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*

November 15, 2018

Abstract

We study a continuous-time, heterogeneous agent economy with aggregate shocks and financial frictions. Households cannot invest directly in risky capital. Instead, they lend to a leveraged expert through risk-free bonds. To solve the model, we extend the Krusell-Smith approach to accommodate a nonlinear law of motion for the aggregate endogenous variables by employing machine learning techniques. We take the model to the data by building the likelihood function associated with the solution described above. Regarding results, we document, first, the strong nonlinearities created by financial frictions and why our solution method is required. Second, we show how the economy displays more leverage and higher wealth inequality in the stochastic steady-state than in the deterministic one. Third, we report how the impulse-response functions are highly state dependent. In particular, we find that the recovery after a negative capital shock hits the economy is more sluggish if experts are more leveraged.

Keywords: Heterogeneous agents; Aggregate shocks; Continuous-time; Machine learning; Neural networks.

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

*We thank Gianluca Violante for 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

The decades before the Great Recession witnessed both an increase in credit and leverage and wealth inequality. See Adrian and Shin (2010), and Nuño and Thomas (2017), for evidence on the increase in debt and leverage and Alvaredo, Chancel, Piketty, Saez, and Zucman (2017), for evidence on the increase in wealth inequality. A natural question is to ask to what extent these two trends were related and which role did they play in bringing about the financial crisis of 2007-8 and the subsequent sluggish output recovery. In this paper, we take a step in the direction of addressing these questions by analyzing the links between financial frictions and the wealth distribution in a stochastic general equilibrium model with heterogeneous agents.

We extend the neoclassical model with heterogeneous households and aggregate capital à la Aiyagari (1994) along several dimensions. First, we introduce aggregate shocks, making capital a risky asset. Second, we consider limited financial markets participation as in Basak and Cuoco (1998) or Brunnermeier and Sannikov (2016, Section 2). There are three agents in the economy: a representative firm, a representative expert, and heterogeneous households subject to idiosyncratic shocks. Only the expert can hold physical capital, which is rented to the firm. The expert cannot issue state-contingent assets (i.e., outside equity) and, thus, she can only finance her capital holdings through debt sold to households plus her wealth (i.e., inside equity). Third, we analyze the model in continuous-time.

A major technical challenge is how to efficiently compute a solution to this model and take it to the data. The two state variables of the model are the expert's equity and households' income-wealth distribution. As the latter is an infinite-dimensional object, standard dynamic programming techniques cannot be employed. Several numerical techniques have been developed in the last two decades to analyze this kind of problems, being the most popular the approach introduced by Krusell and Smith (1998). These authors approximate the cross-sectional distribution with a finite set of moments. To compute the law of motion of these moments, they propose an iterative scheme that starts with an initial guess to compute individual policies and then simulate the economy based on these policies to obtain a time series for the moments. The law of motion is obtained from a regression problem of the simulated time series.

In the original paper by Krusell and Smith (1998) and most of the subsequent literature, the perceived law of motion (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). Hence a standard least-squares method can be employed to estimate the coefficients based on the simulated data conditional on the realization of the exogenous states. This is not the case 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. To overcome this problem, we rely on the machine learning literature and employ a neural network to obtain a flexible parametric function of the PLM. The *universal approximation theorem* (Hornik, Stinchcombe, and White 1989; Cybenko 1989) states that a neural network can approximate any Borel measurable function. More importantly, perhaps, the neural network breaks the curse of dimensionality for a large class of approximated functions. The parameters of the neural networks can be efficiently estimated using a combination of the *stochastic gradient descent* and the *back-propagation* algorithms. Not only is this an efficient and easy to code algorithm, but also one that is particularly amenable to massive parallelization in graphic or tensor processing units.

We take our model to match some features of the US economy, such as the leverage of the corporate sector and solve the model numerically. We find that the introduction of aggregate risk modifies the wealth distribution in the stochastic steady state, SSS(s) (i.e., a fixed-point of the equilibrium conditions of the model when the realization of the aggregate shock is zero) in comparison to the deterministic steady state, DSS (i.e., a fixed-point of the equilibrium conditions of the model when the volatility of the aggregate shock is zero). More concretely, the most-visited SSS has a higher share of wealthy households because of precautionary savings. Consequently, this SSS displays more debt, leverage, and wealth inequality than the DSS. We also analyze the aggregate ergodic distribution of debt and equity based on the simulations of the model. The economy spends most of the time around the most-visited SSS, but may occasionally travels either to a high-leverage and a low-leverage region, generating transient dynamics with different features from the average ones.

Next, we study whether different wealth distributions modify the transmission of aggregate shocks. To this end, we consider three different pairs of distributions and equity levels, corresponding to the most-visited SSS and the high- and low-leverage regions visited in the paths pertaining to the ergodic distribution. We compute the generalized impulse response functions and the distributional impulse response functions to a negative capital shock. The responses are very similar on impact, but the ensuing recession is more persistent if the initial distribution is located in the high-leverage region. This higher persistence is due to the dynamics of aggregate household consumption. In a high-leverage economy, the decline in consumption of wealthy households is less severe than in the most-visited SSS. This milder decline produces a slower path of capital accumulation and, hence, creates a slow recovery path. From an individual perspective, the attenuation in the decline of consumption is consistent with the expected path of interest rates, which is more persistent in the high-leverage economy due to capital dynamics.

Overall, our findings shed light on how the increase in leverage and wealth inequality before a financial crisis may explain why the effects of the ensuing recession are more persistent and why nonlinearities matter for this mechanism.

2 Related literature

Our work relates to several important threads of recent work in macroeconomics. First, we build on the recent 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 (for a more comprehensive survey, see Brunnermeier and Sannikov (2016)). This literature has emphasized the importance of the wealth distribution in the economy. However, most of these papers only consider a basic form of heterogeneity between two representative groups, typically households and experts/intermediaries (or just dealt with models where the heterogeneity is trivially collapsed into an economic-wide average of leverage). Instead, in our paper, we can also deal with a more complex –and empirically relevant– form of heterogeneity, namely a non-trivial heterogeneous wealth distribution across households. The use of continuous time allows 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.

Related to this latter point, our paper also makes a technical contribution to the literature on global solution methods for heterogeneous agent models with aggregate shocks such as Den Haan (1996); Den Haan (1997), Algan, Allais, and Den Haan (2008), Reiter (2009, 2010), Den Haan and Rendahl (2010), Maliar, Maliar, and Valli (2010), Sager (2014), and Pröhl (2015) (a recent survey of the field can be found in Algan, Allais, Den Haan, and Rendahl (2014)). The presence of a nonlinear PLM allows us to investigate how aggregate risk affects the wealth distribution and in turn how different distributions modify the transmission of aggregate shocks.

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. As explained above, we employ a neural network to this end as it provides a flexible approach to the problem of estimation of an unknown nonlinear function. Naturally, other machine learnings schemes may also be proposed (or, for the matter, other nonlinear universal approximators such as series expansions or splines). We will explain in detail, however, why our approach is particularly convenient, both regarding theoretical properties and practical considerations. A further advantage of our approach is that it reflects in a fairly transparent way the self-confirming equilibrium (SCE) nature of the "bounded rationality" solution. 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.

The use of continuous-time methods is widespread in the macro-finance literature described above. In the case of models with heterogeneous agents, it is becoming more popular due to its advantages when performing numerical computations, as discussed in Achdou, Han, Lasry, Lions, and Moll (2017) or Nuño and Thomas (2017). Ahn, Kaplan, Moll, Winberry, and Wolf (2017) introduced a method to compute the solution to heterogeneous agent models with aggregate shocks. However, theirs is a local solution method, based on first-order perturbation around the deterministic steady state and, thus, unable to analyze the class of nonlinearly-related questions posed in our paper.

Finally, our paper contributes to the nascent literature on the application of machine learning techniques to improve the computation of dynamic general equilibrium models. Their methods have so far been concerned with the solution of high dimensional dynamic programming (DP) 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) instead employs a reinforcement learning algorithm together with a neural network to solve a two-sector model with 11 state variables. In contrast to these papers our machine learning algorithm is not used to overcome the curse of dimensionality in DP, but to provide a nonlinear forecast of aggregate variables within the model itself. In this respect, our paper is loosely related to an early literature using neural networks to model bounded rationality and learning, such as Barucci and Landi (1995), In-Koo (1995), Cho and Sargent (1996), or Salmon (1995).

Our methodological work may be useful to analyze other heterogeneous agents models with aggregate shocks. An obvious candidate is the analysis of the zero lower bound (ZLB) in heterogeneous agent New Keynesian (HANK) models models such as Auclert (2016), Gornemann, Kuester, and Nakajima (2012), Kaplan, Moll, and Violante (2018), Luetticke (2015), or McKay, Nakamura, and Steinsson (2016). The ZLB naturally 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. Another potential candidate is models in which the "quasi-aggregation" result does not hold, that is, models in which the higher-order moments are necessary to provide an accurate characterization. There is a priori no reason to assume that the law of motion of higher order moments should be linear and hence nonlinear techniques are required.

3 Model

We postulate a continuous-time, infinite-horizon model in the tradition of Basak and Cuoco (1998) and Brunnermeier and Sannikov (2014). Three types of agents populate our economy: a representative firm, a representative 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 the financial intermediaries in the economy, 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 economy has a risky asset in positive net supply, capital, and a risk-free asset in zero net supply, bonds. As it will be 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 at the cost of little additional notation.

3.1 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:

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

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

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

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 according to:

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

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.

3.2 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. And, 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 neither can issue state-contingent claims (i.e., outside equity) against $\widehat{K}_t$ nor have net savings in the risk-free asset. 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 defined as 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 $\widehat{K}_t$:

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

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

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

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:

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

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

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

3.3 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 notationally cumbersome.

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:

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

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 generalize the CRRA felicity function to more general classes 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.[1]

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).

[1] This property of our economy stands in contrast with models à la Bernanke, Gertler, and Gilchrist (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. Note that 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.

3.4 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:

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

We must also have $dB_t = d\widehat{B}_t$.

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

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

we get:

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

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

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

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

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

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.

3.5 Density

The households distribution $G_t(a, z)$ 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 circulate into state j ($\lambda_j g_{jt}(a)$).

4 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 $\{g_{it}(\cdot)\}_{t\geq 0}$ for $i \in \{1, 2\}$ such that:

1. Given w_t, r_t, and g_t, the solution of household m's problem (29) 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 $N_t = K_t - B_t$.

4.1 Some properties of the equilibrium

Several elements 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:

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

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 and expression for 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:

$$r_t = \alpha K_t^{\alpha-1} - \delta - \sigma^2 \frac{K_t}{N_t}. \tag{15}$$

Since $K_t = N_t + B_t$, these three equations depends only on the expert's net wealth N_t and debt B_t.

Fourth, we can describe the evolution of N_t:

$$\begin{aligned} dN_t &= \left[\left(r_t + \omega_t \left(rc_t - \delta - r_t \right) \right) N_t - \widehat{C}_t \right] dt + \sigma \omega_t N_t dZ_t \\ &= \left(\alpha K_t^{\alpha-1} - \delta - \widehat{\rho} - \sigma^2 \left(1 - \frac{K_t}{N_t} \right) \frac{K_t}{N_t} \right) N_t dt + \sigma K_t dZ_t \end{aligned} \tag{16}$$

that depend only on N_t and B_t plus dZ_t. Equation (16) 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):

$$\begin{aligned} dB_t &= (w_t + r_t B_t - C_t)\, dt \\ &= \left((1-\alpha) K_t^{\alpha} + \left(\alpha K_t^{\alpha-1} - \delta - \sigma^2 \frac{K_t}{N_t}\right) B_t - C_t\right) dt, \end{aligned} \tag{17}$$

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:

$$w_t = (1-\alpha) K_t^{\alpha} \tag{18}$$

$$rc_t = \alpha K_t^{\alpha-1} \tag{19}$$

$$r_t = \alpha K_t^{\alpha-1} - \delta - \sigma^2 \frac{K_t}{N_t} \tag{20}$$

$$dr_t^k = (rc_t - \delta)\, dt + \sigma dZ_t \tag{21}$$

$$dN_t = \left(\alpha K_t^{\alpha-1} - \delta - \widehat{\rho} - \sigma^2 \left(1 - \frac{K_t}{N_t}\right) \frac{K_t}{N_t}\right) N_t dt + \sigma K_t dZ_t. \tag{22}$$

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

$$C_t \equiv \sum_{i=1}^{2} \int c\,(a_t, z_t, K_t, G_t)\, g_{it}\,(a)\, da \tag{23}$$

$$dB_t = \left((1-\alpha) K_t^{\alpha} + \left(\alpha K_t^{\alpha-1} - \delta - \sigma^2 \frac{K_t}{N_t}\right) B_t - C_t\right) dt \tag{24}$$

$$dK_t = dN_t + dB_t \tag{25}$$

$$\frac{\partial g_{it}}{\partial t} = -\frac{\partial}{\partial a} \left(s\,(a_t, z_t, K_t, G_t)\, g_{it}(a)\right) - \lambda_i g_{it}(a) + \lambda_j g_{jt}(a),\ i \neq j = 1, 2. \tag{26}$$

The second block shows us 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. Thus, in practice, computing the equilibrium of this economy is equivalent to finding C_t. Once C_t is known, all other aggregate variables follow easily.

