Financial Frictions and the Wealth Distribution

Jesús Fernández-Villaverde

University of Pennsylvania, NBER, and CEPR

Samuel Hurtado — Banco de España

Galo Nuño — Banco de España*

June 24, 2019

Abstract

We postulate a continuous-time model of financial frictions *à la* Basak and Cuoco (1998) and Brunnermeier and Sannikov (2014), but with a non-trivial distribution of wealth among households. We show how such a model can be efficiently computed, despite its substantial non-linearities and time-varying wealth distributions, using tools from machine learning. We illustrate how the model can be structurally estimated with a likelihood function with aggregate and micro observations, using tools from inference with diffusions. Regarding results, we document, first, the strong non-linearities created by financial frictions. Second, we report the existence of multiple stochastic steady-states with properties that differ from the deterministic steady state along important dimensions. Third, we illustrate how the generalized impulse-response functions of the model are highly state-dependent. In particular, we find that the recovery after a negative aggregate shock is more sluggish when the financial expert is more leveraged. Fourth, we prove that wealth heterogeneity matters in this economy because of the asymmetric responses of household consumption decisions to aggregate shocks.

Keywords: Heterogeneous agents; aggregate shocks; continuous-time; machine learning; neural networks; structural estimation; likelihood functions.

JEL codes: C45, C63, E32, E44, G01, G11.

*We thank Emmanuel Farhi, Xavier Gabaix, Mark Gertler, Ben Moll, Gianluca Violante, Ivan Werning, and participants at numerous seminars for pointed comments. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the European Central Bank or the Bank of Spain.

1 Introduction

This paper argues that, in a standard heterogeneous agents model with financial frictions, idiosyncratic individual shocks interact with exogenous aggregate shocks to generate endogenously time-varying levels of leverage and aggregate risk. In particular, such an economy displays highly nonlinear behavior, with i) multiple stochastic steady states; ii) multimodal and skewed ergodic distributions of endogenous variables; and iii) strong state-dependence on the responses of endogenous variables to aggregate shocks.

To do so, we build, compute, and estimate using the likelihood approach a continuous-time neoclassical growth model with heterogeneous households subject to labor productivity shocks as in Aiyagari (1994) and Krusell and Smith (1998). We enrich the model with a financial expert (a stand-in for banks or financial intermediaries), limited financial markets participations as in Basak and Cuoco (1998) and Brunnermeier and Sannikov (2014), and an aggregate shock to physical capital. Households save in a noncontingent bond to self-insure against their labor productivity shocks and the variations in wages and interest rates induced by aggregate risk. The financial expert cannot issue state-contingent assets (i.e., outside equity), but it can issue bonds to leverage its equity and accumulate capital that is rented to a representative firm. Only the expert can hold capital.

The interaction between the demand for bonds by the households and the supply of bonds by the financial expert begets, for parameter values that match important aspects of the U.S. data and maximize the likelihood function, multiple stochastic steady states (SSSs). This multiplicity occurs even despite the model having a unique deterministic steady state (DSS).[1] In particular, we will have a high-leverage SSS and a low-leverage SSS, each of them with their basin of attraction and endogenous levels of aggregate risk (there is a third, unstable SSS that we do not need to discuss).

The intuition for the existence of multiple stable SSSs is as follows. In the basin of attraction of the high-leverage SSS, a negative aggregate shock to capital has dire effects. The financial expert's net wealth is greatly eroded since her relatively small equity must absorb all the losses to capital. The economy suffers a deep and prolonged recession as the expert struggles to rebuild its equity and accumulate capital. This recession translates into persistently low wages. Since the households want to self-insure against the risk of low wages, their demand for bonds is high: this SSS has a higher share of wealthy households and more wealth and income inequality than the DSS. The high demand for bonds translates into a low risk-free interest rate for the bond and a high expected excess return for capital, which, in turn, induce the financial expert to lever aggressively. In other words, a high-leverage region of the economy is a region with

[1] A SSS, also known as the risky steady state, is a fixed-point of the equilibrium conditions of the model when the *realization* of the aggregate shock is zero. A DSS is a fixed-point of the equilibrium conditions of the model when the *volatility* of the aggregate shock (but not of idiosyncratic shocks) is zero.

high aggregate volatility, even when the variance of the aggregate shock is the same as in the low-leverage region.

To document the higher persistence of shocks when leverage is high, we compute the generalized impulse response functions and the distributional impulse response functions to a negative capital shock in different points of the state space. The responses are very similar on impact, but the ensuing recession is more persistent when leverage is high due to the dynamics of aggregate household consumption and capital accumulation. In a high-leverage economy, the expected path of interest rates is more persistent as the financial expert slowly restores her net wealth. This persistence induces a less severe decline in consumption among wealthy households and sluggish capital accumulation by the expert because the more persistent interest rates justify less intertemporal substitution.

In comparison, when the financial expert is lowly levered, the recessions after a negative aggregate shock are mild. Thus, the demand for bonds by the household is low, the risk-free rate high, the expected excess return low, and these prices sustain the low leverage by the expert. With low leverage, the endogenous aggregate risk of the economy is smaller: the economy responds more mutedly to the same aggregate shock than when leverage is high.

For our baseline parameter values, the economy will spend more time, on average, around the high-leverage SSS than around the low-leverage SSS. We will show, however, how this finding varies as we change the volatility of idiosyncratic and aggregate shocks.

The multiplicity of SSSs does not imply a multiplicity of equilibria: in our model, we find a unique stochastic equilibrium. Whether the economy has high or low leverage is a consequence of the past aggregate shocks. Sometimes, while the economy is traveling in the basin of attraction of the high-leverage SSS, a sequence of aggregate shocks will move it to the basin of attraction of the low-leverage SSS (and vice versa). Thus, the interaction of financial frictions with the consumption-saving decisions made by the agents generates endogenous time-variation in aggregate risk: after some sequences of shocks, the economy will be fragile and prone to severe recessions due to high leverage and, after other sequences, the economy will be more resilient and less volatile thanks to low leverage.

The argument above also explains why heterogeneity at the household level is crucial for our argument and why the "quasi-aggregation result" of Krusell and Smith (1998) breaks down in our environment. As we reduce idiosyncratic labor productivity risk, the precautionary saving motive becomes smaller and the high-leverage SSS disappears: the households do not demand enough bonds to sustain the high-leverage SSS. That is, the limit version of the model with a representative household behaves very differently from our baseline economy.

Conversely, when we increase the level of idiosyncratic labor productivity risk, the low-leverage SSS disappears and only the high-leverage SSS survives. In this case, households are so concerned about their own idiosyncratic labor risk that they demand enough bonds as to push

the risk-free rate sufficiently low that only a high level of leverage can be a SSS (paradoxically, increasing the aggregate risk they face). In practical terms, higher micro turbulence (e.g., more volatile labor markets) translates into higher aggregate volatility even when the variance of aggregate shocks is constant. This result suggests that the rise in wealth inequality among capital owners documented by Alvaredo et al. (2017) before the financial crisis of 2007 is linked with i) the increase in financial debt and leverage witnessed during the same period (Adrian and Shin, 2010, and Nuño and Thomas, 2017); ii) low risk-free interest rates; and ii) the heightened fragility of the economy to adverse shocks.

This mechanism is, as far as we know, new in the literature. It is different from the "paradox of volatility" in Brunnermeier and Sannikov (2014). In their model, a low volatility of aggregate shocks leads to higher leverage by the financial expert and, thus, deep recessions when a large shock hits the economy. In our model, a high volatility of idiosyncratic shocks leads to higher leverage and, thus, larger recessions on average, even when the aggregate shocks are not particularly large. The mechanism is also different from Kumhof et al. (2015) because, in our model, the change in the wealth distribution is endogenous and not the consequence of an exogenous shock to the income received by top earners.

The documentation of the importance of individual heterogeneity in models with financial frictions is a novel contribution of our paper. The insight that, in the presence of financial frictions, the wealth distribution is a state of the economy is not new. The idea is already discussed by Bernanke et al. (1999) and Kiyotaki and Moore (1997). However, the previous literature has focused on the case where there is *between-agents* heterogeneity, but no *within-agents* heterogeneity. *Between-agents* heterogeneity means that capital owners are different from experts. No *within-agents* heterogeneity means there is just one capital owner (a representative household) and a representative expert (or, perhaps, different capital owners and experts but where the heterogeneity is collapsed into an economic-wide average of leverage). Our paragraphs above describe how *within-agents* heterogeneity is crucial to understand the aggregate consequences of financial frictions.

Researchers have largely avoided studying economies with *within-agents* heterogeneity and financial frictions because characterizing this class of models is challenging: since the wealth distributions is an infinite-dimensional object, standard dynamic programming techniques cannot be employed. To overcome this problem, our paper provides new methodological tools for the global, nonlinear solution and estimation of heterogeneous agent models with aggregate shocks.

More concretely, we rely on the machine learning literature and employ a neural network to obtain a flexible approximation of the perceived law of motion (PLM) of the cross-sectional distributions of assets (the expert's equity and households' bonds) with a finite set of moments in the spirit of Krusell and Smith (1998). Naturally, other machine learnings schemes may also be proposed (or, for the matter, other nonlinear universal approximators such as series expan-

sions or splines). However, our approach is particularly convenient, both regarding theoretical properties and practical considerations.

First, the *universal approximation theorem* (Hornik et al., 1989; Cybenko, 1989) states that a neural network can approximate any unknown Borel measurable function. Second, the neural network breaks the curse of dimensionality for a large class of approximated functions, which allows our method to be extended to much richer environments with many state variables (or more moments). Third, the neural network can be efficiently trained using a combination of the *gradient descent* and the *back-propagation* algorithms. Fourth, not only is our algorithm efficient and easy to code, but also particularly amenable to massive parallelization in graphic or tensor processing units. Finally, our approach reflects in a fairly transparent way the self-justified equilibria nature of the "bounded rationality" solution of most heterogeneous agents models (Kubler and Scheidegger, 2018). The PLM is computed based on the samples drawn in the simulation of paths within the aggregate ergodic distribution. The agents employ the neural network to extrapolate the dynamics outside of the equilibrium region.

In comparison, in Krusell and Smith (1998) and most of the subsequent literature, the PLM of the aggregate variables is approximately linear in the endogenous state variables (but nonlinear in the exogenous states, since the coefficients of the regression are allowed to vary across shocks). This traditional PLM is a poor choice in our model, in which the nonlinearities of the endogenous state variables play a central role and where, because of the use of continuous time, exogenous states are incorporated into the endogenous states instantaneously. In fact, we will document how a naive implementation of the Krusell and Smith (1998) algorithm in our economy, and even of refinements such as Chebyshev polynomials, deliver a much worse numerical performance. Neural networks allow us to be much more flexible and avoiding having to specify, ex-ante, any concrete nonlinear structure of the PLM.

Continuous time helps us to characterize much of the equilibrium dynamics analytically and to worry only about local derivatives (instead of the whole shape of equilibrium functions) even when solving the model globally. However, nothing essential depends on this choice and we could replicate our approach –with somewhat heavier computational costs– in discrete time. Achdou et al. (2017) and Nuño and Thomas (2017) provide a more general presentation of the advantages of continuous-time methods.

We also illustrate how a fully nonlinear model can be structurally estimated with a likelihood function with aggregate and micro observations using tools from inference with diffusions (Lo, 1988). Such likelihood is computationally straightforward once we have solved the model with the approach outlined above: it just amounts to transposing a matrix. In particular, it avoids having to resort to more computationally intensive algorithms such as the particle filter (Fernández-Villaverde and Rubio-Ramírez, 2007). Then, we take our model to the data by matching some features of the U.S. economy, such as the average leverage of the corporate

sector, and estimating the volatility of the aggregate shocks by maximum likelihood.

Our work relates to several important threads in macroeconomics. First, we follow the macro-finance literature pioneered by Basak and Cuoco (1998), Adrian and Boyarchenko (2012), He and Krishnamurthy (2012, 2013), and Brunnermeier and Sannikov (2014), among others. As we mentioned before, most of these papers only consider *between-agents* heterogeneity, but no *within-agents* heterogeneity. Instead, we deal with a non-trivial wealth distribution across households and show why such heterogeneity matters.

Our paper also contributes to the literature on global solution methods for heterogeneous agents models with aggregate shocks such as Den Haan (1996), Den Haan (1997), Algan et al. (2008), Reiter (2009, 2010), Den Haan and Rendahl (2010), Maliar et al. (2010), Sager (2014), and Pröhl (2015) (Algan et al. 2014 is a recent survey of the field). To the best of our knowledge, ours is the first paper to generalize the celebrated algorithm of Krusell and Smith (1998) to accommodate a universal nonlinear law of motion in the endogenous state variables.[2]

Finally, our paper builds on the nascent literature on the application of machine learning techniques to compute dynamic equilibrium models. The proposed methods have so far been concerned with the solution of high dimensional dynamic programming problems. Scheidegger and Bilionis (2017) combine Gaussian process regression with an active subspace method to solve discrete-time stochastic growth models of up to 500 dimensions. Duarte (2018) employs a reinforcement learning algorithm together with a neural network to solve a two-sector model with 11 state variables. In contrast, our machine learning algorithm is used to provide a nonlinear forecast of aggregate variables within the model. In this respect, our paper reconnects with an early literature using neural networks to model bounded rationality and learning, such as Barucci and Landi (1995), Cho (1995), Cho and Sargent (1996), and Salmon (1995).

Our methodology may also be useful to analyze other heterogeneous agents models with aggregate shocks. An obvious candidate is the heterogeneous agent New Keynesian (HANK) model with a zero lower bound (ZLB) on the nominal interest rates, such as Auclert (2016), Auclert and Rognlie (2018), Gornemann et al. (2012), Kaplan et al. (2018), Luetticke (2015), and McKay et al. (2016). The ZLB introduces a nonlinearity in the state space of aggregate variables that cannot be addressed either with local methods or with global methods based on linear laws of motion. Other potential candidates include any model in which the "quasi-aggregation" result breaks down and, thus, we require higher-order moments to provide an accurate characterization of the equilibrium dynamics of the agents' distribution.

[2] Ahn et al. (2017) introduce a related method to compute the solution to heterogeneous agent models with aggregate shocks in continuous time. However, theirs is a local solution, based on first-order perturbation around the DSS and, thus, unable to analyze the class of nonlinearly-related questions posed by our paper.

2 Model

We postulate a continuous-time, infinite-horizon model. Three types of agents populate our economy: a representative firm, a representative financial expert, and a continuum of households. There are two assets: a risky asset, which we identify as capital, and a risk-free one, which we call bonds. Only the expert can hold the risky asset. In the interpretation implicit in our terminology, this is because the expert is the only agent with knowledge in accumulating capital. However, other interpretations, such as the expert standing in for banks or other financial intermediaries, are possible. In contrast, households can lend to the expert at the riskless rate, but cannot hold capital themselves, as they lack the required skill to handle it. The expert cannot issue outside equity, but she can partially finance her holdings of the risky asset by issuing bonds to households. Together with market clearing, our assumptions imply that the economy has a risky asset in positive net supply, capital, and a risk-free asset in zero net supply, bonds. As will become clear below, there is no need to separate between the representative firm and expert, and we could write the model consolidating both agents in a single type. Keeping both agents separate, though, clarifies the exposition. Also, we only introduce heterogeneity on the side of the households (and not among the experts or the firms) because, in our environment, it is the type of heterogeneity that generates the most interesting aggregate consequences.

2.1 The firm

A representative firm rents aggregate capital, K_t, and aggregate labor, L_t, to produce output with a Cobb-Douglas technology:

$$Y_t = F(K_t, L_t) = K_t^{\alpha} L_t^{1-\alpha}.$$

Since input markets are competitive, wages, w_t, are equal to the marginal productivity of labor:

$$w_t = \frac{\partial F\left(K_t, L_t\right)}{\partial L_t} = (1-\alpha)\frac{Y_t}{L_t} \tag{1}$$

and the rental rate of capital, rc_t, is equal to the marginal productivity of capital:

$$rc_t = \frac{\partial F\left(K_t, L_t\right)}{\partial K_t} = \alpha\frac{Y_t}{K_t}. \tag{2}$$

During production, capital depreciates at a constant rate δ and receives a growth rate shock Z_t that follows a Brownian motion with volatility σ. Thus, aggregate capital evolves as:

$$\frac{dK_t}{K_t} = (\iota_t - \delta)\, dt + \sigma dZ_t, \tag{3}$$

where ι_t is the reinvestment rate per unit of capital that we will characterize below. The capital growth rate shock is the only aggregate shock to the economy. Following convention, the rental rate of capital rc_t is defined over the capital contracted, K_t, and not over the capital returned after depreciation and the growth rate shock. Thus, we define the instantaneous return rate on capital dr_t^k as:

$$dr_t^k = (rc_t - \delta)\, dt + \sigma dZ_t.$$

The coefficient of the time drift, $rc_t - \delta$, is the profit rate of capital, equal to the rental rate of capital less depreciation. The volatility σ is the capital gains rate.

2.2 The expert

The expert holds capital $\widehat{K}_t$ (we denote variables related to the expert with a caret). She rents this capital to the representative firm. To finance her holding of $\widehat{K}_t$, the expert issues risk-free debt $\widehat{B}_t$ at rate r_t to the households. The financial frictions in the model come from the fact that the expert cannot issue state-contingent claims (i.e., outside equity) against $\widehat{K}_t$. In particular, the expert must absorb all the risk from holding capital.

The net wealth (i.e., inside equity) of the expert, $\widehat{N}_t$, is the difference between her assets (capital) and her liabilities (debt):

$$\widehat{N}_t = \widehat{K}_t - \widehat{B}_t.$$

We allow $\widehat{N}_t$ to be negative, although this would not occur along the equilibrium path.

Let $\widehat{C}_t$ be the consumption of the expert. Then, the dynamics of $\widehat{N}_t$ are given by:

$$\begin{aligned} d\widehat{N}_t &= \widehat{K}_t dr_t^k - \widehat{B}_t r_t dt - \widehat{C}_t dt \\ &= \widehat{\omega}_t \widehat{N}_t dr_t^k + \left[(1 - \widehat{\omega}_t)\, \widehat{N}_t r_t - \widehat{C}_t \right] dt \\ &= \left[(r_t + \widehat{\omega}_t\, (rc_t - \delta - r_t))\, \widehat{N}_t - \widehat{C}_t \right] dt + \sigma \widehat{\omega}_t \widehat{N}_t dZ_t, \end{aligned} \tag{4}$$

where $\widehat{\omega}_t \equiv \frac{\widehat{K}_t}{\widehat{N}_t}$ is the leverage ratio of the expert. The term $r_t + \widehat{\omega}_t\, (rc_t - \delta - r_t)$ is the deterministic return on net wealth, equal to the return on bonds, r_t, plus $\widehat{\omega}_t$ times the excess return on leverage, $rc_t - \delta - r_t$. The term $\sigma \widehat{\omega}_t \widehat{N}_t$ reflects the risk of holding capital induced by the capital growth rate shock.

The previous expression allows us to derive the law of motion for $\widehat{K}_t$:

$$d\widehat{K}_t = d\widehat{N}_t + d\widehat{B}_t = \left[(r_t + \widehat{\omega}_t\, (rc_t - \delta - r_t))\, \widehat{N}_t - \widehat{C}_t \right] dt + \sigma \widehat{\omega}_t \widehat{N}_t dZ_t + d\widehat{B}_t.$$

The expert's preferences over $\widehat{C}_t$ are representable by:

$$\widehat{U}_j = \mathbb{E}_j \left[\int_j^\infty e^{-\widehat{\rho}(t-j)} \log(\widehat{C}_t) dt \right], \tag{5}$$

where $\widehat{\rho}$ is her discount rate. Using a log utility function will make our derivations below easier, but could be easily generalized to the class of recursive preferences introduced by Duffie and Epstein (1992).

The expert decides her consumption levels and leverage ratio to solve the problem:

$$\max_{\{\widehat{C}_t, \widehat{\omega}_t\}_{t \geq 0}} \widehat{U}_0, \tag{6}$$

subject to evolution of her net wealth (4), an initial level of net wealth N_0, and the No-Ponzi-game condition:

$$\lim_{T \to \infty} e^{-\int_0^T r_\tau d\tau} B_T = 0. \tag{7}$$

2.3 Households

There is a continuum of infinitely-lived households with unit mass. Households are heterogeneous in their wealth a_m and labor supply z_m for $m \in [0, 1]$. The distribution of households at time t over these two individual states is $G_t(a, z)$. To save on notation, we will drop the subindex m when no ambiguity occurs.

Each household supplies z_t units of labor valued at wage w_t. Idiosyncratic labor productivity evolves stochastically following a two-state Markov chain: $z_t \in \{z_1, z_2\}$, with $0 < z_1 < z_2$. The process jumps from state 1 to state 2 with intensity λ_1 and vice versa with intensity λ_2. The ergodic mean of z is 1. As in Huggett (1993), we identify state 1 with unemployment (where z_1 is the value of leisure and home production) and state 2 with working. We will follow this assumption when the model faces the data, but nothing essential depends on it. Also, increasing the number of states of the chain is trivial, but two points will suffice for our purposes.

Households can save an amount a_t in the riskless debt issued by the expert at interest rate r_t. Hence, a household's wealth follows:

$$da_t = (w_t z_t + r_t a_t - c_t) dt = s(a_t, z_t, K_t, G_t) dt, \tag{8}$$

where the short-hand notation $s(a_t, z_t, K_t, G_t)$ denotes the drift of the wealth process. The first two variables, a_t and z_t, are the household individual states, the next two, K_t and G_t, are the aggregate state variables that determine the returns on its income sources (labor and bonds). All four variables pin down the optimal choice, $c_t = c(a_t, z_t, K_t, G_t)$, of the control.

The households also faces a borrowing limit that prevents them from shorting bonds:

$$a_t \geq 0. \tag{9}$$

Households have a CRRA instantaneous felicity function over consumption flows $c(\cdot)$:

$$u(c_t) = \frac{c_t^{1-\gamma} - 1}{1-\gamma}$$

discounted at rate $\rho > 0$. As before, we could substitute the CRRA felicity function with a more general class of recursive preferences.

Two points are worth discussing here. First, we have a CRRA felicity function to allow different risk aversions in the households and the expert. Second, we make the households less patient than the expert, $\rho > \widehat{\rho}$. We will show later how the risk-free rate in the DSS (recall, the deterministic steady state) is pinned down by the discount factor of the expert, i.e., $r = \widehat{\rho}$ (we drop the subindex when we denote a variable evaluated at the DSS). But, if $\rho \leq r = \widehat{\rho}$, the households would want to accumulate savings without bounds to self-insure against idiosyncratic labor risk (Aiyagari, 1994). Hence, we can only have a DSS –and an associated ergodic distribution of individual endogenous variables– if we increase the households' discount rate above the expert's.[3]

In summary, households maximize

$$\max_{\{c_t\}_{t\geq 0}} \mathbb{E}_0 \left[\int_0^\infty e^{-\rho t} \frac{c_t^{1-\gamma} - 1}{1-\gamma} dt \right], \tag{10}$$

subject to the budget constraint (8), initial wealth a_0, and the borrowing limit (9).

2.4 Market clearing

There are three market clearing conditions. First, the total amount of debt issued by the expert must equal the total amount of households' savings:

$$B_t \equiv \int a dG_t(da, dz) = \widehat{B}_t, \tag{11}$$

[3]This property of our economy stands in contrast with models à la Bernanke et al. (1999), where borrowers are more impatient than lenders to prevent the former from accumulating enough wealth as to render the financial friction inoperative. But in these models, borrowers are infinitesimal and subject to idiosyncratic risk, and the lenders' discount rate determines the DSS risk-free rate. In our model, the situation is reversed, with the lenders being infinitesimal and subject to idiosyncratic risk and the borrower's discount rate controlling the DSS risk-free rate. We have framed our discussion for the case without aggregate shocks since we want to ensure the existence of a DSS. The characterization of the admissible region for ρ in relation with $\widehat{\rho}$ when we only care about the properties of the economy with aggregate shocks is beyond the scope of our paper.

which also implies $dB_t = d\widehat{B}_t$.

