h^{-1}}{1-\varepsilon_c^{-1}} c^{\frac{\varepsilon_c - 1}{\varepsilon_c}} + \omega n^{\frac{\varepsilon_h - 1}{\varepsilon_h}}\right)^{\frac{\varepsilon_h}{\varepsilon_h - 1}} \tag{3}$$

where ω captures the relative preference of housing services relative to non-housing consumption.

When $\varepsilon_h = \varepsilon_c$, equation (3) nests the constant elasticity of substitution (CES) case; however, whenever $\varepsilon_h \neq \varepsilon_c$, preferences are non-homothetic. In particular, the assumption that $\varepsilon_h < \varepsilon_c$ implies that housing is a necessary good and non-housing consumption is a luxury good.[6] As we will see later, the assumption that housing is a necessary good is well supported by data on consumption expenditure and has significant implications for portfolio choice. In fact, the non-homotheticity assumption implies that the composition of the consumption basket changes with total expenditure. Because households are less willing to tolerate fluctuations in consumption when the share of housing consumption is higher, households with different total expenditure will care differently about fluctuations in their consumption basket. This implies that both relative risk aversion and elasticity of intertemporal substitution vary along the expenditure distribution, which will in turn affect

[5]The assumption that households face a constant death rate will obviously imply a stylized characterization of households' portfolio choice over the life-cycle. While being a main departure from most of the literature on optimal portfolio choice, this is unlikely to have major implications for wealth inequality and its evolution. In particular, given that the model aim is a realistic characterization of portfolio choice as a function of *wealth* (and the strong correlation in the data between age and wealth), this is a simplification that allows me to retain numerical tractability without giving away too much in terms of model realism.

[6]While the assumption that housing consumption is the only source of non-homotheticities in the model is clearly a simplification (as there are likely different sources of non-homotheticity in the data), it does capture the main mechanism through which non-homotheticities work on households' portfolio choice. The model is therefore broadly consistent with the empirical evidence suggesting that households' preferences are indeed non-homothetic (Atkeson and Ogaki 1996; Blundell, Browning, and Meghir 1994; Meeuwis 2020; Pakoš 2011; Straub 2019, among others)

optimal allocations along the expenditure distribution and (as long as total expenditure varies with wealth) along the wealth distribution. We return to this in section 2.2.

Earnings — At each instant, households receive an endowment of earnings subject to idiosyncratic risk. In particular, I assume log-earnings, z, follow an Ornstein-Uhlenbeck process:

$$\mathrm{d}z_{i,t} = \eta(\bar{z} - z_{i,t})\,\mathrm{d}t + \sigma_z\,\mathrm{d}\tilde{W}_{i,t} \tag{4}$$

where $\tilde{W}$ is a standard Brownian motion idiosyncratic to each household. In the above process $\bar{z}$ governs the average level of earnings; η is a measure of autocorrelation, and σ_z controls the dispersion around the mean.

The process in equation (4) is simple enough to capture the main characteristics of earnings in the data. There are, however, at least two moments of the distribution of earnings in the model that will fall short of the data and that are worth some discussion: First, given that the ergodic distribution associated to equation (4) is log-normal, the model will not be able to generate a fat tail in the earnings distribution. Because earnings are a crucial determinant of the wealth distribution, I deliberately ignore this feature of the data to highlight the strength of households' portfolio choice as a mechanism for generating wealth inequality. Second, Guvenen, Ozkan, and Song (2014) shows that higher order moments of the earnings distribution are correlated with the business cycle. In fact Catherine (2020) shows that modeling such correlation has important implications for portfolio choice and helps explain equity shares along the wealth distribution. Excluding such correlation is therefore also a conservative choice that allows me to focus the role of non-homotheticity in households' preferences as a source of portfolio heterogeneity.

Asset Structure — Households can smooth consumption over time by trading in financial assets (bonds and stocks), and housing. Denote by $\mathrm{d}r_{j,t}$ the asset j's instantaneous return over a time interval of length $\mathrm{d}t$, with $j \in \{B, S, H\}$ and collect them in a vector $\mathrm{d}\mathbf{r}_t = (\mathrm{d}r_{B,t}, \mathrm{d}r_{S,t}, \mathrm{d}r_{H,t})$.[7] Assuming return innovations are correlated across assets and expected returns are constant, we can write instantaneous returns as:

$$\mathrm{d}\mathbf{r}_t = \mathbf{r}\,\mathrm{d}t + \boldsymbol{\sigma}\,\mathrm{d}\mathbf{W}_t \tag{5}$$

where $\mathbf{r} = (r_B, r_S, r_H)$ is the vector of expected returns, $\mathbf{W}_t = (W_{1,t}, W_{2,t}, W_{3,t})$ is a vector of three mutually independent standard Brownian motions, and $\boldsymbol{\sigma}$ is the 3×3 matrix of sensitivities of returns to the three independent shocks (which therefore collects all the information related to variances and covariances of returns).

[7] Notice that, consistent with the fact that households directly get utility from how much housing they own, we can ignore the rental yield component in the returns to housing. $\mathrm{d}r_{H,t}$ therefore only captures the capital gain component of housing returns.

Frictions — Households face a participation decision in both housing and equity markets. When it comes to housing, they can be owners – in which case they derive utility directly from the amount of housing they own h – or renters – in which case they are free to choose every period how much housing services to consume. With respect to the equity market on the other hand, households can participate in the market – in which case they can freely adjust their portfolio of financial assets between bonds and stocks – or not – in which case they will have 100% of their financial assets in bonds. I will henceforth refer to households who participate in the equity market as *participants* and everybody else as *non-participants*.

The entry and exit decisions in both the equity and housing markets have a similar structure and have two components: a stochastic component that governs movements across participation states, and a transaction-cost component. In particular, renters only get a chance to buy a house with constant rate λ^h, at which point they can decide either to become homeowners or to continue renting; if they decide to buy, they are free to choose their optimal housing size h and pay a transaction cost $\kappa^h(h) = \kappa_0^h + \kappa_1^h h$. Similarly, owners also get a chance to sell their house at rate λ^h and, to do so, they have to pay cost $\kappa^h(h)$. Infrequent trading in the housing market as captured by λ^h allows us to model the illiquid nature of housing better than a transaction cost alone would do. We can in fact interpret such stochastic component as a reduced-form way of modeling search and matching frictions in the housing market

Similarly, in the equity markets, non-participants get an opportunity to enter the equity market with constant probability λ_0^p, at which point they can choose to pay an entry cost κ_0^p and become a participant. Then, to stay in the market, participants have to pay a participation fee κ_1^p at a constant rate λ_1^p; if they don't, they exit the equity market and become non-participants.[8]

Finally, households can borrow in bonds up to an exogenous borrowing limit ϕ but, when they do so, face a wedge κ^b that increases the interest rate on borrowing.[9] Notice that, while households are allowed to borrow in bonds, they cannot borrow in stocks and are not allowed to borrow against their house. Due to its short-term nature and the fact that it is unrelated to the amount of housing owned, borrowing should therefore not be interpreted as mortgages. While mortgage choice is clearly an important feature of the housing market, its introduction would significantly complicate the model; because I do not strive to achieve a comprehensive model of housing, but rather of the interaction between households' portfolio shares and wealth inequality, I therefore take this as a necessary shortcut that can be relaxed in future work.

[8]Notice that the stochastic component to the entry and exit decision in the equity market has a much different interpretation than in the housing market. In fact, while some authors do model search and matching frictions in the equity market (most notably McKay 2013), these frictions are usually associated to getting higher rates of return rather than to the entry and exit decision. However, as discussed in the appendix, the stochastic components λ_1^p makes the numerical solution of the model significantly simpler while allowing for minimal economics implications once appropriately calibrated.

[9]Just as in Kaplan, Moll, and Violante (2018), this will allow me to match the fraction of households with zero financial assets.

Household Problem — Given the environment just described, all households choose the optimal level of non-housing consumption c. Homeowners derive utility directly from the amount of housing they own with multiplier $\chi > 1$ (i.e. for owners $n = \chi h$). On the other hand, renters get to choose how much housing services to consume each period, n, at the market rental rate r_n; participants additionally choose the optimal share θ of financial wealth to allocate in equity.

Because participants can freely choose how much to hold in bonds and stocks at every instant, and non-participants can only hold bonds, we can collapse financial assets into a single state variable a and a participation state in the equity market $p \in \{0, 1\}$. Households also differ based on their (log) earnings, z, and their holding of housing h (where we denote renters as households with $h = 0$).

Financial wealth, a, and housing, h, therefore evolve according to:

$$\mathrm{d}a_{i,t} = \Big(e^{z_{i,t}} + r_B a_{i,t} + (r_S - r_B)\theta_{i,t} p_{i,t} a_{i,t} - x_{i,t} - \kappa^b(a_{i,t})\Big)\,\mathrm{d}t + a_{i,t}\boldsymbol{\sigma}_a\,\mathrm{d}\mathbf{W}_t \tag{6}$$

$$\mathrm{d}h_{i,t} = h_{i,t} r_H\,\mathrm{d}t + h_{i,t}\boldsymbol{\sigma}_3\,\mathrm{d}\mathbf{W}_t \tag{7}$$

where $x_{i,t} = c_{i,t} - r_n n_{i,t}\mathbf{1}\{h_{i,t} = 0\}$ is total consumption expenditure, $\kappa^b(a) = \kappa^b a \mathbf{1}\{(1-\theta)a < 0\}$, $\boldsymbol{\sigma}_i$ is the i-th row of the $\boldsymbol{\sigma}$ matrix in equation (5), and $\boldsymbol{\sigma}_a = \big((1 - \theta_{i,t}p_{i,t}), \theta_{i,t}p_{i,t}\big)\,(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)'$ is the vector of sensitivities of financial assets to the three independent shocks, which directly depends on the optimal equity share $\theta_{i,t}$ and the participation state $p_{i,t}$.

With some abuse of notation, it will be useful to collect the continuous states (conditional on the participation state p) in a vector $\mathrm{s}^p = (a, h, z)$, which allows us to more compactly rewrite their evolution as:

$$\mathrm{d}\mathrm{s}^p_{i,t} = \boldsymbol{\mu}(\mathrm{s}^p_{i,t})\,\mathrm{d}t + \boldsymbol{\sigma}_{\mathbf{s}}(\mathrm{s}^p_{i,t})\,\mathrm{d}\mathbf{W}_t + \tilde{\boldsymbol{\sigma}}_{\mathbf{s}}(\mathrm{s}^p_{i,t})\,\mathrm{d}\tilde{W}_{i,t} \tag{8}$$

where $\boldsymbol{\mu}(\mathrm{s}^p_{i,t})$ is the vector of drift components of (a, h, z) as in equations (4), (6) and (7), $\boldsymbol{\sigma}_{\mathbf{s}} = \mathrm{diag}(\mathrm{s}) \cdot (\boldsymbol{\sigma}_a, h\boldsymbol{\sigma}_3, 0)$ and $\tilde{\boldsymbol{\sigma}}_{\mathbf{s}}(\mathrm{s}^p) = (0, 0, \sigma_z)'$.

Denoting by $v(a, h, z, p)$ the household's value function, the household problem can be simply written as:

$$0 = \max\big\{f(u, v) + \mathcal{A}v\big\} \tag{9}$$

where $\mathcal{A}$ is a partial differential operator including all the information about the evolution of households' states.[10,11]

[10]As mentioned above, households choice set depends on their current participation and homeownership state: all households choose non-housing consumption c, renters additionally choose housing consumption n, and participants also choose the portfolio share θ; then, when hit by the stochastic participation shock, participants choose whether to stay in the market or not while non-participants whether to enter or not; finally, when hit by the stochastic buying/selling shock, homeowners choose whether to sell their house and become renters or not, and renters choose if they want to become homeowners or not (and if they do, how much housing to buy).

[11]I leave the derivation and exact definition of $\mathcal{A}$ to the appendix as it is neither particularly useful nor enlightening.

Cross-sectional Distribution — Given the assumption of exogenous returns, characterizing the model economy amounts to describing the evolution of its cross-sectional distribution $g_t(\mathbf{s})$. Proposition 1 therefore presents the main theoretical result which most of the quantitative analysis will hinge on.

Proposition 1. *The distribution of households over individual states, $g_t(\mathbf{s})$, solves the following Kolmogorov Forward Equation (KFE):*

$$\mathrm{d}g_t(\mathbf{s}) = \Big\{\mathcal{A}^* g_t(\mathbf{s}) + \zeta\left(\Psi(\mathbf{s}) - g_t(\mathbf{s})\right)\Big\}\,\mathrm{d}t - \partial_{\mathbf{s}}\Big\{\left[\boldsymbol{\sigma}_{\mathbf{s}}(\mathbf{s})\,\mathrm{d}\mathbf{W}_t\right] g_t(\mathbf{s})\Big\} \tag{10}$$

where $\mathcal{A}^*$ is the adjoint of the differential operator in the HJB equation (9).