4.2 The DSS of the model

Before we explain our general solution in the next section, we use the previous equations to analyze three interesting cases. In this subsection, we describe the DSS of the model. We will look below at the SSS and at the representative household version of the model.

In particular, we look now at the situation where there are no capital growth rate shocks, but we still have idiosyncratic household shocks. In the next subsection, we describe the representative household version of the model.

To do so, we can go to the law of motion for the expert net wealth (16), set $\sigma = 0$, and get:

$$dN_t = \left(\alpha K_t^{\alpha-1} - \delta - \widehat{\rho}\right) N_t dt. \tag{27}$$

Since the drift of N_t, $\mu^N(B, N) = \left(\alpha K^{\alpha-1} - \delta - \widehat{\rho}\right) N$, must be zero in a DSS (recall 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 (15) equals the return on capital and the rental rate of capital less depreciation:

$$r = r_t^k = rc_t - \delta = \alpha K_t^{\alpha-1} - \delta = \widehat{\rho}. \tag{28}$$

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.

4.3 The SSS of the model

A SSS (recall, a stochastic steady state) in this model is formally defined as a density $g^{SSS}(\cdot)$ and equity N^{SSS} that remain invariant in the absence of aggregate shocks. Let $\Gamma_\sigma(g(\cdot), N, W)$ the law of motion of the economy given an aggregate capital volatility σ and a realization of the Brownian motion W. It is an operator that maps income-wealth densities

$g(\cdot)$ and equity levels N into changes in these variables:

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

The SSS, therefore, solves:

$$\Gamma_\sigma(g^{SSS}(\cdot), N^{SSS}, 0) = \left[\begin{array}{c} 0 \\ 0 \end{array} \right].$$

In general, there is no reason why we could not have multiple SSSs. In fact, we will document below the existence of several of them for our model.

Notice that 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 (31)− but no shock arrives, whereas in the latter agents live in an economy without aggregate risks ($\sigma = 0$) and take their decisions accordingly. The DSS is then formally defined, using our notation above, as

$$\Gamma_0(g^{DSS}(\cdot), N^{DSS}) = \left[\begin{array}{c} 0 \\ 0 \end{array} \right].$$

4.4 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 (there is still an expert and a household), but we do not have any *within-agents* heterogeneity (there is one representative expert and one representative household).

Equations (18) to (22) of the equilibrium, as well as equations (24) and (25), remain unchanged. Instead of equation (23), 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:

$$\max_{\{C_t\}_{t\geq 0}} \mathbb{E}_0 \left[\int_0^\infty e^{-\rho t} \frac{C_t^{1-\gamma} - 1}{1-\gamma} dt \right], \tag{29}$$

subject to the budget constraint (24) (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 (26) in this version of the model.

5 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 start by following 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 $n > 1$.

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

$$dB_t = h(B_t, N_t)\, dt, \tag{30}$$

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

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

instead of the exact law of motion (17). We borrow the term PLM from the learning literature (Evans and Honkapohja, 2001). Our choice of words accentuates that we allow $h(\cdot,\cdot)$ be a general function, and not just a linear or polynomial function of its arguments (perhaps with state-dependent coefficients). In fact, our methodology will let the PLM approximate, arbitrarily well, equation (17).

In the original Krusell and Smith (1998) methodology, the PLM is assumed to be log-linear in the endogenous state variables and nonlinear in the exogenous state by making the coefficients of the log-linear specification dependent on the realization of the aggregate shock. As explained in Subsection 5.3, we propose a more flexible methodology, in which the functional form $h(\cdot,\cdot)$ is not specified, but obtained from simulated data by employing machine learning techniques. This extra flexibility is key given the complex nonlinearities present in laws of motion of N_t, equation (16), and B_t, equation (17).

Given the PLM, the household's problem has an associated Hamilton-Jacobi-Bellman

(HJB) equation:

$$\rho V_i(a,B,N) = \max_c \frac{c^{1-\gamma}-1}{1-\gamma} + s\frac{\partial V_i}{\partial a} + \lambda_i\left[V_j(a,B,N) - V_i(a,B,N)\right]$$
$$+h(B,N)\frac{\partial V_i}{\partial B} + \mu^N(B,N)\frac{\partial V_i}{\partial N} + \frac{\left[\sigma^N(B,N)\right]^2}{2}\frac{\partial^2 V_i}{\partial N^2}, \tag{31}$$

$i \neq j = 1, 2$, and where we use the shorthand notation $s = s(a, z, N + B, G)$. Note how the HJB incorporates $h(B, N)$. Equation (31) complements the equilibrium conditions (18)-(26) by making the problem of the household explicit.

5.1 An overview of the algorithm

The algorithm to find $h(B, N)$ in (30) 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 (31). This solution can be obtained by using an upwind finite differences scheme described in Appendix A (although other numerical algorithms 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 (16) and (25).

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.

5.2 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 with respect to discrete time (Achdou, Han, Lasry, Lions, and Moll, 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:

$$c_i^{-\gamma} = \frac{\partial V_i}{\partial a}. \tag{32}$$

Third, it is easier to capture occasionally binding constraints such as equation (9) in continuous time than in discrete time as the optimality condition (32) 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). Appendix A provides further details.

Regarding the generation of data, we simulate T periods of the economy with a constant time step Δt. The initial income-wealth distribution can be the one 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.

5.3 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.[2] But, as shown in equations (16) and (17), 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 5.1 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, Stinchcombe, and White 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.[3]

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 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 approxi-

[2]Recall that the SSS (stochastic steady state; also called the risky steady state) is the fixed-point of the law of motion for capital and the household distribution when the realization of the aggregate shock is zero, but agents still consider the shock can occur. In comparison, in the DSS (deterministic steady state), the volatility of the aggregate shock has been driven to zero.

[3]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.

mated. 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). 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.[4]

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 15 and 16 in Appendix B 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.

5.4 More on the neural network approximator

Here we briefly describe our neural network approximator of the PLM in more detail. For excellent introductory treatments of this material, see Bishop (2006) and Goodfellow, Bengio, and Courville (2016).

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

$$h(\mathbf{s}; \theta) = \theta_0^2 + \sum_{q=1}^{Q} \theta_q^2 \phi\left(\theta_{0,q}^1 + \sum_{i=1}^{2} \theta_{i,q}^1 s^i\right), \tag{33}$$

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

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

Note how we call θ "coefficients," as they represent a numerical entity, in comparison with the structural parameters, Ψ, that have a sharp economic interpretation. Different alternatives are available for the activation function. For our model, we choose a *softplus* function,

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

$\phi(x) = \log(1 + e^x)$ for a given input x. why?

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.

The neural network (33) 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 $\left(\mathbf{S}, \widehat{\mathbf{h}}\right)$:

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

A standard approach to perform this minimization in neural networks is the *stochastic gradient descent* algorithm. The stochastic gradient descent begins by drawing a random initialization of θ_0 from a known distribution Θ, typically a Gaussian or uniform:

$$\theta_0 \sim \Theta. \tag{34}$$

Then, 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, Hinton, and Williams (1986), which builds on the chain rule of differential calculus. In our case, this results in:

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

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

One potential concern with neural networks is that the algorithm might converge to a local minimum. A way to coping with it is to implement a Monte Carlo multi-start. We select P initial vectors θ_0^p, with $p = 1, ..., P$ from (34). For each of these vectors, we run the stochastic 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 acceptable solutions to the approximation problem.

6 Estimation

Once we have solved the model given some structural parameter values Ψ, the next step is to take it to the data and let observations determine the values in Ψ and perform inference on them. We will proceed in two stages. First, we will discuss the simple case where the econometrician has access to only aggregate data and wants to build the likelihood associated with them. Once we understand how to construct the likelihood in this situation, we move to describe how to add microeconomic observations from the cross-sectional distribution of assets and evaluate a much more sophisticated likelihood.

Let $X_t \equiv [B_t; N_t]'$ be a vector of aggregate state variables, debt and net wealth of the expert. Imagine that we have $D+1$ observations of X_t at fixed time intervals $[0, \Delta, 2\Delta, .., D\Delta,]$:

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

We could also handle the more general case where the states are not observed (or they are observed given measurement error). In that situation, we would require 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).

We are interested in estimating Ψ using the likelihood:

$$\mathcal{L}\left(X_0^D|\Psi\right) = \prod_{d=1}^{D} p_X\left(X_{d\Delta}|X_{(d-1)\Delta};\Psi\right)$$

where

$$p_X\left(X_{d\Delta}|X_{(d-1)\Delta};\Psi\right) = f_{d\Delta}(B_{d\Delta}, N_{d\Delta})$$

is the conditional density of $X_{d\Delta}$ given $X_{(d-1)\Delta}$.

The key to evaluate the likelihood is to note that $f_t(B, N)$ follows the KF equation in the interval $[(d-1)\Delta, d\Delta]$:

$$\begin{aligned}\frac{\partial f_t}{\partial t} = -\frac{\partial}{\partial B}\left[h(B,N)f_t(B,N)\right] - \frac{\partial}{\partial N}\left[\mu_t^N(B,N)f_t(B,N)\right] + \\ \frac{1}{2}\frac{\partial^2}{\partial N^2}\left[\left(\sigma_t^N(B,N)\right)^2 f_t(B,N)\right]\end{aligned} \tag{35}$$

where

$$f_{(d-1)\Delta} = \delta\left(B - B_{(d-1)\Delta}\right)\delta\left(N - N_{(d-1)\Delta}\right)$$

and $\delta(\cdot)$ is the Dirac delta function (Lo, 1988).

A fundamental property of the operator in the KF equation (35) is that it is the adjoint of the infinitesimal generator employed in the HJB. 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 once the model is solved.[5]

A promising avenue to improve the estimation is to add micro observations, which bring much additional information and help to integrate different levels of aggregation to assess the empirical validity of the model. This route has the additional advantage that we can substitute B_t as an observable (which is not a state variable of the model, but only an element of the Krusell-Smith-type of solution we adopt) for the cross-sectional distribution of assets.

More concretely, we consider $X_t \equiv [g_t(a,z); N_t]'$, that is, we replaced aggregate debt by the complete income-wealth distribution. As before, X_t^D stacks the whole sequence of observations.

At this moment, we need to assume –as typically 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 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 $\log\left[f_{d\Delta}^d(a_{d\Delta}^i, z_{d\Delta}^i, B_{d\Delta}, N_{d\Delta})\right]$. Note that this log-likelihood is a function of i. The conditional aggregate log-likelihood across all agents is

$$\log p_X\left(X_{d\Delta}|X_{(d-1)\Delta};\Psi\right) = \int \log\left[f_{d\Delta}^d(a_{d\Delta}^i, z_{d\Delta}^i, B_{d\Delta}, N_{d\Delta})\right]\Phi(di),$$

[5]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).

and taking into account the CNAU condition

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

where in the second line we have applied the definition of the Radon-Nikodym derivative.

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

$$\begin{aligned}\frac{\partial f^d_t}{\partial t} = &-\frac{\partial}{\partial a}\left(s_t\left(a, z_i\right) f^d_t(a, z_i, B, N)\right) - \lambda_i f^d_t(a, z_i, B, N) + \lambda_j f^d_t(a, z_j, B, N) \\ &-\frac{\partial}{\partial B}\left[h(B, N) f^d_t(a, z_i, B, N))\right] - \frac{\partial}{\partial N}\left[\mu^N_t(B, N) f^d_t(a, z_i, B, N)\right] \\ &+\frac{1}{2}\frac{\partial^2}{\partial N^2}\left[\left(\sigma^N_t(B, N)\right)^2 f^d_t(B, N)\right], \;\; (i \neq j = 1, 2)\end{aligned} \tag{36}$$

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, again 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 (36) 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 $\mathbf{f}^d_t$ is defined as

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

and δ is the Kronecker delta.