Second, the total amount of labor rented by the firm is equal to labor supplied:

$$L_t = \int z dG_t.$$

Due to the assumption about the ergodic mean of z, we have that $L_t = 1$. Then, total payments to labor are given by w_t. If we define total consumption by households as

$$C_t \equiv \int c\left(a_t, z_t, K_t, G_t\right) dG_t\left(da, dz\right),$$

we get:

$$d\widehat{B}_t = dB_t = \left(w_t + r_t B_t - C_t\right) dt, \tag{12}$$

which tells us that the evolution of aggregate debt is the labor income of households (w_t) plus its debt income ($r_t B_t$) minus their aggregate consumption C_t.

Third, the total amount of capital in this economy is owned by the expert, $K_t = \widehat{K}_t$, and, therefore, $dK_t = d\widehat{K}_t$ and $\widehat{\omega}_t = \frac{K_t}{N_t}$, where $N_t = \widehat{N}_t = K_t - B_t$. With these results, we derive

$$\begin{aligned} dK_t &= \left(\left(r_t + \widehat{\omega}_t\left(rc_t - \delta - r_t\right)\right)\widehat{N}_t - \widehat{C}_t\right) dt + \sigma\widehat{\omega}_t\widehat{N}_t dZ_t + d\widehat{B}_t \\ &= \left(Y_t - \delta K_t - C_t - \widehat{C}_t\right) dt + \sigma K_t dZ_t, \end{aligned} \tag{13}$$

where the last line uses the fact that, from competitive input markets and constant-returns-to-scale, $Y_t = rc_t K_t + w_t$. Recall, from equation (3), that

$$dK_t = \left(\iota_t - \delta\right) K_t dt + \sigma K_t dZ_t.$$

Then, equating (13) and (3) and cancelling terms, we get

$$\iota_t = \frac{Y_t - C_t - \widehat{C}_t}{K_t},$$

i.e., the reinvestment rate is output less aggregate consumption divided by aggregate capital.

2.5 Density

The households distribution $G_t\left(a, z\right)$ has a density on assets a, $g_{it}(a)$, conditional on the labor productivity state $i \in \{1, 2\}$. The density satisfies the normalization

$$\sum_{i=1}^{2} \int_0^\infty g_{it}(a) da = 1.$$

The dynamics of this density conditional on the realization of aggregate variables are given by the Kolmogorov forward (KF) equation:

$$\frac{\partial g_{it}}{\partial t} = -\frac{\partial}{\partial a}\left(s\left(a_t, z_t, K_t, G_t\right) g_{it}(a)\right) - \lambda_i g_{it}(a) + \lambda_j g_{jt}(a), \;\; i \neq j = 1, 2. \tag{14}$$

Reading equation (14) is straightforward: the density evolves according to the optimal consumption-saving choices of each household plus two jumps corresponding to households that circulate out of the labor state i ($\lambda_i g_{it}(a)$) and the households that move into state j ($\lambda_j g_{jt}(a)$).

3 Equilibrium

An equilibrium in this economy is composed by a set of prices $\left\{w_t, rc_t, r_t, r_t^k\right\}_{t\geq 0}$, quantities $\left\{K_t, N_t, B_t, \widehat{C}_t, c_{mt}\right\}_{t\geq 0}$ and a density $\left\{g_{it}\left(\cdot\right)\right\}_{t\geq 0}$ for $i \in \{1, 2\}$ such that:

1. Given w_t, r_t, and g_t, the solution of household m's problem (31) is $c_{mt} = c\left(a_t, z_t, K_t, G_t\right)$.
2. Given r_t^k, r_t, and N_t, the solution of the expert's problem (6) is $\widehat{C}_t$, K_t, and B_t.
3. Given K_t, the firm maximizes their profits and input prices are given by w_t and rc_t and the rate of return on capital by r_t^k.
4. Given w_t, r_t, and c_t, g_{it} is the solution of the KF equation (14).
5. Given r_t, g_{it}, and B_t, the debt market (11) clears and $N_t = K_t - B_t$.

3.1 Equilibrium characterization

Several properties of the equilibrium are characterized with ease. We proceed first with the expert's problem. The use of log-utility implies that the expert consumes a constant share $\widehat{\rho}$ of her net wealth and chooses a leverage ratio proportional to the difference between the expected return on capital and the risk-free rate:

$$\widehat{C}_t = \widehat{\rho} N_t \tag{15}$$

$$\omega_t = \widehat{\omega}_t = \frac{1}{\sigma^2}\left(rc_t - \delta - r_t\right). \tag{16}$$

Second, rewriting the latter result, we get that the excess return on leverage,

$$rc_t - \delta - r_t = \sigma^2 \frac{K_t}{N_t},$$

depends positively on the variance of the aggregate shock, σ^2, and the leverage of the economy $\frac{K_t}{N_t}$. The higher the volatility or the leverage ratio in the economy, the higher the excess return that the expert requires to isolate households from dZ_t. A positive capital growth rate shock, by increasing N_t relatively to K_t, lowers the excess return. Analogously, a higher volatility of the aggregate shock increases the excess return.

Third, we can use the values of rc_t, L_t, and ω_t in equilibrium to get the wage $w_t = (1-\alpha) K_t^{\alpha}$, the rental rate of capital $rc_t = \alpha K_t^{\alpha-1}$, and the risk-free interest rate:

$$r_t = \alpha K_t^{\alpha-1} - \delta - \sigma^2 \frac{K_t}{N_t}. \tag{17}$$

This equation will play a key role, below, in explaining our quantitative results. Since $K_t = N_t + B_t$, equations (15)-(17) depends only on the expert's net wealth N_t and debt B_t.

Fourth, we can describe the evolution of N_t:

$$\begin{aligned} dN_t &= \left[\left(r_t + \omega_t \left(rc_t - \delta - r_t \right) \right) N_t - \widehat{C}_t \right] dt + \sigma \omega_t N_t dZ_t \\ &= \left(\alpha K_t^{\alpha-1} - \delta - \widehat{\rho} - \sigma^2 \left(1 - \frac{K_t}{N_t} \right) \frac{K_t}{N_t} \right) N_t dt + \sigma K_t dZ_t \end{aligned} \tag{18}$$

as a function only of N_t, B_t, and dZ_t. Equation (18) shows the nonlinear dependence of dN_t on the leverage level $\frac{K_t}{N_t}$. We will stress this point in the next pages repeatedly. For convenience, some times we will write

$$dN_t = \mu^N(B_t, N_t)dt + \sigma^N(B_t, N_t)dZ_t,$$

where $\mu^N(B_t, N_t) = \left(\alpha K_t^{\alpha-1} - \delta - \widehat{\rho} - \sigma^2 \left(1 - \frac{K_t}{N_t} \right) \frac{K_t}{N_t} \right) N_t$ is the drift of N_t and $\sigma^N(B_t, N_t) = \sigma K_t$ its volatility.

Fifth, we have from equation (12):

$$dB_t = \left(w_t + r_t B_t - C_t \right) dt = \left(\left(1 - \alpha \right) K_t^{\alpha} + \left(\alpha K_t^{\alpha-1} - \delta - \sigma^2 \frac{K_t}{N_t} \right) B_t - C_t \right) dt, \tag{19}$$

which depends nonlinearly on N_t and B_t and linearly on aggregate household consumption, C_t, an endogenous variable we must determine.

We can stack all the equilibrium conditions (except the optimality condition for households, which we will discuss in the next section) in two blocks. The first block includes all the variables

that depend directly on N_t, B_t, and dZ_t:

$$w_t = (1-\alpha)\, K_t^{\alpha} \tag{20}$$

$$rc_t = \alpha K_t^{\alpha-1} \tag{21}$$

$$r_t = \alpha K_t^{\alpha-1} - \delta - \sigma^2 \frac{K_t}{N_t} \tag{22}$$

$$dr_t^k = (rc_t - \delta)\, dt + \sigma dZ_t \tag{23}$$

$$dN_t = \left(\alpha K_t^{\alpha-1} - \delta - \widehat{\rho} - \sigma^2 \left(1 - \frac{K_t}{N_t}\right) \frac{K_t}{N_t}\right) N_t dt + \sigma K_t dZ_t. \tag{24}$$

The second block includes the equations determining the aggregate consumption of the households, dB_t, dK_t, and $\frac{\partial g_{it}}{\partial t}$:

$$C_t \equiv \sum_{i=1}^{2} \int c\,(a_t, z_t, K_t, G_t)\, g_{it}\,(a)\, da \tag{25}$$

$$dB_t = \left((1-\alpha)\, K_t^{\alpha} + \left(\alpha K_t^{\alpha-1} - \delta - \sigma^2 \frac{K_t}{N_t}\right) B_t - C_t\right) dt \tag{26}$$

$$dK_t = dN_t + dB_t \tag{27}$$

$$\frac{\partial g_{it}}{\partial t} = -\frac{\partial}{\partial a}\left(s\,(a_t, z_t, K_t, G_t)\, g_{it}(a)\right) - \lambda_i g_{it}(a) + \lambda_j g_{jt}(a),\ i \neq j = 1, 2. \tag{28}$$

The second block shows i) how the density $\{g_{it}\,(\cdot)\}_{t\geq 0}$ for $i \in \{1, 2\}$ matters to determine C_t, ii) that C_t pins down dB_t, and iii) that once we have dB_t, we can calculate dK_t. Hence, computing the equilibrium of this economy is equivalent to finding C_t. Once C_t is known, all other aggregate variables follow directly.

3.2 The DSS of the model

Before solving for the general case of the model, we use the previous equations to analyze three interesting cases. In this subsection, we describe the DSS of the model where there are no capital growth rate shocks, but we still have idiosyncratic household shocks. In the next two subsections, we will look at the SSS of our economy and the representative household version of the model.

To study the DSS, we set $\sigma = 0$ in the law of motion for the expert net wealth (18) and get:

$$dN_t = \left(\alpha K_t^{\alpha-1} - \delta - \widehat{\rho}\right) N_t dt. \tag{29}$$

Since the drift of N_t, $\mu^N(B, N) = (\alpha K^{\alpha-1} - \delta - \widehat{\rho})\, N$, must be zero in a DSS (remember that

we drop the t subindex to denote the DSS value of a variable), we get

$$K = \left(\frac{\widehat{\rho} + \delta}{\alpha} \right)^{\frac{1}{\alpha - 1}}.$$

With this result, the DSS risk-free interest rate (17) equals the return on capital and the rental rate of capital less depreciation:

$$r = r_t^k = rc_t - \delta = \alpha K_t^{\alpha - 1} - \delta = \widehat{\rho}. \tag{30}$$

As mentioned above, this condition forces us to have $\widehat{\rho} < \rho$. Otherwise, the households would accumulate too many bonds and the DSS would not be well-defined.

Finally, the dispersion of the idiosyncratic shocks determines the DSS expert's net wealth:

$$N = K - B = K - \int a dG\left(da, dz\right),$$

a quantity that, unfortunately, we cannot compute analytically.

3.3 The SSS of the model

A SSS in our model is formally defined as a density $g^{SSS}(\cdot)$ and equity N^{SSS} that remain invariant in the absence of aggregate shocks. Let $\Gamma_\sigma(g(\cdot), N, W)$ be the law of motion of the economy given an aggregate capital volatility σ and a realization of the Brownian motion W. More precisely, $\Gamma_\sigma(\cdot, \cdot, \cdot)$ is an operator that maps income-wealth densities $g(\cdot)$ and equity levels N into changes in these variables:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \begin{bmatrix} g_{t+\Delta t}(\cdot) - g_t(\cdot) \\ N_{t+\Delta t} - N_t \end{bmatrix} = \Gamma_\sigma(g_t(\cdot), N_t, W_t).$$

The SSS, therefore, solves:

$$\Gamma_\sigma(g^{SSS}(\cdot), N^{SSS}, 0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

In general, we will have multiple SSSs that solve the previous functional equation. Indeed, several of them will appear in our quantitative exercise.

The difference between the SSS and the DSS is that the former is the steady state of an economy where individual agents make their decisions taking into account aggregate risks ($\sigma > 0$) −using equation (33)− but no shock arrives along the equilibrium path, whereas in the latter agents live in an economy without aggregate risks ($\sigma = 0$) and arrange their consumption paths

accordingly. The DSS is then formally defined as:

$$\Gamma_0(g^{DSS}(\cdot), N^{DSS}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

3.4 The representative household version of the model

In this subsection, we eliminate the idiosyncratic labor risk of households. In that way, we come up with a representative household version of the model where we keep the *between-agents* heterogeneity (we still have an expert and a household), but we do not have any *within-agents* heterogeneity (there are one representative expert and one representative household).

Equations (20) to (24) of the equilibrium, as well as equations (26) and (27), remain unchanged. Instead of equation (25), the consumption rule of the household $C_t = c(B_t, N_t)$ depends now just on aggregate debt and capital and it solves the problem:

$$\max_{\{C_t\}_{t\geq 0}} \mathbb{E}_0 \left[\int_0^\infty e^{-\rho t} \frac{C_t^{1-\gamma} - 1}{1-\gamma} dt \right], \tag{31}$$

subject to the budget constraint (26) (where input prices have already been substituted in), initial wealth B_0, and the borrowing limit $B_t \geq 0$. Because of the absence of *within-agents* heterogeneity, there is no need for an analogous of equation (28) in this version of the model.

4 Solution

Our previous discussion highlighted the role of finding, in the general version of the model, the households aggregate consumption, C_t, to compute the equilibrium of the economy given some structural parameter values:

$$\Psi = \{\alpha, \delta, \sigma, \widehat{\rho}, \rho, \gamma, z_1, z_2, \lambda_1, \lambda_2\}.$$

To do so, we follow Krusell and Smith (1998) and assume that, when forming their expectations, households only use a finite set of n moments of the cross-sectional distribution of assets instead of the complete distribution. In contrast to Krusell and Smith (1998), in which the income-wealth distribution is the only endogenous state variable, here the expert's net wealth N_t is also a state variable. At the same time, we do not have any exogenous state variable, as $K_t = N_t + B_t$ instantaneously incorporates the capital growth rate shocks. For ease of exposition, we discuss the case with $n = 1$. All the techniques can be trivially extended to the case with $n > 1$ at the cost of heavier notation.

More concretely, households consider a *perceived law of motion* (PLM) of aggregate debt:

$$dB_t = h\left(B_t, N_t\right) dt, \tag{32}$$

where $h\left(B, N\right)$ is the conditional expectation of dB_t given available information (B_t, N_t):

$$h\left(B_t, N_t\right) = \frac{\mathbb{E}\left[dB_t | B_t, N_t\right]}{dt},$$

instead of the exact law of motion (19). We borrow the term PLM from the learning literature (Evans and Honkapohja, 2001). Our choice of words accentuates that we allow $h\left(\cdot, \cdot\right)$ be a general function, and not just a linear or polynomial function of its arguments, perhaps with state-dependent coefficients, as in Krusell and Smith (1998). In fact, our methodology will let the PLM approximate, arbitrarily well, equation (19). More precisely, and as explained in Subsection 4.3, we propose a methodology in which the functional form $h\left(\cdot, \cdot\right)$ is obtained by employing machine learning techniques on simulated data. This extra flexibility is key given the complex nonlinearities present in laws of motion of N_t, equation (18), and B_t, equation (19).

Given the PLM, the household's problem has an associated Hamilton-Jacobi-Bellman (HJB) equation:

$$\begin{aligned} \rho V_i\left(a, B, N\right) &= \max_c \frac{c^{1-\gamma}-1}{1-\gamma} + s\frac{\partial V_i}{\partial a} + \lambda_i\left[V_j(a, B, N) - V_i(a, B, N)\right] \\ &\quad + h\left(B, N\right)\frac{\partial V_i}{\partial B} + \mu^N\left(B, N\right)\frac{\partial V_i}{\partial N} + \frac{\left[\sigma^N\left(B, N\right)\right]^2}{2}\frac{\partial^2 V_i}{\partial N^2}, \end{aligned} \tag{33}$$

$i \neq j = 1, 2$, and where we use the shorthand notation $s = s\left(a, z, N + B, G\right)$. Notice how the HJB incorporates $h\left(B, N\right)$. Equation (33) complements the equilibrium conditions (20)-(28) by making the problem of the household explicit.

4.1 An overview of the algorithm

Our algorithm to find $h(B, N)$ in (32) proceeds according to the following iteration:

1) Start with h_0, an initial guess for h.

2) Using current guess for h, solve for the household consumption, c_m, in the HJB equation (33). This solution can be obtained by using an upwind finite differences scheme described in Appendix A (although other numerical algorithms such as a meshfree method can be applied).

3) Construct a time series for B_t by simulating the cross-sectional distribution over time. Given B_t, we can find N_t and K_t using equations (18) and (27).

4) Use a universal nonlinear approximator to obtain h_1, a new guess for h.

5) Iterate steps 2)-4) until h_n is sufficiently close to h_{n-1} given some pre-specified norm and tolerance level.

Steps 1)-5) show that our solution has two main differences with respect to the original Krusell-Smith algorithm: the use of continuous time and our employment of a universal nonlinear approximator to update the guess of the PLM. Both differences deserve some explanation.

4.2 Continuous time

Krusell and Smith (1998) wrote their model in discrete time. Our continuous-time formulation, while not changing any fundamental feature of the model, enjoys several numerical advantages (Achdou et al., 2017). First, continuous time naturally generates sparsity in the matrices characterizing the transition probabilities of the discretized stochastic processes. Intuitively, continuously moving state variables such as wealth only drift an infinitesimal amount in an infinitesimal unit of time. Therefore, in an approximation that discretizes the state space, households reach only states that directly neighbor the current state. Second, the optimality characterizing consumption has a simpler structure than in discrete time:

$$c_i^{-\gamma} = \frac{\partial V_i}{\partial a}. \tag{34}$$

Third, it is easier to capture occasionally binding constraints such as equation (9) in continuous time than in discrete time as the optimality condition (34) for consumption holds with equality everywhere in the interior of the state space. Fourth, the dynamics of the cross-sectional wealth distribution are characterized by the KF equation (14). The discretization of this equation yields an efficient way to simulate a time series of the cross-sectional distribution (although this can also be performed in discrete time, as in Ríos-Rull 1997, Reiter 2009, and Young 2010, at some additional cost).

Regarding the generation of data, we simulate T periods of the economy with a constant time step Δt. We start from the initial income-wealth distribution at the DSS (although we could pick other values). A number of initial samples is discarded as a burn-in. If the time step is small enough, we have

$$B_{t_j+\Delta t} = B_{t_j} + \int_{t_j}^{t_j+\Delta_t} dB_s = B_{t_j} + \int_{t_j}^{t_j+\Delta t} h\left(B_s, N_s\right) ds \approx B_{t_j} + h\left(B_{t_j}, N_{t_j}\right) \Delta t.$$

Our simulation $\left(\mathbf{S}, \widehat{\mathbf{h}}\right)$ is composed by a vector of inputs $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_J\}$, where $\mathbf{s}_j = \left\{s_j^1, s_j^2\right\} = \left\{B_{t_j}, N_{t_j}\right\}$ are samples of aggregate debt and expert's net wealth at J random

times $t_j \in [0, T]$, and a vector of outputs $\widehat{\mathbf{h}} = \left\{\widehat{h}_1, \widehat{h}_2 ..., \widehat{h}_J\right\}$, where

$$\widehat{h}_j \equiv \frac{B_{t_j+\Delta t} - B_{t_j}}{\Delta t}$$

are samples of the growth rate of B_t. The evaluation times t_j should be random and uniformly distributed over $[0, T]$ as, ideally, samples should be independent.

4.3 Neural networks: A universal nonlinear approximator

In the original Krusell-Smith algorithm, the law of motion linking the mean of capital tomorrow and the mean of capital today is log-linear, with the coefficients in that function depending on the aggregate shock. This approximation is highly accurate due to the near log-linearity of their models in the vicinity of the DSS. Indeed, in such a model, the DSS and SSS almost coincide. But, as shown in equations (18) and (19), this linearity of the law of motion of the endogenous variables with respect to other endogenous variables does not carry out to our model.

This nonlinear structure causes two problems. First, we face the *approximation* problem: we need an algorithm that searches for an unknown nonlinear functional instead of a simple linear regression with aggregate-state-dependent coefficients. Second, we need to tackle the *extrapolation* problem. While the theoretical domain of B_t and N_t is unbounded, practical computation requires to limit it to a compact subset of $\mathbb{R}^2$ large enough as to prevent boundary conditions from altering the solution in the subregion where most of the ergodic distribution accumulates. But precisely because we deal with such a large area, the simulation in step 3) of the algorithm in Subsection 4.1 never visits an ample region of the state space. Thus, the approximation algorithm should not only provide an accurate nonlinear approximation in the visited region, but also a "reasonable" extrapolation to the rest of the state space. We will return to what "reasonable" means in this context momentarily.

To address these two problems, we propose to employ a nonlinear approximation technique based on neural networks. Our approach has four crucial advantages. First, the *universal approximation theorem* (Hornik et al. 1989; Cybenko 1989) states that a neural network with at least one hidden layer can approximate any Borel measurable function mapping finite-dimensional spaces arbitrarily well. In particular, the theorem does not require that the approximated function be differentiable and can handle cases with kinks and occasionally binding constraints.[4]

Second, the neural network coefficients can be efficiently estimated using gradient descent methods and back-propagation. This allows for an easier coding and shorter implementation

[4]Recall that Lusin's theorem tells us that every measurable function is a continuous function almost everywhere. Thus, we can approximate jumps in a finite number of points, but not functions with extremely intricate shapes. Those complex functions, however, are unlikely to be of much relevance in solving standard dynamic equilibrium models.

time than other approaches.

Third, neural networks are more economical, for middle and high dimensions, than other approximators. More concretely, Barron (1993) shows that a one-layer neural network achieves, for functions on the first moment of the magnitude distribution of the Fourier transform, integrated square errors of order $O(1/n)$, where n is the number of nodes. In comparison, for series approximations (polynomials, spline, and trigonometric expansions), the integrated square error is of order $O(1/(n^{2/d})$ where d is the dimensions of the function to be approximated. In other words: the "curse of dimensionality" does not apply to neural networks that approximate functions of the very wide class considered by Barron (1993).[5] This advantage is not present in our baseline model, with $d = 2$, but will appear in any extension with additional aggregate state variables. Even going to $d = 3$ or $d = 4$ saturates alternatives such as Chebyshev polynomials.[6]

Fourth, neural networks extrapolate outstandingly. This is, in practice, key. Neural networks have well-behaved shapes outside their training areas. In contrast, Chebyshev polynomials (or other series) more often than not display explosive behaviors outside the fitted area that prevent the algorithm from converging. Figures D.1 and D.2 in Appendix D show this disappointing behavior of an approximation to the PLM in our model with Chebyshev polynomials. The two figures document how, within the area of high density of the ergodic distribution, Chebyshev polynomials approximate the law of motion for aggregate debt fairly (compare them with panel c) in Figure 2, obtained with our neural network). But Chebyshev polynomials start oscillating as soon as we abandoned the well-traveled area of the simulation.

We can now briefly describe our neural network approximator of the PLM. For excellent introductory treatments of this material, see Bishop (2006) and Goodfellow et al. (2016).

A single hidden layer neural network $h(\mathbf{s};\theta)$ is a linear combination of Q fixed nonlinear basis (i.e., activation) functions $\phi(\cdot)$:

$$h(\mathbf{s};\theta) = \theta_0^2 + \sum_{q=1}^{Q} \theta_q^2 \phi\left(\theta_{0,q}^1 + \sum_{i=1}^{2} \theta_{i,q}^1 s^i\right), \tag{35}$$

where $\mathbf{s}$ is an two-dimensional input and θ a vector of coefficients (i.e., weights):

$$\theta = \left(\theta_0^2, \theta_1^2, ..., \theta_Q^2, \theta_{0,1}^1, \theta_{1,1}^1, \theta_{2,1}^1, ..., \theta_{0,Q}^1, \theta_{1,Q}^1, \theta_{2,Q}^1\right).$$

We call θ "coefficients," as they represent a numerical entity, in comparison with the structural

[5] In fact, we can rely on the more general theorems shown by Bach (2017) that cover non-decreasing positively homogeneous activation functions like the rectified linear unit and that show, beyond the break of the curse dimensionality, approximation and the estimation errors.