The proof of proposition 1, as well as the proofs of all other theoretical results can be found in appendix A.1. Proposition 1 makes clear that, unlike in models of idiosyncratic return risk, the cross-sectional distribution g varies along the ergodic distribution. In fact, the KFE (10) is composed of two terms: a drift term – which is present in all continuous-time models *a lá* Achdou et al. (2017) – that essentially determines the average level of inequality, and a diffusion term – which is only present in models with aggregate risk (e.g. Gomez 2019; 2021; Schaab 2020) – that specifies how the evolution of the cross-sectional distribution g_t depends on both the specific path of shocks $\mathrm{d}\mathbf{W}_t$ and on the exposure of households portfolios to aggregate shocks (through $\boldsymbol{\sigma}_{\mathbf{s}}(\mathbf{s})$).[12]

While it might not be particularly enlightening on its own, equation (10) is a crucial input of the model: it in fact encompasses all the relevant information about households' choices, the distribution of wealth, and, most importantly, their evolution. To gain some more intuition about the role of heterogeneous exposure for the evolution of wealth inequality, in section 2.3 I therefore solve a simplified version of the model which allows us to get an analytical expression for the evolution of wealth shares.

2.2. The Role of Non-homothetic Preferences

As I briefly hinted at above, non-homothetic preferences imply that ψ and γ do not exactly coincide with the elasticity of intertemporal substitution (EIS) and relative risk aversion (RRA); rather, RRA and EIS are a function of total consumption expenditure and therefore vary along the wealth distribution. To see how RRA and EIS change with expenditure, recall that total consumption expenditure in housing services and non-housing consumption is given by $x = c + r_n n$. Proposition 2 characterizes how expenditure shares change along the expenditure distribution and shows that, if $\varepsilon_h < \varepsilon_c$, the expenditure share in housing declines in total expenditure, implying that housing is a

[12]The HJB and KF equations (9) and (10) together constitute a forward-backward system with common noise. For an introduction to the theory of mean field games with common noise see Carmona and Delarue (2018).

necessary good.[13,14]

Proposition 2. *If $\varepsilon_h < \varepsilon_c$, the expenditure share of non-housing consumption rises in total expenditure. Define $\alpha(x) \equiv c(x)/x$ the expenditure share in non-housing consumption, where $x = c + r_n n$, then:*

$$\frac{\partial \alpha(x)}{\partial x} > 0 \quad \forall\, x$$

Furthermore:

$$\alpha_0 \equiv \lim_{x \to 0} \alpha(x) = 0; \qquad\qquad \alpha_\infty \equiv \lim_{x \to \infty} \alpha(x) = 1$$

Why does the assumption that housing is a necessary good imply EIS and RRA vary with wealth? The first formal argument for the implications of non-homothetic preferences for the EIS can be found in Browning and Crossley (2000) and is particularly intuitive: because wealthier households have a larger share of their consumption basket in luxury goods (non-housing consumption in my model), and because households do not care as much about fluctuations in the consumption of luxuries, the elasticity of intertemporal substitution increases with wealth. Browning and Crossley actually present their argument for a model without any intratemporal risk, which therefore does not have any implications for risk aversion. Nonetheless, because households do not care as much about fluctuations in luxury-good consumption both across states and over time, the argument works exactly the same way for risk aversion. Wachter and Yogo (2010) in fact use this exact mechanism to generate variation in risk aversion and generate portfolio shares of equity that are increasing in wealth. On the other hand, because Wachter and Yogo have time-separable preferences, RRA and EIS are obviously linked and they only need to make a statement about one of the two. Corollaries 2.1 and 2.2 therefore extend the results in Browning and Crossley (2000) and Wachter and Yogo (2010) for RRA and EIS in a model with recursive preferences.

Corollary 2.1. *Relative risk aversion is given by:*

$$\mathit{RRA}(x) = \frac{\alpha(x)\varepsilon_c\left(1-\varepsilon_h^{-1}\right)\left(1+\frac{1-\varepsilon_c^{-1}}{1-\varepsilon_h^{-1}}(\gamma-1)\right)+(1-\alpha(x))\varepsilon_h\left(1-\varepsilon_c^{-1}\right)\gamma}{\left[\alpha(x)\varepsilon_c+(1-\alpha(x))\varepsilon_h\right]\left[\alpha(x)\left(1-\varepsilon_h^{-1}\right)+(1-\alpha(x))\left(1-\varepsilon_c^{-1}\right)\right]}$$

[13] Aside from its theoretical implications, one also needs to assess if the assumption that $\varepsilon_h < \varepsilon_c$ makes sense empirically. In the calibration section I will be more specific on how to pin down ε_h and ε_c using data on expenditure. For now let me just state that $\varepsilon_h < \varepsilon_c$ is indeed the empirically relevant assumption.

[14] Because the proofs of the theoretical results in this section make use of the intra-temporal Euler equation (which generally does not hold for homeowners), these results only apply to the renters' problem. Nonetheless, as households have the opportunity to infrequently adjust their housing consumption, they still provide a useful device to understand the main mechanism behind households' portfolio choice.

Then, if $\varepsilon_h < \varepsilon_c < 1 < \gamma$, we have that $\frac{\partial RRA}{\partial x} < 0$ and

$$\gamma_0 \equiv \lim_{x\to 0} RRA(x) = \gamma; \qquad \gamma_\infty \equiv \lim_{x\to\infty} RRA(x) = 1 + \frac{1-\varepsilon_c^{-1}}{1-\varepsilon_h^{-1}}(\gamma - 1) < \gamma_0$$

Corollary 2.2. *The elasticity of intertemporal substitution is given by:*

$$EIS(x) \equiv -\frac{\partial\left(\frac{\mathrm{d}\ln c_t}{\mathrm{d}t}\right)}{\partial\left(\frac{\mathrm{d}\ln \Lambda_t}{\mathrm{d}t}\right)} = \frac{\varepsilon_c}{1 - \frac{u_c c + u_n n}{u}\left(\alpha(x)\varepsilon_c + (1-\alpha(x))\varepsilon_h\right)\left(\frac{1}{\varepsilon_h} - \frac{1}{\psi}\right)}$$

where $\Lambda_t = e^{\int_0^t f_v(u_s,v_s)\,\mathrm{d}s} f_u(u_t,v_t)u_c(c_t,n_t)$ is the utility gradient. Then , if $\psi < \varepsilon_h < \varepsilon_c < 1$, we have:

$$\psi_0 \equiv \lim_{x\to 0} EIS(x) = \frac{\varepsilon_c}{1-\left(1-\frac{\varepsilon_h}{\psi}\right)}; \qquad \psi_\infty \equiv \lim_{x\to\infty} EIS(x) = \frac{\varepsilon_c}{1-\frac{1-\varepsilon_c^{-1}}{1-\varepsilon_h^{-1}}\left(1-\frac{\varepsilon_h}{\psi}\right)} > \psi_0$$

Corollaries 2.1 and 2.2 tell us that, under reasonable parameter values for γ and ψ, $\varepsilon_h < \varepsilon_c$ implies risk aversion is declining in total expenditure and intertemporal elasticity is increasing in it. Notice also that the relative distance between both γ_0 and γ_∞ and ψ_0 and ψ_∞ depends directly on the ratio $\frac{1-\varepsilon_c^{-1}}{1-\varepsilon_h^{-1}}$. This ratio, which varies between 0 and 1, can therefore be interpreted as an inverse measure of the non-homotheticity in the model.

As we will see shortly, the implications of the non-homotheticity for RRA and EIS are key to replicate the portfolio shares we observe in the data (i.e. that poorer households mostly invest in bonds; the middle-class in housing; and only the wealthiest in equity). In the model, because poorer households have a large share of their consumption basket in housing services, they will have a high RRA and low EIS: they really need to smooth consumption but the high level of risk associated with stocks and the illiquidity of housing imply they would much rather invest in low-return bonds. Viceversa wealthier households do not care about consumption smoothing that much and have a lower risk aversion, hence they will heavily invest in stocks to reap the higher return.[15]

2.3. The Role of Heterogeneous Exposure

We saw in proposition 1 that the evolution of the wealth distribution g_t depends on both the specific path of aggregate shocks $\mathrm{d}\mathbf{W}_t$ and on households' exposure to such shocks (through $\boldsymbol{\sigma}_\mathbf{s}$). That aggregate risk is asset returns affects the distribution of wealth is particularly intuitive: if everybody's returns are exposed to movements in the prices of the same assets, such movements will obviously

[15]This mechanism implicitly relies on the assumption that as wealth increases total expenditure does as well. However, this is the case in almost all models of consumption smoothing and certainly the case in the model just presented.

affect everybody's wealth holdings. However, the KFE (10) does not tell us much about how *heterogeneity* in risk exposure influences the evolution of wealth inequality.

In this section I try to shed light on this mechanism by showing that, first, in the general case of arbitrary exposure to aggregate risk in asset returns, wealth inequality depends on the specific path of aggregate shocks and, second, that if exposure to aggregate risk is instead homogeneous, inequality is independent of aggregate shocks. To do so, consider a (very) special case of the model presented above. In particular, assume that households live forever ($\zeta = 0$) and face no earnings risk ($z_{i,t} = 0$). Further assume they can invest in riskless bonds ($\boldsymbol{\sigma}_1 = \mathbf{0}$) and risky stocks but not in housing ($\lambda^h = 0$, i.e. all households are renters). Finally, assume households face no other frictions: that is, all households participate in the stock market and there are no restrictions on borrowing (essentially a slightly more general version of the model in Merton (1971) in which we allow for different preferences).

If without any further loss of generality we also assume that $\boldsymbol{\sigma}_2 = (0,\ \sigma_S,\ 0)$, the evolution of households total wealth $a_{i,t}$ solves the following stochastic differential equation:

$$\mathrm{d}a_{i,t} = \left[r_B + (r_S - r_B)\theta(a_{i,t}) - x(a_{i,t})\right] a_{i,t}\,\mathrm{d}t + \theta(a_{i,t})a_{i,t}\sigma_S\,\mathrm{d}W_{2,t} \tag{11}$$

where $x(a)$ is total consumption expenditure as a function of wealth.

Denoting by $\mu(a)$ and $\sigma(a)$ the drift and diffusion terms in equation (11), the KFE (10) simplifies to:

$$\mathrm{d}g_t(a) = \left\{-\partial_a\left[\mu(a)g_t(a)\right] + \frac{1}{2}\partial_{aa}\left[\sigma(a)^2 g_t(a)\right]\right\}\mathrm{d}t - \partial_a\left[\sigma(a)g_t(a)\right]\mathrm{d}W_{2,t}. \tag{12}$$

While being much simpler than equation (10), equation (12) still does not tell us much about the role of heterogeneous exposure to aggregate risk. Its simplicity, however, makes it much more manageable, and we can in fact use it to look directly at the evolution of wealth inequality. Define aggregate wealth in the economy as $\bar{A}_t = \int_{-\infty}^{+\infty} a g_t(a)\,\mathrm{d}a$ and the share of aggregate wealth held by the top p percent of the wealth distribution as $S_{p,t}$. That is,

$$p = \int_{q_t}^{+\infty} g_t(a)\,\mathrm{d}a \tag{13}$$

$$S_{p,t} = \frac{1}{\bar{A}_t}\int_{q_t}^{+\infty} a g_t(a)\,\mathrm{d}a \tag{14}$$

The next proposition, and its associated corollary, present a simple yet powerful theoretical result that is at the heart of the full model presented in this paper.

Proposition 3. *If the distribution of wealth evolves according to equation* (12)*, top shares* $S_{p,t}$ *as defined* *in equation* (14) *evolve according to:*

$$
\begin{aligned}
\mathrm{d}S_{p,t} = & \frac{1}{\bar{A}_t}\left\{\int_{q_t}^{+\infty} \mu(a) g_t(a)\,\mathrm{d}a - S_{p,t}\int_{-\infty}^{+\infty} \mu(a) g_t(a)\,\mathrm{d}a + \frac{1}{2}\sigma(q_t)^2 g_t(q_t)\right\}\mathrm{d}t + \\
& + \frac{1}{\bar{A}_t}\left\{\int_{q_t}^{+\infty} \sigma(a) g_t(a)\,\mathrm{d}a - S_{p,t}\int_{-\infty}^{+\infty} \sigma(a) g_t(a)\,\mathrm{d}a\right\}\mathrm{d}W_{2,t}
\end{aligned}
\tag{15}
$$

The second term on the right-hand side of equation (15) clarifies that top shares, just like the overall wealth distribution, also depend on aggregate shocks and do so via exposure along the wealth distribution $\sigma(a)$. We already know from Merton (1971) that if preferences are CRRA and homothetic, portfolio shares are constant. Corollary 3.1 then formalizes the effects of this assumption on wealth inequality and brings us back to the usual result that, if exposure to aggregate risk is homogeneous, wealth inequality *does not* depend on aggregate shocks.

Corollary 3.1. *If households exposure to aggregate risk is proportional to wealth, i.e.* $\sigma(a) \propto a$, *top shares* $S_{p,t}$ *as defined in equation* (14) *are independent of aggregate shocks.*

The corollary's proof is simple enough to be included here. Notice in fact that, if $\sigma(a) = \theta^* a$, the second term in brackets on the right hand side of equation (15) simplifies to

$$
\theta^* \int_{q_t}^{+\infty} a g_t(a)\,\mathrm{d}a - \theta^* S_{p,t}\int_{-\infty}^{+\infty} a g_t(a)\,\mathrm{d}a = \theta^* S_{p,t}\bar{A}_t - \theta^* S_{p,t}\bar{A}_t = 0
$$

where in the first equality we have simply used the definition of $S_{p,t}$ and $\bar{A}_t$.

The above two results plainly tell us that aggregate risk in asset returns generates fluctuations in wealth inequality and that portfolio heterogeneity is crucial to such result. That is, if we want to understand inequality dynamics, we ought to understand households' portfolios decisions too.