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 $\mathbf{g}_{d\Delta}$.

6.1 Results

Once we have evaluated the likelihood, we can either maximize it or perform Bayesian inference relaying on a posterior sampler.

To illustrate that our approach, indeed, works, we implement the following Monte Carlo experiment. We calibrate the model replicate some relevant features of the U.S. economy. In our presentation of the parameters, we will report rates at an annual term. The capital share parameter, α, is taken to be 0.35 and the depreciation rate of capital, δ, is 0.1. The discount rate ρ, is set to 0.05. The intertemporal elasticity of substitution of the households $\frac{1}{\gamma}$ is set to 0.5 so that the risk aversion is 2.

Table 1. Baseline calibration for Monte Carlo experiment

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
y_1	0.72	income in unemployment state	Hall and Milgrom (2008)
y_2	1.015	income in employment state	$E(y) = 1$
$\widehat{\rho}$	0.0497	experts' discount rate	$K/N = 2$
σ	0.015	volatility of shocks	-

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.[6] 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 y_1 equal to 71 percent of y_2, as in Hall and Milgrom (2008). Both targets allow us to solve for y_1 and y_2. We set experts' discount rate $\widehat{\rho}$ such that the leverage ratio K/N in the most-visited

[6] Analogously to Blanchard and Galí (2010, see their 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.

SSS is 2, which is roughly the average leverage from a Compustat sample of non-financial corporations. We set volatility $\sigma = 0.01$ and will conduct some robustness analysis below. Table 1 summarizes our baseline calibration.

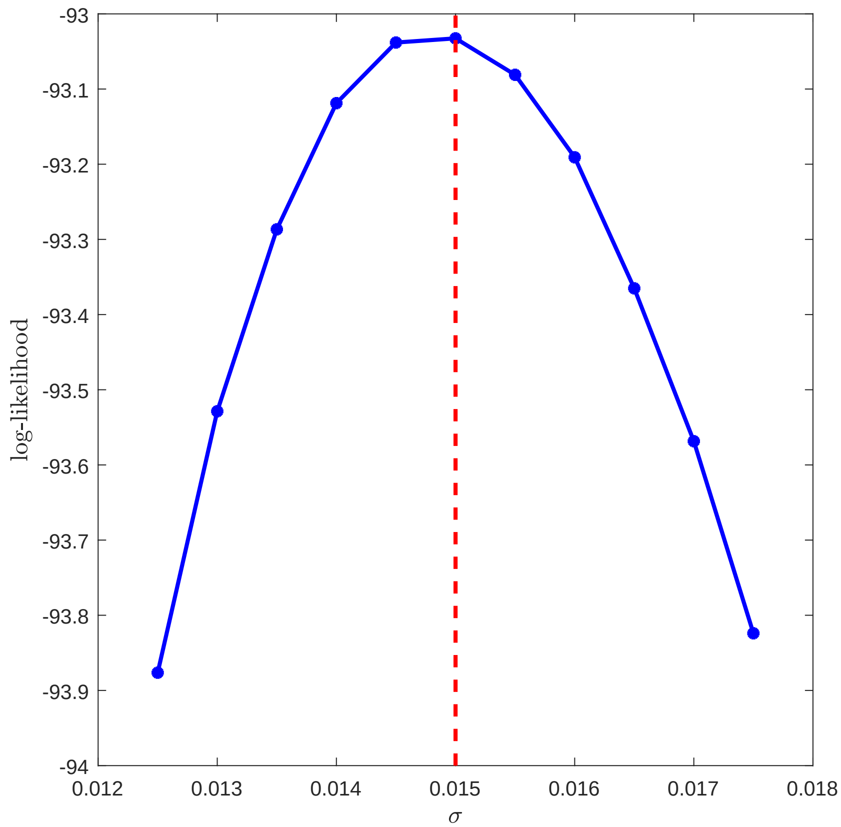

Figure 1: The log-likelihood for different values of σ with simulated data.

We solve the model according to the algorithm in Section 5 and sample of 5,000 years at monthly frequency, using a minibatch comprising 64 observations. Then, we evaluate the likelihood of those observations for different values of σ, where all the other parameter values are fixed at their calibrated quantities. We use a grid between 0.008 and 0.018 with step 0.00025. We plot the resulting loglikelihood in Figure 1. The pseudo-true parameter value is drawn as a vertical discontinuous red line. The loglikelihood picks, indeed, at this point. The smoothness of the plot confirms, also, that our algorithm has successfully converged.

7 Results

7.1 The PLM

The resulting PLM generated by our solution algorithm for aggregate debt, $h(B,N)$, is reported in Figure 2. Panel (a), at the top left, displays three transversal cuts of $h(B,N)$ along a range of values of equity (N). The first cut is fixes B at its baseline SSS value ($B = 1.9437, N = 1.7694$, with $K = 3.7131$), the second cut fixes B at a selected high-leverage point ($B = 2.15, N = 1.5$, with $K = 3.65$), and the third cut fixes B at a SSS with low-leverage point ($B = 1.0868, N = 2.6115$, with $K = 3.6983$). The thicker part of the lines indicate the regions of the state space in which the ergodic distribution of aggregates variables, to be described below, is nonzero. The white point indicates, in the first cut, the baseline SSS; in the second cut, the high leverage point described above; in the third cut, the SSS with low leverage point. 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. 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 demonstrates the non-linearity of $h(B,N)$ even within the area of the ergodic distribution that has positive mass. The agents in our economy expect very different growth rates of B_t in each region of the state space, with the function switching from concavity to convexity along the state space. While this argument is clear from the general shape of panel (c), it encodes rich dynamics that deserve further explanation. For example, panel (b) shows how, as leverages increases, $h(B,N)$ becomes stepper. 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 comparatively more risk. This result will resurface when we document the dynamics of the economy after a shock.

The general shape of panel (c) also documents that $h(B,N)$ is generally decreasing in debt and equity (although with some non-monotonicities).

Figure 3 reproduces the same three panels of Figure 2, except for $dh(B,N)$ once we have used equation (30). Similar comments regarding the non-linear structure of the solution apply here. For example, now, $dh(B,N)$ becomes less steep as a function of equity as the level of leverage falls.

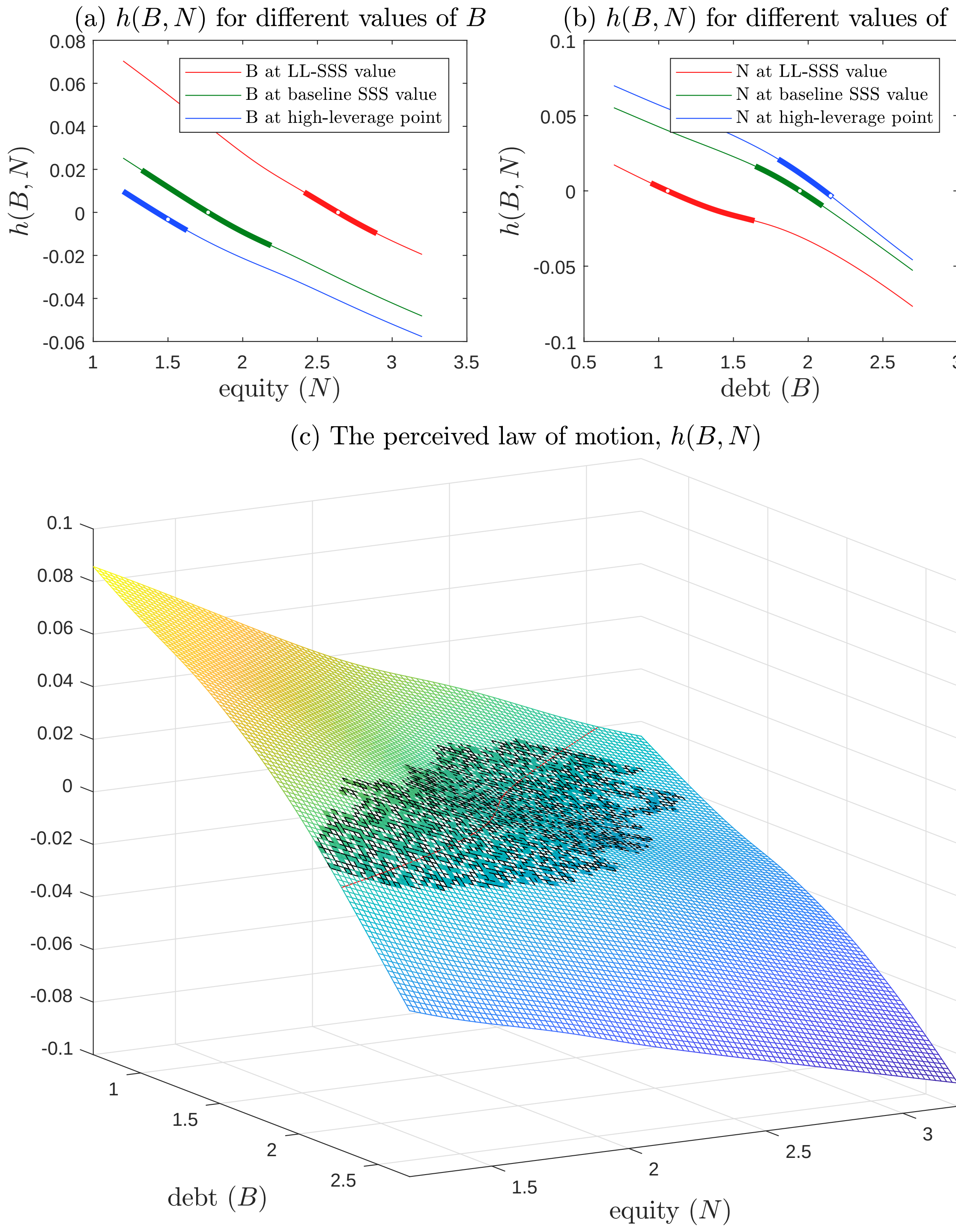

Figure 2: The PLM $h(B,N)$.

Note: White points in panels (a) and (b) indicate the most-visited SSS and selected points in the high- and low-leverage regions of the ergodic distribution. 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.

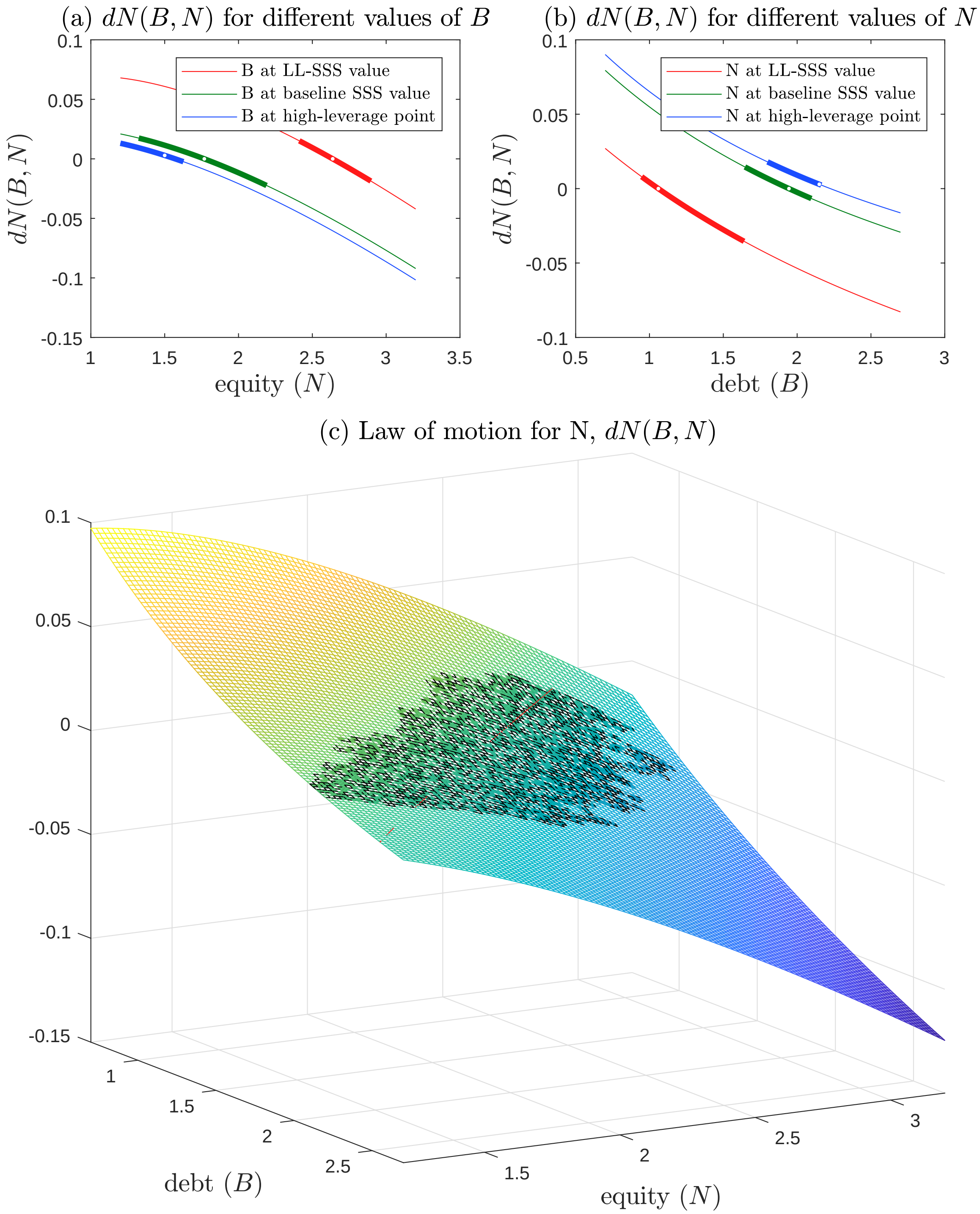

Figure 3: The law of motion $dh(B,N)$.