[6] Similarly, approaches, such as Smolyak interpolation, that alleviate the "curse of dimensionality" in standard problems are harder to apply here because we deal with shapes of the ergodic distribution are hard to characterize ex-ante. Neural networks are more resilient to sparse initial information regarding such shapes.

parameters, Ψ, that have a sharp economic interpretation.

The neural network provides a flexible parametric function h that determines the growth rate of aggregate debt:

$$\widehat{h}_j = h\left(\mathbf{s}_j; \theta\right), \ j = 1, .., J,$$

and that satisfies the properties of universal approximation, breaking of the curse of dimensionality, good extrapolative behavior, and easy implementation we discussed above.

Different alternatives are available for the activation function. For our model, we choose a *softplus* function, $\phi(x) = \log(1 + e^x)$ for a given input x. The softplus function has a simple sigmoid derivative, which avoids some of the problems caused by the presence of a kink in rectified linear units, while keeping an excellent efficient computation and gradient propagation.

The neural network (35) can be generalized to include many hidden layers, stacked one after the other. In that case, the network is called a *deep* neural network. However, for the particular problem of approximation a two-dimensional function a single layer is enough. The size of the hidden layer is determined by Q. This hypercoefficient can be set by regularization or simply by trial-and-error in relatively simple problems, such as the one presented here. In our case, we set $Q = 16$ because the cost of a larger hidden layer is small.

The vector of coefficients θ is selected to minimize the quadratic error function $\mathcal{E}\left(\theta; \mathbf{S}, \widehat{\mathbf{h}}\right)$ given a simulation $\left(\mathbf{S}, \widehat{\mathbf{h}}\right)$:

$$\theta^* = \arg\max_{\theta} \mathcal{E}\left(\theta; \mathbf{S}, \widehat{\mathbf{h}}\right) = \arg\max_{\theta} \frac{1}{2} \sum_{j=1}^{J} \left\| h\left(\mathbf{s}_j; \theta\right) - \widehat{h}_j \right\|^2.$$

A standard approach to perform this minimization in neural networks is the *batch gradient descent* algorithm. Appendix B describes the training of the network in detail in detail

One concern with neural networks is that the algorithm might converge to a local minimum. A way of coping with it is to implement a Monte Carlo multi-start in the first iteration of the algorithm. We select P initial vectors θ_0^p, with $p = 1, ..., P$ from (**??**). For each of these vectors, we run the batch gradient descent until convergence. Once we achieve convergence, we select the θ_m^p that yields the minimum error across all the trials . Furthermore, since we are interested in approximating an unknown function, not clearing a market or satisfying an optimality condition, local minima that are close to a global minimum are often acceptable solutions to the approximation problem.

Finally, notice that the algorithm is massively parallel, either in CPUs or GPUs (and, in the middle-run, in the new generation of Tensor Processing Units or TPUs), a most convenient feature for scaling and estimation.

5 Estimation

Once we have solved the model given some structural parameter values Ψ, the next step is to take the model to the data, to let observations determine the values of Ψ. We will proceed in two stages. First, we will discuss the simple case where the econometrician has access to output data and wants to build the likelihood associated with it. Second, we will show the results of our estimation with real data.

5.1 Building the likelihood function

Let us assume that the econometrician has access to $D+1$ observations of output, Y_t, at fixed time intervals $[0, \Delta, 2\Delta, .., D\Delta]$:

$$Y_0^D = \{Y_0, Y_\Delta, Y_{2\Delta}, ..., Y_D\}.$$

The derivations below would be very similar for observables other than output. More important, though, is that since we have one aggregate shock in the model (to capital), we can only use one observable in our likelihood. Otherwise, we would suffer from stochastic singularity. If we wanted to have more observables, such other aggregate variables or the states of model, we would need to either enrich the model with more shocks or to introduce measurement shocks in the observables. In those more complex situations, we might need to resort to a sequential Monte Carlo approximation to the filtering problem described by the associated Kushner-Stratonovich equation of our dynamic system (see, for a related approach in discrete time, Fernández-Villaverde and Rubio-Ramírez, 2007).

The likelihood function $\mathcal{L}_D\left(Y_0^D|\Psi\right)$ for our observations of output has the form:

$$\mathcal{L}_D\left(Y_0^D|\Psi\right) = \prod_{d=1}^{D} p_Y\left(Y_{d\Delta}|Y_{(d-1)\Delta};\Psi\right),$$

where $p_Y\left(Y_{d\Delta}|Y_{(d-1)\Delta};\Psi\right)$, the conditional density function of $Y_{d\Delta}$ given $Y_{(d-1)\Delta}$, is equal to:

$$p_Y\left(Y_{d\Delta}|Y_{(d-1)\Delta};\Psi\right) = \int f_{d\Delta}(Y_{d\Delta}, B)dB$$

given a density function for output and debt, $f_{d\Delta}(Y_{d\Delta}, B)$, implied by the solution of the model.

Our task is, therefore, to compute the sequences of conditional densities $p_Y\left(Y_{d\Delta}|Y_{(d-1)\Delta};\Psi\right)$ at the fixed time intervals $[0, \Delta, 2\Delta, .., D\Delta,]$. To do so, first, we obtain the diffusion of $Y_t = (B_t + N_t)^\alpha$.

Applying Itô's lemma, we get:

$$\begin{aligned} dY_t &= \frac{\partial (B+N)^{\alpha}}{\partial B} dB_t + \frac{\partial (B+N)^{\alpha}}{\partial N} dN_t + \frac{1}{2} \frac{\partial^2 (B+N)^{\alpha}}{\partial N^2} \sigma^2 (B+N)^2 dt \\ &= \alpha K_t^{\alpha-1} \left[\left(h(B_t, N_t) + \mu_t^N \right) dt + \sigma K_t dZ_t \right] + \frac{1}{2} \alpha (\alpha-1) K_t^{\alpha-2} \sigma^2 K_t^2 dt \\ &\quad + \frac{\sigma^2}{2} \alpha (\alpha-1) Y_t dt + \sigma \alpha Y_t dZ_t \\ &= \mu^Y (B_t, Y_t) dt + \sigma_t^Y (Y_t) dZ_{t,} \end{aligned} \tag{36}$$

where:

$$\begin{aligned} \mu^Y (B_t, Y_t) &= \alpha Y_t^{\frac{\alpha-1}{\alpha}} * \Bigg[h(B_t, Y_t^{\frac{1}{\alpha}} - B_t) + \alpha Y_t + \left[\frac{(\alpha-1)\sigma^2}{2} - \delta \right] Y_t^{\frac{1}{\alpha}} \\ &\quad - \left(\alpha Y_t^{\frac{\alpha-1}{\alpha}} - \delta - \sigma^2 \frac{Y_t^{\frac{1}{\alpha}}}{Y_t^{\frac{1}{\alpha}} - B_t} \right) B_t - \hat{\rho} \left(Y_t^{\frac{1}{\alpha}} - B_t \right) \Bigg], \end{aligned}$$

and $\sigma^Y (Y_t) = \alpha \sigma Y_t$.

With equation (36), the density $f_t^d(Y, B)$ follows the KF equation in the interval $[(d-1)\Delta, d\Delta]$:

$$\begin{aligned} \frac{\partial f_t}{\partial t} &= -\frac{\partial}{\partial Y} \left[\mu^Y (Y, B) f_t(Y, B) \right] - \frac{\partial}{\partial B} \left[h(B, Y^{\frac{1}{\alpha}} - B) f_t^d(Y, B) \right] \\ &\quad + \frac{1}{2} \frac{\partial^2}{\partial Y^2} \left[\left(\sigma^Y (Y) \right)^2 f_t(Y, B) \right]. \end{aligned} \tag{37}$$

The initial condition at the beginning of the interval is

$$f_{(d-1)\Delta}(Y, B) = \delta \left(Y - Y_{(d-1)\Delta} \right) f_{(d-2)\Delta}(B | Y_{(d-1)\Delta}),$$

where $f_{(d-1)\Delta}(B|Y_{(d-1)\Delta})$ is the probability of B conditional on $Y = Y_{(d-1)\Delta}$:

$$f_{(d-2)\Delta}(B|Y_{(d-1)\Delta}) = \frac{f_{(d-2)\Delta}(Y_{(d-1)\Delta}, B)}{f_{(d-2)\Delta}(Y_{(d-1)\Delta})} = \frac{f_{(d-2)\Delta}(Y_{(d-1)\Delta}, B)}{\int f_{(d-2)\Delta}(Y_{(d-1)\Delta}, B) dB)},$$

if $d \geq 2$, $f_{-1}(B) = f(B)$ is the ergodic distribution of B, and $\delta(\cdot)$ is the Dirac delta function.

Lo (1988) pioneered the estimation of the likelihood of a continuous-time stochastic process on discrete-time samples using the KF equation to characterize the transition density functions. That paper provides a number of technical assumptions that has to be satisfied for the estimation. In our model these conditions are met provided that $h(B, N)$ is twice continuously differentiable in B and N and three times continuously differentiable in Ψ, which is guaranteed if it is computed using a neural network with activation functions that meet this requirements

–as it happens in our case– and that Ψ lies in the interior of a finite- dimensional closed and compact parameter space.

A fundamental property of the operator in the KF equation (37) is that it is the adjoint of the infinitesimal generator employed in the HJB. The intuition for that result is that one can think about the dynamic choices of the agents implied by the HJB as a probability distribution of their future choices, and, hence, a distribution on observables, such as output, induced by the stochastic shocks of the model. There is, in other words, an intimate link between optimal choices and likelihood functions.

This result is remarkable since it means that the solution of the KF equation amounts to transposing an inverting a sparse matrix that has already been computed when we solved the HJB. This provides a highly efficient way of evaluating the likelihood after the model is solved.[7]

In Appendix C, we show how to build the likelihood function of the model when, in addition, we add add microeconomic observations from the cross-sectional distribution of assets.

5.2 Maximizing the likelihood

Once we have evaluated the likelihood, we can either maximize it or perform Bayesian inference relaying on a posterior sampler. In this paper, for simplicity, we follow the former approach. Also, since we are dealing with a rather novel approach to solution and estimation of models with heterogeneous agents, we want to keep the estimation relatively simple, and we fix most of the structural parameters at conventional calibrated values for the U.S. economy and use only aggregate data.

As our data, we will rely on U.S. quarterly output observations for 1984Q1-2017Q4, with bandpass filter keeping frequencies between 20 and 60 quarters (between 5 and 15 years). We start in 1984, as often done in the literature, to focus in capturing the dynamics that have governed aggregate fluctuations in the U.S. after the arrival of the great moderation (Galí and Gambetti, 2009).[8] We bandpass the data as to eliminate long-run trends and to skip the business cycle frequencies caused by productivity shocks and monetary policy shocks our model is not designed to account for. Note, however, that our methodology does not depend on this filtering and that a richer model could be estimated with raw data without theoretical problems.

With respect to the fixed parameters, the capital share parameter, α, is taken to be 0.35 and the depreciation rate of capital, δ, is 0.1 (we report all rates at an annual term). The discount rate ρ, is set to 0.05. The risk aversion of the households γ is set to 2. These are standard values

[7] If the KF would become numerically cumbersome in more general models, we could construct Hermite polynomials expansions of the (exact but unknown) likelihood as in Aït-Sahalia (2002). We could also consider methods of moments in continuous time such as those pioneered by Andersen and Lund (1997) and Chacko and Viceira (2003).

[8] See also the updated evidence in Liu et al. (2018), who document how the great moderation has clearly survived the financial crisis of 2008.

in the business cycle literature to match the investment-output ratio and the rate of return on capital.

The idiosyncratic income process parameters are calibrated following our interpretation of state 1 as unemployment and state 2 as employment. The transition rates between unemployment and employment (λ_1, λ_2) are chosen such that (i) the unemployment rate $\lambda_2/(\lambda_1+\lambda_2)$ is 5 percent and (ii) the job finding rate is 0.3 at monthly frequency or $\lambda_1 = 0.986$ at annual frequency.[9] These numbers describe the 'US' labor market calibration in Blanchard and Galí (2010). We normalize average income $\bar{y} = \frac{\lambda_2}{\lambda_1+\lambda_2} y_1 + \frac{\lambda_1}{\lambda_1+\lambda_2} y_2$ to 1. We also set z_1 equal to 71 percent of z_2, as in Hall and Milgrom (2008). Both targets allow us to solve for y_1 and y_2. We set the experts' discount rate $\widehat{\rho}$ to ensure that the leverage ratio K/N in the High-leverage SSS below is 2, which is roughly the average leverage from a Compustat sample of non-financial corporations. Table 1 summarizes our baseline calibration.

Parameter	Value	Description	Source/Target
α	0.35	capital share	standard
δ	0.1	capital depreciation	standard
γ	2	risk aversion	standard
ρ	0.05	households' discount rate	standard
λ_1	0.986	transition rate unemp.-to-employment	monthly job finding rate of 0.3
λ_2	0.052	transition rate employment-to-unemp.	unemployment rate 5 percent
z_1	0.72	income in unemployment state	Hall and Milgrom (2008)
z_2	1.015	income in employment state	$\mathbb{E}(y) = 1$
$\widehat{\rho}$	0.0497	experts' discount rate	$K/N = 2$

Table 1: Baseline parametrization.

We solve the model according to the algorithm in Section 4. Then, we evaluate the likelihood of the observations on U.S. output for different values of σ, the volatility of the aggregate shock and, therefore, the most interesting parameters in terms of the properties of the model. While doing so, we keep all the other parameter values fixed at their calibrated quantities. We maximize the likelihood function by searching on a grid between 0.008 and 0.018 with a step 0.00025.

We plot the resulting log-likelihood in Figure 1. The point estimate, 0.0142, is drawn as a vertical discontinuous red line, with a standard error of 0.00011342, computed by the local derivative of the function. The smoothness of the plot confirms, also, that our algorithm has successfully converged and that changes in one parameter value do not lead us into substantially different numerical solutions.

[9] Analogously to Blanchard and Galí (2010, footnote 20), we compute the equivalent annual rate λ_1 as $\lambda_1 = \sum_{i=1}^{12} (1-\lambda_1^m)^{i-1} \lambda_1^m$, where λ_1^m is the monthly job finding rate.

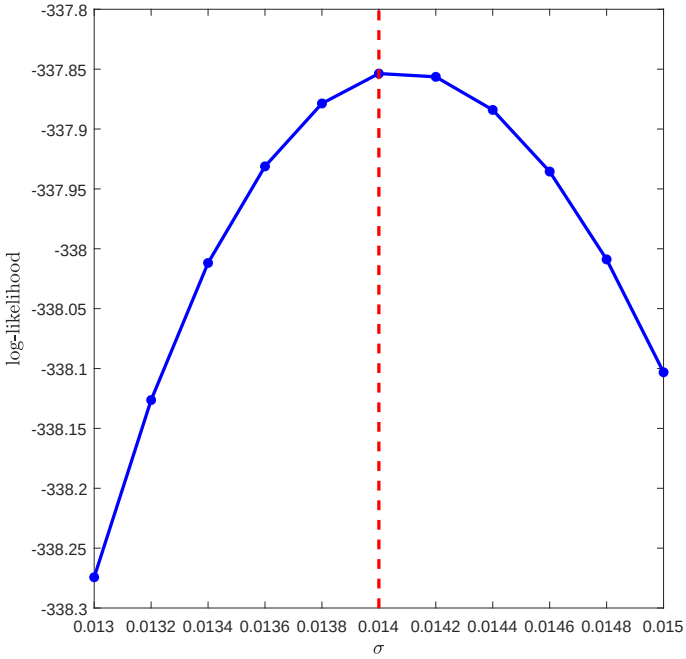


Figure 1: Log-likelihood for different values of σ and point estimate.

6 Quantitative results

This section presents some quantitative results generated by our solution algorithm with the parameter values from Section 5. We will report, first, the PLM for aggregate debt, $h(B, N)$. We will also assess our approximation for $dh(B, N)$ and evaluate its accuracy. Next, we will explore the phase diagram of the model and explain, through the dynamic responses of the model to an aggregate shock, why we find several SSS(s). After having discussed the convergence properties of the SSS(s), the random fluctuations around those, and documented the presence of time-varying aggregate risk in our economy, we will analyze the role of the value of σ in determining the multiplicity of SSS(s). We will close by looking at the aggregate ergodic distribution of debt and equity.

6.1 The PLM

Figure 2 reports the resulting PLM for aggregate debt, $h(B, N)$. Panel (a), at the top left, displays three transversal cuts of $h(B, N)$ along with a range of values of equity (N). The first cut is fixes B at the high-leverage SSS (HL-SSS) value ($B^{HL} = 1.9641, N^{HL} = 1.7470$, with $K^{HL} = 3.7111$ and $\frac{K^{HL}}{N^{HL}} = 2.1243$), the second cut fixes B at a selected high-leverage point ($B^* = 2.15, N^* = 1.5$, with $K^* = 3.65$ and $\frac{K^*}{N^*} = 2.4333$), and the third cut fixes B at the low-leverage SSS (LL-HSSS; $B^{LL} = 1.0967, N^{LL} = 2.6010$, with $K^{LL} = 3.6977$ and $\frac{K^{LL}}{N^{LL}} = 1.4216$). The thicker part of the lines indicates the regions of the state space in which the ergodic distribution of aggregates variables is nonzero. The white point indicates, in the

first top cut, the LL-SSS; in the second cut, the HL-SSS; in the third cut, the high-leverage point described above. Panel (b), at the top right, follows the same pattern that panel (a), but switching the roles of equity (N) and debt (B). Finally, panel (c), at the bottom, shows the complete three-dimensional representation of the PLM. The shaded area in this bottom panels highlights the region of the PLM visited in the ergodic distribution with (non-trivial) positive probability. The thin red line is the "zero" level intersected by the PLM: to the right of the line, aggregate debt falls, and to the left, it grows.

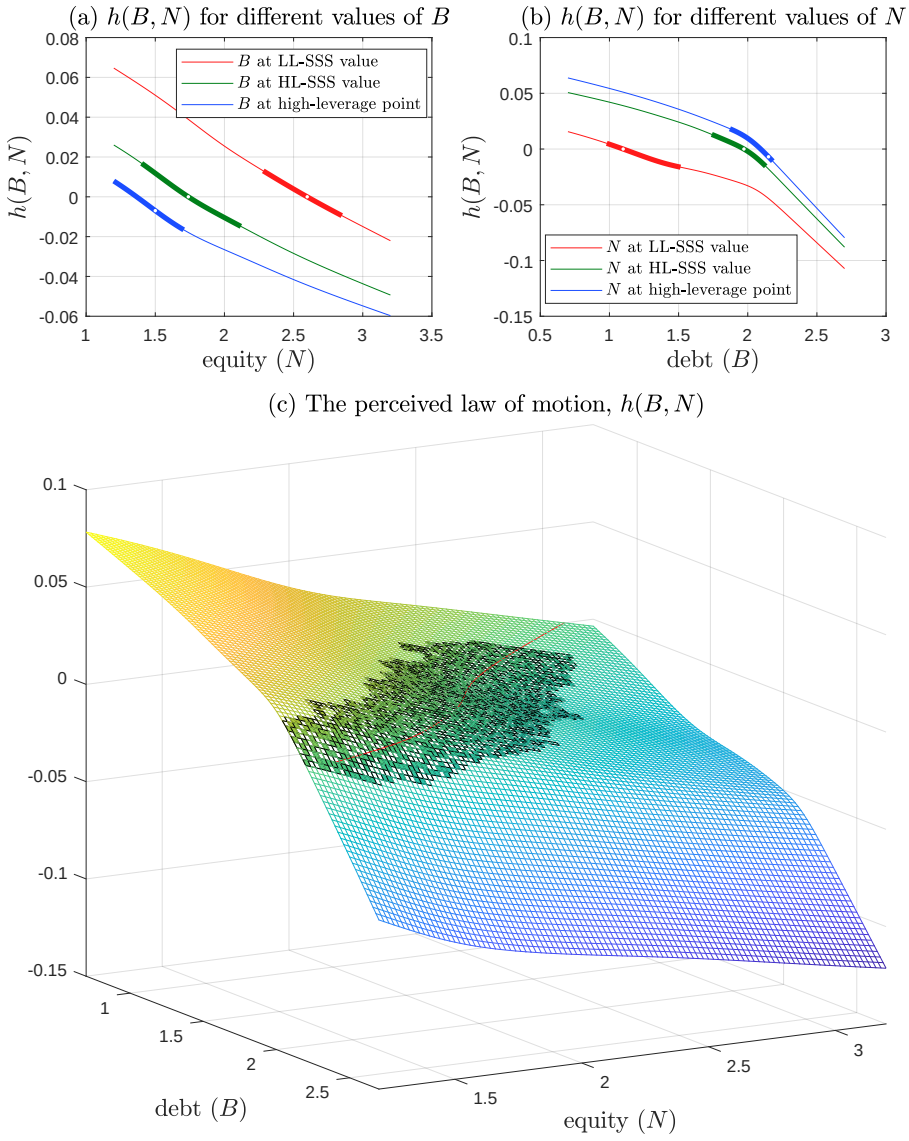


Figure 2: The PLM $h(B,N)$.

Note: White points in panels (a) and (b) indicate the LL-SSS, the HL-SSS, and a selected high-leverage point. The thicker part of the lanes in panels (a) and (b) and the shaded area in panel (c) displays the region of the PLM visited in the ergodic distribution. The thin red line is the "zero" level intersected by the PLM.

Figure 2 demonstrates the nonlinearity of $h(B,N)$ even within the area of the ergodic dis-

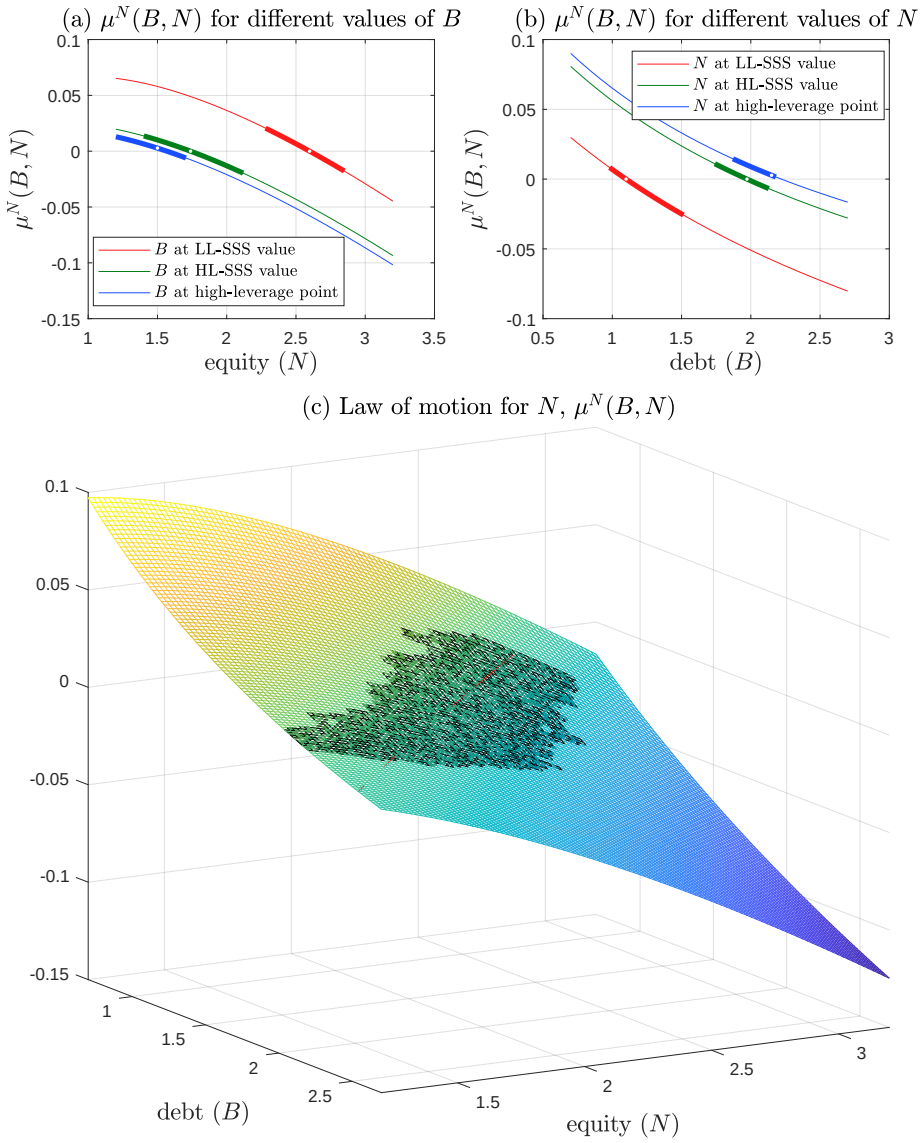


Figure 3: The law of motion $dh(B,N)$.