2.4. Numerical Solution

The model is solved by discretizing the state vector $\mathbf{s}$ and using finite difference approximations to both the HJB (9) and KFE (10) (the numerical procedure is described more in detail in the appendix and is similar to the one in Achdou et al. 2017). The solution to the household problem in equation (9) essentially boils down to writing a discretization of the infinitesimal operator $\mathcal{A}$. Once we have that, the drift term of the kolmogorov forward equation (10) comes for free, as the discretization of $\mathcal{A}^*$, only involves computing a matrix transpose. The diffusion term in (10) can then also be expressed in its discretized form, and solving the KFE therefore simply requires moving forward in time starting from an initial distribution g_0.

Unless stated otherwise all model solutions are solved by simulating a random sequence of prices from equation (5) on a quarterly time grid. The simulation starts in period 0 and runs for 11,000 years, of which the first 1,000 years are then discarded to remove dependence on the initial condition. Running the simulation on a finer time grid or for longer does not change the results in any meaningful way. The initial distribution g_0 is the solution of the deterministic portion in equation (10) (i.e. the solution to a model with $\mathbf{W}_t = \mathbf{0} \, \forall \, t$). Finally, for the birth distribution Ψ, because the model includes neither voluntary nor unvoluntary bequests, I assume every newborn agent starts as a non-participant with no financial wealth, no housing and an endowment of earnings drawn from the ergodic distribution associated to equation (4). Because bequests are often used as a mechanism to increase wealth concentration, this is a conservative choice that allows me to focus directly on the portfolio-choice channel.

3. Model Calibration

The model is calibrated to match some key features of the average cross-sectional distribution starting from the 1980s. Some parameters are calibrated externally, in particular the processes for asset prices and earnings, while others are calibrated internally by directly comparing moments from the model to the data. In particular, for the internally-calibrated parameters, one key moment is the schedule of households portfolio holdings along the wealth distribution. All model moments are calculated using time averages along the ergodic distribution.[16]

Table 1 collects all the calibrated parameters with the exception of asset prices. One unit of time is normalized to be one year and parameter values are expressed accordingly (e.g. a death rate of $\zeta = 1/45$ implies an average lifespan of 45 years). When expressed in percentage terms, values are meant to be relative to average earnings in the model.

3.1. External

Asset Prices — Asset prices are calibrated using data from 1970 to 2010 taken both from the MacroHistory Database of Jordà et al. (2019) for bonds and housing and from Kartashova (2014) for equity. The database provides data for the rate of return on safe assets (taken to be the rate of return on treasury bills) and on housing. For the latter, given that the model separates between renters and owners, I only use the series on capital gains. Then, because I will calibrate portfolio shares on households' net home equity position, capital gains are also adjusted for leverage and costs of mortgages by using the average loan-to-value ratio and the average 30-year fixed rate mortgage series from FRED. The rental rate r_n is set to equal the average rental returns over the period.[17]

[16] As mentioned above, unlike in model of idiosyncratic return risk, the cross-sectional distribution g varies over time. To avoid any confusion between the distribution of households at a given point in time g_t and the distribution of g over time, I will henceforth always use the term "ergodic" when referring to the time-series dimension of g.

[17] This is consistent with the fact that most of the observed variation in the aggregate rent-to-price ratio comes from movements in prices rather than rents.

Table 1: Calibration

Parameter	Value	Target	Source
Preferences			
ζ	1/45	avg. working life	internal
ρ	0.06	wealth-to-income ratio	”
γ	2	portfolio shares	”
ψ	0.5	time-separable preferences	”
χ	1.5	avg. homeownership rate	”
ω	0.31	avg. expenditure share	”
ε_h	0.75	expenditure shares	”
ε_c	0.91	avg. expenditure elasticity	Aguiar and Bils (2015)
Earnings Process			
$\bar{z}$	0	normalization	-
η_z	0.03	autocorrelation	-
σ_z	0.3	earnings inequality	Kuhn and Rios-Rull (2016)
Frictions			
ϕ	75%	external	Heathcote, Storesletten, and Violante (2020)
κ^b	0.06	”	Kaplan, Moll, and Violante (2018)
λ_0^p, λ_1^p	12, 365	-	-
κ_0^p	0	external	Vissing-Jorgensen (2002)
κ_1^p	1.5%	avg. participation	internal
λ^h	5.2	external	Garriga and Hedlund (2020)
κ_0^h	5%	”	Favilukis, Ludvigson, and Van Nieuwerburgh (2016)
κ_1^h	0.055	”	”

Equity prices come directly from Kartashova (2014) and are a weighted average of returns on private equity and on a public-equity index. All returns are adjusted for inflation.

The estimated values for the vector of average returns $\mathbf{r}$ and the matrix of sensitivities $\boldsymbol{\sigma}$ in equation (5) are:

$$\mathbf{r} = \begin{bmatrix} 0.019 \\ 0.111 \\ 0.003 \end{bmatrix} \quad \boldsymbol{\sigma} = \begin{bmatrix} 0.022 & . & . \\ -0.025 & 0.095 & . \\ -0.006 & 0.030 & 0.053 \end{bmatrix}$$

Earnings — I calibrate the earnings process in equation (4) to match inequality in the earnings distribution; in particular I directly target a Gini coefficient of 0.67 from Kuhn and Rios-Rull (2016) and an autocorrelation coefficient of 0.97 (which are matched exactly). The process also closely matches other (untargeted) moments of the distribution such as the mean to median ratio and the 90-50 ratio. However, as mentioned above, because the process in equation (4) generates a log-normal distribution of earnings, it does not feature a fat tail and will therefore miss some of the earnings concentration at the very top (top 1% and above).

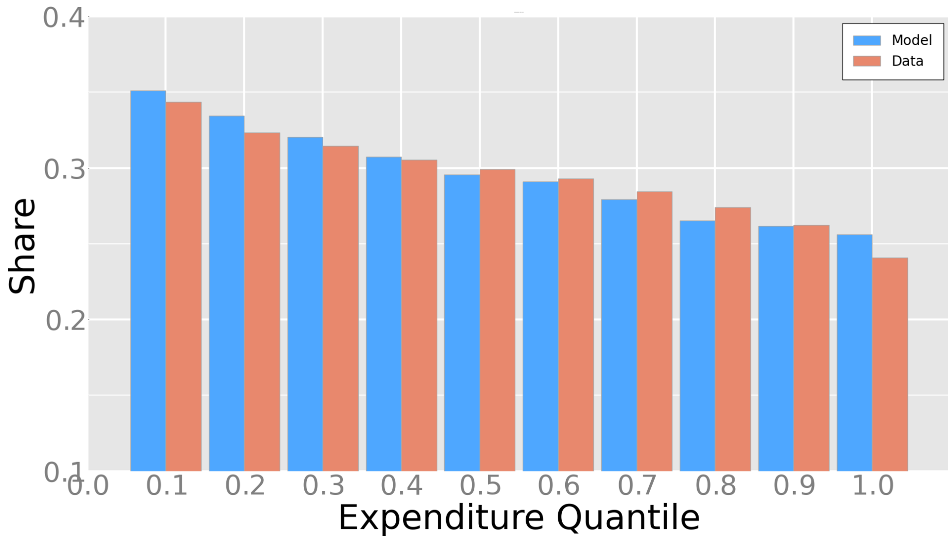
Borrowing and Housing Frictions — The exogenous borrowing limit ϕ is set to 75% of average earnings, which approximately corresponds to the median credit limit (from Heathcote, Storesletten, and Violante 2020), while the borrowing wedge κ^b is set at 6% as in Kaplan, Moll, and Violante (2018). Transaction costs in the housing markets are taken directly from Favilukis, Ludvigson, and Van Nieuwerburgh (2016) who have a similar model of portfolio choice: variable costs κ_1^h are set at 5.5% of the housing value, while the fixed cost is set to 5% of average earnings (which approximately corresponds to their 3.2% of average consumption). The buying and selling frequency λ^h is set such that the average search time in the housing market is of 10 weeks (from Garriga and Hedlund 2020).

3.2. *Internal*

Preferences — Aside from providing intuition for why non-homothetic preferences help us match portfolio shares along the wealth distribution, the theoretical results in section 2.2 also give some useful guidance for the calibration of the preference parameters. We already mentioned that, if $\varepsilon_h = \varepsilon_c$, the intratemporal consumption aggregator in equation (3) is of the CES form. Constant elasticity would have two main implications: first, expenditure shares would be constant and, second, all goods would have unit expenditure elasticity. On the other hand, if $\varepsilon_h < \varepsilon_c$, proposition 2 tells us that the expenditure share of housing is declining in total expenditure; then, it is trivial to show that the expenditure elasticity of housing is given by:

$$\frac{\partial \log n}{\partial \log x} = \frac{\varepsilon_h}{\alpha(x)\varepsilon_c + (1-\alpha(x))\varepsilon_h}. \tag{16}$$

Figure 1: Housing Expenditure Shares



which, as households get wealthiers, gradually decreases from 1 to $\varepsilon_h/\varepsilon_c < 1$.

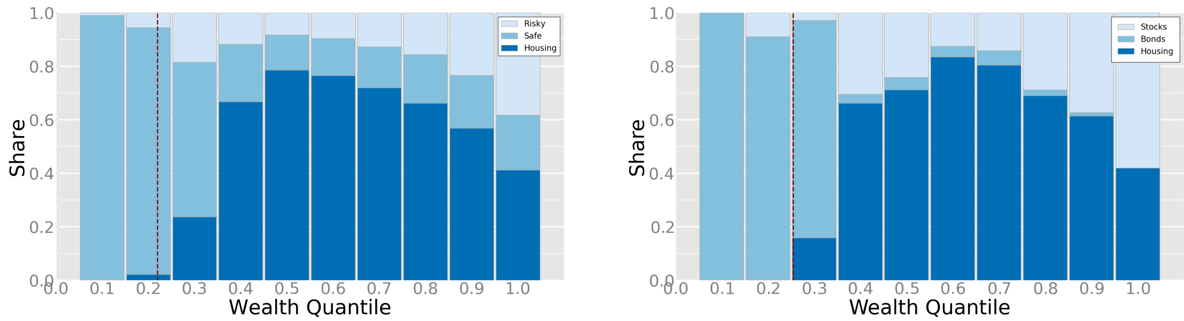
This suggests the change in expenditure shares along the expenditure distribution and the average expenditure elasticity of housing will be informative for the non-homotheticity parameters. I therefore ask the model to match both the decline in housing expenditure shares along the expenditure distribution (measured as deviations in the average expenditure share for each expenditure decile), and the average expenditure elasticity of housing (which is readily available from Aguiar and Bils 2015, Table 2). Using data from the Consumer Expenditure Survey (CEX) they estimate an average elasticity of 0.92, which my model matches exactly. Then, figure 1 shows that the model also matches the declining pattern of expenditure shares along the expenditure distribution. The data series in figure 1 is also obtained from the CEX using the same methodology and sample as in Aguiar and Bils (2015), which allows for consistency between the data on expenditure shares and the average elasticity.[18]

Finally, we also ask the model to match the wealth to earnings ratio and the average homeownership rate. The calibrated values of ε_h and ε_c are of 0.82 and 0.92, respectively, with an implied value for the non-homotheticity parameter $\frac{1-\varepsilon_c^{-1}}{1-\varepsilon_h^{-1}}$ of approximately 0.38 suggesting that, although the average expenditure elasticity of housing is close to 1, preferences exhibit a high degree of non-homotheticity. Given the calibrated values for γ and ψ, corollaries 2.1 and 2.2 also give us implied values for γ_0 and γ_∞ of 2 and 1.38, while those for ψ_0 and ψ_∞ are 0.56 and 0.745, which are well within the range of estimated values from Calvet et al. (2021).

[18]Housing expenditure is computed using actual rent paid for renters and (self-reported) rental equivalence for home owners (see Aguiar and Bils (2015) for more details about data definitions).

Figure 2: Portfolio Shares

(a) SCF 1989-2019 (b) Model



Portfolio Shares — When it comes to households' portfolio choice, I specifically target the distribution of portfolio shares along wealth deciles. In particular, using data from the Survey of Consumer Finances (SCF), I aggregate assets into three different asset-classes which can be immediately mapped to the three assets in the model. Risky assets include all direct and indirect holdings of both public and private equity. Public equity comprises of stocks, stock mutual funds, IRAs/Keoghs invested in stocks, and other managed assets with equity interest. Private equity instead refers to the total value of businesses in which the household has either an active or nonactive interest. Safe assets are then computed as a residual of total financial assets minus equity holdings, they include transaction accounts (money market, checking, savings accounts etc.), certificates of deposit, bonds and other managed assets. Finally, housing is computed as the total value of the primary residence and of other residential real estate, and the value of net equity in nonresidential real estates other than the principal residence. As discussed in section 2, since the model does not feature mortgages but only short-term borrowing, I subtract mortgage debt from the total value of housing and directly calibrate on households' net housing position.