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

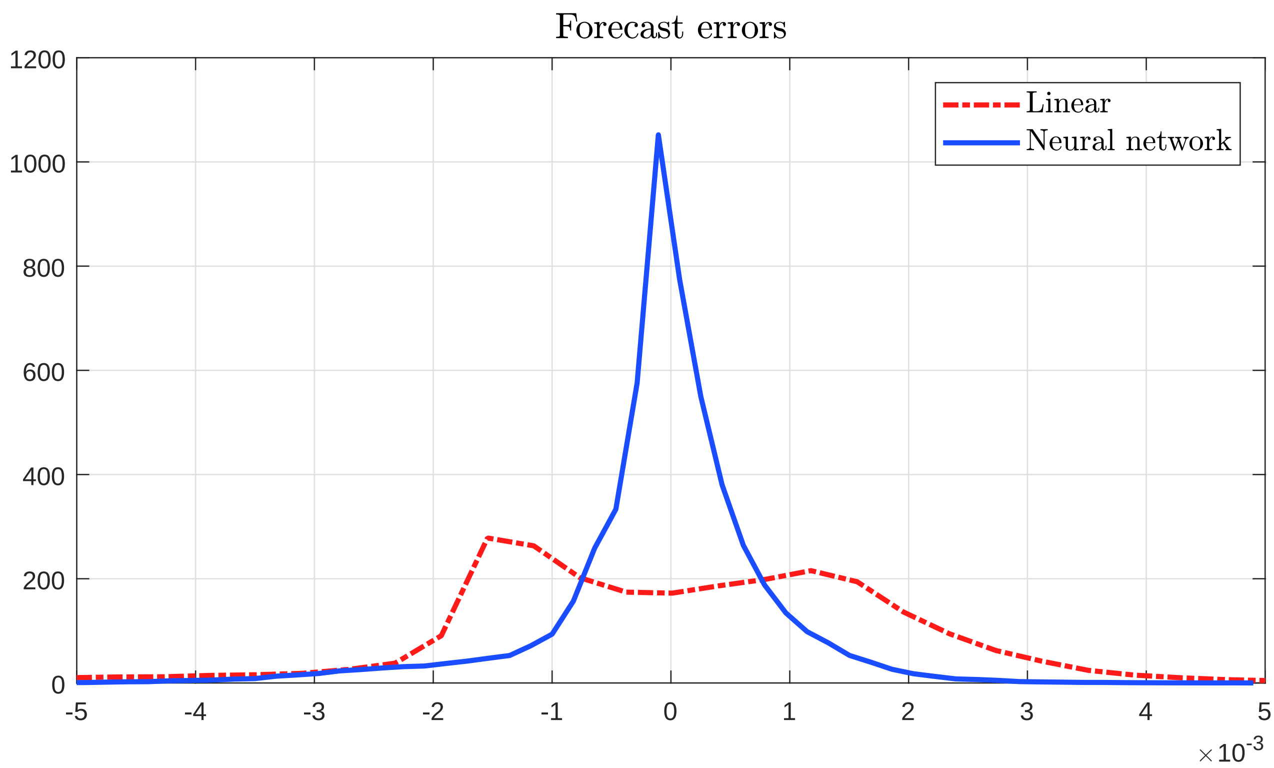

Figure 4: Forecasting error distribution at a 1-month horizon, linear PLM (left) and nonlinear (right).

The non-linearity of the PLM confirms our conjecture that more traditional solution methods that rely on linear structures might not be appropriate for solving this class of models. We can document this argument more formally by looking at the forecasting capability of our PLM. The R^2 associated to the PLM we compute using our neural network is 0.9704, with an RMSE of 0.00088. The forecasting errors, furthermore, are nicely clustered around zero. To compare it 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 0.8137, considerably lower than typical values reported in the literature, and with an RMSE of 0.00220. Figure 4 plots the histogram of forecasting errors at a 1-month horizon (the time step selected in the simulation step). The linear Krusell-Smith algorithm produces more volatile forecasting errors, which are also skewed to the right and without a mode at zero. These results confirm the importance of taking into account the nonlinearities of the model when computing the PLM.[7]

[7]In the Appendix, 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.

7.2 The phase diagram

Figure 5 plots the model phase diagram along the aggregate debt (B) axis. The blue line represents the PLM when it takes zero value, i.e., $h(B, N) = 0$. The discontinuous red line represents the loci of zero changes in aggregate debt, $dN(B, N) = 0$.[8] The arrows indicate the movement of debt, B, and equity, N, when we are away from the blue and red lines. For completeness, we also plot the point of high leverage that we use in Figures 2 and 3 (and later, when computing GIRFs) to illustrate the behavior of the economy at a high leverage situation.

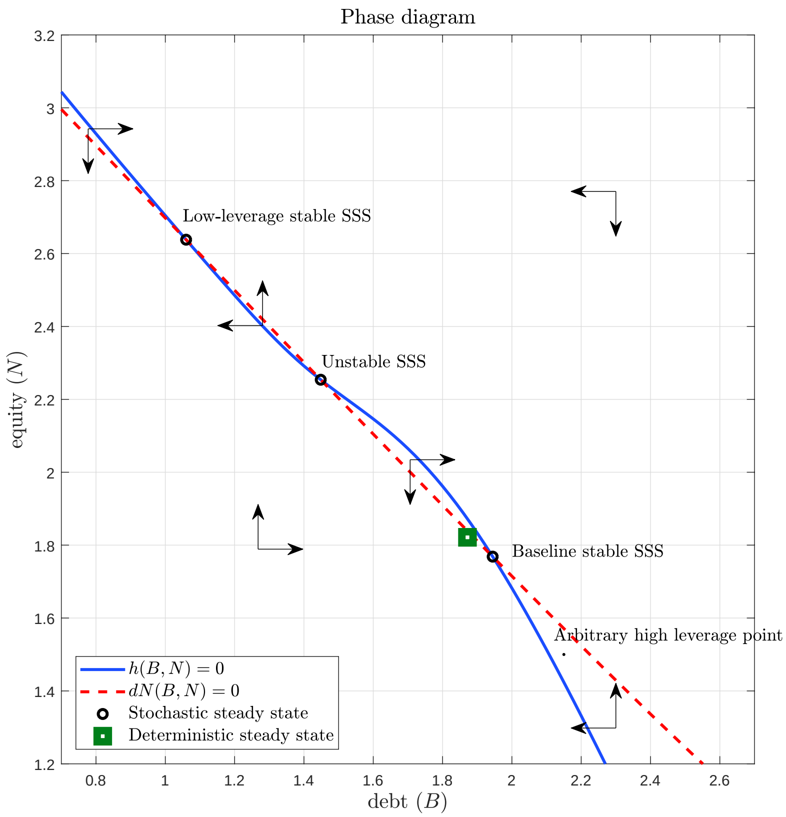

Figure 5: Phase diagram and SSS.

The lines intersect three times, defining three SSS. From the bottom right, the first intersection is the baseline, stable SSS (recall, with $B = 1.9437, N = 1.7694$, and $K = 3.7131$). We call this SSS "baseline" because of two reasons. First, this SSS (and its neighborhood) is the most visited one in the ergodic distribution (see Figure 8 below). Second, this SSS is the closest to the DSS (green square). In comparison with the DSS ($B = 1.8718, N = 1.8215$,

[8]Note that these two lines are the intersections of the zero level with the PLM and $dN(B, N)$ in Figures 2 and 3, which we represented –in those figures– with a thin red line.

with $K = 3.6933$), the baseline SSS has more debt and less equity. Furthermore, this SSS has slightly more capital than the DSS.

The mechanism behind this shift between the baseline SSS and the DSS is as follows. The presence of aggregate and idiosyncratic uncertainty motivates households to undertake precautionary behavior to smooth consumption intertemporally. This precautionary behavior increases, everything else equal, aggregate debt. But for the debt market to clear, the financial expert must be induced to demand more debt, which can only happen if the risk-free rate falls. A lower risk-free rate persuades the financial expert to finance herself more with debt and less with equity, as to gather the larger excess returns from leverage. In other words, at this SSS, the financial experts absorbs much of the aggregate risk in the economy.

Mechanically, the risk-free interest rate in the SSS is given by equation (15):

$$r^{SSS} = \alpha \left(K^{SSS}\right)^{\alpha-1} - \delta - \sigma^2 \frac{K^{SSS}}{N^{SSS}},$$

whereas in the DSS is given by equation (28):

$$r^{DSS} = \alpha \left(K^{DSS}\right)^{\alpha-1} - \delta.$$

From these two equations and the fact that, for our parameter values, $K^{SSS} > K^{DSS}$, we get $r^{DSS} > r^{SSS}$. First, a higher capital lowers the marginal productivity of capital. Second, aggregate uncertainty pushes the risk-free rate even further.

The second intersection is at a middle SSS with less debt and more equity ($B = 1.4485$, $N = 2.2538$, with $K = 3.7023$). This SSS is, however, unstable, and the dynamics of the economy quickly move away from it. Thus, we will not spend much time discussing it.

Finally, the third SSS, at the top left, is a state SSS with low leverage: debt is much lower ($B = 1.0868$) and equity much higher ($N = 2.6115$) than in the basic stable SSS, yielding $K = 3.6983$. In this SSS, we have the reverse situation from the first SSS. The low leverage translates into a high risk-free rate, inducing the financial expert to finance herself through equity and not debt. On the other hand, the households have an incentive to save more, which will mean that the economy will move, on average from this SSS. As we will document below, this low-leverage SSS occurs, indeed, much less often than the baseline SSS.