Note: White points in panels (a) and (b) indicate the LL-SSS, the HL-SSS, and a selected high-leverage point. The thicker part of the lanes in panels (a) and (b) and the shaded area in panel (c) displays the region of $dh(B,N)$ in the ergodic distribution. The thin red line is the "zero" level intersected by $dh(B,N)$.

tribution that has positive mass. The agents in our economy expect different growth rates of B_t in each region of the state space, with the function switching from concave to convex along the way. While this argument is clear from the shape of panel (c), it encodes rich dynamics. For instance, panel (b) shows how, as leverages increases, $h(B,N)$ becomes stepper and, in the ergodic distribution, more concave. Given the same level of debt, a higher level of leverage induces larger changes in the level of aggregate debt as the financial expert is exposed to more capital risk. This result will resurface several times in future paragraphs. Panel (c) also shows that, as intuition suggests, $h(B,N)$ is generally decreasing in debt and equity.

Figure 3 replicates Figure 2, except for $dh(B, N)$ once we have used equation (32). Similar comments regarding the nonlinear structure of the solution apply here. For example, now, $dh(B, N)$ becomes higher as a function of equity as the level of leverage falls.

The nonlinearity of the PLM confirms our conjecture that more traditional solution methods that rely on linear structures (conditional on aggregate shocks) might not be appropriate for solving this model. We illustrate this argument more formally by looking at the forecasting capability of our PLM. The R^2 associated with the PLM we compute using our neural network is 0.9922, with an RMSE of 0.0004. The forecasting errors, furthermore, are nicely clustered around zero, with a mode roughly equal to zero. To compare these errors with the standard Krusell and Smith (1998) algorithm of finding an OLS over a linear regression on endogenous state variables, we have recomputed our model using the latter approach. The R^2, in that case, is 0.8275, considerably lower than typical values reported in the literature for more standard heterogeneous agents models, and with an RMSE of 0.0021.

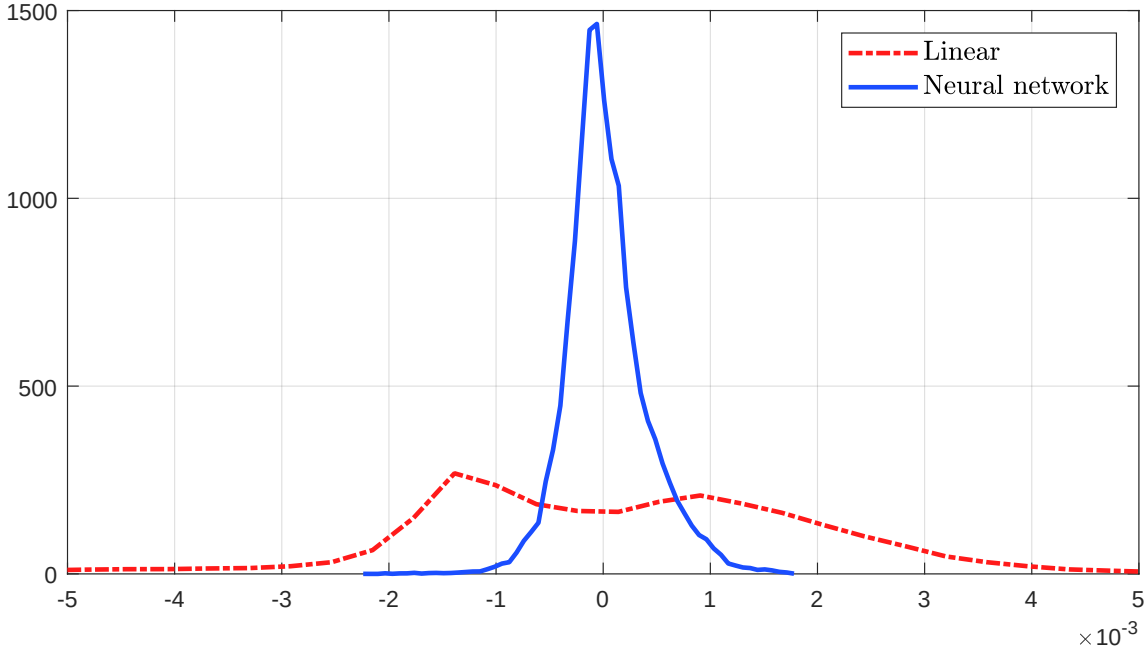


Figure 4: Forecasting error distribution at a 1-month horizon, linear PLM (left) and nonlinear (right).

Figure 4 plots the histogram of forecasting errors at a 1-month horizon (the time step selected in the simulation step), with a continuous blue line representing the errors from our algorithm and the discontinuous red line the errors from a linear-in-endogenous variables Krusell-Smith algorithm. The latter approach produces more volatile forecasting errors, which are also skewed to the right and without a mode at zero. These results back the importance of taking into account the nonlinearities of the model when computing the PLM.

In Appendix D, we discuss other alternatives to the standard Krusell and Smith (1998)

algorithm and argue that our method has advantages over them as well. Also, we checked that adding additional moments to the OLS regression do not help much concerning accuracy.

6.2 The phase diagram

Figure 5 plots the phase diagram of our model along with the aggregate debt (B) on the x-axis and equity (N) on the y-axis. The blue line in Figure 5 represents the loci of zero changes in debt, $h(B, N) = 0$. The line inherits the nonlinear dependence on B and N of the right-hand side of equation (19), the object $h(B, N)$ approximates. There is a convex segment for low levels of debt and a concave segment for high levels of debt. The discontinuous red line represents the loci of zero changes in the equity of the expert, $dN(B, N) = 0$. These two lines are the intersections of zero with $h(B, N)$ and $dN(B, N)$ in Figures 2 and 3, which we represented with a thin red line. The arrows in Figure 5 indicate the movement of debt, B, and equity, N, when we are away from the blue and red lines.

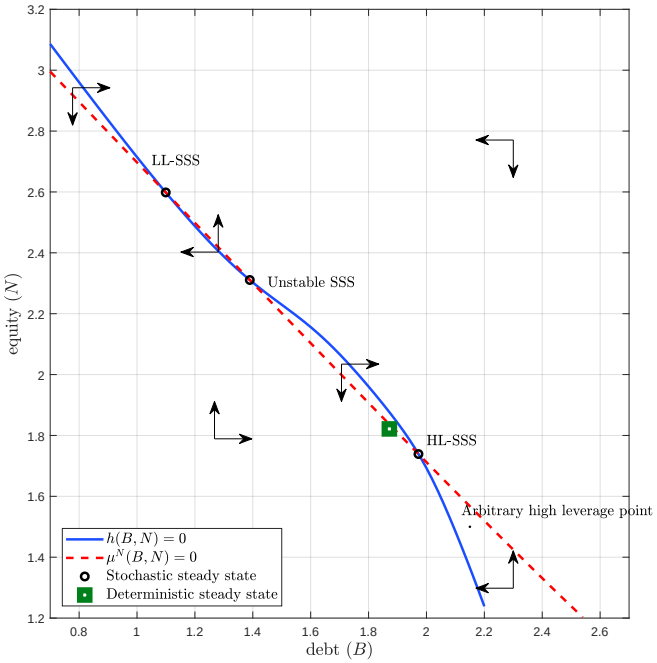


Figure 5: Phase diagram, DSS, and SSS(s).

The two lines intersect three times, defining three SSS(s). From the bottom right, the first intersection is the stable HL-SSS (recall, with $B^{HL} = 1.9641, N^{HL} = 1.7470$, and $K^{HL} = 3.7111$). This SSS is the closest to the DSS (green square). In comparison with the DSS ($B = 1.8718, N = 1.8215$, and $K = 3.6933$), the HL-SSS has 0.5% more capital, 4.9% more debt, and 4.1% less equity. The second intersection is at a middle-leverage SSS with less debt and more equity ($B^{ML} = 1.3800, N^{ML} = 2.3204$, and $K^{ML} = 3.7004$). This SSS is, however, unstable, and the dynamics of the economy quickly move away from it. Thus, we will not discuss

it further. The final third intersection, at the top left, is the stable LL-SSS. Here, debt is much smaller ($B^{LL} = 1.0967$) and equity considerably higher ($N^{LL} = 2.6010$) than in the HL-SSS, yielding however a similar capital, $K^{LL} = 3.6977$. For completeness, we also plot the point of high leverage that we use in Figures 2 and 3 to show, later, the behavior of the economy when leverage is high.

6.3 Why do we have two stable SSS(s)?

To understand why we have two stable SSS(s) in Figure 5, first, recall the diffusion for the expert net wealth in equation (24) and rewrite it in rates as:

$$\frac{dN_t}{N_t} = \left(\alpha K_t^{\alpha-1} - \delta - \widehat{\rho}\right) dt + \sigma^2 B_t \frac{K_t}{N_t^2} dt + \sigma \frac{K_t}{N_t} dZ_t. \tag{38}$$

The volatility term of (38), $\sigma \frac{K_t}{N_t} dZ_t$, reflects how the proportional effect on the expert's equity of a shock to capital is higher when her leverage level is high. Given that the expert must absorb all the gains or losses from a shock to capital, the smaller the base of her equity in relation to total capital, the larger the proportional drop in N_t.

The two drift terms of (38) explain why the rate of recovery of N_t from a negative shock to capital is roughly independent of leverage. The first drift term, $\left(\alpha K_t^{\alpha-1} - \delta - \widehat{\rho}\right) dt$, does not depend directly on N_t. The second drift term, $\sigma^2 B_t \frac{K_t}{N_t^2} dt$, indicates that the rate of accumulation of equity will be higher when N_t is small relative to K_t. However, since the term is multiplied by σ^2, a small quantity, the total effect on $\frac{dN_t}{N_t}$ is muted.

Next, let us inspect the law of motion for debt, (19), and rearrange it as:

$$dB_t = \left((1-\alpha)K_t^{\alpha} - \sigma^2 \frac{K_t}{N_t} B_t\right) dt + \left((\alpha K_t^{\alpha-1} - \delta) B_t - C_t\right) dt. \tag{39}$$

The first drift term of (39), $\left((1-\alpha)K_t^{\alpha} - \sigma^2 \frac{K_t}{N_t} B_t\right) dt$, becomes smaller after a negative capital shock: less capital lowers labor income, $(1-\alpha)K_t^{\alpha}$, and $\sigma^2 \frac{K_t}{N_t} B_t$ rises because the expert must be compensated with a larger excess return for her higher leverage to be an equilibrium choice. How much smaller this drift term is depends positively on leverage.

The second drift term of (39), $\left((\alpha K_t^{\alpha-1} - \delta) B_t - C_t\right) dt$, is bigger after a negative capital shock, as the marginal productivity of capital, $\alpha K_t^{\alpha-1}$, goes up and households reduce their consumption, C_t, to compensate for lower wages. However, when leverage is high, the drop of household consumption is moderate. The risk-free interest rate is more persistent when leverage is high because of the slow recovery of N_t and the associated compensation of the financial expert for her higher leverage. A more persistent risk-free rate induces less intertemporal substitution by households. We will see, when we report the time-varying consumption decision rules in

Figure 19, how wealthy households drive this phenomenon.

In summary: equations (38) and (39) show that, when leverage is high, i) the financial expert's equity falls more; ii) the recovery of this equity is roughly at the same rate than when leverage is low; iii) debt increases less; and iv) consequently, capital recovery is slow.

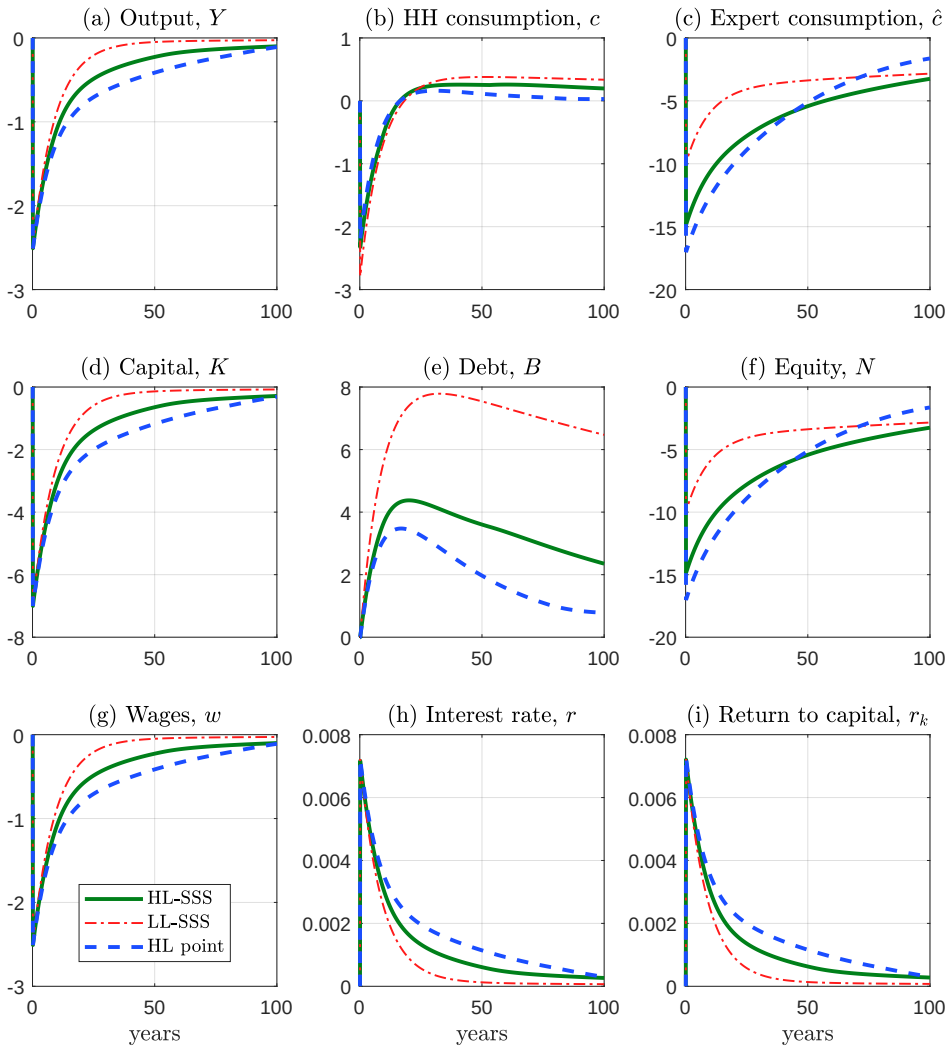


Figure 6: GIRFs for different initial states.

We can gauge the quantitative size of these four points in Figure 6, where we display the generalized impulse response functions (GIRFs) to a one-standard-deviation negative capital shock. The GIRF is defined as the difference between the transition path if an initial shock hits the economy and that if no shock arrives.[10]

[10]The word "generalized" comes because, in comparison with linear models, impulse responses in nonlinear models are i) state dependent, that is, they depend on the point where the economy is when the shock arrives; ii) their shape is a nonlinear function of the size of the shock; iii) and they are not symmetric, i.e., the response to a negative shock is not the mirror image of the response to a positive shock. Thus, we need to specify the size and sign of the shock and when this shock occurs.

We plot three GIRFs: to a one-standard-deviation negative capital shock that hits the economy at the HL-SSS (continuous green lines); to a one-standard-deviation negative capital shock that hits the economy at the LL-SSS (discontinuous red lines); and to a one-standard-deviation negative capital shock that hits the economy at the arbitrary high-leverage point defined in Figure 5 (discontinuous blue lines). Time units are years.

In all three cases, the shock destroys the same amount of capital (left panel in the central row) and output falls, at impact, by the same quantity (left panel in the top row). However, the higher the leverage, the larger the reduction of equity at impact (right panel in the central row). As time passes, the higher the leverage, the longer it takes for capital to recover as the expert requires more time to rebuild her equity and accumulates less debt (central panel in the central row). The lower debt is an equilibrium choice for households because of the more persistent risk-free rate (central panel in the bottom row) induces a mild reduction in household consumption (central panel in the top row).

The state-dependence of the GIRFs demonstrate the critical nonlinearities of our model. Furthermore, the one-standard-deviation shock is not large enough to send the economy away from the basins of attraction of each SSS. An even larger shock or a shock closer to the frontier between the two basins, by inducing a basins switch, will have even acuter nonlinear consequences (we will return to this point below).

A central property of the GIRFs is their persistence. When leverage is high, even after 40 years, the economy is still around half a point percentage below its pre-shock level. The dynamics of equity and debt accumulation propagate aggregate shocks in ways that are not present in models without financial frictions.

The latter point is particularly salient for the next step of our argument explaining the existence of multiple SSS: the evolution of wages (left panel in the bottom row). When leverage is high, wages fall for a much longer period than when leverage is low. This means, for example, that the standard deviation of wages when the economy fluctuates around the basin of attraction of the HL-SSS is 0.0125 while the same statistic around the basin of attraction of the LL-SSS is 0.0108.

The persistent fall in wages induces a stronger precautionary behavior by the households. Since they want to smooth their consumption, households demand for debt sifts to the right when leverage is high. At the same time, the supply of debt is given by equation (16) after some term rearrangement:

$$B_t = \left(1 + \frac{1}{\sigma^2}(rc_t - \delta - r_t)\right) N_t,$$

Thus, the debt market will clear with a reduction in r_t. At the HL-SSS, this reduction in r_t will need to compensate, also, for the fact that the higher excess returns induces a lower financial expert's equity, N_t, and a slightly higher capital, with a subsequent reduction in rc_t.

Hence, we have shown, why we have a fixed point at the HL-SSS. High leverage makes wages persistently lower after a negative capital shock due to the large reduction in the expert's equity and the associated capital dynamics. Persistently lower wages create a precautionary motive that lowers the risk-free interest rate and such low risk-free interest rate justifies the high leverage of the financial expert.

We only need to reverse the argument to show why we also have a fixed point at the LL-SSS. Low leverage makes wages recover quickly after a negative capital shock as the expert rebuilds her equity briskly. The lower associated precautionary behavior means the risk-free interest rate is relatively high and that sustains the low leverage of the financial expert.

Mechanically, note the high sensitivity of leverage to excess returns induced by $\frac{1}{\sigma^2}$ in equation (16). The excess return in basis points and our parameter values for the HL-SSS is:

$$rc_t^{HL} - \delta - r_t^{HL} = 4.1636$$

and for the LL-SSS:

$$rc_t^{LL} - \delta - r_t^{LL} = 2.7864.$$

This high sensitivity accounts for the large quantitative differences between the two SSS(s) in terms of how a roughly equal total wealth ($K^{HL} = 3.7111$ vs. $K^{LL} = 3.6977$) is allocated between experts ($N^{HL} = 1.7470$ vs. $N^{LL} = 2.6010$) and households ($B^{HL} = 1.9641$ vs. $B^{LL} = 1.0967$). The impossibility of issuing state-contingent debt by the expert and the high sensitivity of leverage to excess returns eliminates the "concavity" forces that lead to a unique SSS in a setup without financial frictions.

6.4 Convergence to the SSS(s) and random fluctuations around them

How do we know that the two SSS(S) described above are stable? The state space $(g(\cdot), N)$ is infinite dimensional and, hence, we cannot check convergence numerically for all possible initial states. Instead, we analyze convergence for densities visited in the ergodic distribution.

Figure 7 considers an array of different initial income-wealth densities and equity levels $(g_0(\cdot), N_0)$, selected from the simulations used to compute the aggregate ergodic distribution and analyze the transitional dynamics in the absence of aggregate shocks (agents continue forming their expectations assuming $\sigma > 0$). We plot, in red, the paths converging toward the LL-SSS and, in green, the paths converging toward the HL-SSS. In all cases, the economy converges to the SSS(s) (which one depending on the basin of attraction of the initial condition). The convergence path is plodding: for initial values in the high- or low-leverage regions convergence may take several centuries.

Notwithstanding, we cannot rule out that, for other initial conditions, the model would

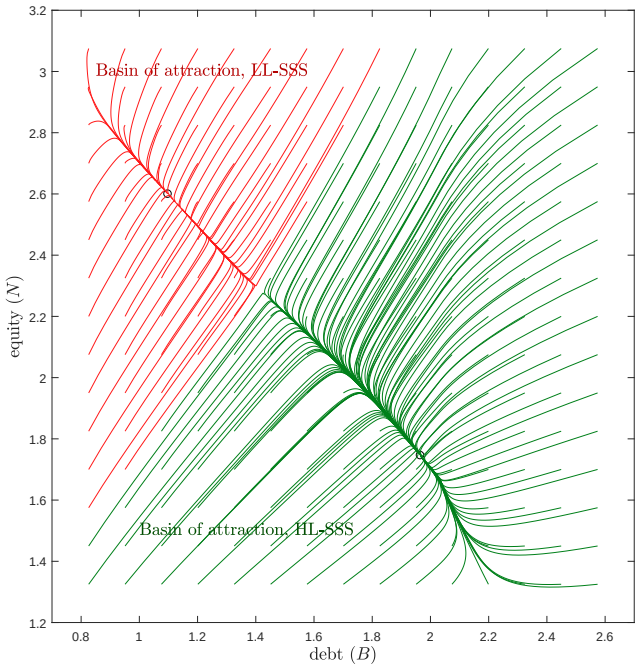


Figure 7: Convergence paths.

not converge to a SSS. This limitation is related to the self-justified nature of the solution. The PLM is computed based on the income-wealth distributions visited along the paths of the ergodic distribution. One could potentially find a distribution that would lead to alternative dynamics. Similarly, one could also find other equilibria beyond the one we compute (although, despite our efforts, we failed to do so). Recall, too, that multiplicity of SSS(s) is different from multiplicity of equilibria: in our model, we are in one basin of attraction or another depending on the sequence of shocks that the economy has experienced, but the equilibrium we compute is unique given the initial condition and sequence of shocks.

An interesting feature of the transitional dynamics is the curvature of the paths that lie in regions with relatively low levels of capital. These paths are bent because debt adjusts faster than equity. If equity and aggregate capital are below their long-run values, households accumulate wealth, increasing the financial expert's leverage. Once the economy reaches a point with low equity, but a capital level closer to that in the SSS, there is a progressive redistribution of wealth from households to the expert as the latter accumulate excess returns undisturbed, in the absence of aggregate shocks, by capital losses. The converse phenomenon occurs in the case of initial high equity and capital.