The decision to bundle together public and private equity as a single asset class is not entirely without consequence, and is therefore worth to discuss more at length. With the exception of the period from 1990 to 2000, which exhibited exceptionally high public equity returns, private equity generally earns a premium over public equity (Kartashova 2014; Moskowitz and Vissing-Jørgensen 2002). Given private equity is disproportionately concentrated at the very top of the wealth distribution, it is not surprising the literature on wealth inequality has often turned to models that try to replicate the private equity concentration at the top. However, one thing these models miss by modeling private equity as an asset subject to purely idiosyncratic risk is that private-equity returns are generally highly correlated with public equity returns over time. Bundling together public and private equity, while missing features associated to private equity such as idiosyncratic risk and illiquidity, allows me to capture the common fluctuations in the return to both assets. While definitely important, including such additional features of private equity in a model of endogenous portfolio choice is not trivial and beyond the scope of the paper.

Figure 2a plots portfolio holdings in the SCF pooling all waves from 1989 to 2019 and shows that, while poor households mostly hold safe assets, housing is *the* asset of the middle class and that equity only starts to have a predominant role in household portfolios above the 8th decile. Figure 2b shows that the model captures the main characteristics of households' portfolio holdings along the wealth distribution as well as the fraction of households with negative net worth (represented with the vertical red dashed line) which equal 25% vs. 22% in the model vs. the data.

Equity-market Frictions — I calibrate equity market frictions to match the average fraction of equity holders in the SCF from 1989 to 2019, where equity is defined as in figure 2. Unfortunately there is no clear data moment that can help setting the stochastic components governing entry and exit in the equity market. I therefore set them so as to minimize their economic implications: the frequency of the entry shock λ_0^p is set such that no more than 1% of households would ever want to be a stockholder but is forbidden to do so because of the entry friction, while the frequency for the per-period participation cost λ_1^p is simply set so as to arrive on average once a day (therefore mimicking a continuous-time cost structure).

When it comes to the participation costs, the model still has two parameters that are both acting on households participation. Consistent with evidence from Vissing-Jorgensen (2002) in favor of per-period transaction costs only, I set the entry cost κ_0^p to zero and only calibrate κ_1^p. The calibrated value of 2.5% is consistent with results from Favilukis, Ludvigson, and Van Nieuwerburgh (2016) who set it at about one third of the fixed housing transaction cost.

4. Results

Now that we have shown that the model does a good job at matching households' portfolio shares along the wealth distribution, we are ready to move to the main objective of the paper: quantitatively assess the role of heterogeneous exposure for wealth inequality and its dynamics. In particular, we are interested in characterizing how the dynamics of wealth inequality are affected by changes in asset returns and to what extent the model can help us understand the sharp rise observed from 1980 to 2010. Before doing so, however, it is useful to understand some key properties of the ergodic distribution. Hence, I first look at how well the model matches the average distribution of wealth and perform a series of exercises to understand how much each of the model ingredients contributes to inequality; then, I look at the ergodic distribution and at what is the effect of changes in asset returns; finally, I feed the realized sequence of prices into the model to show that it can indeed replicate the observed increase in top wealth shares and that the latter was mostly the result of abnormal equity returns.

Table 2: Wealth Inequality

	Bottom 60%	Next 30%	Top 10%	Top 1%	**Gini**
SCF (1989-2019)	7.0%	22.7%	70.3%	33.9%	0.827
Model	6.0%	31.1%	62.9%	27.1%	0.759

4.1. The Level of Wealth Inequality

In this section I show first, that the model generates a significant degree of wealth concentration and, second, that the main driver of such concentration is heterogeneity in returns.

Average Inequality — Table 2 compares average wealth inequality along the ergodic distribution with average inequality in the U.S. from 1989 to 2019. The model roughly matches the data, with a top 10% wealth share of 62.9% (vs. 70.3% in the data), a top 1% wealth share of 27.1% (vs. 33.9% in the data), and a Gini coefficient of 0.76 (vs. 0.83 in the data).[19]

Notice also that, while average inequality in the model is slightly lower than in the data, it should at this point be clear that, in the presence of portfolio heterogeneity, the model's cross-sectional distribution is a stochastic object. In section 4.2 I will in fact show that values for the top 10% wealth share and Gini coefficient of 77.9% and 0.86 (as in the 2019 SCF) are well within the ergodic distribution.

Decomposition — The model presented in section 2 includes several features that contribute to generate wealth inequality: earnings risk, non-homothetic preferences, return heterogeneity, and aggregate risk in asset returns. What exactly is the contribution of each of these mechanisms to generate wealth inequality? To answer this question I run a series of counterfactual exercises in which, starting from a single asset model with idiosyncratic income risk and homothetic preferences (essentially a version of the classic Bewley (1980) and Huggett (1993) models), I sequentially introduce non-homothetic preferences, return heterogeneity, and aggregate risk in asset returns. One problem that arises when solving the model with return heterogeneity only (i.e. without risk) and risk only (i.e. without return heterogeneity) is that – in a model of optimal portfolio choice – no agent would ever choose bonds over stocks if both are riskless (and $r_S > r_B$), and no agent would ever choose equity over bonds if they have the same returns (and equity is riskier than bonds). To circumvent this problem I solve all models here by shutting off agents' portfolio *choices*; that is, agents get directly assigned a portfolio of bonds, stocks, and housing based on their wealth level. The

[19]That the model is capable of generating large levels of wealth inequality once it can match heterogeneity in portfolios (and therefore returns) should not come as too much of a surprise, especially given the recent work that shows return heterogeneity plays a crucial role in generating wealth inequality (see Hubmer, Krusell, and Smith 2021; Xavier 2020, among others).

Table 3: Wealth Inequality - Decomposition

	Parameters		Gini	
Model	Rates of return	Risk	Mean	Std. dev.
Baseline	$\mathbf{r}$	$\boldsymbol{\sigma}$	0.759	0.049
Earnings	-	-	0.662	-
No Ret. Heterogeneity, Homothetic	$r_j = 5.3\%$	$\sigma_{i,j} = 0$	0.684	-
Non-Homothetic Preferences	$r_j = 5.3\%$	$\sigma_{i,j} = 0$	0.711	-
Only Risk	$r_j = 5.3\%$	$\boldsymbol{\sigma}$	0.707	0.031
Only Ret. Heterogeneity	$\mathbf{r}$	$\sigma_{i,j} = 0$	0.746	-
All	$\mathbf{r}$	$\boldsymbol{\sigma}$	0.743	0.040

Notes: The first two rows report the Gini coefficient for the baseline model in section 2, and that for the *earnings* distribution, respectively. All other rows report results for a model in which households get directly assigned portfolio shares based on their wealth level (the exogenous shares are taken directly from SCF data). Models without return heterogeneity set the average rate of return to be the same as in the baseline economy. Each model is recalibrated to match the wealth to earnings ratio and, for models with non-homothethic preferences, expenditure shares.

exogenous portfolio shares are taken directly from SCF data, and will therefore perfectly replicate figure 2a.[20]

Notice also that, in the data, at any given level of wealth there will obviously be both homeowners and renters. Allowing for such separation in the model with exogenous portfolio shares would therefore require me to take a stance on how to assign households across homeownership states, as well as on how households move from one state to the other. In these exercises I therefore simply separate the investment dimension of housing from the consumption one: namely, investment in housing is directly determined by the portfolio share, while the consumption choice is left unconstrained.

Table 3 presents the results of this exercise: in the first two rows I report the Gini coefficient for the baseline model in section 2, and that for the *earnings* distribution (which is a useful benchmark point all models start from), respectively. The third row then shows that a single-asset model with homothetic preferences generates very little wealth inequality above and beyond earnings inequality, a result consistent with the theoretical analysis in Benhabib and Bisin (2018) which tells us that "standard" models without any heterogeneity in returns have a hard time generating high wealth inequality.[21] In the fourth row we see that adding non-homothetic preferences does slightly increase wealth inequality (mainly through its effect on the optimal consumption-saving choice) but gets us nowhere close to the level of inequality observed in the data. The last three rows then look at the roles

[20]For models without return heterogeneity we set the average rate of return to be the same as the average return in the baseline economy. Each model is then recalibrated to match the wealth to earnings ratio and, for models with non-homothetic preferences, expenditure shares.

[21]The theoretical results in Benhabib and Bisin (2018) actually applies only to the tail coefficient of the distributions of earnings and wealth.

of heterogeneity in the risk-return profile: the fourth row – only risk – keeps the return to all assets at its average value of 5.3%, but reintroduces the $\boldsymbol{\sigma}$ matrix in the price equation equation (5). We can immediately see that, while the average level of inequality is essentially unchanged, the introduction of aggregate risk obviously makes the wealth distribution stochastic, as shown by a positive standard deviation of the Gini coefficient along the ergodic distribution. Return heterogeneity, on the other hand, increases wealth inequality far more than any other channel, and gets us essentially to the same level as the baseline model; a result very much in line with the recent literature stressing the role of return heterogeneity for wealth inequality Hubmer, Krusell, and Smith (2021); Kuhn, Schularick, and Steins (2020); Xavier (2020). Finally, the last row – exogenous shares – reintroduces aggregate risk in asset returns and reports inequality for a model with all the same features as the baseline but without optimal portfolio choice. Both the level of inequality and its standard deviation here are not much smaller than the baseline model which suggests that, aside from generating the correct schedule of returns, optimal portfolio choice does not play a major role for inequality itself.

4.2. The Dynamics of Wealth Inequality

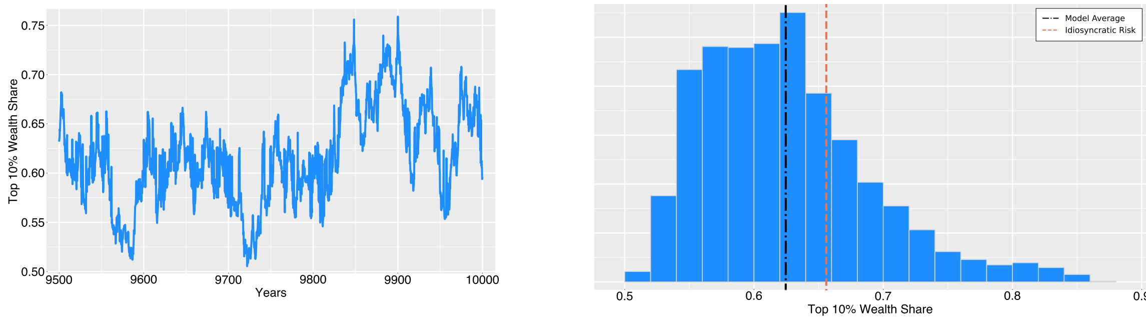
We have seen that, through the effect of heterogeneity in returns to wealth, the model does a good job at matching the average level of inequality. In this section we look at the role such heterogeneity plays for the dynamics of wealth inequality. In particular, I show that top wealth shares exhibit large and long-lasting fluctuations and that, as a consequence wealth inequality takes a wide range of values along the ergodic distribution. Then, I show that most of these effects are due to the economy's response to temporary shocks to equity returns.

Ergodic Distribution — To look at the how wealth inequality changes over the ergodic distribution, in figure 3 I plot both the time series over a "short" time interval and the histogram over the whole simulation. Panel 3a plots the evolution of the top 10% wealth shares over the last 500 years of the simulation: while the top share fluctuates mostly around its mean value, it also experiences some sporadic large and long-lasting fluctuations with changes as large as 20% in less than a century. Panel 3b then confirms what we see in the panel 3a: the bulk of the distribution is concentrated around a top 10% share of approximately 60% but over time it ranges from about 50% to well over 80% with some – very infrequent – peaks of almost 90%. The ergodic distribution in figure 3b tells us that, although a realization of the top 10% wealth share of 76% (as in the 2019 SCF) is indeed high, it is still well within the ergodic distribution.

The Effect of Returns — The results in figure 3 suggest that prices must have large effects on wealth inequality and that these effects must be persistent, or we would not see such long-lasting periods of high inequality as in panel 3a. To understand what are the main determinants of such large and persistent effects, we first look at the effects of shocks to the returns of different

Figure 3: The Dynamics of Wealth Inequality

(a) Sample Path (b) Ergodic Distribution



assets, and then assess if the temporary nature of the shocks has anything to do with it. We start by computing impulse response functions (IRF) of the top 10% wealth share to the various components of $\mathbf{W}_t$. In particular, figure 4 plots the IRF to a one-time shock in $\mathbf{W}_t$ corresponding to a 1% excess return in each asset.[22,23] The shock is entirely temporary and starts from the "steady-state" distribution g_0; that is, starting from a situation in which $\mathbf{W}_t = \mathbf{0}\ \forall\, t < 1$, $\mathbf{W}_t$ suddenly increases at $t = 1$ and then goes back to $\mathbf{0}$ forever.[24] We are interested both in the impact effect of the shock, measured at its peak, and in its duration, measured as the time it takes for inequality to go from its peak to half of it (which I will refer to as a shock's half-life).