We can explore the role of σ in determining the different SSS in Figure 6. Each panel plots, for a different value of σ, the phase diagram of the economy following the same convention than Figure 5. To understand the panels in this figure, let us recall that, in our economy, the financial intermediary is exposed to the capital risk created by positive σ. The households are exposed to the risk created by positive σ on their wages and risk-free returns on their assets plus the idiosyncratic labor productivity shocks. For low values of σ, the financial

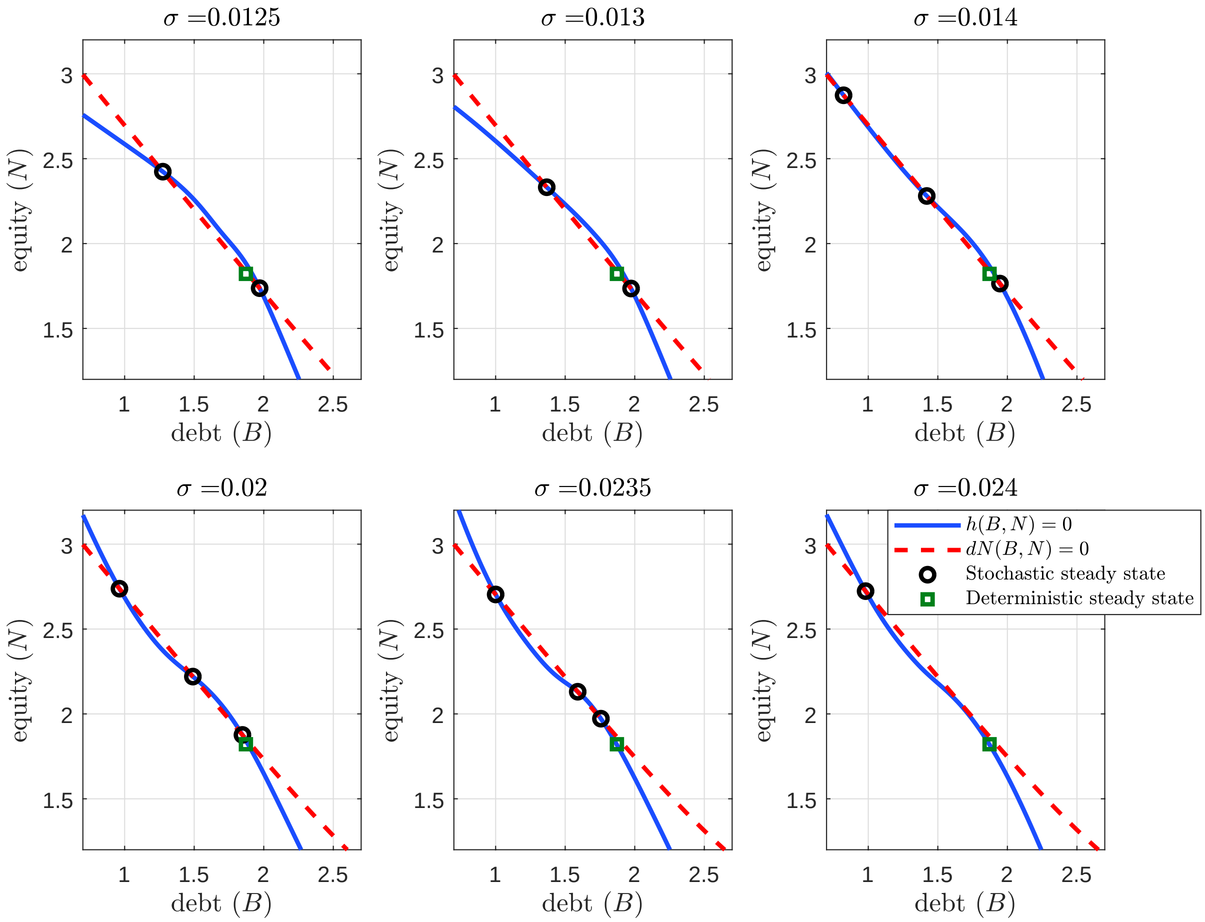

Figure 6: Phase diagram as a function of σ.

intermediary is exposed to less total risk than the household and, therefore, the baseline SSS is to the right of the DSS, i.e., the baseline SSS has more debt and less equity than the DSS (we still have, nevertheless, the second stable SSS with much more equity and less debt).

However, as σ increases, the situation changes and the baseline SSS gets closer to the DSS, until it actually crosses it for σ around 0.019 and we end up with a baseline SSS with less debt and more equity. Also, as σ increases, $h(B, N) = 0$ becomes more curvy and the unstable SSS appears around $\sigma = 0.014$.

Figure 7 complements the previous results by plotting the values of the low-leverage SSS, the unstable SSS, and the baseline SSS (plus, for reference, the deterministic SS) for a range of values of σ. We can see how the leverage in the baseline SSS is a negative function of σ, a positive function in the unstable SSS, and roughly constant in the low-leverage SSS.

Does the economy converge the any of the SSS(s)? The state space $(g(\cdot), N)$ is infinite dimensional and hence we cannot numerically analyze convergence for any possible initial state. Instead we analyze convergence for densities visited in the aggregate ergodic distribution. We

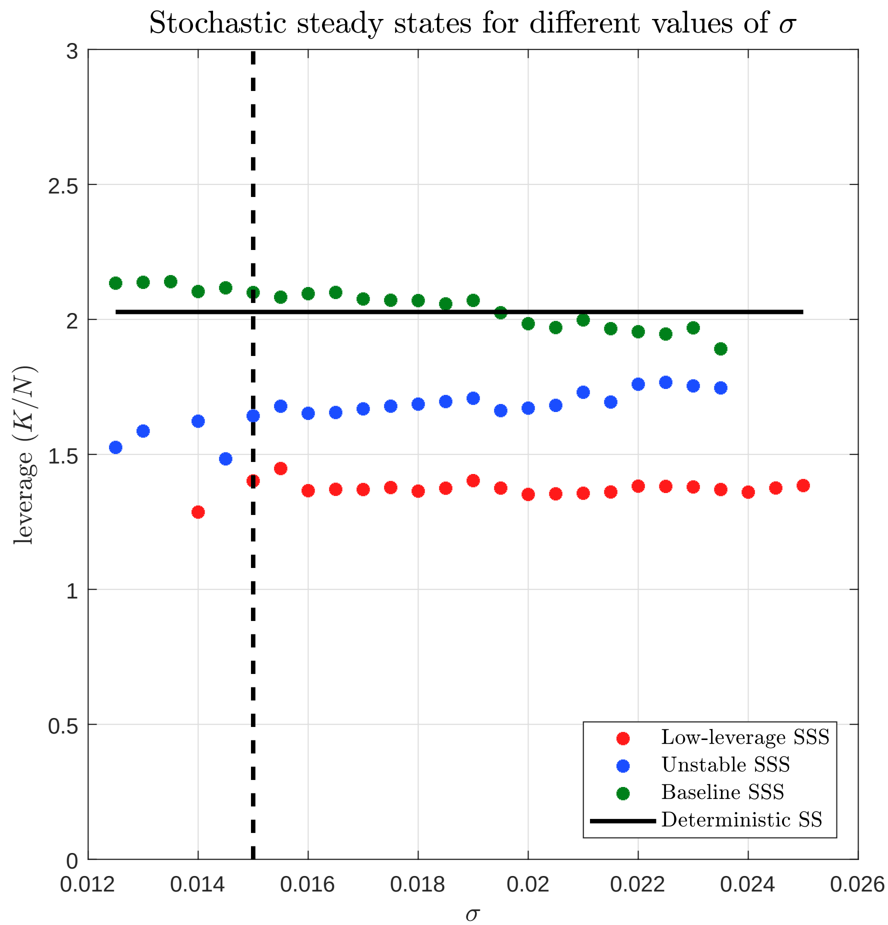

Figure 7: SSS as a function of σ.

consider different initial values, i.e.,income-wealth densities and equity levels $(g_0(\cdot), N_0)$, selected from the simulations employed to compute the aggregate ergodic distribution above and analyze the transitional dynamics in the absence of aggregate shocks.[9] In all the cases considered the economy converges to the SSS. The convergence path is very slow: for initial values in the high- or low-leverage regions convergence may take several centuries. Notwithstanding, we cannot rule out that for other densities different to the ones in the aggregate ergodic distribution the model would not converge to the SSS. This is again related to the SCE nature of the solution. The PLM is computed based on the income-wealth distributions visited along the paths of the ergodic distribution. In the case of a different distribution agents will necessarily extrapolate the dynamics only based on the aggregate first moment B. This, however, may lead to different dynamics.

An interesting feature of the transitional dynamics is the curvature of the paths that lie

[9]Despite the fact that agents form their expectations assuming the continuous arrival of shocks with $\sigma > 0$.

in regions with relatively low of high capital levels, in these cases the paths look "curved" due to the fact that debt adjusts much faster than equity, the latter mainly depending on the leverage ratio. If equity and aggregate capital are below (above) their long-run values households (de-)accumulate wealth increasing their leverage. Once they reach a point with still low (high) equity but a capital level closer to that in the SSS then there is a progressive redistribution of wealth from households to experts (or the other way around in the case of initial high equity and capital). In the next sunsection, we will discuss more about the model ergodic distribution.

7.3 The aggregate ergodic distribution

Panel (a) of Figure 8 displays the aggregate ergodic joint distribution of debt and equity $F(B,N)$. This distribution is defined as

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

for any subset Ω of the state space. This distribution is not obtained directly from the PLM, but from the simulation of the paths of the income-wealth distribution and equity processes (naturally, the PLM is employed in the HJB equation (31) to obtain the optimal consumption policies of individual agents). A Monte Carlo simulation of 5,400 years of the economy at monthly frequency has been employed to compute it, in which the model is initialized at the deterministic steady state and the first 400 samples are discarded.

Panels (b) and (c) of Figure 8 plot the marginal distributions for debt and equity. These 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 2.5. Note how the marginal distribution of debt is much more concentrated than the marginal distribution of equity and how the ergodic distribution only presents one region of substantial mass, although it has a substantial tail in areas of high equity and low debt. Finally, note how the total amount of capital ($K=B+N$) is on average larger in the low leverage region than in the high leverage one.

We complement the previous discussion with Figure 9, which plots the histogram of duration of spells of the economy around the baseline SSS and the low-leverage SSS. The average duration a spell of the economy around the baseline SSS is 48.69 years, a long period of relatively high leverage. On the other hand, the average duration of a spell of the economy around the low-leverage SSS is 4.75 years, a much shorter length, reflecting the strong push effect of high risk-free rates on household accumulation decisions (and, therefore, on the re-building of leverage). Nevertheless, the accumulated effects of repeated shocks also mean that the spells around this low-leverage SSS can be much longer than 4.75 years, with some

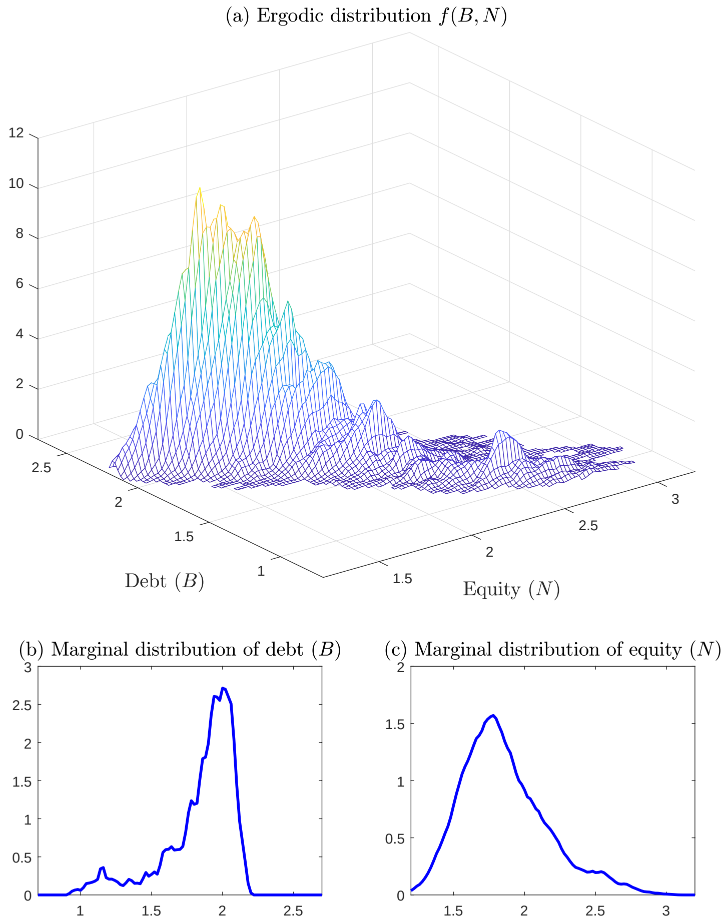

Figure 8: Ergodic distribution $F(B, N)$. Lighter colors indicate higher probability.

lasting as long as a century.