Figure 8 documents how the economy evolves around each SSS by plotting several random equilibrium paths corresponding to different sequences of shocks. In red, we plot the paths when they are in the basin of attraction of the LL-SSS. In green, we plot the paths when they are in the basin of attraction of the HL-SSS. A path in the basin of attraction of the LL-SSS can be pushed the basin of attraction of the HL-SSS by a sequence of shocks that reduce equity and

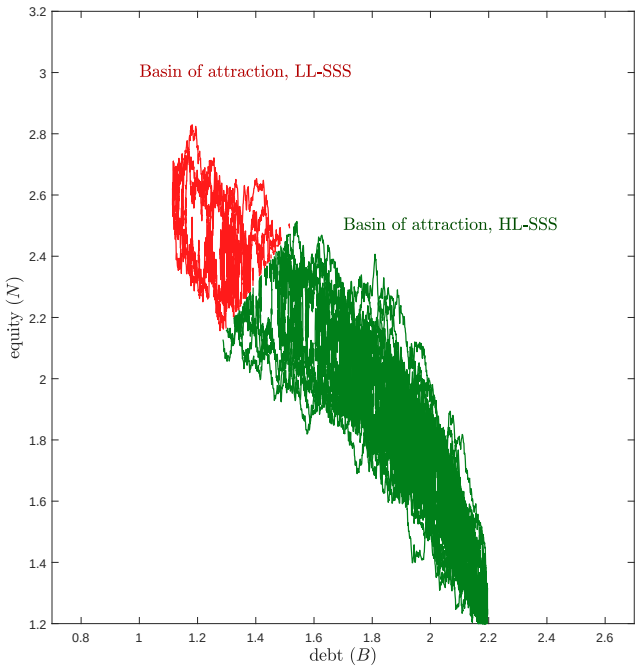


Figure 8: Simulation of equilibrium paths.

increase debt. Conversely, a series of shocks that increase equity and reduce debt can push the equilibrium path from the basin of attraction of the HL-SSS toward the basin of attraction of the LL-SSS. These shocks will induce a sharp nonlinearity, as the GIRFs do not return to their origin, but to the new SSS.

6.5 Time-varying endogenous aggregate risk

Table 2 reports the moments of the economy conditional on the basin of attraction at which the variables fluctuate (note that the ergodic distribution of the economy over time is the combination of the distributions at both basins; see Subsection 6.7 below).

	Mean	Standard deviation	Skewness	Kurtosis
$Y^{\text{basin } HL}$	1.5807	0.0193	-0.0831	2.8750
$Y^{\text{basin } LL}$	1.5835	0.0166	0.16417	3.1228
$r^{\text{basin } HL}$	4.92	0.3360	0.1725	2.8967
$r^{\text{basin } LL}$	4.88	0.2896	-0.0730	3.0905
$w^{\text{basin } HL}$	1.0274	0.0125	-0.0831	2.875
$w^{\text{basin } LL}$	1.0293	0.0108	0.1642	3.1228

Table 2: Moments conditional on basin of attraction.

The main takeaways from Table 2 are as follows. First, the mean of output is higher at the LL-SSS basin than at the HL-SSS. This is despite output being higher at the HL-SSS itself

than at the LL-SSS ($Y^{HL} = 1.5824 > Y^{LL} = 1.5804$). The reversal is due to a distribution of output at the LW-SSS basin that is more skewed to the right. Second, the standard deviation of output is higher in the basin of the HL-SSS. This observation is the sense in which financial frictions create time-varying endogenous aggregate risk: when leverage is high, the economy fluctuates more even when the variance, σ, of aggregate shocks is the same. The mechanism, outlined a few paragraphs back, is the higher persistence of capital in the HL-SSS basin after a shock. This higher volatility will be even bigger when leverage is higher than at the HL-SSS (something that often occurs, as seen in Figure 8). Third, there is a mild excess kurtosis of output at the HL-SSS. Similar statements can be made about the risk-free rate and the wage, the latter being nearly 16% more volatile in the basin of attraction of the HL-SSS. Figure 9 plots the distributions behind the moments in Table 2.

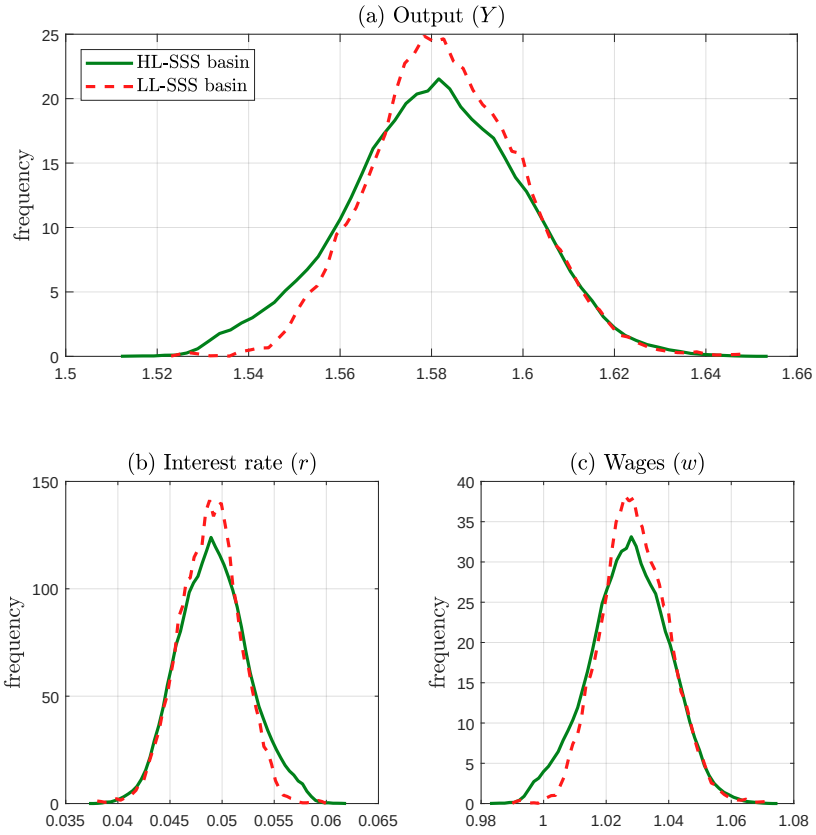


Figure 9: Ergodic distributions conditional on basin of attraction.

Figure 10 plots the histogram of the duration of spells of the economy around each SSS. The figure shows how the interaction of financial frictions and heterogeneity can lead to complex middle- and long-run dynamics that are not present in more conventional business cycle models. The average duration of a spell of the economy around the HL-SSS is 55.40 years, a long period of high leverage. On the other hand, the average duration of a spell of the economy around the LL-SSS is 9.60 years, a much shorter length. Consequently, the economy spends 91.40% of its

time in the basin of attraction of the HL-SSS. However, the distribution of spells has long tails: some spells around the LS-SSS can last a century.

Thus, the model suggests that an economy that has suffered a significant loss of capital, for example, during a war, will pass through a prolonged period of steady capital accumulation, possibly lasting several decades, with little debt and low volatility. In the absence of further shocks, this era will culminate with the economy moving endogenously to a period of higher leverage and larger responses to capital shocks. As we will see in Section 7, this transition will also be accompanied by a sharp increase in the Gini coefficient of wealth inequality. As a rough first approximation, this process resembles the experience of many Western European countries after 1945.

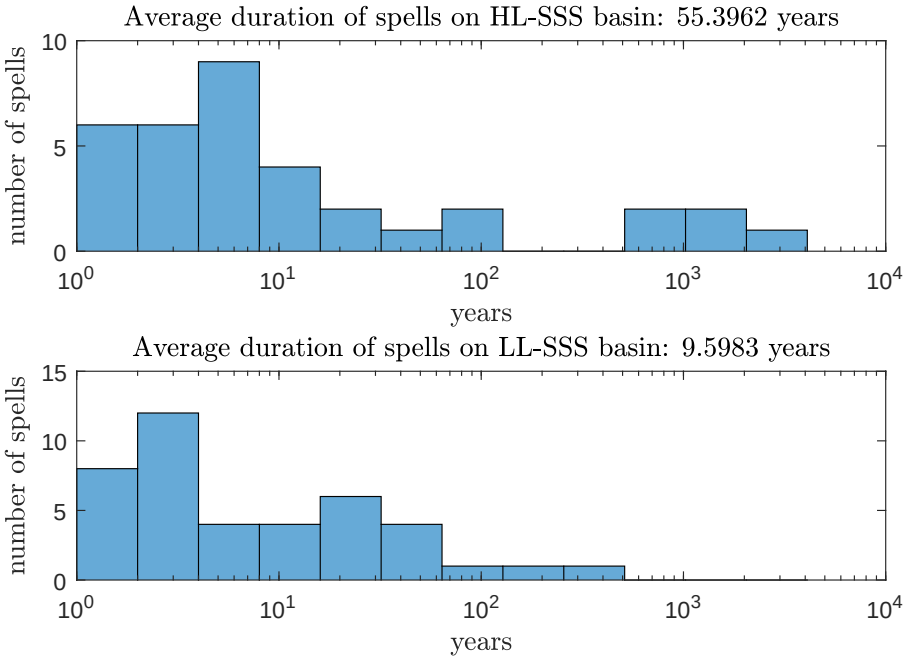


Figure 10: Spell durations at each SSS.

Why does the economy stay longer, on average, around the HL-SSS? For our parameter values, i) the state variables values where the financial expert wants to lever aggressively cover a large region; and ii) the level of exogenous aggregate and idiosyncratic risk sufficiently high as to induce the households to demand enough debt.

6.6 The role σ in determining the different SSS(s)

Our previous discussion highlighted how the equilibrium excess return sends the economy toward a high or a low-leverage region. Figure 11 pushes this argument further by exploring the role of σ in determining the different SSS(s). Each panel plots, for a different value of σ, the

phase diagram of the economy following the same convention than Figure 5.[11]

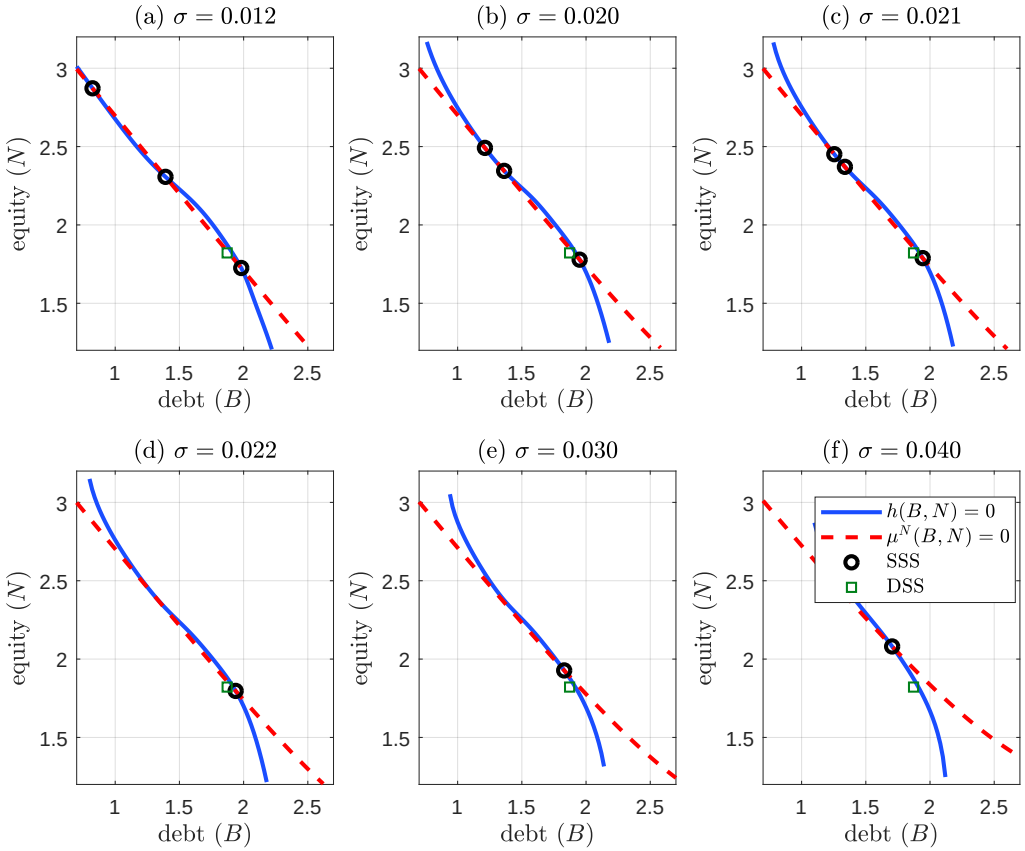


Figure 11: Phase diagram as a function of σ.

For low values of σ, the precautionary motive of households is mild and, therefore, we can sustain the LL-SSS (in addition to the HL-SSS). However, as we increase σ, the precautionary motive becomes stronger, $h(B, N) = 0$ becomes curvier, and the LL-SSS disappears. Even with high risk-free interest rates, households do not demand enough bonds. Similarly, as σ grows the HL-SSS moves to the left (i.e., less debt and more equity) as the financial expert is exposed to additional capital risk. This effect becomes sufficiently strong that the HL-SSS, instead of being to the right of the DSS (i.e., more debt and more equity than the DSS because of the higher excess return induced by precautionary savings), crosses to the left of the DSS.

Figure 12 plots the values of the LL-SSS, the unstable SSS, and the HL-SSS (plus, for reference, the DSS) as a function of σ. We can see how the leverage in the HL-SSS is a negative function of σ, a roughly constant function in the unstable SSS, and an increasing function in the LL-SSS (until the additional SSS(s) disappear). The mechanism for these three slopes is the same than the one discussed above. In the HL-SSS, as σ grows, the expert wants to unload

[11]Here, as in all the other exercises of the paper where we perform comparisons across different values of a parameter, we keep all the other parameter values constant. In that way, we can understand the mechanisms at work better than if we re-parameterized the model.

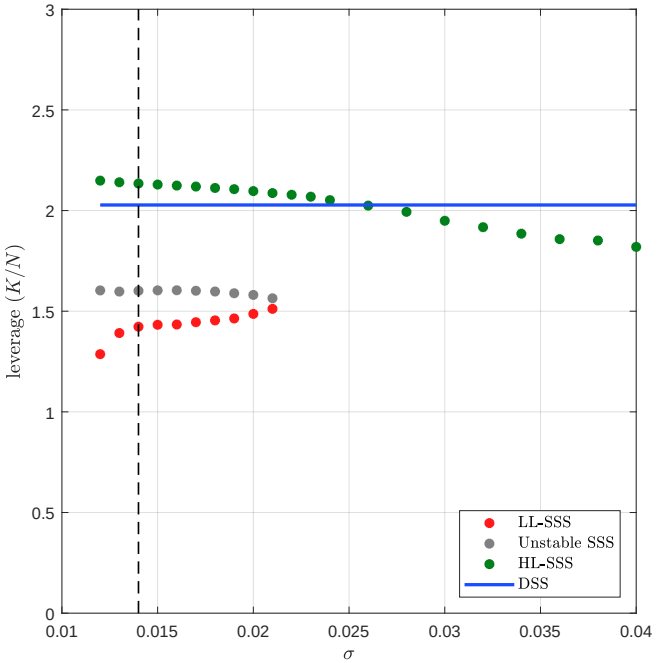


Figure 12: SSS as a function of σ.

some of the capital risk by reducing its leverage. In comparison, in the LL-SSS, the households demand more debt as σ increases.

6.7 The aggregate ergodic distribution of debt and equity

Panel (a) of Figure 13 displays the aggregate ergodic joint distribution of debt and equity $F(B,N)$. This distribution is defined, for any subset Ω of the state space, as:

$$\mathbb{P}\left\{(B,N)\in\Omega\right\}=\int_{\Omega}dF.$$

The ergodic distribution is not obtained directly from the PLM, but from the simulation of the paths of the income-wealth distribution and equity. Naturally, the PLM is employed in the HJB equation (33) to obtain the optimal consumption policies of individual households. We use a Monte Carlo simulation of 5,400 years at a monthly frequency. We initialize the model at the DSS, and we disregard the first 400 samples as a burn-in.[12]

Panels (b) and (c) of Figure 13 plot the marginal ergodic distributions for debt and equity.

[12]We compute the PLM based on simulations on a region of the state space. However, when forming expectations, households evaluate the PLM over the entire state space. Thus, the PLM is extrapolated over the regions of the state space not included in the support of the ergodic distribution. There is no guarantee that the dynamics of the model in the extrapolated region coincide with the ones expected in the PLM. Thus, as in Krusell and Smith (1998), our approximation can be interpreted as a self-justified equilibrium in which households' beliefs about the PLM coincide with the actual law of motion only in the equilibrium paths. Off-equilibrium, the PLM and the actual law of motion may diverge, but households never discover it, as this region is never visited.

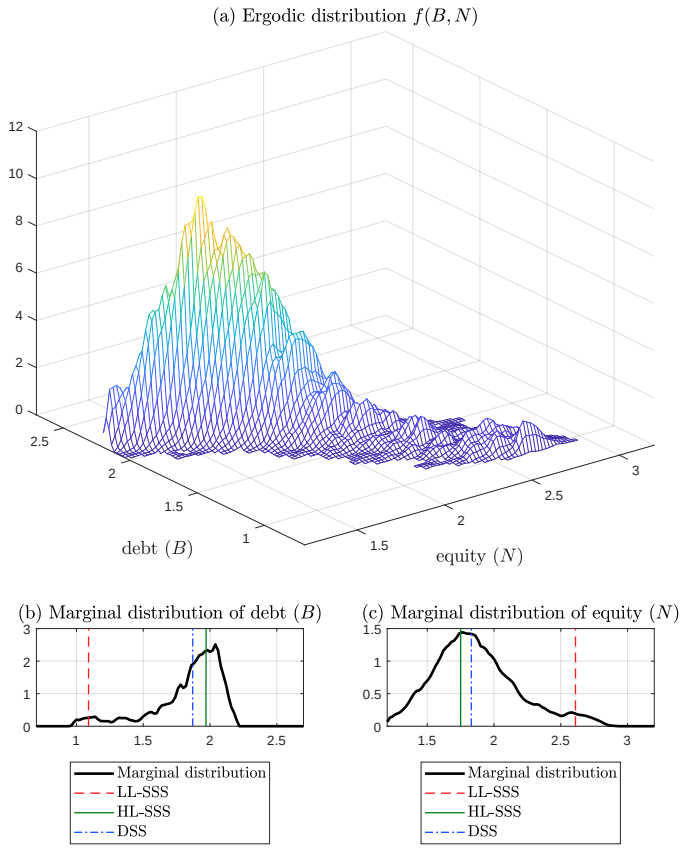


Figure 13: Ergodic distribution $F(B, N)$. Lighter colors indicate higher probability.

These two panels show how the economy spends most of the time in a region with debt levels between 1.5 and 2.2 and equity between 1 and 2.2. Those values of debt and equity correspond to the neighborhood of the HL-SSS. In comparison, the neighborhood of the LL-SSS appears much less often. As we explained above, the desire of households of accumulating more debt lowers the height of this second peak as there is a strong force for reversion toward higher levels of leverage. We will see momentarily how this result changes as we vary, for example, the degree of heterogeneity among households.

Note, as well, how the marginal distribution of debt is much more concentrated than the marginal distribution of equity. The ergodic distribution has a substantial tail in areas of high equity and low debt, creating an intriguing asymmetry in aggregate dynamics.

7 The role of heterogeneity

This section explores how heterogeneity among households drive our results. To do so, we start by comparing, in Figure 14, the wealth distribution in the DSS (discontinuous blue line), the HL-SSS (continuous green line), and the LL-SSS (discontinuous red line). In the left panel, we plot the distributions for low labor productivity households (with circles denoting the mass

at zero assets). In the right panel, we plot the distributions for high productivity households.

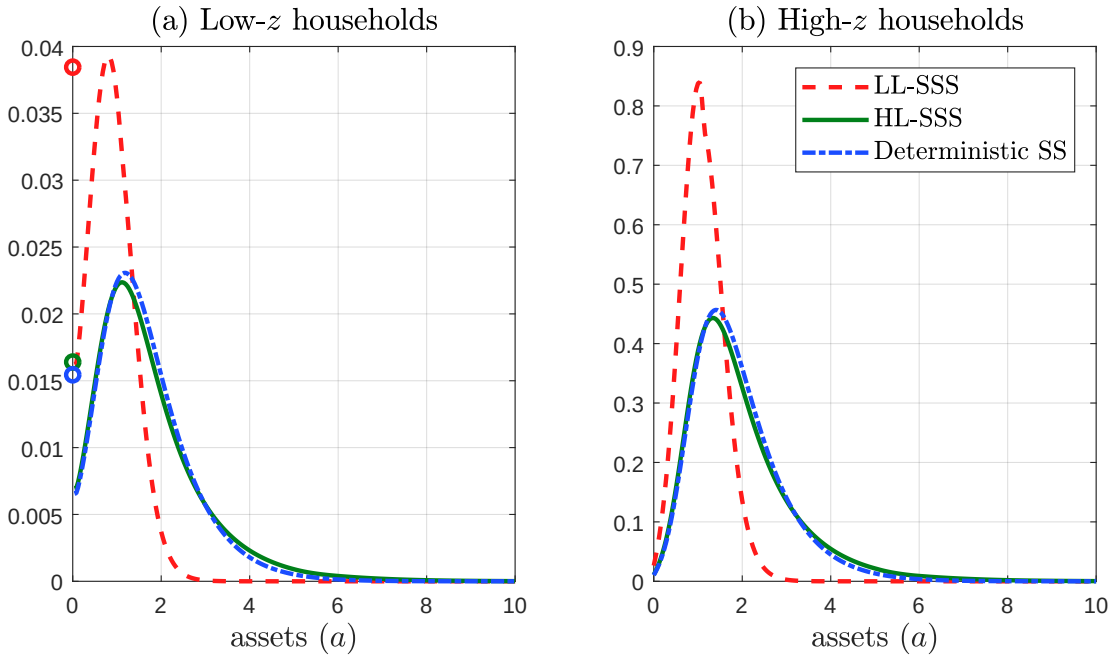


Figure 14: Wealth distribution in the DSS and SSS.

The wealth distribution slightly shifts to the left as we move from the DSS to the HL-SSS, but shifts rather dramatically, also to the left, as we travel from the HL-SSS to the LL-SSS. The higher volatility and persistence of wages in the HL-SSS pushes the assets of the households to the left, in particular, the richer ones, with a thicker right tail. This leads to substantial differences in the Gini coefficient of wealth: 0.28977 in the DSS, 0.24398 in the DSS, and 0.31968 in the HL-SSS.

Next, we show how the results of Section 6 change as we modify the forces driving the wealth distributions in Figure 14. We will answer this question in three steps. First, we will vary the level of idiosyncratic labor risk among households. Second, we will show how the GIRFs of the economy depends on the degree of heterogeneity among households. Third, we will explore how the differences in consumption decisions and the variation in distributions over time account for the differences in the GIRFs.

7.1 Varying idiosyncratic labor risk

Figure 15 plots the phase diagram of the model for nine different values of z_1 (still keeping the ergodic mean of z equal to 1) from 0.67 to 0.95. Each panel follows the same convention that in Figure 5. In particular, we only plot the segments of $h(B, N)$ and μ^N (B,N visited in the ergodic distribution). Figure 15 is different from Figure 11 because now we keep σ constant. Instead, Figure 15 shows the consequences of having more or less idiosyncratic labor risk.