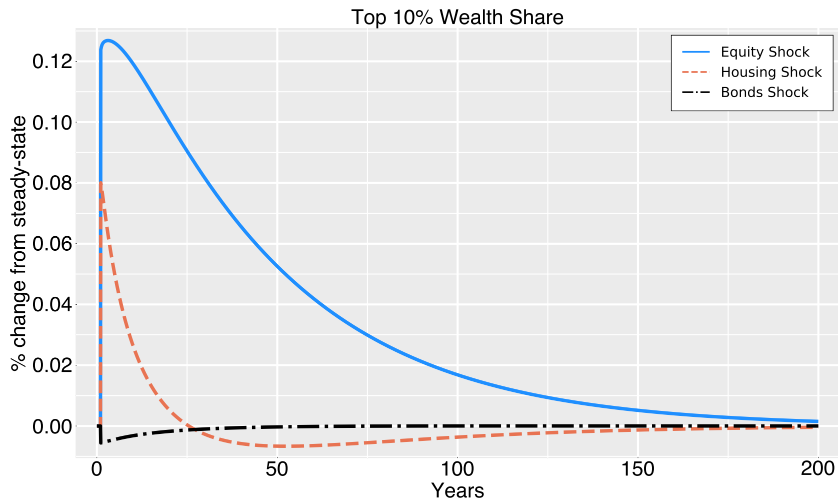
Figure 4 shows that the three shocks imply very different responses in the top 10% wealth share: First, the response to a 1% excess return in bonds is very small and negative on impact (less than a 0.01% change from steady state). On the other hand, a shock to house prices generates a much more substantial increase in inequality (0.08% on impact), reverting fairly quickly (with a half-life of 6 years) and has a negative but small effect on inequality from 25 years onwards. From the response to the equity shock, however, we see that it is movements in equity prices that generate by far the largest effect on inequality. The increase in the top 10% share reaches its peak 3 years after the initial shock and is more than 50% larger than the effect of the corresponding housing shock with a peak response of 12.7%. In fact, the equity shock is also the most persistent, with a half-life of almost 40 years. To get a sense of how large and important equity shocks are for inequality fluctuations, recall that the estimated standard deviation of equity returns is about 9.5%. That means in steady state a one standard deviation shock to equity returns is going to generate an increase in the top 10% wealth

[22] Notice that, because the variance-covariance matrix $\mathbf{\Sigma}(\mathbf{P}_t)$ is non-diagonal, to get a 1% excess return in only one asset the entire vector $\mathbf{W}_t$ might have to change (i.e. there is not a one-to-one mapping between the components of $\mathbf{W}_t$ and returns to each asset).

[23] Given $r_B < r_H < r_S$, a 1% excess return in equity will clearly be smaller (in relative terms) than a 1% excess return in housing which is itself smaller than the same-sized shock to bonds. Given equity is by far the largest driver of inequality, normalizing the shocks to be 1% for each asset is a conservative choice (rather than normalizing the shock either to each asset's standard deviation or to each asset's average returns).

[24] Notice that, because households believe prices evolve according to equation (5), this is not technically an "unexpected" shock.

Figure 4: IRF to one-time 1% excess return



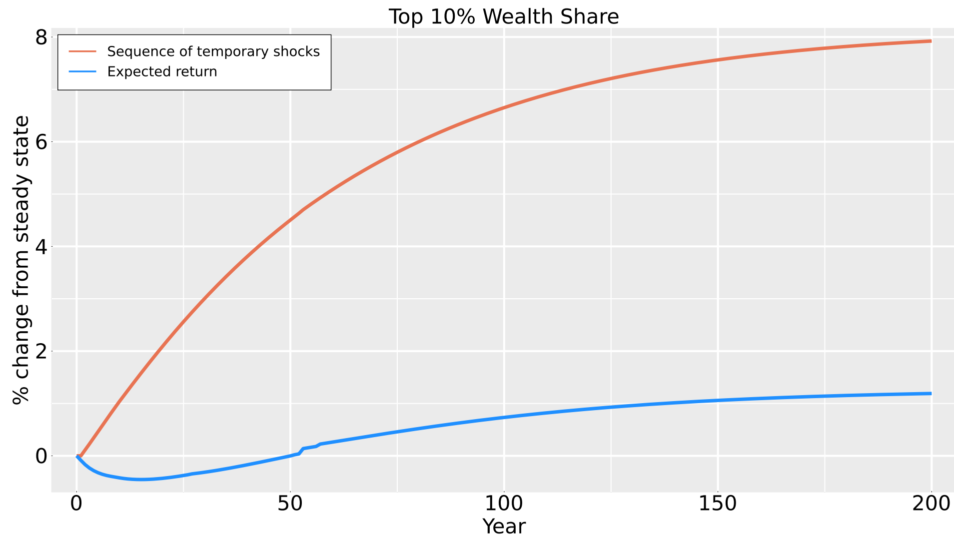
share of almost 1 percentage point.

A second question relates to whether the temporary nature of movements in returns has any impact on the size and persistence of the shocks. To answer this question we can compare the response of the model to two different sequences of shocks that generate the exact same sequence of returns. In particular, in figure 5 I plot the evolution of the top 10% wealth share in response to: first, a continuous sequence of shocks to $\mathbf{W}_t$ such that equity returns are always 1% above r_S (in orange) and, second, a one time increase of 1% in expected returns r_S (in blue).[25] Figure 5 shows that the difference in the economy's response to the two identical sequences of returns is extremely different. In fact, in response to a 1% change in expected returns, inequality decreases in the short run and eventually only increases by about 1% versus a total increase of about 8% in response to the sequence of temporary shocks.

An important implication of this exercise is that, if we want to understand the dynamics of wealth inequality, we need to carefully model the nature of aggregate risk in asset returns. That is, there are two possible ways of modeling fluctuations in rates of returns: in one case we can interpret all movements in returns as fluctuations around a constant expected rate of return, which is the interpretation taken in this paper; in the other case we can interpret all movements in returns as changes in *expected returns*. Figure 5 tells us that these two extreme cases have extremely different implications for the wealth distribution. Hence, given reality is likely in between these two extremes, if we want to carefully account for the role of changing asset prices for wealth inequality we should carefully account for the nature of price changes.

[25]While the model does not allow for time variation in expected returns, we can easily solve for the response to a one-time change by applying the households' policy functions after the shock to the wealth distribution before the shock and let the distribution move forward in time.

Figure 5: IRF to constant 1% excess return to equity



4.3. *The Rise in Wealth Inequality*

Given aggregate movements in asset returns have such large and persistent effects on wealth inequality, how much of the observed increase in inequality can the model explain? In this section I first show that movements as large and as fast as observed in U.S. data are a perfectly normal outcome of the model; then, by feeding the realized sequence of returns into the model, I show that the model can in fact replicate the increase in observed top wealth shares. Third, by running a series of counterfactual experiments, I show that most of the rise in inequality was due to abnormal equity returns in the late 1990s and early 2000s, with housing playing a very minor role only during the housing boom in the mid 2000s. Finally, I also show that the model predicts a slow reversal of inequality towards its mean.

The Distribution of Changes — Saez and Zucman (2016) and Smith et al. (2021) (henceforth SZ and SZZ) estimate the increase in the top 10% wealth share from 1986 to 2012 to be 13.6 and 10.6 percentage points, respectively.[26] We can compare these two numbers with the distribution of 25-years changes in the same share along the ergodic distribution, plotted in figure 6. Although the observed changes in inequality are indeed at the right end of the changes produced by the model – that is, the model predicts such to be somewhat unlikely events – the estimated increases are well within the ergodic distribution. In fact, in the model the probability that the top 10% wealth share increases by more than 10.6 p.p. over the span of 25 years, $\mathbb{P}(\Delta_{25} S_{0.1} \geq 0.106)$ equals 6.1%,

[26] Depending on the methodology used to compute top shares, the literature is somewhat in disagreement about the exact level of top shares across the sample period. Given the assumption of heterogeneous returns, my model is more directly comparable to results from Smith et al. (2021). However, for sake of completeness I also include results from Saez and Zucman (2016).

Figure 6: Distribution of Changes

while $\mathbb{P}(\Delta_{25}S_{0.1} \geq 0.136) = 0.023$.

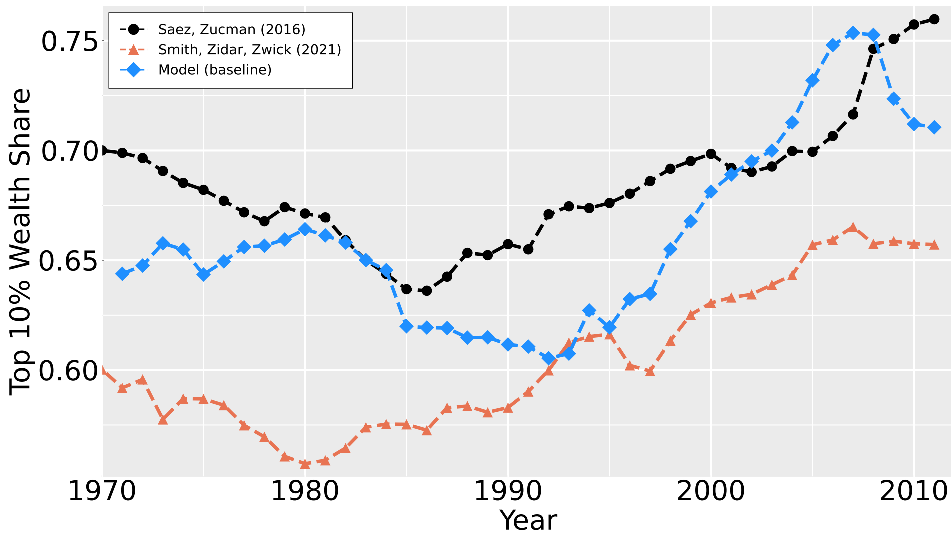
As a matter of fact, the probabilities computed above are *unconditional* probabilities; however, accounting for the fact that the starting level of inequality was quite below average in 1986 significantly increases the implied probabilities. To account for mean reversion in inequality, we can then compute the probability of observing the same increases conditional on the starting value being below the values estimated by SZ and SZZ in 1986 (i.e. 63% and 56%, respectively). Unsurprisingly, the conditional probabilities implied by the model are significantly higher than the unconditional ones at 14.3% for the change estimated in SZZ and 3.5% for SZ. That is, large positive changes in wealth inequality are a perfectly reasonable outcome of a model in which inequality fluctuates over time in response to asset-price movements.

The Realized Sequence of Returns — Given the model is capable of generating large and positive changes in inequality, can it also replicate the observed increase in inequality if we feed into it the realized sequence of asset returns directly from the data? In figure 7 I feed the realized sequence of returns into the model and compare the evolution of the top 10% wealth shares across model and data. The figure shows that the model is capable of closely replicating the U-shape in the evolution of wealth inequality and that it indeed generates essentially all of the realized increase since the 1980s. [27]

The exercise is similar in spirit to the one performed by Hubmer, Krusell, and Smith (2021)

[27]Given the above discussion about the different model response to temporary vs. permanent changes, it is worth noticing that all movements in returns are here assumed to be temporary. That is, to feed in the realized sequence of prices I simply extract a sequence of shocks for $\mathbf{W}_t$ using the calibrated process in equation (5) and feed it into the model.

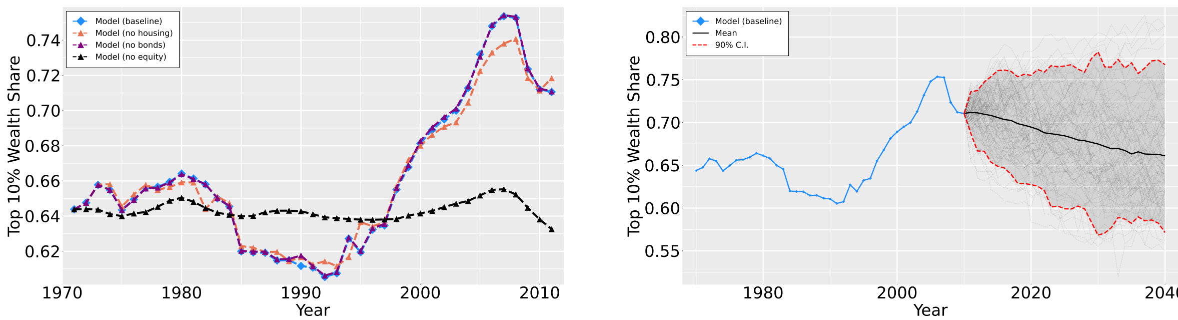
Figure 7: Top 10% Wealth Share Dynamics - Model vs. Data



but their conclusions are, quite strikingly, almost exactly the opposite. In fact, first, they claim that changes in taxes (rather than changes in returns) are the main driver of the increase in inequality. Most importantly, however, when feeding in *only* the realized sequence of prices, their model is capable of explaining the initial dip in inequality but generates almost none of the subsequent increase. They in fact argue that, while changes in asset returns are key to explain the U-shape of top wealth shares, they "have also dampened the increase in wealth concentration on net, in particular explaining the initial dip" (Hubmer, Krusell, and Smith 2021, p. 430). The reason why they reach such different conclusion is likely to be found in the difference between the effect of temporary versus permanent changes in returns explained in the last section. In fact, they assume all changes in returns are entirely permanent and only feed the model with the "10-year moving averages of realized aggregate returns" (p. 418). Given the overall response of inequality to permanent shocks in figure 5, it is therefore not surprising that changes in expected returns have a hard time generating the increase in wealth inequality observed in the data.[28] This is neither to say that my assumption about movements in asset returns being temporary is right, nor that theirs about such movements being permanent is wrong. Rather, I simply hope to further highlight that it is not an inconsequential assumption and that, if we want to understand fluctuations in wealth inequality, future research should do a better job at understanding the sources of fluctuations in asset prices in the first place.

[28]Notice also that the model in Hubmer, Krusell, and Smith (2021) is quite different from the one presented here, in that it focuses more on a realistic characterization of the earnings process rather than on households' portfolio choice – which are instead assumed to be completely exogenous. Nonetheless, the way in which fluctuations in asset return affect wealth inequality is similar across the two models.