When discussing the challenges of a nonlinear model with heterogenous agents and aggregate shocks in Section 3, we mentioned the problem of extrapolation. In order to compute the PLM based on simulated paths only a limited region of the state space is visited, that of the ergodic distribution $F(B, N)$ displayed in Figure 2 in a shaded area. However, when forming expectations, households evaluate the PLM over the entire state space. Therefore 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 −were it to be ever visited− coincide with the ones expected in the PLM. Thus, the approximation employed both in Krusell and Smith (1998) and in this paper should be understood as a particular instance of a self-confirming equilibrium (SCE). See also Brumm and Scheidegger (2017) and Piazzesi and Schneider (2016) for related discussions. In a SCE households' beliefs about the PLM coincide with the actual law of motion only in the equilibrium paths. Off–equilibrium they may diverge, but households never discover it as this

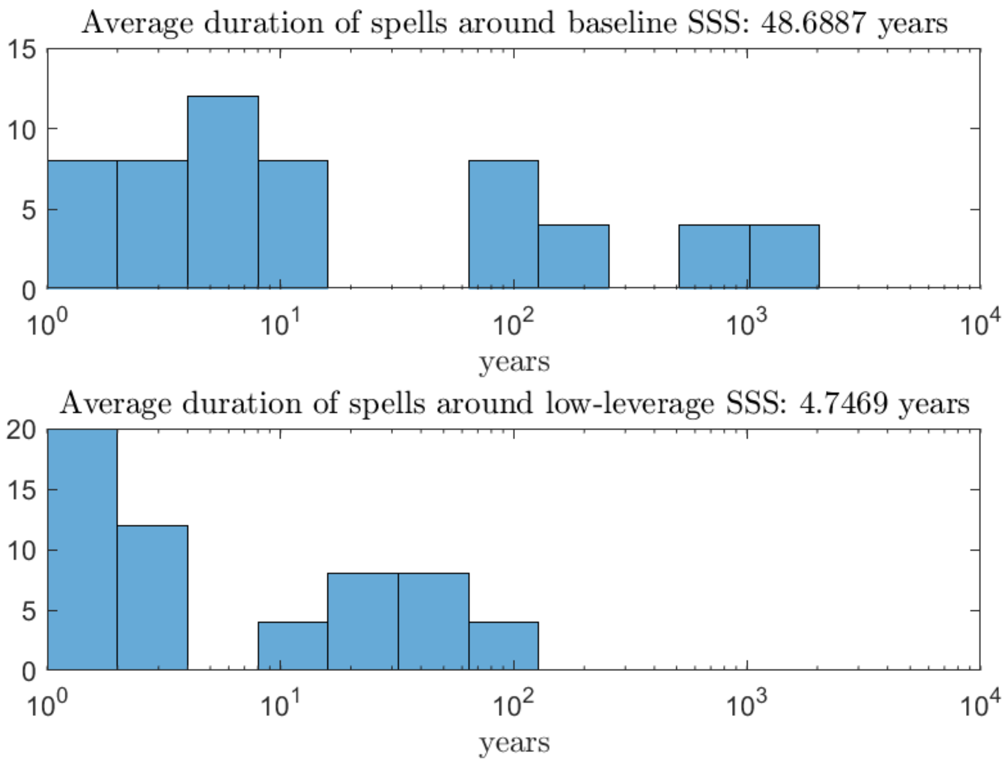

Figure 9: Spell durations at each DSS.

region is never visited.

7.4 Wealth distributions in the DSS and SSS(s)

Figure 10 compares the wealth distribution in the DSS, the baseline SSS, and the low-leverage SSS. The distribution shift to the left as we move from the DSS to the baseline SSS, and from the baseline SSS to the low-leverage SSS.

As we discussed before, everything else equal, the presence of $\sigma > 0$ lowers the risk-free interest rate, and thus pushes the assets of the household to the left in the baseline SSS with respect to the DSS. However, this move is small as the baseline SSS displays higher wages and lower interest rates due to the higher capital level and the higher leverage ratio. The net effect is that savings almost coincide for low-wealth households in the DSS and baseline SSS, but are higher for high-wealth households in the DSS. There is, however, a thicker right tail in the baseline SSS. Therefore, the presence of aggregate risk in this economy produces an allocation with a higher level of aggregate savings and more wealth inequality, together with higher output and more leverage, compared to the allocation without aggregate risk.

When, in addition, $\frac{K^{SSS}}{N^{SSS}}$ is high, the risk-free is considerable lower, as the financial expert

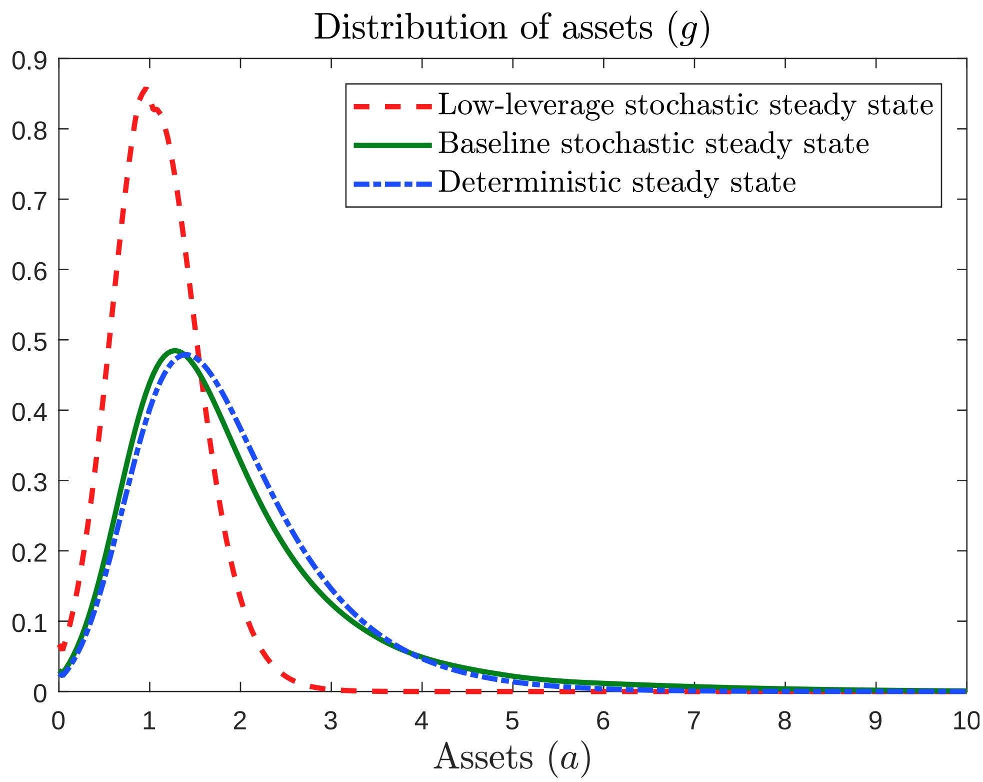

Figure 10: Wealth distribution in the DSS and SSS.

must be induced to take on the risk of holding a large amount of equity (baseline SSS vs. low-leverage SSS). The effect now is much larger because, in addition, total capital (and with it wages) is lower in the low-leverage SSS. The combination of a lower risk-free with a lower wage reduces the savings of both high- and low-wealth households.

The bounded rationality solution introduce by Krusell and Smith (1998) assumes that households only employ a finite set of moments of the income-wealth distribution to forecast the dynamics of aggregate variables. In our case, as in the original Krusell and Smith paper, we only consider the first moment, the total amount of wealth, aggregate debt in our case. This 'approximate aggregation' holds well in our model as well as in theirs due to the linearity of the consumption policy. Only agents close to the borrowing limit face a nonlinear consumption policy but, being close to zero assets, they contribute relatively little to the aggregate dynamics of capital.[10]

[10]Note, however, that the quasi-linearity of the consumption function is only 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.

7.5 Dynamics

The analysis above has shown how, in this model, the introduction of aggregate fluctuations modifies the distribution of wealth across households and between households and experts. In this section we analyze instead whether different wealth distributions modify the transmission of aggregate shocks.

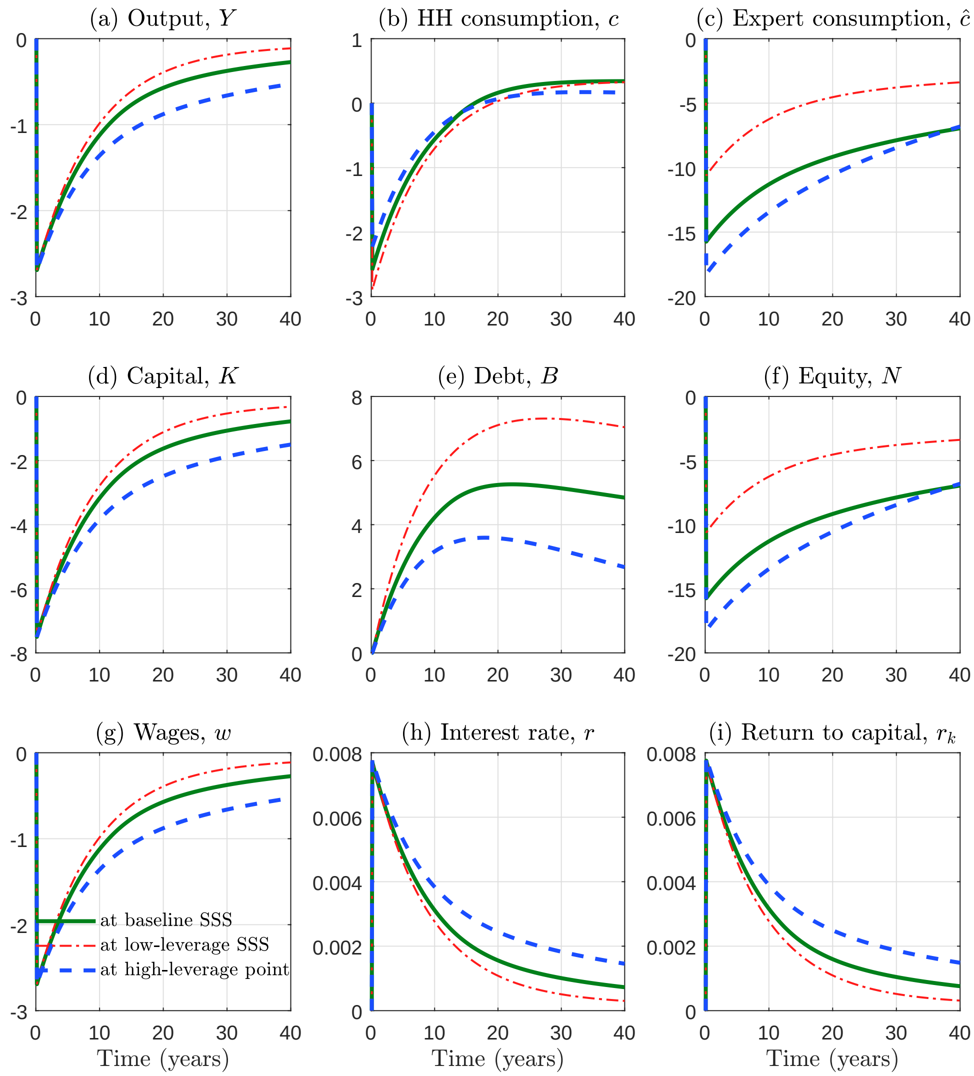

Figure 11: Generalized impulse response functions for different initial states.

Figure 11 displays the generalized impulse response functions to a negative shock. The generalized impulse response is defined as the difference between the transition path if an initial shock hits the economy and that if no shock arrives. We consider initial states $(g_0(\cdot), N_0)$ in the high- and low-leverage regions and the SSS.

We analyze first the response to a 1 standard deviation shock that hits the economy at the SSS. The shock generates a destruction in capital and equity, as experts absorb all the risk. The reduction in capital produces a contraction in output and a decline in wages. The lower level of capital increase the return on capital and the risk-free interest rate. Despite the fact that the output contraction necessarily reduces households' total income, its effects are going to be asymmetric due to the heterogeneity in asset holdings. Poor households will undoubtedly loose, as they see their wages decline. However, the increase in interest rates benefits wealthy households by raising their capital gains.

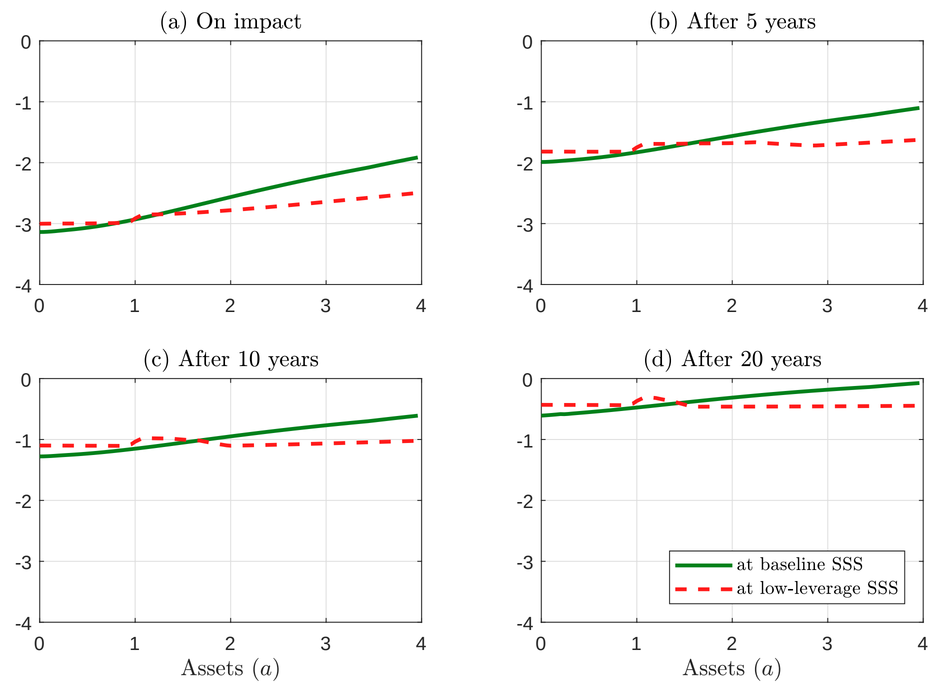

Figure 12: Consumption decision rules at different points in time after the shock.