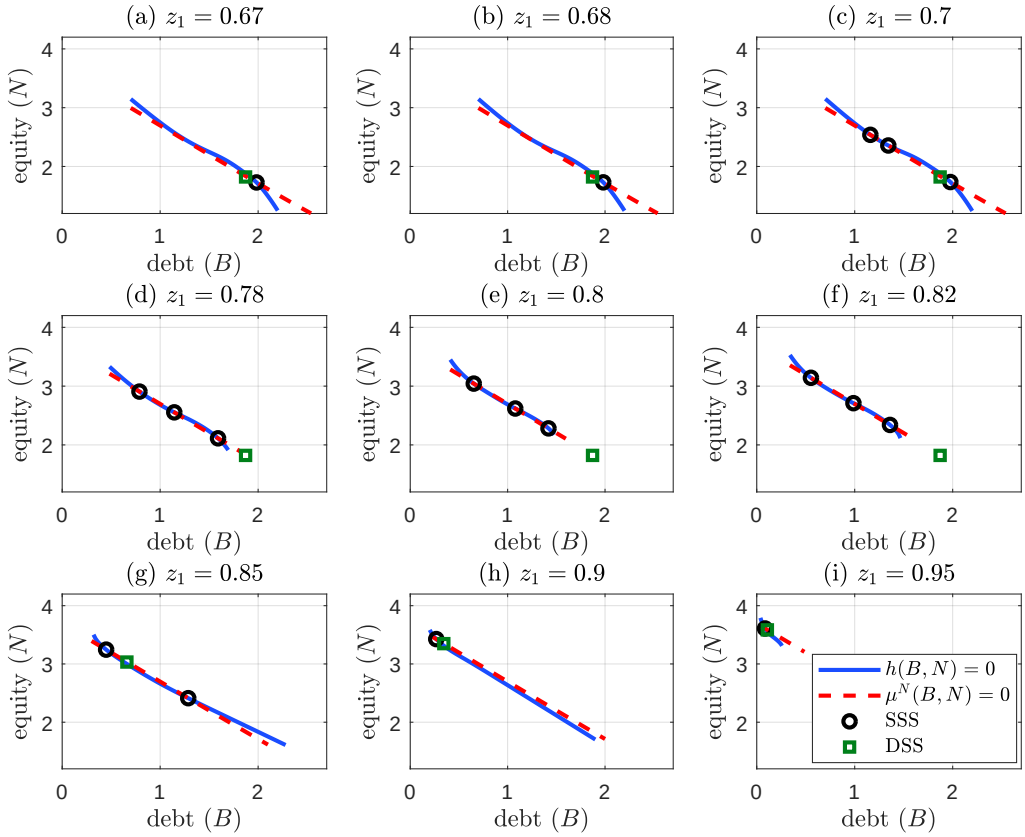


Figure 15: Phase diagram as a function of z_1.

For low levels of z_1, the only SSS is the HL-SSS (see, for instance, the left and center panels at the top row). Idiosyncratic labor risk is so high that households demand enough debt to sustain only one SSS, which, besides, has more debt than the DSS. As z_1 increases and we cross a threshold around 0.69, and we find three SSS(s), with the same interpretation than in our baseline case of Figure 15. As z_1 continues rising, the HL-LLL moves to the left of the DSS since households demand less debt to self-insure against idiosyncratic labor risk. See, for instance, the three panels of the center row. By the time z_1 reaches 0.9, the HL-SSS has disappeared: the precautionary demand for debt by household is now so weak that only the LL-SSS survives.

Figure 16 draws the ergodic distributions of equity and debt as z_1 varies. For low levels of z_1, most of the ergodic distribution accumulates in the region of high debt and low equity. As z_1 increases, the ergodic distribution spreads toward the upper left corner; first slowly, but gathering steam by the time we reach $z_1 = 0.85$. At this level, there is a bifurcation and region around the LL-SSS becoming the predominant one and the higher leverage region eventually disappearing. This change in the ergodic distribution is crucial for aggregate fluctuations since, as we saw in Figure 6, the responses of the economy to a capital shock is very different with low leverage than with leverage and with them, the moments of the economy.

A different view of the same results appears in Figure 17, where we plot the marginal

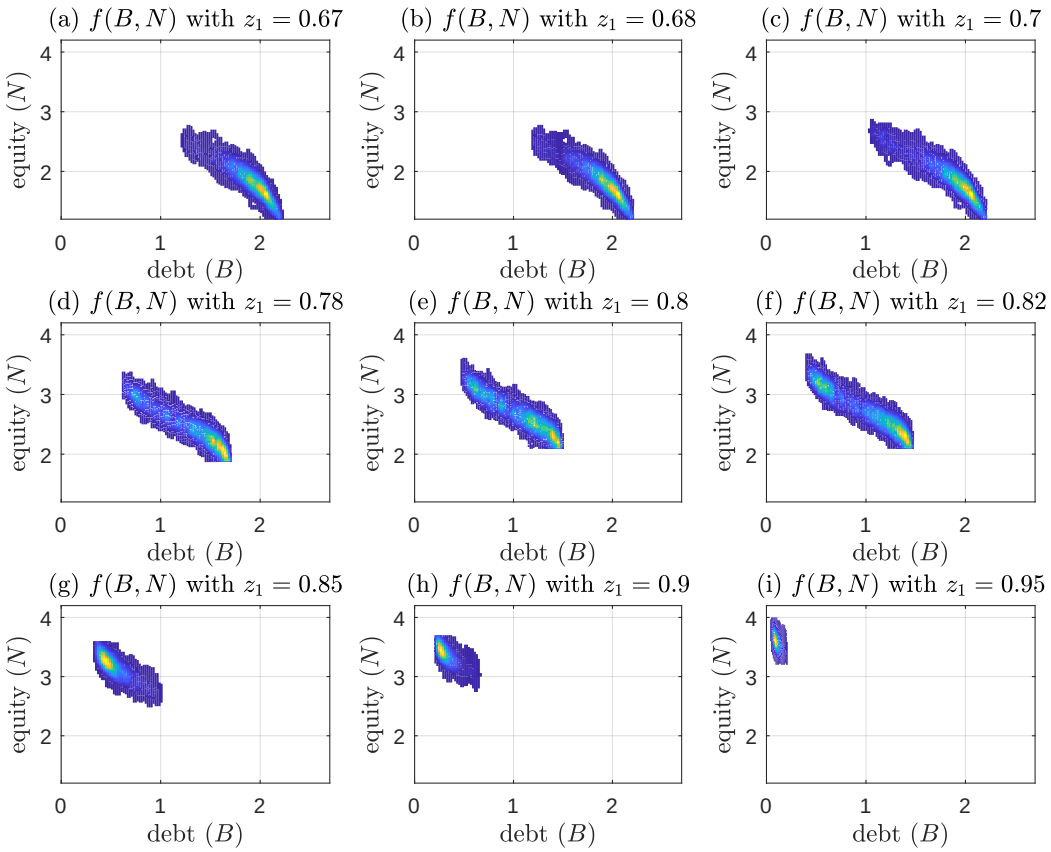


Figure 16: Ergodic distribution $F(B, N)$ as a function of z_1. Lighter colors indicate higher probability.

distributions of debt, equity, and capital as we change z_1.

7.2 High vs. low household heterogeneity

Figure 18 reports the GIRFs of the baseline, high household heterogeneity version of the model (continuous green line) to a one-standard-deviation negative capital shock when the economy is at the HL-SSS. These GIRFs are, by construction, identical to the GIRFs (also in continuous green lines) in Figure 6. Figure 18 also plots the GIRFs when $z_1 = 0.98$, a version of the model with very low household heterogeneity at its unique SSSS. While this is not strictly the representative household version of the model, idiosyncratic risk is so small (low productivity is just 2% below the ergodic mean, and only 5% of households are in such a situation), that we approximate well the limit case of a representative household.[13]

Both sets of GIRFs vary considerably. The fall in output (left panel in the top row) is

[13]We do not compute the exact representative household version of the model because it would require a different algorithm than the one we use for the heterogeneous version of the model. Thus, there would be some numerical chatter between both solutions that might complicate the evaluation of differences. Therefore, we prefer to set $z_1 = 0.98$ (as high as our algorithm can go before breaking down).

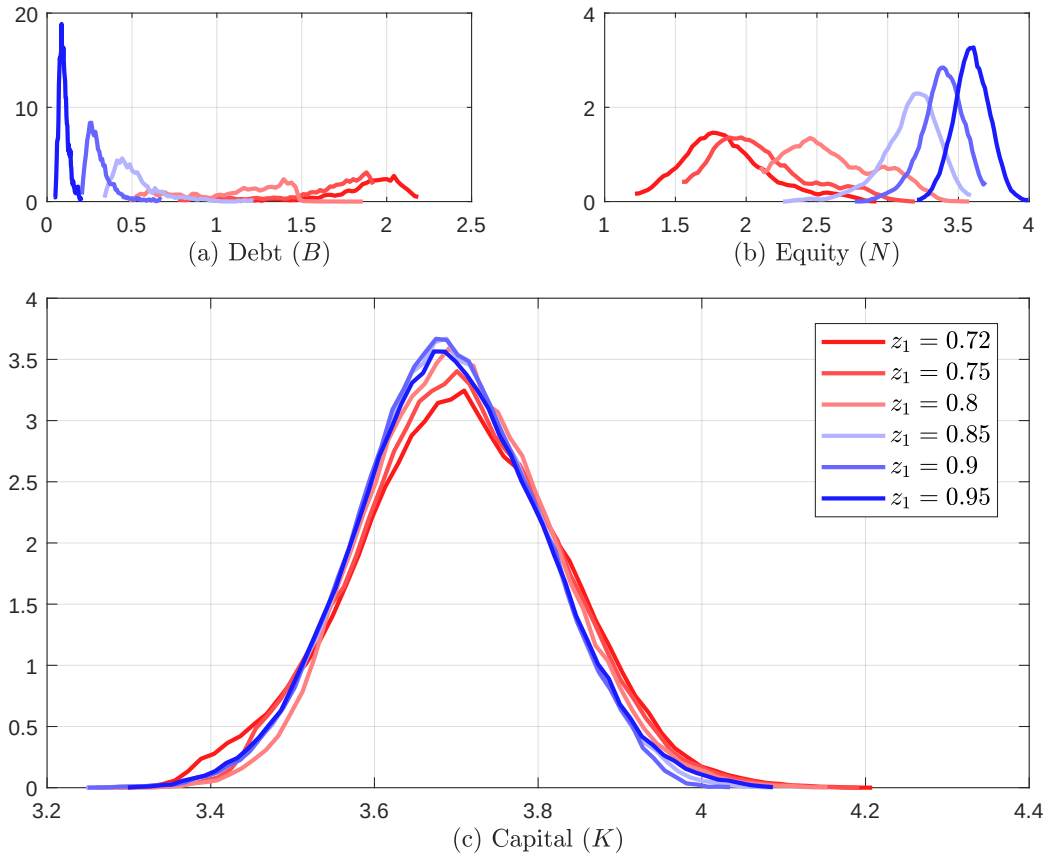


Figure 17: Ergodic distribution debt, equity, and capital as a function of z_1.

significantly more persistent when we have more heterogeneity. We will show below how this is related to the dynamics of consumption and wealth accumulation by wealthy households. The reduction of the total consumption of the households (central panel in the top row) is lower at impact when heterogeneity is high. In comparison, the consumption of the expert drops much more (right panel in two rows). This is because the expert starts with higher leverage (right panel of the central row the large decrease in equity) and it cannot issue much additional debt for a long time (central panel of the central row).

Figures 15-18 also connect changes in microeconomic conditions with aggregate outcomes. Imagine an economy that, due to technological change or structural transformation, starts having a more turbulent labor market, with households rotating in and out of unemployment more often (or suffering, in an equivalent interpretation of z, more changes to their wages while employed). The increased precautionary saving lowers the risk-free interest rate and rises, on average, leverage. Thus, the economy becomes, due to magnified idiosyncratic risks, more volatile even when the volatility of the aggregate shocks remains constant.

This result suggests the existence of a link between more wealth inequality among capital owners before the financial crisis of 2007 and more debt, lower risk-free interest rates, and

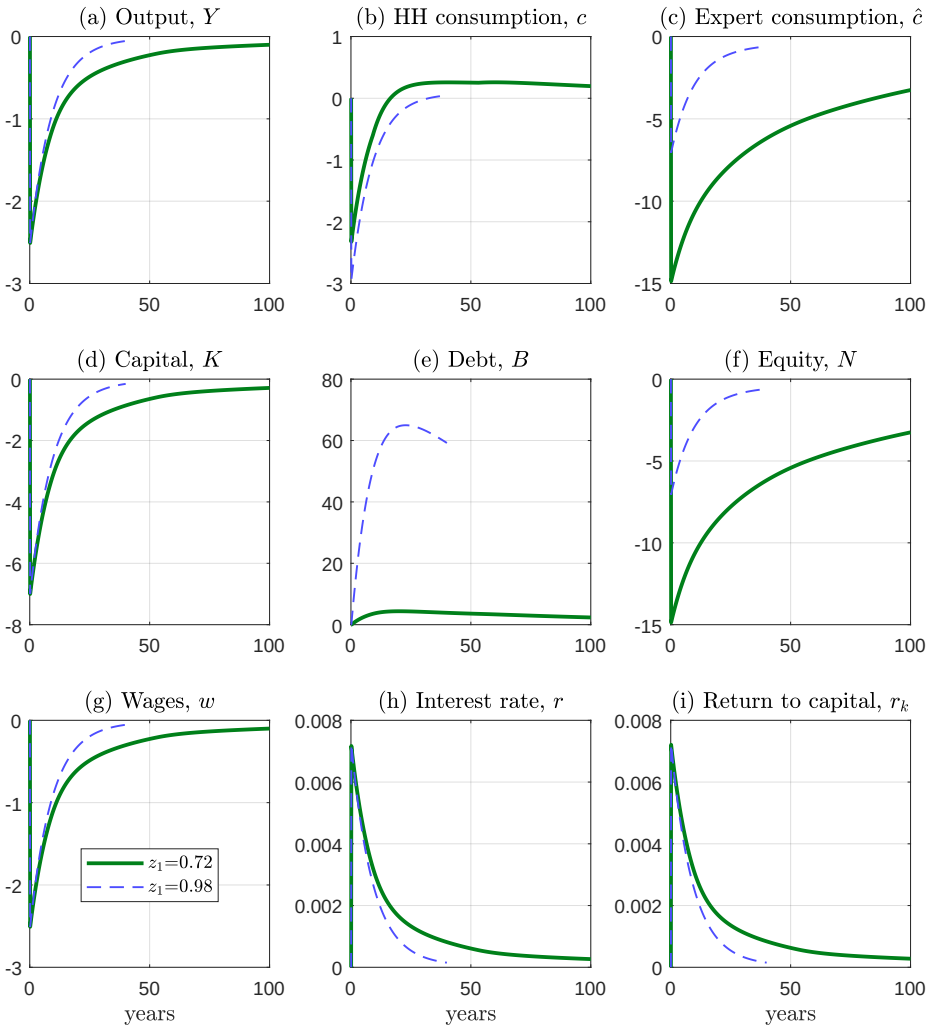


Figure 18: GIRFS, different levels of heterogeneity.

heightened fragility of the economy to negative capital shocks.[14]

7.3 Mechanism for heterogeneity: Differences across households

The key to understanding the differences in GIRFs from Figure 18 is to explore how the reduction in households' total consumption (see the central panel in the top row of Figure 18) is distributed among households.

After a negative aggregate shock, wages decrease and the risk-free rate increases. Poorer households, mainly dependent on labor earnings, loose much. In comparison, richer households, which depend more on income from their assets, lose less or, if they are sufficiently wealthy, can

[14]Note that, in the data, the economy is buffeted by plenty of aggregate shocks. It is plausible to argue that some of these shocks, such as monetary policy shocks, have become much less volatile during the last few decades, thus lowering the overall volatility of the economy despite the increase in microeconomic turbulence. Our statement in the main text is *with respect to* capital (or equivalent) shocks.

even gain. Even more, these rich households can take advantage of a higher risk-free rate to accumulate wealth fast. They reduce their consumption not as much as a consequence of lower income (wealth effect), but as a consequence of better rewards for savings (substitution effect). Moreover, the sluggish aggregate dynamics of the model means that wealthy households expect a higher risk-free rate to last for a long time. The integral of these consumption responses across households is different, because of the marginal decreasing utility of consumption than the response of a household with average wealth. Hence, the responses of total consumption and the evolution of total debt are also different.

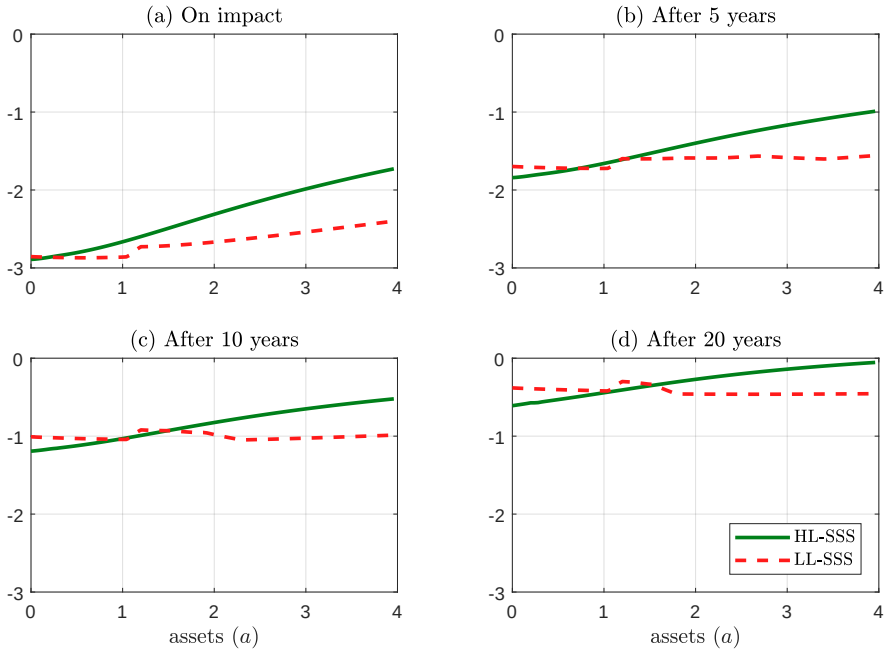


Figure 19: Difference in consumption decision rules at different points in time after the shock.

Figure 19 documents the heterogeneity of consumption effects. The figure plots the consumption decision rules for high-productivity households ($z = z_2$) along the asset axis (results for $z = z_1$ are qualitatively similar) when a one-standard-deviation negative capital shock that hits the economy at the HL-SSS (continuous green line). To facilitate interpretation, we plot the *difference* in the consumption decision rules with respect to the case without the shock, instead of the level. Also, we plot the difference in the consumption decision rules at different points in time: at impact (left panel at the top row); after 5 years (right panel at the top row); after 10 years (left panel at the bottom row); and after 20 years (right panel at the bottom row).

At impact, all households reduce their consumption, but poorer households do so by a larger amount, reflecting their lower income. For instance, richer households (level of assets of 4) reduce their consumption around one-third less than poor households (level of assets of 0). The difference in consumption reduction survives over time.

The asymmetry in the consumption response is much smaller when we have a one-standard-deviation negative capital shock at the LL-SSS. Since in this case the risk-free interest rate is less persistent, even relatively rich households reduce their consumption as to substitute intertemporally and accumulate bonds. Consequently, the expert can issue additional debt and the economy recovers fast.

The asymmetric consumption responses have a direct impact on how the wealth distribution evolves. To illustrate this point, Figure (20) draws what we call the *distributional impulse response functions* or DIRFs. A DIRF is the natural analog of a GIRF except that, instead of plotting the evolution of an aggregate variable such as output or wages, we plot the evolution of the wealth density $g_t(\cdot)$. More concretely, Figure (20) plots the difference between the density before and after the shock. Time, in years, is plotted in the z-axis, assets in the x-axis, and the DIFRs in the y-axis. A positive value of the DIRF at a given asset level and point of time should be read as the density is higher at that asset level and point of time than it would have been in the absence of a shock. A negative value has the opposite interpretation.

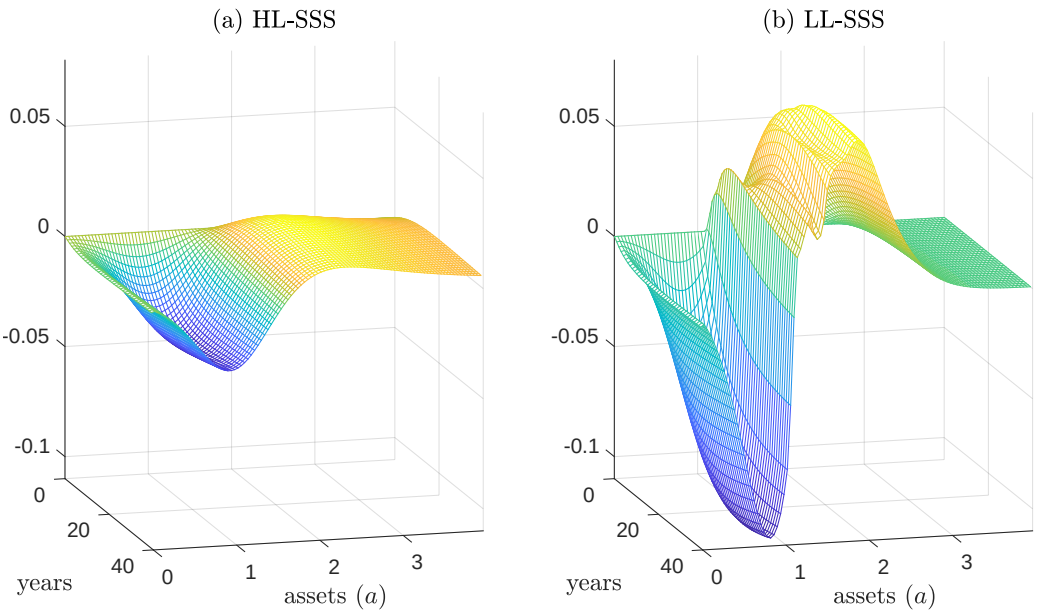


Figure 20: DIRFs at the HL-SSS.

In the left panel of Figure (20), we plot the DIRF to a one-standard-deviation negative capital shock when the economy is at the HL-SSS. In the right panel, we plot the DIRF to a one-standard-deviation negative capital shock when the economy is at the LL-SSS.

In the left panel, we can see how households with low assets must draw from their wealth to smooth consumption (even if consumption still drops) to compensate for lower income. This mechanism makes the DIRF negative in that region. In comparison, households with higher assets are reducing their consumption to respond to a temporarily higher risk-free rate and

accumulate wealth. Thus, the DIRF is positive in the region of high assets. These effects are more pronounced in the right panel. In the LL-SSS, poor households have too little debt to smooth consumption, and wealthy households accumulate much additional debt as the risk-free interest rate changes.

7.4 Taking stock

The previous three subsections summarize why household heterogeneity matters in this economy. The "quasi-aggregation result" of Krusell and Smith (1998) breaks down in our environment because, as we change idiosyncratic labor risk, precautionary savings and the leverage levels vary and, with them, the responses of aggregate variables, even creating or destroying SSS(s). The mechanism is through the heterogeneity in the consumption response of poor vs. wealthy households and the disparate impact of the wealth and substitution effect when wages fall, and the risk-free interest rate increases. Depending on how many of those wealthy vs. poor households we have (for instance, because of different degrees of idiosyncratic labor risk), and thus depending on the mean levels of bond holdings, the economy will behave differently.