Figure 8: Top 10% Wealth Share Dynamics



(a) Counterfactuals (b) Future

Counterfactuals — In order to further understand what exactly drove the increase in wealth inequality since the 1980s we can also perform a series of counterfactual exercises in which we fix returns to each asset at its historical average. Shutting down variation in returns to each asset will then tell us how much returns to that one asset contributed to the evolution of inequality. Figure 8 plots the results of these counterfactual experiments, together with the baseline from figure 7 (the solid blue line). We can immediately see that, when keeping returns to bonds at the historical average (purple line), the top 10% wealth share is virtually unchanged from the baseline simulation. That is, variation in bond returns had very little impact on wealth inequality. In fact, most of the observed increase in wealth inequality was actually driven by abnormal returns to equity in the late 90s and early 2000s, with housing only accounting for a minor part of the increase at the onset of the housing bubble in the mid 2000s.

Finally, given the model predicts inequality to fluctuate over time, what should we expect the future to look like? Figure 8 plots the average and 90% confidence intervals for the to 10% wealth shares over 100 simulations starting from the final point of the simulation in figure 7. As it was to be expected, because the current level of inequality is much higher than the model's average, inequality is expected to slowly revert back on average. However, due to the presence of aggregate risk, the realized paths of inequality are very disperse and values for the top 10% wealth share of 75% would still be a perfectly reasonable outcome.

5. Conclusions

Existing explanations for the observed increase in U.S. wealth inequality mostly rely on the economy's response to changes in fundamentals. Examples of such changes include – but are not limited to – increased income inequality, higher volatility in returns to wealth, and changes in the progressivity of taxes. In this paper I instead propose a novel theory based only on heterogeneous exposure to aggregate risk along the wealth distribution. I show that large increases in wealth inequality are natural outcomes of a model in which households are differently exposed to the same shocks. One key conclusion is that the observed increase in top wealth shares is perfectly compatible with a stationary economy in which the realized history happened to be especially favorable to the portfolios of the wealthy.

I propose a model of households' portfolio choice that builds on the novel observation that housing is a necessary good, which generates non-homotheticity in preferences. Such non-homotheticity is crucial for the model to generate the right schedule of portfolio shares: just like in the data – poor households mostly hold safe assets, the middle class is heavily invested in housing, and equity is *the* asset of the wealthy. This generates both scale dependence in rates of return and differential exposure to aggregate risk, which allows me to match both the level of wealth inequality and its dynamics.

I use the model to show that temporary shocks to returns have large and persistent effects on top wealth shares and ask how much of the observed rise in U.S. wealth inequality can be explained by changes in asset prices alone. The model is perfectly capable of replicating the evolution of top wealth shares in the data, which increased mostly due to the effect of abnormal returns to equity.

One of this paper's main implications is that, to have a better understanding of the evolution of wealth inequality we should turn our attention to prices. In particular, in this paper I take a very reduced-form approach to asset price determination. However, asset prices are equilibrium objects that are likely to be influenced by changes in the wealth distribution; that is, in this paper I show that asset price movements matter for wealth inequality but the opposite is likely true as well.

In particular, consider a model – such as the one proposed here – in which equity demand is increasing in wealth. In such model an increase in inequality generates additional demand for equity which, as long as supply cannot immediately adjust, reflects into higher equity prices. Such higher prices further increase stockholders' wealth and therefore wealth inequality, generating an amplification mechanism from wealth inequality to equity prices. This is exactly the approach I take in Cioffi (2021) where I show that, in a model with increasing equity demand, approximate aggregation does not hold and wealth inequality matters for asset-price determination. That is to say, asset prices matter for wealth inequality and inequality matters for prices.

References

Achdou, Y., J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll (2017). "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach". *Working Paper*.

Aguiar, M. and M. Bils (2015). "Has Consumption Inequality Mirrored Income Inequality?" *American Economic Review* 105(9), pp. 2725–2756.

Aiyagari, S. R. (1994). "Uninsured Idiosyncratic Risk and Aggregate Saving". *The Quarterly Journal of Economics* 109(3), pp. 659–684.

Atkeson, A. and M. Ogaki (1996). "Wealth-varying intertemporal elasticities of substitution: Evidence from panel and aggregate data". *Journal of Monetary Economics* 38(3), pp. 507–534.

Bach, L., L. E. Calvet, and P. Sodini (2020). "Rich Pickings? Risk, Return, and Skill in Household Wealth". *American Economic Review* 110(9), pp. 2703–2747.

Benhabib, J. and A. Bisin (2018). "Skewed Wealth Distributions: Theory and Empirics". *Journal of Economic Literature* 56(4), pp. 1261–1291.

Bewley, T. (1980). "Interest bearing money and the equilibrium stock of capital". *Mimeo*.

Blundell, R., M. Browning, and C. Meghir (1994). "Consumer Demand and the Life-Cycle Allocation of Household Expenditures". *The Review of Economic Studies* 61(1), pp. 57–80.

Browning, M. and T. F. Crossley (2000). "Luxuries Are Easier to Postpone: A Proof". *Journal of Political Economy* 108(5), pp. 1022–1026.

Calvet, L. E., J. Y. Campbell, F. Gomes, and P. Sodini (2021). "The Cross-Section of Household Preferences". *Working Paper*, p. 76.

Carmona, R. and F. Delarue (2018). *Probabilistic Theory of Mean Field Games with Applications II: Mean Field Games with Common Noise and Master Equations*. Vol. 84. Springer.

Catherine, S. (2020). "Countercyclical Income Risk and Portfolio Choices over the Life-Cycle". *Working Paper*.

Cioffi, R. A. (2021). "When Does Wealth Inequality Matter for Asset Prices?" *Working Paper*.

Cocco, J. F. (2005). "Portfolio Choice in the Presence of Housing". *The Review of Financial Studies* 18(2), pp. 535–567.

Duffie, D. and L. G. Epstein (1992). "Stochastic Differential Utility". *Econometrica* 60(2), pp. 353–394.

Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri (2020). "Heterogeneity and Persistence in Returns to Wealth". *Econometrica* 88(1), pp. 115–170.

Favilukis, J. (2013). "Inequality, stock market participation, and the equity premium". *Journal of Financial Economics* 107(3), pp. 740–759.

Favilukis, J., S. C. Ludvigson, and S. Van Nieuwerburgh (2016). "The Macroeconomic Effects of Housing Wealth, Housing Finance, and Limited Risk Sharing in General Equilibrium". *Journal of Political Economy* 125(1), pp. 140–223.

Flavin, M. and S. Nakagawa (2008). "A Model of Housing in the Presence of Adjustment Costs: A Structural Interpretation of Habit Persistence". *American Economic Review* 98(1), pp. 474–495.

Flavin, M. and T. Yamashita (2002). "Owner-Occupied Housing and the Composition of the Household Portfolio". *American Economic Review* 92(1), pp. 345–362.

Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016). "The Dynamics of Inequality". *Econometrica* 84(6), pp. 2071–2111.

Garriga, C. and A. Hedlund (2020). "Mortgage Debt, Consumption, and Illiquid Housing Markets in the Great Recession". *American Economic Review* 110(6), pp. 1603–1634.

Gomes, F. and A. Michaelides (2008). "Asset Pricing with Limited Risk Sharing and Heterogeneous Agents". *The Review of Financial Studies* 21(1), pp. 415–449.

Gomez, M. (2019). "Asset Prices and Wealth Inequality". *Working Paper*, p. 62.

Gomez, M. (2021). “Decomposing the Growth of Top Wealth Shares”. *Working Paper*.

Gomez, M. and E. Gouin-Bonenfant (2020). “A Q-Theory of Inequality”. *Working Paper*, p. 37.

Greenwald, D. L., M. Leombroni, H. Lustig, and S. Van Nieuwerburgh (2021). “Financial and Total Wealth Inequality with Declining Interest Rates”. *Working Paper*.

Guvenen, F. (2009). “A Parsimonious Macroeconomic Model for Asset Pricing”. *Econometrica* 77(6), pp. 1711–1750.

Guvenen, F., S. Ozkan, and J. Song (2014). “The Nature of Countercyclical Income Risk”. *Journal of Political Economy* 122(3), pp. 621–660.

Hanoch, G. (1977). “Risk Aversion and Consumer Preferences”. *Econometrica* 45(2), pp. 413–426.

Heathcote, J., K. Storesletten, and G. L. Violante (2020). “How Should Tax Progressivity Respond to Rising Income Inequality”. *Working Paper*.

Hubmer, J., P. Krusell, and A. A. Smith (2021). “Sources of U.S. Wealth Inequality: Past, Present, and Future”. *NBER Macroeconomics Annual* 35(1), pp. 391–455.

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

Jordà, Ò., K. Knoll, D. Kuvshinov, M. Schularick, and A. M. Taylor (2019). “The Rate of Return on Everything, 1870–2015”. *The Quarterly Journal of Economics* 134(3), pp. 1225–1298.

Kaplan, G., B. Moll, and G. L. Violante (2018). “Monetary Policy According to HANK”. *American Economic Review* 108(3), pp. 697–743.

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

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

Kuhn, M. and J.-V. Rios-Rull (2016). “2013 update on the U.S. earnings, income, and wealth distributional facts : a view from macroeconomics”. *Federal Reserve Bank of Minneapolis Quarterly Review* 37(1), pp. 2–73.

Kuhn, M., M. Schularick, and U. I. Steins (2020). “Income and Wealth Inequality in America, 1949–2016”. *Journal of Political Economy* 128(9), pp. 3469–3519.

Martinez-Toledano, C. (2020). “House Price Cycles, Wealth Inequality and Portfolio Reshuffling”. *Working Paper*.

McKay, A. (2013). “Search for financial returns and social security privatization”. *Review of Economic Dynamics* 16(2), pp. 253–270.

Meeuwis, M. (2020). “Wealth Fluctuations and Risk Preferences: Evidence from U.S. Investor Portfolios”. *Working Paper*, p. 100.

Merton, R. C. (1971). “Optimum consumption and portfolio rules in a continuous-time model”. *Journal of Economic Theory* 3(4), pp. 373–413.

Moskowitz, T. J. and A. Vissing-Jørgensen (2002). “The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?” *American Economic Review* 92(4), pp. 745–778.

Pakoš, M. (2011). "Estimating Intertemporal and Intratemporal Substitutions When Both Income and Substitution Effects Are Present: The Role of Durable Goods". *Journal of Business & Economic Statistics* 29(3), pp. 439–454.

Piazzesi, M. and M. Schneider (Jan. 2016). "Chapter 19 - Housing and Macroeconomics". In: *Handbook of Macroeconomics*. Ed. by J. B. Taylor and H. Uhlig. Vol. 2. Elsevier, pp. 1547–1640. DOI: `10.1016/bs.hesmac.2016.06.003`. URL: `http://www.sciencedirect.com/science/article/pii/S1574004816300167` (visited on 10/16/2019).

Piazzesi, M., M. Schneider, and S. Tuzel (2007). "Housing, consumption and asset pricing". *Journal of Financial Economics* 83(3), pp. 531–569.

Saez, E. and G. Zucman (2016). "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data". *The Quarterly Journal of Economics* 131(2), pp. 519–578.

Schaab, A. (2020). "Micro and Macro Uncertainty". *Working Paper*.

Schroder, M. and C. Skiadas (1999). "Optimal Consumption and Portfolio Selection with Stochastic Differential Utility". *Journal of Economic Theory* 89(1), pp. 68–126.

Smith, M., O. Zidar, E. Zwick, and C. Booth (2021). "Top Wealth in the United States: New Estimates and Implications for Taxing the Rich". *Working Paper*, p. 68.

Straub, L. (2019). "Consumption, Savings, and the Distribution of Permanent Income". *Working Paper*.

Vestman, R. (2019). "Limited Stock Market Participation Among Renters and Homeowners". *The Review of Financial Studies* 32(4), pp. 1494–1535.

Vissing-Jorgensen, A. (2002). "Towards an Explanation of Household Portfolio Choice Heterogeneity: Nonfinancial Income and Participation Cost Structures". *Working Paper* (w8884), w8884.

Wachter, J. A. and M. Yogo (2010). "Why Do Household Portfolio Shares Rise in Wealth?" *The Review of Financial Studies* 23(11), pp. 3929–3965.

Xavier, I. (2020). "Wealth Inequality in the US: the Role of Heterogeneous Returns". *Working Paper*.

Appendix

A. Theoretical Appendix

A.1. Proofs

Proof of Proposition 1

Incomplete, coming soon.

Proof of Proposition 2

Given the definition of $\alpha(x) = c(x)/x$, $x = c + r_n \cdot n$ is total consumption expenditure, and n and r_n are consumption of housing services and its user-cost, respectively, we need to prove that $\partial\alpha(x)/\partial x > 0$ and that

$$\lim_{x \to 0} \alpha(x) = 0$$
$$\lim_{x \to \infty} \alpha(x) = 1.$$

From the definition of $\alpha(x)$, we can immediately see that:

$$\alpha'(x) = \frac{1}{x}\left(c'(x) - \alpha(x)\right)$$

and, given x is always positive, the proof that $\alpha'(x) > 0$ boils down to showing that $c'(x) > \alpha(x)$.