Figure 12 displays the consumption decision rules for high-productivity households ($z = z_2$) (results for low-income households are qualitatively similar). It shows how after the shock has hit the economy all households reduce their consumption –compared to the transition path–, but poorer households do so by a larger amount, reflecting their lower income. This explains the decline in aggregate consumption displayed in Figure 11. As all households reduce their consumption, those wealthy enough will experience an increase in their asset holdings. This can also be seen in Figure 13. Households with assets levels around 1 decline as they increase their wealth in the around 2. This helps households to be less dependent on reduced wage income and to enjoy the rise in interest rates. At the aggregate level, this pattern explain the dynamics of aggregate debt.

Figure 13 displays what we have called the "distributional impulse response" (DIRF). It is just the generalized impulse response of the wealth density $g_t(\cdot)$, that is, the difference

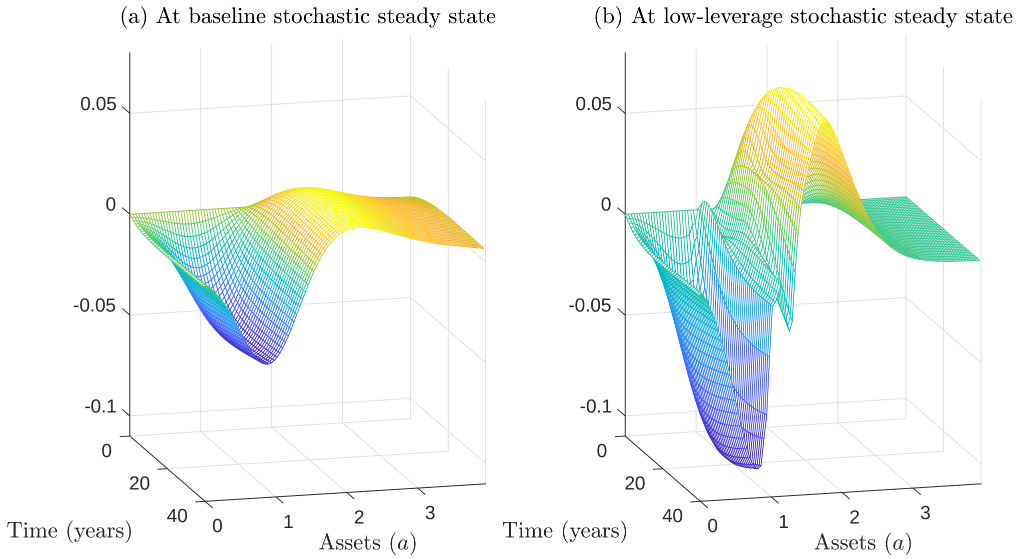

Figure 13: Distributional impulse response at the SSS.

between the density after the shock and that in the transition path (this is the aggregate density across income levels). The DIRF displays the progressive transition of households with wealth levels between 0 and 1 to a new level above $a > 1.5$.

The generalized impulse responses in Figure 14 also show how the response of the economy to the same aggregate shock may be strongly modified depending on the initial wealth allocation. In particular, in the case of a high-leverage economy, the persistence of the shock is greatly amplified. This is because in a high-leverage economy the decline in consumption in the first decade after the shock is less pronounced. The comparison of the time-varying consumption policies in Figure 12 shows how this is mainly driven by wealthy households. As a consequence the path of wealth accumulation is slower, beginning to decline after the initial decade, approximately.

The lower level of household's debt compared to the SSS case discussed above implies a lower capital level and thus less output. Therefore households with assets below 1 are forced to consume less that in the SSS response after the initial decade, as displayed in Figure (13). This also explains the progressive increase in the share of low-wealth households after three decades, as the recovery is sluggish compared to the SSS case.

Notice that aggregate dynamics help to explain why wealthy households reduce less their consumption. it is precisely because they expect a sluggish recovery, and thus higher interest rates that they decide to consume relatively more.

The previous analysis confirms the strong nonlinearity of the model responses in terms of its initial states. However, the model is approximately linear for small perturbations in

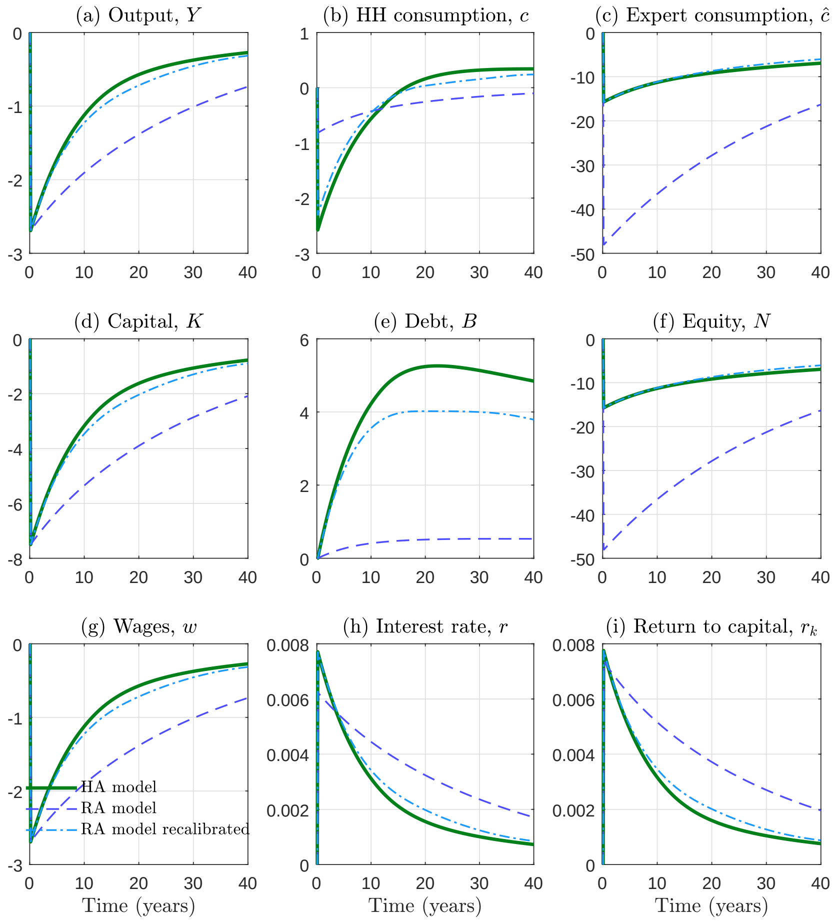

Figure 14: GIRFS, heterogeneous agent vs. representative agent.

the vicinity of the SSS (not shown). Besides, the impulse response in the case of a low-leverage economy is almost identical to that in the SSS as individual policies do not change significatively compared to the SSS.

8 Conclusion

In this paper we have presented a "proof of concept" of how to efficiently compute and estimate a continuous-time model of financial frictions and the wealth distribution. For the computation, we have exploited tools borrowed from machine learning. For the estimation, we have built on contributions from inference with diffusions.

References

Achdou, Y., J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll (2017): "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach," Working Paper 23732, National Bureau of Economic Research.

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

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

Ahn, S., G. Kaplan, B. Moll, T. Winberry, and C. Wolf (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., O. Allais, and W. J. Den Haan (2008): "Solving Heterogeneous-Agent Models with Parameterized Cross-Sectional Distributions," *Journal of Economic Dynamics and Control*, 32(3), 875–908.

Algan, Y., O. Allais, W. J. Den Haan, and P. Rendahl (2014): "Solving and Simulating Models with Heterogeneous Agents and Aggregate Uncertainty," in *Handbook of Computational Economics*, ed. by K. Schmedders, and K. L. Judd, vol. 3, pp. 277–324. Elsevier.

Alvaredo, F., L. Chancel, T. Piketty, E. Saez, and G. Zucman (2017): "Global Inequality Dynamics: New Findings from WID.world," Working Paper 23119, National Bureau of Economic Research.

Andersen, T. G., and J. Lund (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.

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 L. Landi (1995): "Non-Parametric versus Linear Learning Devices: a Procedural Perspective," Discussion paper, University of Florence.

Basak, S., and D. Cuoco (1998): "An Equilibrium Model with Restricted Stock Market Participation," *Review of Financial Studies*, 11(2), 309–341.

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

Bernanke, B. S., M. Gertler, and S. Gilchrist (1999): "The financial accelerator in a quantitative business cycle framework," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1, pp. 1341–1393. Elsevier.

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

Blanchard, O., and J. Galí (2010): "Labor Markets and Monetary Policy: A New Keynesian Model with Unemployment," *American Economic Journal: Macroeconomics*, 2(2), 1–30.

Brumm, J., and S. Scheidegger (2017): "Using Adaptive Sparse Grids to Solve High-Dimensional Dynamic Models," *Econometrica*, 85(5).

Brunnermeier, M., and Y. Sannikov (2016): "Macro, Money, and Finance: A Continuous-Time Approach," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and H. Uhlig, vol. 2, pp. 1497–1545. Elsevier.

Brunnermeier, M. K., and Y. Sannikov (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 *Computational Methods for the Study of Dynamic Economies*, ed. by R. Marimón, and A. Scott, pp. 172–194. Oxford University Press.

Chacko, G., and L. M. Viceira (2003): "Spectral GMM Estimation of Continuous-Time Processes," *Journal of Econometrics*, 116(1), 259–292.

Cho, I.-K., and T. J. Sargent (1996): "Neural Networks for Encoding and Adapting in Dynamic Economies," in *Handbook of Computational Economics*, ed. by H. A. D. K. J. Rust, vol. 1, pp. 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 P. Rendahl (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 L. Epstein (1992): "Stochastic Differential Utility," *Econometrica*, 60(2), 353–94.

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

Fernández-Villaverde, J., and J. F. Rubio-Ramírez (2007): "Estimating Macroeconomic Models: A Likelihood Approach," *Review of Economic Studies*, 74(4), 1059–1087.

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

Gornemann, N., K. Kuester, and M. Nakajima (2012): "Monetary Policy with Heterogeneous Agents," Working Papers 12-21, Federal Reserve Bank of Philadelphia.

Hall, R. E., and P. R. Milgrom (2008): "The Limited Influence of Unemployment on the Wage Bargain," *American Economic Review*, 98(4), 1653–74.

He, Z., and A. Krishnamurthy (2012): "A Model of Capital and Crises," *Review of Economic Studies*, 79(2), 735–777.

——— (2013): "Intermediary Asset Pricing," *American Economic Review*, 103(2), 732–70.

Hornik, K., M. Stinchcombe, and H. White (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.

In-Koo, C. (1995): "Perceptrons Play the Repeated Prisoner's Dilemma," *Journal of Economic Theory*, 67(1), 266–284.

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

Krusell, P., and A. A. Smith (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106(5), 867–896.

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., S. Maliar, and F. Valli (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., E. Nakamura, and J. Steinsson (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 C. Thomas (2017): "Bank Leverage Cycles," *American Economic Journal: Macroeconomics*, 9(2), 32–72.

Piazzesi, M., and M. Schneider (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.

——— (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., G. E. Hinton, and R. J. Williams (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 *Learning and Rationality in Economics*, ed. by A. P. Kirman, and M. Salmon, pp. 236–275. Blackwell Publishing.

Scheidegger, S., and I. Bilionis (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.