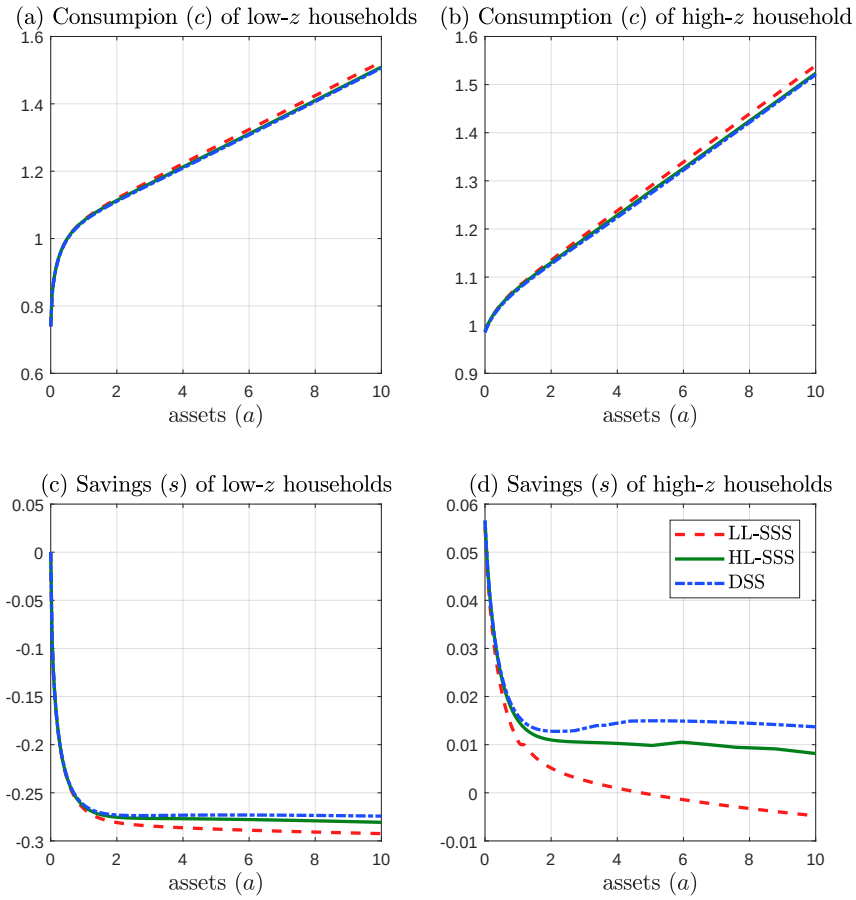


Figure 21: Consumption and saving functions.

However, a final point deserves discussion. As in Krusell and Smith (1998), our solution only considers the first set of moments of the wealth distribution to approximate the PLM. This approximation holds well in our model due to the linearity of the consumption decision rule of the households (see Figure 21). Only agents close to the borrowing limit face a nonlinear consumption decision rule but, being close to zero assets, they contribute relatively little to the aggregate dynamics of capital (and, as shown by Figure 14, there are not that many of them in any case). However, and this point is vital to understand our results, the quasi-linearity of the consumption decision rule is with respect to the household state variables, not with respect to the aggregate state variables. This is why we can still have highly nonlinear behavior for the economy as a whole while we can approximate well the PLM as a nonlinear function of only first moments.

8 Conclusion

In this paper, we have postulated, computed, and estimated a continuous-time model of financial frictions *à la* Basak and Cuoco (1998) and Brunnermeier and Sannikov (2014), but with a non-trivial distribution of wealth among households. Such exercise has allowed us to uncover critical non-linear features of the model, such as the presence of multiple SSS(s) and the strong state-dependence of the GIRFs and to document the importance of household heterogeneity in the presence of financial frictions.

While we stand by the importance *per se* of our quantitative results, this paper can also be understood as a "proof of concept" of how to efficiently compute and estimate a continuous-time model with heterogeneous agents. For the computation, we have exploited tools borrowed from machine learning. For the estimation, we have built on contributions from inference with diffusions. The importance of heterogeneity for many questions in macroeconomics suggests that there are multiple potential applications for the methodological tools that we presented and we hope to show, in the close future, how to extend our current results to richer environments.

References

Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., and Moll, B. (2017). Income and wealth distribution in macroeconomics: A continuous-time approach. Working Paper 23732, National Bureau of Economic Research.

Adrian, T. and Boyarchenko, N. (2012). Intermediary Leverage Cycles and Financial Stability. Staff Reports 567, Federal Reserve Bank of New York.

Adrian, T. and Shin, H. S. (2010). Liquidity and leverage. *Journal of Financial Intermediation*, 19(3):418–437.

Ahn, S., Kaplan, G., Moll, B., Winberry, T., and Wolf, C. (2017). When inequality matters for macro and macro matters for inequality. In *NBER Macroeconomics Annual 2017, Vol. 32*. University of Chicago Press.

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70(1):223–262.

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

Algan, Y., Allais, O., and Den Haan, W. J. (2008). Solving Heterogeneous-Agent Models with Parameterized Cross-Sectional Distributions. *Journal of Economic Dynamics and Control*, 32(3):875–908.

Algan, Y., Allais, O., Den Haan, W. J., and Rendahl, P. (2014). Solving and simulating models with heterogeneous agents and aggregate uncertainty. In Schmedders, K. and Judd, K. L., editors, *Handbook of Computational Economics*, volume 3, pages 277–324. Elsevier.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., and Zucman, G. (2017). Global inequality dynamics: New findings from wid.world. Working Paper 23119, National Bureau of Economic Research.

Andersen, T. G. and Lund, J. (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics*, 77(2):343–377.

Auclert, A. (2016). Monetary policy and the redistribution channel. Stanford University.

Auclert, A. and Rognlie, M. (2018). Inequality and aggregate demand. Stanford University.

Bach, F. (2017). Breaking the curse of dimensionality with convex neural networks. *Journal Machine Learning Research*, 18(1):629–681.

Barron, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, 39(3):930–945.

Barucci, E. and Landi, L. (1995). Non-parametric versus linear learning devices: a procedural perspective. Technical report, University of Florence.

Basak, S. and Cuoco, D. (1998). An equilibrium model with restricted stock market participation. *Review of Financial Studies*, 11(2):309–341.

Bergin, J. and Bernhardt, D. (1992). Anonymous sequential games with aggregate uncertainty. *Journal of Mathematical Economics*, 21(6):543–562.

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

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer-Verlag.

Blanchard, O. and Galí, J. (2010). Labor markets and monetary policy: A new keynesian model with unemployment. *American Economic Journal: Macroeconomics*, 2(2):1–30.

Brumm, J. and Scheidegger, S. (2017). Using adaptive sparse grids to solve high-dimensional dynamic models. *Econometrica*, 85(5).

Brunnermeier, M. K. and Sannikov, Y. (2014). A macroeconomic model with a financial sector. *American Economic Review*, 104(2):379–421.

Candler, G. V. (1999). Finite difference methods for continuous time dynamic programming. In Marimón, R. and Scott, A., editors, *Computational Methods for the Study of Dynamic Economies*, pages 172–194. Oxford University Press.

Chacko, G. and Viceira, L. M. (2003). Spectral gmm estimation of continuous-time processes. *Journal of Econometrics*, 116(1):259–292.

Cho, I.-K. (1995). Perceptrons Play the Repeated Prisoner's Dilemma. *Journal of Economic Theory*, 67(1):266–284.

Cho, I.-K. and Sargent, T. J. (1996). Neural networks for encoding and adapting in dynamic economies. In Rust, H. A. D. K. J., editor, *Handbook of Computational Economics*, volume 1, pages 441–470. Elsevier.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314.

Den Haan, W. J. (1996). Heterogeneity, Aggregate Uncertainty, and the Short-Term Interest Rate. *Journal of Business & Economic Statistics*, 14(4):399–411.

Den Haan, W. J. (1997). Solving Dynamic Models with Aggregate Shocks and Heterogeneous Agents. *Macroeconomic Dynamics*, 1(02):355–386.

Den Haan, W. J. and Rendahl, P. (2010). Solving the Incomplete Markets Model with Aggregate Uncertainty Using Explicit Aggregation. *Journal of Economic Dynamics and Control*, 34(1):69–78.

Duarte, V. (2018). Sectoral reallocation and endogenous risk-aversion: Solving macro-finance models with machine learning. MIT Sloan School of Management.

Duffie, D. and Epstein, L. (1992). Stochastic differential utility. *Econometrica*, 60(2):353–94.

Evans, G. W. and Honkapohja, S. (2001). *Learning and Expectations in Macroeconomics*. Princeton University Press.

Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2007). Estimating macroeconomic models: A likelihood approach. *Review of Economic Studies*, 74(4):1059–1087.

Galí, J. and Gambetti, L. (2009). On the sources of the great moderation. *American Economic Journal: Macroeconomics*, 1(1):26–57.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

Gornemann, N., Kuester, K., and Nakajima, M. (2012). Monetary policy with heterogeneous agents. Working Papers 12-21, Federal Reserve Bank of Philadelphia.

Hall, R. E. and Milgrom, P. R. (2008). The limited influence of unemployment on the wage bargain. *American Economic Review*, 98(4):1653–74.

He, Z. and Krishnamurthy, A. (2012). A model of capital and crises. *Review of Economic Studies*, 79(2):735–777.

He, Z. and Krishnamurthy, A. (2013). Intermediary asset pricing. *American Economic Review*, 103(2):732–70.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.

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

Kaplan, G., Moll, B., and Violante, G. L. (2018). Monetary policy according to hank. *American Economic Review*, 108(3):697–743.

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

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

Kubler, F. and Scheidegger, S. (2018). Self-justified equilibria: Existence and computation. Working paper, DBF, University of Zurich.

Kumhof, M., Ranciére, R., and Winant, P. (2015). Inequality, leverage, and crises. *American Economic Review*, 105(3):1217–45.

Liu, P., Theodoridis, K., Mumtaz, H., and Zanetti, F. (2018). Changing macroeconomic dynamics at the zero lower bound. *Journal of Business & Economic Statistics*, 0(0):1–14.

Lo, A. (1988). Maximum likelihood estimation of generalized itô processes with discretely sampled data. *Econometric Theory*, 4(2):231–247.

Luetticke, R. (2015). Transmission of monetary policy and heterogeneity in household portfolios. UCL.

Maliar, L., Maliar, S., and Valli, F. (2010). Solving the incomplete markets model with aggregate uncertainty using the krusell-smith algorithm. *Journal of Economic Dynamics and Control*, 34(1):42–49.

McKay, A., Nakamura, E., and Steinsson, J. (2016). The power of forward guidance revisited. *American Economic Review*, 106(10):3133–58.

Miao, J. (2006). Competitive equilibria of economies with a continuum of consumers and aggregate shocks. *Journal of Economic Theory*, 128(1):274–298.

Nuño, G. and Thomas, C. (2017). Bank leverage cycles. *American Economic Journal: Macroeconomics*, 9(2):32–72.

Piazzesi, M. and Schneider, M. (2016). Housing and macroeconomics. Working Paper 22354, National Bureau of Economic Research.

Pröhl, E. (2015). Approximating equilibria with ex-post heterogeneity and aggregate risk. Research Paper 17-63, Swiss Finance Institute.

Reiter, M. (2009). Solving Heterogeneous-Agent Models by Projection and Perturbation. *Journal of Economic Dynamics and Control*, 33(3):649–665.

Reiter, M. (2010). Solving the Incomplete Markets Model with Aggregate Uncertainty by Backward Induction. *Journal of Economic Dynamics and Control*, 34(1):28–35.

Ríos-Rull, J.-V. (1997). Computation of equilibria in heterogeneous agent models. Staff Report 231, Federal Reserve Bank of Minneapolis.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323:533–536.

Sager, E. (2014). Solving the incomplete markets model with aggregate uncertainty: The method of mixtures. Bureau of Labor Statistics, Price Research Division.

Salmon, M. (1995). Bounded rationality and learning: Procedural learning. In Kirman, A. P. and Salmon, M., editors, *Learning and Rationality in Economics*, pages 236–275. Blackwell Publishing.

Scheidegger, S. and Bilionis, I. (2017). Machine learning for high-dimensional dynamic stochastic economies. UCL.

Young, E. R. (2010). Solving the Incomplete Markets Model with Aggregate Uncertainty Using the Krusell-Smith Algorithm and Non-Stochastic Simulations. *Journal of Economic Dynamics and Control*, 34(1):36–41.

Online appendix

A Numerical algorithm

We describe the numerical algorithm used to jointly solve for the equilibrium value function, $v(a, z, B, N)$, the density $g(a, z, B, N)$ and the aggregate debt B and equity N. The algorithm proceeds in 3 steps. We describe each step in turn.

Step 1: Solution to the Hamilton-Jacobi-Bellman equation

The HJB equation is solved using an *upwind finite difference* scheme similar to Candler (1999) and Achdou et al. (2017). It approximates the value function $v_i(a, B, N)$, $i = 1, 2$ on a finite grid with steps Δa, ΔB, ΔN : $a \in \{a_1, ..., a_J\}$, $B \in \{B_1, ..., B_L\}$, $N \in \{N_1, ..., N_M\}$, where

$$\begin{aligned} a_j &= a_{j-1} + \Delta a = a_1 + (j-1)\Delta a, \ 2 \leq j \leq J, \\ B_l &= B_{l-1} + \Delta B = B_1 + (l-1)\Delta L, \ 2 \leq l \leq L, \\ N_m &= N_{m-1} + \Delta N = N_1 + (m-1)\Delta N, \ 2 \leq m \leq M. \end{aligned}$$

The lower bound in the wealth space is $a_1 = 0$, such that $\Delta a = a_J/(J-1)$. We use the notation $v_{i,j,l,m} \equiv v_i(a_j, B_l, N_m)$, and similarly for the policy function $c_{i,j,l,m}$. The derivatives are evaluated according to

$$\begin{aligned} \frac{\partial_i v(a_j, B_l, N_m)}{\partial a} &\approx \partial_f v_{i,j,l,m} \equiv \frac{v_{i,j+1,l,m} - v_{i,j,l,m}}{\Delta a}, \\ \frac{\partial_i v(a_j, B_l, N_m)}{\partial a} &\approx \partial_b v_{i,j,l,m} \equiv \frac{v_{i,j,l,m} - v_{i,j-1,l,m}}{\Delta a}. \\ \frac{\partial_i v(a_j, B_l, N_m)}{\partial B} &\approx \partial_B v_{i,j,l,m} \equiv \frac{v_{i,j,l+1,m} - v_{i,j,l,m}}{\Delta B}, \\ \frac{\partial_i v(a_j, B_l, N_m)}{\partial Z} &\approx \partial_N v_{i,j,l,m} \equiv \frac{v_{i,j,l,m+1} - v_{i,j,l,m}}{\Delta N}, \\ \frac{\partial_i^2 v(a_j, B_l, N_m)}{\partial N^2} &\approx \partial_{NN}^2 v_{i,j,l,m} \equiv \frac{v_{i,j,l,m+1} + v_{i,j,l,m-1} - 2v_{i,j,l,m}}{(\Delta N)^2}. \end{aligned}$$

Note that at each point of the grid, the first derivative with respect to a can be approximated with a forward (f) or a backward (b) approximation. In an upwind scheme, the choice of forward or backward derivative depends on the sign of the *drift function* for the state variable, given by

$$s_{i,j,l,m} \equiv w_{l,m} z_i + r_{l,m} a_j - c_{i,j,l,m}, \tag{40}$$

where

$$
\begin{aligned}
c_{i,j,l,m} &= \left[\frac{\partial v_{i,j,l,m}}{\partial a}\right]^{-1/\gamma}, && (41)\\
w_{l,m} &= (1-\alpha)\, Z\left(B_l + N_m\right)^{\alpha}, && (42)\\
r_{l,m} &= \alpha Z\left(B_l + N_m\right)^{\alpha-1} - \delta - \sigma^2 \frac{(B_l + N_m)}{N_m}. && (43)
\end{aligned}
$$

Let superscript n denote the iteration counter. The HJB equation is approximated by the following upwind scheme,

$$
\begin{aligned}
\frac{v_{i,j,l,m}^{n+1} - v_{i,j,l,m}^{n}}{\Delta} + \rho v_{i,j,l,m}^{n+1} &= \frac{(c_{i,j,l,m}^{n})^{1-\gamma}}{1-\gamma} + \partial_f v_{i,j,l,m}^{n+1} s_{i,j,l,m,f}^{n} \mathbf{1}_{s_{i,j,n,m,f}^{n}>0} + \partial_B v_{i,j,l,m}^{n+1} s_{i,j,l,m,b}^{n} \mathbf{1}_{s_{i,j,l,m,b}^{n}<0}\\
&\quad + \lambda_i \left(v_{-i,j,l,m}^{n+1} - v_{i,j,l,m}^{n+1}\right) + h_{l,m}\partial_B v_{i,j,l,m} + \mu_{l,m}^{N}\partial_N v_{i,j,l,m}\\
&\quad + \frac{\left[\sigma_{l,m}^{N}\right]^2}{2}\partial_{NN}^2 v_{i,j,l,m}
\end{aligned}
$$

for $i = 1, 2$, $j = 1, ..., J$, $l = 1, .., L$, $m = 1, ..., M$, where where $\mathbf{1}(\cdot)$ is the indicator function and

$$
\begin{aligned}
h_{l,m} &\equiv h\left(B_l, N_m\right),\\
\mu_{l,m}^{N} &\equiv \mu^{N}\left(B_l, N_m\right) = \alpha Z\left(B_l + N_m\right)^{\alpha} - \delta\left(B_l + N_m\right) - r_{l,m}B_l - \widehat{\rho} N_m,\\
\sigma_{l,m}^{N} &\equiv \sigma^{N}\left(B_l, N_m\right) = \sigma\left(B_l + N_m\right),\\
s_{i,j,l,m,f}^{n} &= w_{l,m} z_i + r_{l,m} a_j - \left[\frac{1}{\partial_f^n v_{i,j,l,m}}\right]^{1/\gamma},\\
s_{i,j,l,m,b}^{n} &= w_{l,m} z_i + r_{l,m} a_j - \left[\frac{1}{\partial_b^n v_{i,j,l,m}}\right]^{1/\gamma}
\end{aligned}
$$

Therefore, when the drift is positive ($s_{i,j,l,m,f}^{n} > 0$), we employ a forward approximation of the derivative, $\partial_f^n v_{i,j,l,m}$; when it is negative ($s_{i,j,l,m,b}^{n} < 0$), we employ a backward approximation, $\partial_b^n v_{i,j,l,m}$. The term $\frac{v_{i,j,l,m}^{n+1} - v_{i,j,l,m}^{n}}{\Delta} \to 0$ as $v_{i,j,l,m}^{n+1} \to v_{i,j,l,m}^{n}$.

Moving all terms involving v^{n+1} to the left hand side and the rest to the right hand side, we obtain:

$$
\begin{aligned}
\frac{v_{i,j,l,m}^{n+1} - v_{i,j,l,m}^{n}}{\Delta} + \rho v_{i,j,l,m}^{n+1} &= \frac{(c_{i,j,n,m}^{n})^{1-\gamma} - 1}{1-\gamma} + v_{i,j-1,l,m}^{n+1}\alpha_{i,j,l,m}^{n} + v_{i,j,l,m}^{n+1}\beta_{i,j,l,m}^{n} + v_{i,j+1,l,m}^{n+1}\xi_{i,j,l,m}^{n}\\
&\quad + \lambda_i v_{-i,j,l,m}^{n+1} + v_{i,j,l+1,m}^{n+1}\frac{h_{l,m}}{\Delta B} + v_{i,j,l,m+1}^{n+1}\varkappa_{l,m} + v_{i,j,l,m-1}^{n+1}\varrho_{l,m} \qquad (44)
\end{aligned}
$$

where

$$\begin{aligned}
\alpha^n_{i,j} &\equiv -\frac{s^n_{i,j,B}\mathbf{1}_{s^n_{i,j,B}<0}}{\Delta a},\\
\beta^n_{i,j,l,m} &\equiv -\frac{s^n_{i,j,l,m,f}\mathbf{1}_{s^n_{i,j,n,mF}>0}}{\Delta a}+\frac{s^n_{i,j,l,m,b}\mathbf{1}_{s^n_{i,j,l,m,b}<0}}{\Delta a}-\lambda_i-\frac{h_{l,m}}{\Delta B}-\frac{\mu^N_{l,m}}{\Delta N}-\frac{\left(\sigma^N_{l,m}\right)^2}{\left(\Delta N\right)^2},\\
\xi^n_{i,j} &\equiv \frac{s^n_{i,j,F}\mathbf{1}_{s^n_{i,j,F}>0}}{\Delta a},\\
\varkappa_{l,m} &\equiv \frac{\mu^N_{l,m}}{\Delta N}+\frac{\left(\sigma^N_{l,m}\right)^2}{2\left(\Delta N\right)^2}=\frac{\left[\alpha Z\left(B_l+N_m\right)^\alpha-\delta\left(B_l+N_m\right)-r_{l,m}B_l-\widehat{\rho}N_m\right]}{\Delta N}+\frac{\sigma^2\left(B_l+N_m\right)^2}{2\left(\Delta N\right)^2},\\
\varrho_{l,m} &\equiv \frac{\left(\sigma^N_{l,m}\right)^2}{2\left(\Delta N\right)^2}=\frac{\sigma^2\left(B_l+N_m\right)^2}{2\left(\Delta N\right)^2}.
\end{aligned}$$

for $i=1,2$, $j=1,...,J$, $l=1,..,L$, $m=1,...,M$. We consider boundary state constraints in a ($s^n_{i,1,B}=s^n_{i,J,F}=0$). The boundary conditions in B and N are reflections.

In equation (44), the optimal consumption is set to

$$c^n_{i,j,n,m}=\left(\partial v^n_{i,j,l,m}\right)^{-1/\gamma}. \tag{45}$$

where

$$\partial v^n_{i,j,l,m}=\partial_f v^n_{i,j,l,m}\mathbf{1}_{s^n_{i,j,n,mF}>0}+\partial_b v^n_{i,j,l,m}\mathbf{1}_{s^n_{i,j,l,m,b}<0}+\partial\bar{v}^n_{i,j,l,m}\mathbf{1}_{s^n_{i,j,n,mF}\leq 0}\mathbf{1}_{s^n_{i,j,l,m,b}\geq 0}.$$

In the above expression, $\partial\bar{v}^n_{i,j,l,m}=(\bar{c}^n_{i,j,n,m})^{-\gamma}$ where $\bar{c}^n_{i,j,n,m}$ is the consumption level such that the drift is zero :

$$\bar{c}^n_{i,j}=w_{l,m}z_i+r_{l,m}a_j.$$

We define

$$\mathbf{A}^n_{l.m} = \begin{bmatrix} \beta^n_{1,1,l,m} & \xi^n_{1,1,l,m} & 0 & 0 & \cdots & 0 & \lambda_1 & 0 & \cdots & 0 \\ \alpha^n_{1,2,l,m} & \beta^n_{1,2,l,m} & \xi^n_{1,2,l,m} & 0 & \cdots & 0 & 0 & \lambda_1 & \ddots & 0 \\ 0 & \alpha^n_{1,3,l,m} & \beta^n_{1,3,l,m} & \xi^n_{1,3,l,m} & \cdots & 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha^n_{1,J-1,l,m} & \beta^n_{1,J-1,l,m} & \xi^n_{1,J-1,l,m} & 0 & \cdots & \lambda_1 & 0 \\ 0 & 0 & \cdots & 0 & \alpha^n_{1,J,l,m} & \beta^n_{1,J,l,m} & 0 & 0 & \cdots & \lambda_1 \\ \lambda_2 & 0 & \cdots & 0 & 0 & 0 & \beta^n_{2,1,l,m} & \xi^n_{2,1,l,m} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \lambda_2 & 0 & \cdots & \alpha^n_{2,J,l,m} & \beta^n_{2,J,l,m} \end{bmatrix},$$

$$\mathbf{v}^{n+1}_{l,m} = \begin{bmatrix} \mathbf{v}^{n+1}_{1,1,l,m} \\ \mathbf{v}^{n+1}_{1,2,l,m} \\ \vdots \\ \mathbf{v}^{n+1}_{1,J,l,m} \\ \mathbf{v}^{n+1}_{2,1,l,m} \\ \vdots \\ \mathbf{v}^{n+1}_{2,J,l,m} \end{bmatrix}$$

and

$$\mathbf{A}^n_m = \begin{bmatrix} \mathbf{A}^n_{1,m} & \frac{h_{1,m}}{\Delta K}\mathbf{I}_{2J} & \mathbf{0}_{2J} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} \\ \mathbf{0}_{2J} & \mathbf{A}^n_{2,m} & \frac{h_{2,m}}{\Delta K}\mathbf{I}_{2J} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} \\ \mathbf{0}_{2J} & \mathbf{0}_{2J} & \mathbf{A}^n_{3,m} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ & & & \mathbf{0}_{2J} & \mathbf{A}^n_{L-1,m} & \frac{h_{L-1,m}}{\Delta K}\mathbf{I}_{2J} \\ \mathbf{0}_{2J} & \mathbf{0}_{2J} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} & \left(\mathbf{A}^n_{L,m} + \frac{h_{L,m}}{\Delta K}\mathbf{I}_{2J}\right) \end{bmatrix}, \quad \mathbf{v}^{n+1}_m = \begin{bmatrix} \mathbf{v}^{n+1}_{1,m} \\ \mathbf{v}^{n+1}_{2,m} \\ \vdots \\ \mathbf{v}^{n+1}_{L,m} \end{bmatrix},$$

where $\mathbf{I}_n$ and $\mathbf{0}_n$ are the identity matrix and the zero matrix of dimension $n \times n$, respectively.