The intratemporal Euler equation has the usual form:

$$r_n = \frac{u_n}{u_c} = \frac{\omega}{1-\omega}\frac{n^{-1/\varepsilon_h}}{c^{-1/\varepsilon_c}}$$

where we have used the definition of $u(c, n)$ in equation (3), which implies $u_c = (1-\omega)u^{1/\varepsilon_h}c^{-1/\varepsilon_c}$ and $u_n = \omega u^{1/\varepsilon_h}h^{-1/\varepsilon_h}$. We can then use the intratemporal Euler equation and the definition of x to find $\alpha(x)$:

$$\frac{c}{x} = \frac{1}{1 + r_n\left(\frac{1-\omega}{\omega}r_n\right)^{-\varepsilon_h} c^{\frac{\varepsilon_h}{\varepsilon_c}-1}}. \tag{17}$$

Then, to find $c'(x)$, we use the fact that $c(x)$ solves:

$$\begin{aligned} x &= c(x) + r_n\, n(c(x)) \\ &= c(x) + r_n \left(\frac{1-\omega}{\omega} r_n\right)^{-\varepsilon_h} c(x)^{\varepsilon_h/\varepsilon_c} \end{aligned}$$

where $n(c(x))$ came directly from the intratemporal Euler equation. Hence, we find:

$$1 = \frac{dc}{dx} + r_n \left(\frac{1-\omega}{\omega} r_n\right)^{-\varepsilon_h} \frac{\varepsilon_h}{\varepsilon_c} c(x)^{\frac{\varepsilon_h}{\varepsilon_c}-1} \frac{dc}{dx}$$

which immediately gets us $c'(x)$:

$$\frac{dc}{dx} = \frac{1}{1 + r_n \left(\frac{1-\omega}{\omega} r_n\right)^{-\varepsilon_h} \frac{\varepsilon_h}{\varepsilon_c} c(x)^{\frac{\varepsilon_h}{\varepsilon_c}-1}}.$$

Showing that $c'(x) > \alpha(x)$ therefore requires that

$$1 + r_n \left(\frac{1-\omega}{\omega} r_n\right)^{-\varepsilon_h} \frac{\varepsilon_h}{\varepsilon_c} c(x)^{\frac{\varepsilon_h}{\varepsilon_c}-1} < 1 + r_n \left(\frac{1-\omega}{\omega} r_n\right)^{-\varepsilon_h} c(x)^{\frac{\varepsilon_h}{\varepsilon_c}-1}$$

which is satisfied as long as $\varepsilon_h < \varepsilon_c$.

Now we have left to prove that $\alpha_0 = 0$ and $\alpha_\infty = 1$. Before starting, in the same way we obtained equation (17) above, we get

$$\frac{n}{x} = \frac{1}{r_n + \left(\frac{1-\omega}{\omega} r_n\right)^{\varepsilon_c} n^{\frac{\varepsilon_c}{\varepsilon_h}-1}}. \tag{18}$$

To find α_0 notice that x can go to 0 if and only if both c and n also go to 0. Because $\varepsilon_h < \varepsilon_c$, from equation (18) have that $n \to 0$ implies $\frac{n}{x} \to 1$. Viceversa, from equation (17), $c \to 0$ implies $\frac{c}{x} \to 0$. It immediately follows that

$$\alpha_0 \equiv \lim_{x \to 0} \frac{c}{c + r_n\, n} = 0.$$

To prove that $\alpha_\infty = 1$ notice that, as x goes to ∞ we have one of three cases:

1. $c \to \infty$ and $h \to \infty$
2. $c \to \infty$ and $h \leq \bar{h} < \infty$
3. $c \leq \bar{c} < \infty$ and $h \to \infty$

In case 1, equations (17) and (18) imply both that $c/x \to 1$ and that $n/x \to 0$. In case 2, again by equation (17), we have that $c/x \to 1$ which, together with $h \leq \bar{h}$, implies $\frac{c}{c + r_n\, n} \to 1$. Finally, in case 3 equation (18) implies that $n/x \to 0$ which, by the definition of x, can only happen if c also grows to infinity, which contradicts $c \leq \bar{c}$. Hence, in all three cases we have that

$$\alpha_\infty \equiv \lim_{x \to \infty} \frac{c}{c + r_n\, n} = 1$$

Proof of Corollary 2.1

Incomplete, coming soon.

The formula for RRA comes directly from applying Hanoch (1977), Theorem 1. Then, to show that $\frac{\partial RRA}{\partial x} < 0$ we simply differentiate $RRA(x)$ with respect to $\alpha(x)$ and use the fact that $\alpha'(x) > 0$.

Proof of Corollary 2.2

The elasticity of intertemporal substitution in the presence of recursive utility is defined as:

$$\text{EIS} \equiv -\frac{\partial \left(\frac{\mathrm{d} \ln c_t}{\mathrm{d}t} \right)}{\partial \left(\frac{\mathrm{d} \ln \Lambda_t}{\mathrm{d}t} \right)}$$

where $\Lambda(t)$ is the utility gradient with respect to non-housing consumption, which is given by:[29]

$$\Lambda(t) = \exp \left[\int_0^t f_v(u_s, v_s)\, \mathrm{d}s \right] f_u(u_t, v_t) u_c(c_t, n_t).$$

To ease notation, whenever it is unambiguous we will denote time derivative with dots, i.e. $\dot{c} = \partial c(t)/\partial t$. In the proof we will also make use of the intratemporal Euler equation from the proof of proposition 2, which also implies

$$\frac{\dot{n}}{n} = \frac{\varepsilon_h}{\varepsilon_c} \frac{\dot{c}}{c}.$$

[29] See Schroder and Skiadas 1999.

From the definition of Λ, it follows that

$$
\begin{aligned}
\dot{\Lambda} &= \Lambda \left\{ f_v + \frac{1}{f_u u_c} \frac{\partial (f_u u_c)}{\partial t} \right\} \\
&= \Lambda \left\{ f_v + \frac{1}{f_u u_c} \left(\dot{f}_u u_c + f_u \dot{u}_c \right) \right\} \\
&= \Lambda \left\{ f_v + \frac{\dot{f}_u}{f_u} + \frac{\dot{u}_c}{u_c} \right\} \\
&= \Lambda \left\{ f_v + \frac{1}{f_u} \left[f_{uu} \left(u_c \dot{c} + u_n \dot{n} \right) + f_{uv} \dot{v} \right] + \frac{1}{u_c} \left[u_{cc} \dot{c} + u_{cn} \dot{n} \right] \right\}
\end{aligned}
$$

Using the intratemporal Euler equation we also know that

$$
\begin{aligned}
u_{cc} &= \frac{1}{\varepsilon_h} \frac{u_c}{u} u_c - \frac{1}{\varepsilon_c} \frac{u_c}{c} \\
u_{cn} &= \frac{1}{\varepsilon_h} \frac{u_c}{u} u_n .
\end{aligned}
$$

Then, using the definition of f in equation (2), it is easy to show that

$$
\begin{aligned}
\frac{f_{uu}}{f_u} &= -\frac{1}{u\psi} \\
\frac{f_{uv}}{f_u} &= \frac{\psi^{-1} - \gamma}{1-\gamma} \frac{1}{v}
\end{aligned}
$$

which allow us to rewrite:

$$
\frac{\dot{\Lambda}}{\Lambda} = f_v + \frac{\psi^{-1} + \gamma}{1-\gamma} \frac{\dot{v}}{v} + \frac{u_c \dot{c} + u_n \dot{n}}{u} \left(\frac{1}{\varepsilon_h} - \frac{1}{\psi} \right) - \frac{1}{\varepsilon_c} \frac{\dot{c}}{c}.
$$

The intratemporal Euler equaiton $u_n = r_n u_c$ implies

$$
\begin{aligned}
u_c \dot{c} + u_n \dot{n} &= u_c x \left(\alpha(x) \frac{\dot{c}}{c} + (1 - \alpha(x)) \frac{\dot{n}}{n} \right) \\
&= u_c x \left(\alpha(x) + (1 - \alpha(x)) \frac{\varepsilon_h}{\varepsilon_c} \right) \frac{\dot{c}}{c} \\
&= (u_c c + u_n n) \left(\alpha(x) + (1 - \alpha(x)) \frac{\varepsilon_h}{\varepsilon_c} \right) \frac{\dot{c}}{c}
\end{aligned}
$$

which gives us

$$\frac{\dot{\Lambda}}{\Lambda} = f_v + \frac{\psi^{-1}+\gamma}{1-\gamma}\frac{\dot{v}}{v} - \left[\frac{1}{\varepsilon_c} - \frac{u_c c + u_n n}{u}\left(\alpha(x) + (1-\alpha(x))\frac{\varepsilon_h}{\varepsilon_c}\right)\left(\frac{1}{\varepsilon_h} - \frac{1}{\psi}\right)\right]\frac{\dot{c}}{c}$$

and, finally:

$$-\frac{\partial\left(\frac{\mathrm{dln}\, c_t}{\mathrm{d}t}\right)}{\partial\left(\frac{\mathrm{dln}\, \Lambda_t}{\mathrm{d}t}\right)} = \frac{\varepsilon_c}{1 - \frac{u_c c + u_n n}{u}\left(\alpha(x)\varepsilon_c + (1-\alpha(x))\varepsilon_h\right)\left(\frac{1}{\varepsilon_h} - \frac{1}{\psi}\right)}.$$

Then, to find ψ_0 and ψ_∞ we are left to find $\lim_{x\to 0}\frac{u_c c+u_n n}{u}$ and $\lim_{x\to\infty}\frac{u_c c+u_n n}{u}$. Using the definitions for u, u_c, and u_n we have:

$$\frac{u_c c + u_n n}{u} = \frac{(1-\omega)c^{1-\varepsilon_c^{-1}} + \omega n^{1-\varepsilon_h^{-1}}}{(1-\omega)A^{-1}c^{1-\varepsilon_c^{-1}} + \omega n^{1-\varepsilon_h^{-1}}} \tag{19}$$

where $A \equiv (1-\varepsilon_c^{-1})/(1-\varepsilon_h^{-1})$. Then, to find the above two limits define the numerator and denominator of equation (19) as $f(x)$ and $g(x)$, respectively. It follows that

$$f'(x) = (1-\omega)(1-\varepsilon_c^{-1})c(x)^{-1/\varepsilon_c}c'(x) + \omega(1-\varepsilon_h^{-1})h(x)^{-1/\varepsilon_h}n'(x).$$

Then, use the fact that $n(x) = B^{-\varepsilon_h}c(x)^{\varepsilon_h/\varepsilon_c}$ where $B = \left(\frac{1-\omega}{\omega}r_n\right)$, to rewrite:

$$\begin{aligned} f'(x) &= (1-\omega)(1-\varepsilon_c^{-1})c(x)^{-1/\varepsilon_c}c'(x) + \omega(1-\varepsilon_h^{-1})Bc(x)^{-1/\varepsilon_c}h'(x) \\ &= (1-\omega)(1-\varepsilon_c^{-1})c(x)^{-1/\varepsilon_c}\left[c'(x) + A^{-1}r_n n'(x)\right]. \end{aligned}$$

Analogously, we also find

$$g'(x) = (1-\omega)A^{-1}(1-\varepsilon_c^{-1})c(x)^{-1/\varepsilon_c}\left[c'(x) + r_n n'(x)\right]$$

and, using L'Hopital rule, we can rewrite:

$$\lim\frac{u_c c + u_n n}{u} = \lim\frac{f'(x)}{g'(x)} = \lim A\frac{c'(x) + A^{-1}r_n n'(x)}{c'(x) + r_n n'(x)}.$$

Using the fact that $x = c(x) + r_n n(x)$ (and therefore $c'(x) + r_n n'(x) = 1$, this means we only need to find

$$\begin{gathered} \lim_{x\to 0} Ac'(x) + r_n n'(x) \\ \lim_{x\to\infty} Ac'(x) + r_n n'(x). \end{gathered}$$

However, recall from the proof of proposition 2 that

$$
\begin{aligned}
\frac{dc}{dx} &= \frac{1}{1 + r_n \frac{\varepsilon_h}{\varepsilon_c} B^{-\varepsilon_h} c(x)^{\frac{\varepsilon_h}{\varepsilon_c}-1}} \\
\frac{dn}{dx} &= \frac{1}{r_n + \frac{\varepsilon_h}{\varepsilon_c} B n(x)^{\frac{\varepsilon_c}{\varepsilon_h}-1}}.
\end{aligned}
$$

Given the assumption of $\varepsilon_h < \varepsilon_c$, we have

$$
\begin{aligned}
\lim_{x\to 0} c'(x) &= 0 \qquad && \lim_{x\to\infty} c'(x) = 1 \\
\lim_{x\to 0} n'(x) &= \frac{1}{r_n} \qquad && \lim_{x\to\infty} n'(x) = 0
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\lim_{x\to 0} \frac{u_c c + u_n n}{u} &= \lim_{x\to 0} \; Ac'(x) + r_n n'(x) = 1 \\
\lim_{x\to\infty} \frac{u_c c + u_n n}{u} &= \lim_{x\to\infty} Ac'(x) + r_n n'(x) = \frac{\varepsilon_h}{1-\varepsilon_h}\frac{1-\varepsilon_c}{\varepsilon_c}.
\end{aligned}
$$

Using the above two limits, and the fact that $\alpha_0 = 0$ and $\alpha_\infty = 1$, finally gives us ψ_0 and ψ_∞.

Proof of Proposition 3

Incomplete, coming soon

The proof here follows similar step to the proof of proposition 3, steps 2 and 3 in Gomez (2021).