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, Han, Lasry, Lions, and Moll (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

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

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

$$s_{i,j,l,m} \equiv w_{l,m} z_i + r_{l,m} a_j - c_{i,j,l,m}, \tag{37}$$

where

$$\begin{aligned}
c_{i,j,l,m} &= \left[\frac{\partial v_{i,j,l,m}}{\partial a}\right]^{-1/\gamma}, &(38)\\
w_{l,m} &= (1-\alpha) Z (B_l + N_m)^{\alpha}, &(39)\\
r_{l,m} &= \alpha Z (B_l + N_m)^{\alpha-1} - \delta - \sigma^2 \frac{(B_l + N_m)}{N_m}. &(40)
\end{aligned}$$

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

$$\begin{aligned}
\frac{v^{n+1}_{i,j,l,m} - v^{n}_{i,j,l,m}}{\Delta} + \rho v^{n+1}_{i,j,l,m} &= \frac{(c^n_{i,j,l,m})^{1-\gamma}}{1-\gamma} + \partial_f v^{n+1}_{i,j,l,m} s^n_{i,j,l,m,f} \mathbf{1}_{s^n_{i,j,n,m,f}>0} + \partial_B v^{n+1}_{i,j,l,m} s^n_{i,j,l,m,b} \mathbf{1}_{s^n_{i,j,l,m,b}<0}\\
&+ \lambda_i \left(v^{n+1}_{-i,j,l,m} - v^{n+1}_{i,j,l,m}\right) + h_{l,m} \partial_B v_{i,j,l,m} + \mu^N_{l,m} \partial_N v_{i,j,l,m}\\
&+ \frac{\left[\sigma^N_{l,m}\right]^2}{2} \partial^2_{NN} v_{i,j,l,m}
\end{aligned}$$

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

$$\begin{aligned}
h_{l,m} &\equiv h(B_l, N_m),\\
\mu^N_{l,m} &\equiv \mu^N(B_l, N_m) = \alpha Z (B_l + N_m)^{\alpha} - \delta (B_l + N_m) - r_{l,m} B_l - \widehat{\rho} N_m,\\
\sigma^N_{l,m} &\equiv \sigma^N(B_l, N_m) = \sigma (B_l + N_m),\\
s^n_{i,j,l,m,f} &= w_{l,m} z_i + r_{l,m} a_j - \left[\frac{1}{\partial^n_f v_{i,j,l,m}}\right]^{1/\gamma},\\
s^n_{i,j,l,m,b} &= w_{l,m} z_i + r_{l,m} a_j - \left[\frac{1}{\partial^n_b v_{i,j,l,m}}\right]^{1/\gamma}
\end{aligned}$$

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

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

$$\begin{aligned}
\frac{v^{n+1}_{i,j,l,m} - v^{n}_{i,j,l,m}}{\Delta} + \rho v^{n+1}_{i,j,l,m} &= \frac{(c^n_{i,j,n,m})^{1-\gamma} - 1}{1-\gamma} + v^{n+1}_{i,j-1,l,m} \alpha^n_{i,j,l,m} + v^{n+1}_{i,j,l,m} \beta^n_{i,j,l,m} + v^{n+1}_{i,j+1,l,m} \xi^n_{i,j,l,m}\\
&+ \lambda_i v^{n+1}_{-i,j,l,m} + v^{n+1}_{i,j,l+1,m} \frac{h_{l,m}}{\Delta B} + v^{n+1}_{i,j,l,m+1} \varkappa_{l,m} + v^{n+1}_{i,j,l,m-1} \varrho_{l,m} \qquad (41)
\end{aligned}$$

where

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

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

In equation (41), the optimal consumption is set to

$$
c_{i,j,n,m}^n = \left(\partial v_{i,j,l,m}^n\right)^{-1/\gamma}. \tag{42}
$$

where

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

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

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

We define

$$
\mathbf{A}_{l.m}^{n} = \begin{bmatrix}
\beta_{1,1,l,m}^{n} & \xi_{1,1,l,m}^{n} & 0 & 0 & \cdots & 0 & \lambda_1 & 0 & \cdots & 0 \\
\alpha_{1,2,l,m}^{n} & \beta_{1,2,l,m}^{n} & \xi_{1,2,l,m}^{n} & 0 & \cdots & 0 & 0 & \lambda_1 & \ddots & 0 \\
0 & \alpha_{1,3,l,m}^{n} & \beta_{1,3,l,m}^{n} & \xi_{1,3,l,m}^{n} & \cdots & 0 & 0 & 0 & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & \alpha_{1,J-1,l,m}^{n} & \beta_{1,J-1,l,m}^{n} & \xi_{1,J-1,l,m}^{n} & 0 & \cdots & \lambda_1 & 0 \\
0 & 0 & \cdots & 0 & \alpha_{1,J,l,m}^{n} & \beta_{1,J,l,m}^{n} & 0 & 0 & \cdots & \lambda_1 \\
\lambda_2 & 0 & \cdots & 0 & 0 & 0 & \beta_{2,1,l,m}^{n} & \xi_{2,1,l,m}^{n} & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & \lambda_2 & 0 & \cdots & \alpha_{2,J,l,m}^{n} & \beta_{2,J,l,m}^{n}
\end{bmatrix},
$$

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

and

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

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

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

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

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

Then, equation (41) 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

$$
\mathbf{B}^n\mathbf{v}^{n+1} = \mathbf{d}^n \tag{43}
$$

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 (42).
2. Find $v_{i,j,l,m}^{n+1}$ solving the linear system of equations (43).
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 Kolmogorov Forward 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:

$$\frac{\partial g}{\partial t} = -\frac{\partial}{\partial a}\left[s_i\left(a, B, N\right) g_{i,t}(a)\right] - \lambda_i g_{i,t}(a) + \lambda_{-i} g_{-i,t}(a), \ i = 1, 2. \quad (44)$$

$$1 = \int_0^\infty g(a)da. \quad (45)$$

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

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

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. Chebyshev polynomials

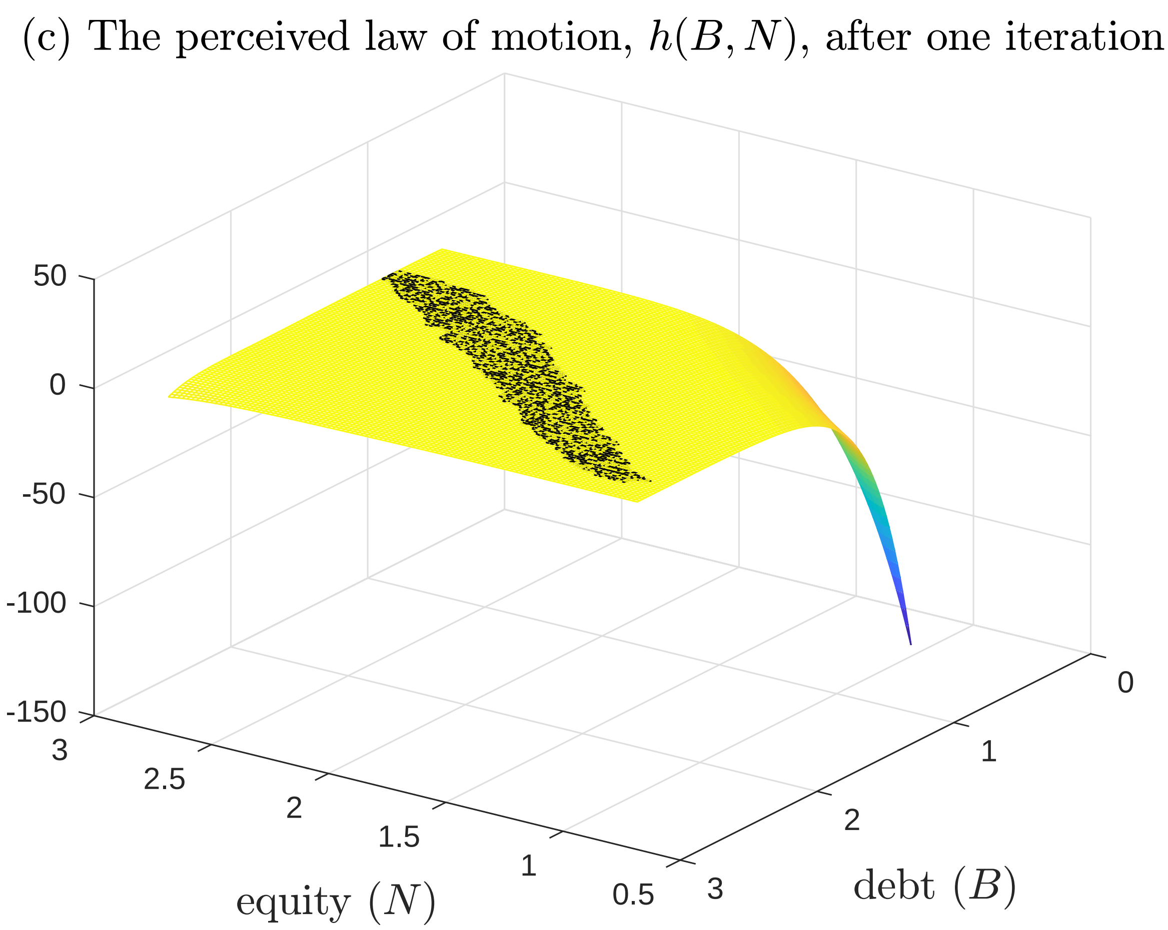

Figure 15: PLM with Chebyshev polynomials.

(c) The perceived law of motion, $h(B, N)$, after one iteration

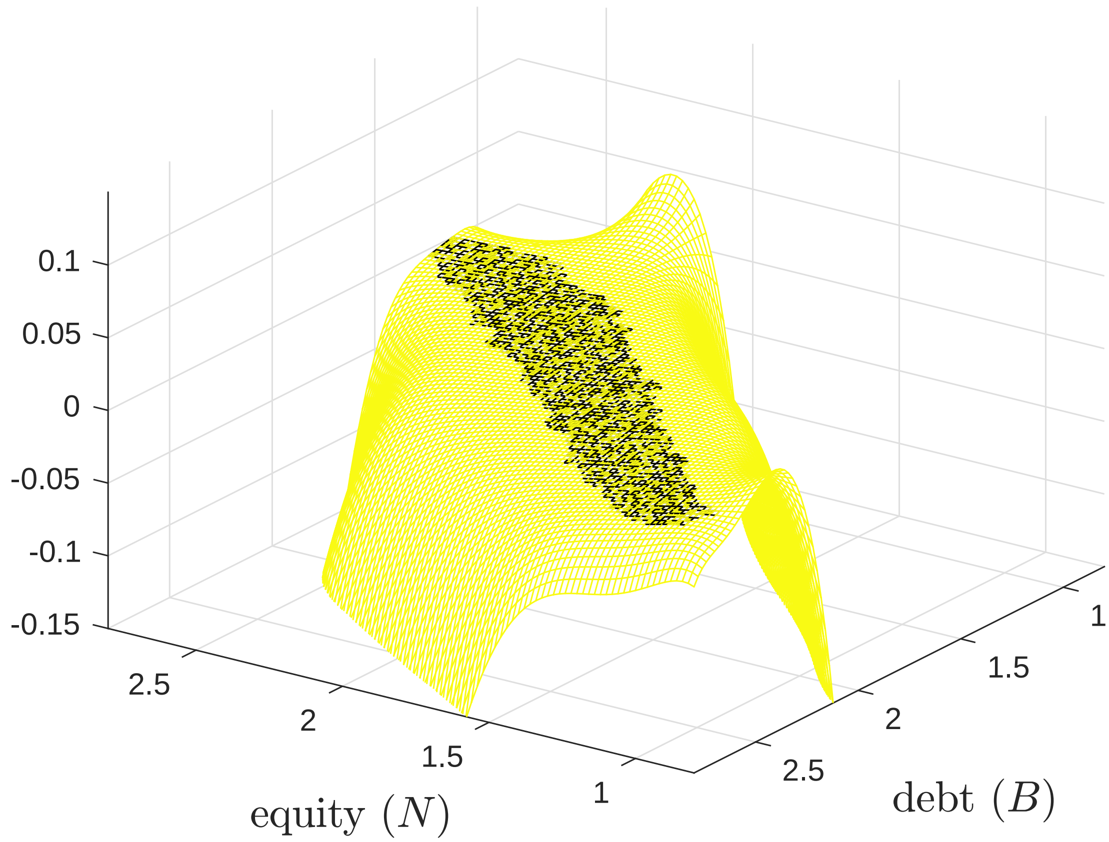

Figure 16: PLM with Chebyshev polynomials (zoom).