We can also define

$$\mathbf{A}^n = \begin{bmatrix} (\mathbf{A}_1^n + \mathbf{P}_1) & \mathbf{X}_1 & \mathbf{0}_{2J\times L} & \cdots & \mathbf{0}_{2J\times L} & \mathbf{0}_{2J\times L} \\ \mathbf{P}_2 & \mathbf{A}_2^n & \mathbf{X}_2 & \cdots & \mathbf{0}_{2J\times L} & \mathbf{0}_{2J\times L} \\ \mathbf{0}_{2J\times L} & \mathbf{P}_3 & \mathbf{A}_3^n & \cdots & \mathbf{0}_{2J\times L} & \mathbf{0}_{2J\times L} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ & & & \mathbf{P}_{M-1} & \mathbf{A}_{M-1}^n & \mathbf{X}_{M-1} \\ \mathbf{0}_{2J\times L} & \mathbf{0}_{2J\times L} & \cdots & \mathbf{0}_{2J\times L} & \mathbf{P}_M & (\mathbf{A}_M^n + \mathbf{X}_M) \end{bmatrix}, \ \mathbf{v}^{n+1} = \begin{bmatrix} \mathbf{v}_1^{n+1} \\ \mathbf{v}_2^{n+1} \\ \vdots \\ \mathbf{v}_M^{n+1} \end{bmatrix},$$

$$\mathbf{X}_m = \begin{bmatrix} \varkappa_{1,m}\mathbf{I}_{2J} & \mathbf{0}_{2J} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} \\ \mathbf{0}_{2J} & \varkappa_{2,m}\mathbf{I}_{2J} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ & & \mathbf{0}_{2J} & \varkappa_{L-1,m}\mathbf{I}_{2J} & \mathbf{0}_{2J} \\ \mathbf{0}_{2J} & \mathbf{0}_{2J} & \mathbf{0}_{2J} & \mathbf{0}_{2J} & \varkappa_{L,m}\mathbf{I}_{2J} \end{bmatrix},$$

$$\mathbf{P}_m = \begin{bmatrix} \varrho_{1,m}\mathbf{I}_{2J} & \mathbf{0}_{2J} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} \\ \mathbf{0}_{2J} & \varrho_{2,m}\mathbf{I}_{2J} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ & & \mathbf{0}_{2J} & \varrho_{L-1,m}\mathbf{I}_{2J} & \mathbf{0}_{2J} \\ \mathbf{0}_{2J} & \mathbf{0}_{2J} & \mathbf{0}_{2J} & \mathbf{0}_{2J} & \varrho_{L,m}\mathbf{I}_{2J} \end{bmatrix}, \mathbf{u}^n = \begin{bmatrix} \frac{(c_{1,1,1,1}^n)^{1-\gamma}-1}{1-\gamma} \\ \frac{(c_{1,2,1,1}^n)^{1-\gamma}-1}{1-\gamma} \\ \vdots \\ \vdots \\ \frac{(c_{2,J,L,M}^n)^{1-\gamma}-1}{1-\gamma} \end{bmatrix}.$$

Then, equation (44) is a system of $2 \times J \times L \times M$ linear equations which can be written in matrix notation as:

$$\frac{1}{\Delta}\left(\mathbf{v}^{n+1} - \mathbf{v}^n\right) + \rho\mathbf{v}^{n+1} = \mathbf{u}^n + \mathbf{A}^n\mathbf{v}^{n+1}.$$

The system in turn can be written as

$$\mathbf{B}^n\mathbf{v}^{n+1} = \mathbf{d}^n \tag{46}$$

where , $\mathbf{B}^n = \left(\frac{1}{\Delta} + \rho\right)\mathbf{I} - \mathbf{A}^n$ and $\mathbf{d}^n = \mathbf{u}^n + \frac{1}{\Delta}\mathbf{v}^n$.

The algorithm to solve the HJB equation runs as follows. Begin with an initial guess $v_{i,j,l,m}^0$. Set $n = 0$. Then:

1. Compute $c_{i,j,l,m}^n$, $i = 1, 2$ using (45).
2. Find $v_{i,j,l,m}^{n+1}$ solving the linear system of equations (46).
3. If $v_{i,j,l,m}^{n+1}$ is close enough to $v_{i,j,l,m}^n$, stop. If not, set $n := n + 1$ and proceed to step 1.

Most programming languages, such as `Julia` or `Matlab`, include efficient routines to handle sparse matrices such as $\mathbf{A}^n$.

Step 2: Solution to the KF equation

The income-wealth distribution conditional on the current realization of aggregate debt $B = B_l$ and equity $N = N_m$ can be characterized by the KF equation:

$$\frac{\partial g}{\partial t} = -\frac{\partial}{\partial a}\left[s_i\left(a, B, N\right) g_{i,t}(a)\right] - \lambda_i g_{i,t}(a) + \lambda_{-i} g_{-i,t}(a), \ i = 1, 2. \tag{47}$$

$$1 = \int_0^\infty g(a) da. \tag{48}$$

If we define a time step Δt we also solve this equation using an finite difference scheme. We use the notation $g_{i,j} \equiv g_i(a_j)$. The system can be now expressed as

$$\begin{aligned}\frac{g_{i,j,t+1} - g_{i,j}}{\Delta t} = & -\frac{g_{i,j,t} s_{i,j,l,m,b} \mathbf{1}_{s_{i,j,l,m,b}>0} - g_{i,j-1,t} s_{i,j-1,l,m,f} \mathbf{1}_{s_{i,j-1,l,m,f}>0}}{\Delta a} \\ & -\frac{g_{i,j+1,t} s_{i,j+1,l,m,b} \mathbf{1}_{s_{i,j+1,l,m,b}<0} - g_{i,j,t} s_{i,j,l,m,b} \mathbf{1}_{s_{i,j,l,m,b}<0}}{\Delta a} - \lambda_i g_{i,j,t} + \lambda_{-i} g_{-i,j,t},\end{aligned}$$

In this case, let us define

$$\mathbf{g}_t = \begin{bmatrix} g_{1,1,t} \\ g_{1,2,t} \\ \vdots \\ g_{1,J,t} \\ g_{2,1,t} \\ \vdots \\ g_{2,J,t} \end{bmatrix},$$

as the density conditional on the current state of B_l and N_m.We assume that g_0 is the density in the deterministic steady state (which coincides with the standard Aiyagari (1994) economy), the update in the next time period is given by the KF equation:

$$\mathbf{g}_{t+1} = \left(\mathbf{I} - \Delta t \mathbf{A}_{l,m}^{\mathbf{T}}\right)^{-1} \mathbf{g}_t,$$

where $\mathbf{A}_{l,m}^{\mathbf{T}}$ is the transpose matrix of $\mathbf{A}_{l,m} = \lim_{n\to\infty} \mathbf{A}_{l,m}^n$, defined above.

Complete algorithm

We can now summarize the complete algorithm. We begin a guess of the PLM $h^0\left(B, N\right)$. Set $s := 1$:

Step 1: Household problem. Given $h^{s-1}\left(B, N\right)$, solve the HJB equation to obtain an estimate of the value function $\mathbf{v}$ and of the matrix $\mathbf{A}$.

Step 2: Distribution. Given **A**, simulate T periods of the economy using the KF equation and obtain the aggregated ebt $\{B_t\}_{t=0}^{T}$ and equity $\{N_t\}_{t=0}^{T}$. The law of motion of equity is

$$N_t = N_{t-1} + \left[\alpha Z \left(B_t + N_t\right)^{\alpha} - \delta \left(B_t + N_t\right) - r_t B_t - \widehat{\rho} N_t\right] \Delta t + \sigma \left(B_t + N_t\right) \sqrt{\Delta t} \varepsilon_t,$$

where $\varepsilon_t \overset{iid}{\sim} N(0,1)$.

Step 3: PLM. Update the PLM using a neural netwok: h^s. If $\|h^s - h^{s-1}\| < \varepsilon$, where ε is a small positive constant, then stop. if not return to Step 1.

B Training the neural network

We tweak and tune the training scheme of our neural network in several ways to meet the specific demands of the problem at hand.

First, it needs to yield a consistently good approximation: a "not-good-enough" approximation in any of the dozens of iterations of the algorithm can make it break and not deliver a result. This is the reason we employ a line-search instead of using a constant or adaptive learning rate: it prevents bad steps in the minimization algorithm.

Second, it cannot introduce big amounts of noise in the progressive approximation to the solution of the model. Otherwise that noise can mask or prevent convergence (strict convergence criteria can only be met by chance, if at all). This is the reason we use batch gradient descent instead of stochastic or mini-batch gradient descent (i.e. all training points are used in every gradient calculation and line-search step): the random choice of training points in each step, common in the machine learning literature, would make progress stochastic and introduces noise.

Reducing stochastic elements in the solution method is also the reason we train the model not with the noisy simulated points, but with a grid approximation to that that clears out noise: we define a 101x101 grid over the (B, N) support, and assign each simulated point to one of the knots in that grid; with all the points assigned to each knot we run a linear regression, and use that to estimate the height of the PLM at that knot. This grid could later be used to solve the model using interpolation (e.g. with splines, or with natural neighbor interpolation) and linear extrapolation (and we do that as a robustness check, finding similar results to those of the solution with the neural network), but on a 2D surface such as our PLM that extrapolation tends to generate ridges (because of the amplification of sample noise in far extrapolations), which could prevent convergence at the HJB step. Instead, we use those knots to train the neural network, which provides a good-enough fit in the visited area and a much smoother extraplation to the non-visited area of the (B, N) support.

Finally, the need to avoid stochastic elements that introduce noise in the algorithm, plus a desire for fast running times, made us change the usual Monte Carlo multi-start initialization

of the neural network parameters for all iterations except the first one. In the first iteration of the algorithm, we do ten random initializations of the parameters of the neural network and ten subsequent training sessions, later choosing the best-performing trained network across those ten training sessions. From the second iteration onwards, the neural network is initialized using weights that were found to be optimal in the previous iteration, and a single training session is carried out. This avoids re-introducing a stochastic element on each step of the algorithm that would greatly reduce its ability to meet a strict convergence criterion (or, more accurately, it would require much longer training sessions in order to clear out that noise and guarantee convergence). Using a small regulaization parameter in the PLM update step (it starts at 0.30 and exponentially decays towards 0.05, reaching 0.20 after 5 iterations and 0.10 after 16) makes the convergence of the full algorithm slower but smoother, with slow updates to the optimal neural network that help this non-random initialization work well.

The vector θ is recursively updated according to $\theta_{m+1} = \theta_m - \epsilon_m \nabla \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)$, where:

$$\nabla \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right) \equiv \left[\frac{\partial \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)}{\partial \theta_0^2}, \frac{\partial \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)}{\partial \theta_1^2}, ..., \frac{\partial \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)}{\partial \theta_{2,Q}^1}\right]^\top$$

is the gradient of the error function with respect to θ evaluated at $\left(\mathbf{s_j}, \widehat{h}_j\right)$.

To improve performance, it is typical to group several points of the simulation in each evaluation step –instead of just one point– in what is known as a *minibatch* algorithm. This takes advantage of the fast convergence of the gradient (think about it as a score of the error function) towards its value with an infinite sample, but at a much lower computational cost. The step size $\epsilon_m > 0$ is selected in each iteration according to a *line-search* algorithm in order to minimize the error function in the direction of the gradient. The algorithm is run until the distance between θ_{m+1} and θ_m is below a threshold ε:

$$\|\theta_{m+1} - \theta_m\| < \varepsilon.$$

An advantage of neural networks is that the error gradient can be efficiently evaluated using a *back-propagation* algorithm, originally developed by Rumelhart et al. (1986), which builds on the chain rule of differential calculus. In our particular case, the corresponding formulae are:

$$
\begin{aligned}
\frac{\partial \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)}{\partial \theta_0^2} &= h\left(\mathbf{s}_j; \theta\right) - \widehat{h}_j \\
\frac{\partial \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)}{\partial \theta_q^2} &= \left(h\left(\mathbf{s}_j; \theta\right) - \widehat{h}_j\right) \phi\left(\theta_{0,q}^1 + \sum_{i=1}^{2} \theta_{i,q}^1 s_j^i\right), \text{ for } q = 1, ..., Q \\
\frac{\partial \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)}{\partial \theta_{0,q}^1} &= \theta_q^2 \left(h\left(\mathbf{s}_j; \theta\right) - \widehat{h}_j\right) \phi'\left(\theta_{0,q}^1 + \sum_{i=1}^{2} \theta_{i,q}^1 s_j^i\right), \text{ for } q = 1, ..., Q \\
\frac{\partial \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)}{\partial \theta_{i,q}^1} &= s_j^i \theta_q^2 \left(h\left(\mathbf{s}_j; \theta\right) - \widehat{h}_j\right) \phi'\left(\theta_{0,q}^1 + \sum_{i=1}^{2} \theta_{i,q}^1 s_j^i\right), \text{ for } i = 1, 2 \text{ and } q = 1, ..., Q,
\end{aligned}
$$

where $\phi'(x) = \frac{1}{(1+e^{-x})}$.

C Evaluating the likelihood function with micro observations

A promising avenue to improve the estimation is to add micro observations, which bring much additional information and help integrating different levels of aggregation to assess the empirical validity of the model. More concretely, let $X_t \equiv [g_t(a,z); N_t]'$, be a vector of observations on the asset holdings of agents in this economy (households, $g_t(a,z)$, and the expert, N_t). Imagine, as before, that we have $D+1$ observations of X_t at fixed time intervals $[0, \Delta, 2\Delta, .., D\Delta,]$:

$$
X_0^D = \{X_0, X_\Delta, X_{2\Delta}, ..., X_D\}.
$$

At this moment, we need to assume –as it is typically done in models with heterogeneous agents and aggregates shocks– that the *conditional no aggregate uncertainty* (CNAU) condition holds. See, for instance, Miao (2006), following Bergin and Bernhardt (1992). This condition implies that if households are distributed on the interval $I = [0,1]$ according to the Lebesgue measure Φ, then

$$
G_t(A \times Z) = \Phi\left(i \in I : \left(a_t^i, z_t^i\right) \in A \times Z\right),
$$

for any subsets $A \subset [0, \infty)$, $Z \subset \{z_1, z_2\}$. That is, the probability under the conditional distribution is the same as the probability according to the Lebesgue measure across the interval I.

The likekihood that an individual agent $i \in I$ at time $t = d\Delta$ is at state $(a_{d\Delta}^i, z_{d\Delta}^i, B_{d\Delta}, N_{d\Delta})$ is $f_{d\Delta}^d(a_{d\Delta}^i, z_{d\Delta}^i, B_{d\Delta}, N_{d\Delta})$. The log-likelihood is then $\log\left[f_{d\Delta}^d(a_{d\Delta}^i, z_{d\Delta}^i, B_{d\Delta}, N_{d\Delta})\right]$. Notice that

this log-likelihood is a function of i.

The conditional aggregate log-likelihood across all agents is

$$\log p_X\left(X_{d\Delta}|X_{(d-1)\Delta};\Psi\right) = \int \log\left[f^d_{d\Delta}(a^i_{d\Delta}, z^i_{d\Delta}, B_{d\Delta}, N_{d\Delta})\right]\Phi(di),$$

and, taking into account the CNAU condition, we get:

$$\begin{aligned}\int \log\left[f^d_{d\Delta}(a^i_{d\Delta}, z^i_{d\Delta}, B_{d\Delta}, N_{d\Delta})\right]\Phi(di) &= \int \log\left[f^d_{d\Delta}(a, z, B_{d\Delta}, N_{d\Delta})\right] G_{d\Delta}(da, dz) \\ &= \sum_{i=1}^{2}\int_0^\infty \log\left[f^d_{d\Delta}(a, z_i, B_{d\Delta}, N_{d\Delta})\right] g_{d\Delta}(a, z)da,\end{aligned}$$

where, in the second line, we have applied the definition of the Radon-Nikodym derivative to get the differential in a.

The density $f^d_t(a, z, B, N)$ follows the KF equation:

$$\begin{aligned}\frac{\partial f^d_t}{\partial t} = &-\frac{\partial}{\partial a}\left(s_t(a, z_i) f^d_t(a, z_i, B, N)\right) - \lambda_i f^d_t(a, z_i, B, N) + \lambda_j f^d_t(a, z_j, B, N) \\ &-\frac{\partial}{\partial B}\left[h(B, N) f^d_t(a, z_i, B, N))\right] - \frac{\partial}{\partial N}\left[\mu^N_t(B, N) f^d_t(a, z_i, B, N)\right] \\ &+\frac{1}{2}\frac{\partial^2}{\partial N^2}\left[\left(\sigma^N_t(B, N)\right)^2 f^d_t(B, N)\right], \quad (i \neq j = 1, 2)\end{aligned} \tag{49}$$

where

$$f^d_{(d-1)\Delta} = g_{(d-1)\Delta}(a, z)\delta\left(B - B_{(d-1)\Delta}\right)\delta\left(N - N_{(d-1)\Delta}\right),$$

which, following the same reasoning than the one in the previous subsection, is easy to evaluate.

More concretely, we use the notation $f^d_{i,j,l,m} \equiv f^d_i(a_j, B_l, N_m)$ and define a time step $\Delta t = \frac{\Delta}{S}$, where $1 << S \in \mathbb{N}$ is a constant. If we solve the KF equation (49) using a finite difference scheme, we have, for $t = (d-1)\Delta$ and $s = 1, .., S-1$, where

$$\begin{aligned}\mathbf{f}^d_{t+s\Delta t} &= \left(\mathbf{I} - \Delta t\mathbf{A}^{\mathbf{T}}\right)^{-1}\mathbf{f}^d_{t+(s-1)\Delta t}, \\ \mathbf{f}^d_t &= \mathbf{g}_t\delta_{N_{(d-1)\Delta}}\delta_{B_{(d-1)\Delta}},\end{aligned}$$

where δ is the Kronecker delta and $\mathbf{f}^d_t$ is defined as

$$\mathbf{f}^d_t = \begin{bmatrix} f_{1,1,1,t} \\ g_{1,1,1,2,t} \\ \vdots \\ g_{2,J,L,M,t} \end{bmatrix}.$$

The conditional density $p_X\left(X_{d\Delta}|X_{(d-1)\Delta};\gamma\right)$ can be approximated by

$$p_X\left(X_{d\Delta}|X_{(d-1)\Delta};\gamma\right) = \sum_{i=1}^{2}\sum_{j=1}^{J}\sum_{l=1}^{L} f^d_{i,j,l^d,m} g^d_{i,j} \Delta a \Delta B,$$

where $f^d_{i,j,l^d,m}$ is the density evaluated at the observed equity point $N_{d\Delta}$, $f^d_i(a_j, B_l, N = N_{d\Delta})$ and $g^d_{i,j}$ are the elements of the observed distribution $\mathbf{g}_{d\Delta}$.

D Chebyshev polynomials

Here we show that the approximation computed with Chebyshev polynomials is not satisfactory.

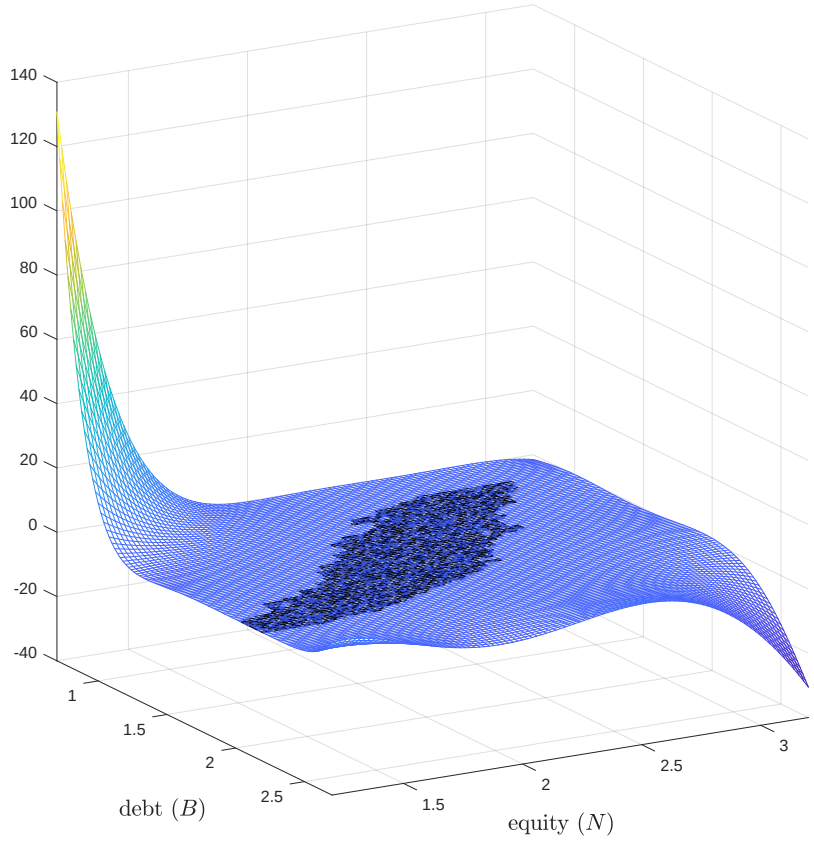


Figure D.1: PLM with Chebyshev polynomials.

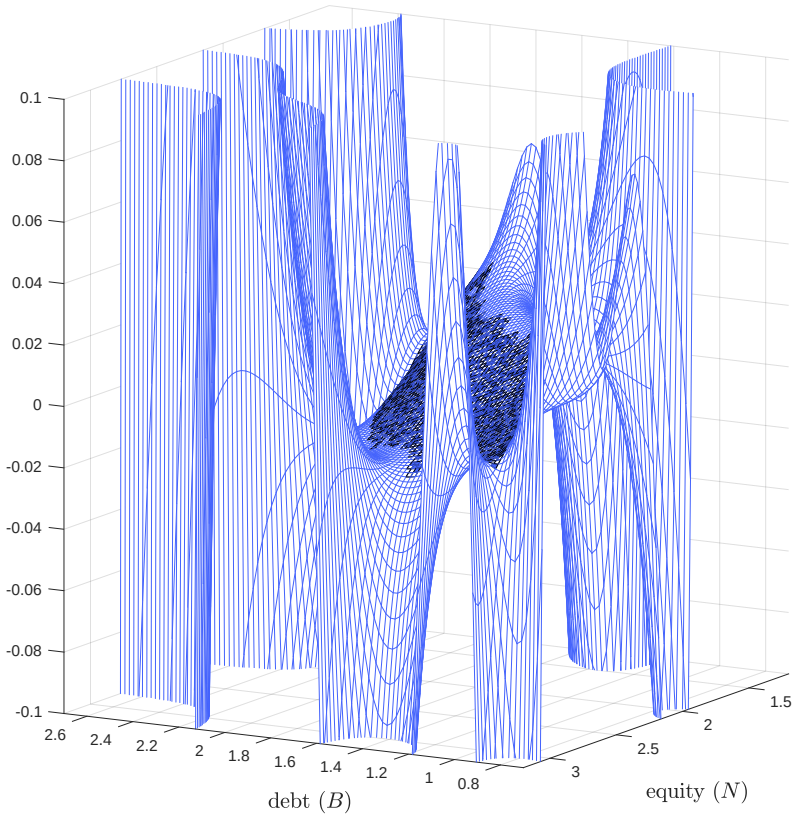


Figure D.2: PLM with Chebyshev polynomials (zoom).