A.2. HJB

The HJB equation is:

$$
0 = \max\left\{ f(u(c,n), v) + \mathcal{A}v \right\}
$$

where

$$
\begin{aligned}
\mathcal{A} &= 1_{\{p=0\}}\left(\mathcal{L}^0+\mathcal{P}^0\right)+1_{\{p=1\}}\left(\mathcal{L}^1+\mathcal{P}^1\right)+1_{\{h=0\}}\mathcal{H}^0+1_{\{h\neq 0\}}\mathcal{H}^+ +\mathcal{Z}\\
\mathcal{L}^0 v &= \mu_a(\mathbf{x})\frac{\partial}{\partial a}v\\
\mathcal{L}^1 v &= \left(\mu_a(\mathbf{x})\frac{\partial}{\partial a}+\sigma_a(\mathbf{x})^2\frac{1}{2}\frac{\partial^2}{\partial a^2}+\rho_z\sigma_a(\mathbf{x})\sigma_z(z)\frac{\partial}{\partial a\partial z}\right)v\\
\mathcal{P}^0 v &= \lambda_0^p\left[\max\left\{v(a-\kappa_0^p,h,z,1-p),v(a,h,z,p)\right\}-v(a,h,z,p)\right]\\
\mathcal{P}^1 v &= \lambda_1^p\left[\max\left\{v(a-\kappa_1^p,h,z,p),v(a,h,z,1-p)\right\}-v(a,h,z,p)\right]\\
\mathcal{H}^0 v &= \lambda_B^h\max\left\{\max_{h'} v(a-h'-\kappa_B^h(h'),h',z,p)-v(a,0,z,p),0\right\}\\
\mathcal{H}^+ v &= \lambda_S^h\max\left\{v\left(a+h-\kappa_S^h(h),0,z,p\right)-v(a,h,z,p),0\right\}+\mu_h(h)\frac{\partial}{\partial h}v(a,h,z,p)\\
\mathcal{Z}v &= \left(\mu_z(z)\frac{\partial}{\partial z}+\sigma_z(z)\frac{1}{2}\frac{\partial^2}{\partial z^2}\right)v
\end{aligned}
$$

B. Numerical Appendix

B.1. Finite Differences

Incomplete, coming soon

Once we have a finite difference approximation $\mathbb{A}$ of the infinitesimal generator $\mathcal{A}$ we can solve the problem. At each iteration n, the value function v needs to solve the (possibly non-linear) system of equations:

$$0 = f(u,v)+\mathbb{A}v$$

Achdou et al. (2017) stress the importance of using an implicit scheme. With recursive preferences the implicit scheme is slightly more complicated. The reason is that, in the separable utility case f is additive in u and v and we can easily solve the system with $0 = f(u^n, v^{n+1})+\mathbb{A}v^n$. Viceversa, with recursive preferences this separation does not occur and solving the implicit scheme in the same fashion would imply solving a non-linear system of equation, which would significantly slow down the computations. However, we can actually split f into a linear and a non-linear component to use a semi-implicit scheme. We can in fact write f as:

$$f(u^n,v^n,v^{n+1}) = \frac{\rho+\zeta}{1-\psi^{-1}}\left[(1-\gamma)v^n\right]^{\frac{\psi^{-1}-\gamma}{1-\gamma}}(u^n)^{1-\psi^{-1}}-(\rho+\zeta)\frac{1-\gamma}{1-\psi^{-1}}v^{n+1}$$

at which point we simply need to solve the system

$$0 = f(u^n, v^n) + (\rho + \zeta)\frac{1-\gamma}{1-\psi^{-1}}v^n + \left(\mathbb{A} - (\rho+\zeta)\frac{1-\gamma}{1-\psi^{-1}}\mathcal{I}\right)v^{n+1}$$

B.2. Participation Costs

To avoid cluttering the problem with variables unrelated to the participation cost problem we here analyze the simplest 2-asset model without any labor income risk, financial wealth a is therefore the only continuous state variable of the household problem. There are two participation states denoted by 0 and 1: households in state 0 only have access to bonds b, while households in state 1 can freely move assets between bonds and equities s. Bonds earn a risk-free rate of return r, while equities earn a risky return R with volatility σ. Households receive an opportunity to jump from state 0 to state 1 (entering the stock market) with intensity λ_0 and the opportunity to jump from state 1 to state 0 (exiting the stock market) with intensity λ_1. In addition, households that get hit by the participation shock can only enter the stock market if they accept to pay an entry cost κ_0^s; viceversa, households who get hit by the exit shock can only stay in the stock market if they accept to pay a staying cost κ_1^s.

The household problem therefore reads:

$$\begin{aligned}
\rho v_0(a) &= \max_c u(c) + v_0'(a)(z + ra - c) + \lambda_0\left[\max\left\{v_1(a-\kappa_0^p), v_0(a)\right\} - v_0(a)\right] \\
\rho v_1(a) &= \max_{c,s} u(c) + v_1'(a)(z + ra + (R-r)s - c) + v_1''(a)\frac{(s\sigma)^2}{2} + \\
&\quad + \lambda_1\left[\max\left\{v_1(a-\kappa_1^p), v_0(a)\right\} - v_1(a)\right]
\end{aligned}$$

The model nests both a model with a fixed cost of participation (with $\lambda_1 = 0$) and a model with a flow cost of participation (when $\lambda_1 \to \infty$). Making the opportunity to enter/exit the stock market stochastic has the merit of making the numerical solution significantly simpler while also being able to arbitrarily approximate the continuous entry/exit decision: in fact, using larger and larger λs allows us to get closer and closer to the continuous entry/exit problem without having to use the (more complicated) methods needed to solve variational inequalities.

The finite-differences approximation of the above HJB is then:

$$\begin{aligned}
\frac{v_{i,p}^{n+1} - v_{i,p}^n}{\Delta} + \rho v_{i,p}^{n+1} &= u(c_{i,p}^n) + v_{i,p}^{a,F,n+1}\cdot\left(\dot{a}_{i,p}^{c,F,n}\right)^+ + v_{i,p}^{a,B,n+1}\cdot\left(\dot{a}_{i,p}^{c,B,n}\right)^- + \\
&\quad + v_{i,p}^{aa,n+1}\cdot\frac{\left(s_{i,p}^n\sigma\right)^2}{2} + \lambda_p\left[\max\left\{v_{i',p}^{n+1}, v_{i,0}^{n+1}\right\} - v_{i,p}^{n+1}\right]
\end{aligned}$$

where $v_{i',p} = v_1(a_i - \kappa_p^s)$.

When specifying the above problem numerically, there are three things that one needs to be especially careful with:

1. The a-state changes after the shock hits
2. $a_i - \kappa_p^s$ does not necessarily fall on a grid point (this is especially true in the presence of non-uniform grids)
3. For any two different $\hat{v}, \tilde{v}$, to compute max $\left\{\hat{v}^{n+1}, \tilde{v}^{n+1}\right\}$ we would need to know their values at iteration $n+1$

Notice that 1) simply implies that there will be mass on some off-diagonal elements of $\mathbb{A}$. Additionally, in the presence of non-uniform grids this implies that the $\mathbb{A}$ matrix will not have a band-matrix structure. This is not a problem if one uses general sparse solvers, but it does exclude the possibility of exploiting the band structure to simplify the construction of $\mathbb{A}$ and/or the solution of the linear system.

To tackle 2) we assign mass to both the two closest points to $a_i - \kappa_p^s$ in proportion to their relative distance: for each grid point a_i, denote by $a_{i',p}$ and $a_{i'+1,p}$ the two grid points closest to $a_i - \kappa_p^s$. That is,

$$a_{i',p} \equiv \sup \left\{ a \in \{a_1, a_2, \dots, a_I\} : a < a_i - \kappa_p^s \right\}.$$

We can then compute the mass assigned to grid-point i' as $\omega_{i',p}^s = \frac{a_{i'+1,p} - (a_i - \kappa_p^s)}{a_{i'+1,p} - a_{i',p}}$ and the mass assigned to grid-point $i'+1$ as $1 - \omega_{i',p}^s$. Notice also that, although in most cases the household would choose to pay (or not) the cost for both $a_{i',p}$ and $a_{i'+1,p}$, there might be edge cases when the household would only choose to pay the cost for $a_{i'+1,p}$ but not for $a_{i',p}$. Because the rows of $\mathbb{A}$ need to always sum to 1, we let the household decide based on the average value and she is forced to accept the "bet" and move to both points.

Finally, concerning 3), we use a semi-implicit scheme consistent with how the other choice variables are determined, that is the choice of whether to enter or exit the stock market is based on the value function at iteration n, but the change in value is based on the value function at iteration $n+1$. Hence, if we define by $\mathcal{I}_{i,p}^n$ the household decision to participate or not, that is:

$$\mathcal{I}_{i,p}^n \equiv \mathbf{1} \left\{ v_1^n(a_i) < \omega_{i',p}^p v_1^n(a_{i',p}) + (1 - \omega_{i',p}^p) v_1^n(a_{i'+1,p}) \right\}$$

we can rewrite the finite difference approximation as

$$
\begin{aligned}
\frac{v_{i,p}^{n+1}-v_{i,p}^{n}}{\Delta}+\rho v_{i,p}^{n+1} &= u(c_{i,p}^{n})+v_{i,p}^{a,F,n+1}\cdot\left(\dot{a}_{i,p}^{c,F,n}\right)^{+}+v_{i,p}^{a,B,n+1}\cdot\left(\dot{a}_{i,p}^{c,B,n}\right)^{-}\\
&+v_{i,p}^{aa,n+1}\cdot\frac{\left(s_{i,p}^{n}\sigma\right)^{2}}{2}-\lambda_p(v_{i,p}^{n+1}-v_{i,0}^{n+1})\\
&+\lambda_p\left[\left(\omega_{i',p}^{p}v_1^{n}(a_{i',p})+(1-\omega_{i',p}^{p})v_1^{n}(a_{i'+1,p})\right)-v_{i,0}^{n+1}\right]\cdot\mathcal{I}_{i,p}^{n}
\end{aligned}
$$

B.3. Kolmogorov Forward Equation

Incomplete, coming soon

The following section covers how to approximate the distribution over states in the presence of non-uniform grid. To simplify notation we here work with a simplified problem with only two states a and z, and no aggregate risk. The results immediately generalize to more states; the presence of aggregate risk instead obviously changes the KFE that needs to be solved, but has no impact on the numerical approximation of the cross-sectional distribution in the presence of non-uniform grids.

Following Achdou et al. (2017) to solve the KFE with non-uniform grids we can work with the rescaled density $\tilde{g}=Dg$ where D is a diagonal matrix with elements $\tilde{\Delta}_\iota$ that appropriately adjust for the presence of non-uniform grids. In particular, if $\mathbb{A}$ is the intensity matrix associated with the HJB problem above we will be solving

$$
\frac{\tilde{g}^{n+1}-\tilde{g}^{n}}{\Delta t}=\mathbb{A}^{T}\tilde{g}^{n+1}
$$

To understand why the above works notice that $\sum_{\iota=1}^{\mathcal{I}} g_\iota\tilde{\Delta}_\iota$ is an approximation of the integral of g using a trapezoidal rule, and we denote by ι the "composite" index of a and z, i.e. $\iota=(i,k)$.

$$
\begin{aligned}
\int g(a,z)dadz &\approx \sum_{i=1}^{I-1}\sum_{k=1}^{K-1}\frac{(a_{i+1}-a_i)}{2}\frac{(z_{k+1}-z_k)}{2}\left(g_{i,k}+g_{i,k+1}+g_{i+1,k}+g_{i+1,k+1}\right)\\
&=\sum_{i=1}^{I}\sum_{k=1}^{K}g_{i,k}\tilde{\Delta}a_i\tilde{\Delta}z_k\\
&=\sum_{\iota=1}^{\mathcal{I}}g_\iota\cdot\tilde{\Delta}_\iota
\end{aligned}
$$

where $\mathcal{I}=I\cdot K$, $\tilde{\Delta}_\iota=\tilde{\Delta}a_i\cdot\tilde{\Delta}z_k$, and the $\tilde{\Delta}$ operator for a generic variable x indexed by j with

$j \in \{1, 2, \ldots, J\}$ is

$$\tilde{\Delta} x_j = \begin{cases} \frac{\Delta^F}{2} x_j & \text{if } j = 1 \\ \frac{\left(\Delta^B + \Delta^F\right)}{2} x_j & \text{if } 1 < j < J \\ \frac{\Delta^B}{2} x_j & \text{if } j = J \end{cases}$$

where $\Delta^F x_j = \Delta^B x_{j+1} = x_{j+1} - x_j$.

After having solved for $\tilde{g}$ we can then back out the true distribution using $g = D^{-1}\tilde{g}$ and, to compute moments, we can use the fact that $\sum_{\iota=1}^{\mathcal{I}} g_\iota \tilde{\Delta}_\iota$ approximates the integral of g. For instance, to find average earnings we would use:

$$\begin{aligned} \mathbb{E}(z) = \int zg(a,z)dadz &\approx \sum_{\iota=1}^{\mathcal{I}} z_\iota g_\iota \tilde{\Delta}_\iota \\ &= \sum_{i=1}^{I} \sum_{k=1}^{K} z_k g_{i,k} \tilde{\Delta} a_i \tilde{\Delta} z_k \end{aligned}$$

C. *Data Appendix*

Incomplete, coming soon.