

# UC San Diego

## Recent Work

### Title

Pareto Extrapolation: Bridging Theoretical and Quantitative Models of Wealth Inequality

### Permalink

<https://escholarship.org/uc/item/90n2h2bb>

### Authors

Gouin-Bonenfant, Emilien  
Toda, Alexis Akira

### Publication Date

2018-10-05

# Pareto Extrapolation: Bridging Theoretical and Quantitative Models of Wealth Inequality\*

Émilien Gouin-Bonenfant<sup>†</sup> and Alexis Akira Toda<sup>‡</sup>

Department of Economics, University of California San Diego

November 13, 2018

## Abstract

We propose a new, systematic approach for analyzing and solving heterogeneous-agent models with fat-tailed wealth distributions. Our approach exploits the asymptotic linearity of policy functions and the analytical characterization of the Pareto exponent to make the solution algorithm more transparent, efficient, and accurate with zero additional computational cost. As an application, we solve a heterogeneous-agent model that features persistent earnings and investment risk, borrowing constraint, portfolio decision, and endogenous Pareto-tailed wealth distribution. We show that relaxing the borrowing limit from 25% of annual income to 250% increases inequality by reducing the bottom 50% wealth share from 11% to 6.7% and decreases welfare by 8.2% in consumption equivalent.

**Keywords:** asymptotic linearity, Bewley-Huggett-Aiyagari model, Pareto exponent, power law, solution accuracy.

**JEL codes:** C63, D31, D58, E21.

## 1 Introduction

Macroeconomic models increasingly incorporate heterogeneity. Doing so allows researchers to identify who gains and who loses from new policies, but also assess how the effectiveness of policies depend on the nature of heterogeneity. Since the first generation of heterogeneous-agent models such as [Huggett \(1993\)](#) and [Aiyagari \(1994\)](#), one challenge has been to generate a realistic wealth distribution. Empirically, it is well known since [Pareto \(1895, 1896, 1897\)](#)'s seminal work that the wealth distributions obey the power law: the fraction of agents with wealth  $w$  or larger decays like a power function  $w^{-\zeta}$ , where  $\zeta$  is called the Pareto exponent. Thus successful economic models of wealth inequality should endogenously generate fat-tailed wealth distributions.

---

\*We thank Mark Bilal, Fedor Iskhahov, Narayana Kocherlakota, Alisdair McKay, John Stachurski, and seminar participants at Australian National University and Rochester for comments.

<sup>†</sup>Email: [egouinbo@ucsd.edu](mailto:egouinbo@ucsd.edu).

<sup>‡</sup>Email: [atoda@ucsd.edu](mailto:atoda@ucsd.edu).

We are now much closer to understanding the economic forces that determine wealth inequality, and many models have been proposed that can account for the extreme concentration of wealth we observe in the data. Two parallel literatures have emerged. The first studies relatively simple models and provides theoretical characterizations of the power law behavior of the wealth distribution, often relying on analytical solutions. The second studies rich general equilibrium models and conducts quantitative analysis and experiments that rely heavily on numerical methods. However, there has always been the trade-off between analytical tractability and the richness of models. The former requires strong (and oftentimes unrealistic) assumptions, while the latter requires numerical methods, which are in general not well-suited for studying the tail behavior of the wealth distribution because models are commonly solved on a finite grid and hence misses the top tail *by definition*. What is lacking in the current literature is a systematic approach for analyzing and solving heterogeneous-agent models that (potentially) generate fat-tailed wealth distributions but do not admit closed-form solutions.

In this paper, we propose a simple, systematic approach for tackling heterogeneous-agent models with fat-tailed wealth distributions numerically. Our approach consists of (i) the “asymptotic analysis” of the individual optimization problem to compute the Pareto exponent of the wealth distribution and (ii) the “Pareto extrapolation” of the wealth grid to compute the equilibrium and wealth distribution accurately. Our approach enables researchers to analyze rich heterogeneous-agent models that feature persistent earnings and investment risk, borrowing constraint, meaningful portfolio choices, recursive utility, and endogenous Pareto wealth distributions, etc., which have been intractable with existing methods.

In the “asymptotic analysis” step we solve, given the candidate equilibrium object (risk-free rate, wage, etc.), a simplified, or “asymptotic” individual optimization problem semi-analytically. Roughly speaking, this problem ignores all additive elements and focuses on proportional elements. For example, consider the income fluctuation problem, which is a building block of Bewley-Huggett-Aiyagari models. The asymptotic problem in this case is one with no income (i.e., consumption is financed only through savings), which can be solved analytically as in [Merton \(1969\)](#) and [Samuelson \(1969\)](#). The benefit of studying the asymptotic problem is threefold. First, its solution determines the behavior of rich agents, which governs the tail property of the wealth distribution. This enables researchers to determine whether the model generates a fat-tailed wealth distribution, and if so, to compute the theoretical Pareto exponent explicitly. This is important because a successful model of inequality should endogenously generate a fat-tailed wealth distribution, and knowing the Pareto exponent allows one to calculate measures of inequality such as the top 1% wealth share. Second, the analysis of the asymptotic problem places parametric restrictions on the equilibrium object through equilibrium considerations such as the existence of a solution to the individual optimization problem and a wealth distribution with a finite mean. This enables researchers to narrow down the set of equilibrium object and search for the equilibrium more efficiently. Third, the solution to the asymptotic problem can be used as an initial guess for solving the actual individual optimization problem, making the algorithm more efficient and stable.

In the “Pareto extrapolation” step, we extrapolate the wealth distribution off the grid using the theoretical Pareto exponent computed from the asymptotic analysis to correct for the

truncation error when constructing the transition probability matrix governing the state variables and computing aggregate quantities from the actual optimization problem. The benefit of Pareto extrapolation is twofold. First, it makes the solution more accurate at no additional computational cost. This is because the correction terms take care of the truncation error, and these terms are introduced only at the largest grid point, which is negligible compared with the total number of grid points. Second, and more importantly, Pareto extrapolation enables researchers to avoid making mistakes. While it is true that we can solve models to any accuracy if we use sufficiently large and fine grids and sufficiently strong computing power, with existing methods one can never be sure whether the truncation error is small enough. As an illustration, suppose some researcher says “I truncate the grid so that there is less than (say)  $10^{-4}$  of the probability mass at the top grid point”. This is not a good idea because (i) the mass at the largest grid point is severely biased downwards with existing methods, and (ii) even with mass  $10^{-4}$  at the largest grid point, there can be substantial amount of wealth held by those agents.

In summary, our approach makes the solution and analysis of heterogeneous-agent models with fat-tailed wealth distributions (i) more transparent (because it exploits economic theory as much as possible), (ii) more efficient (because it narrows down the equilibrium object and uses good initial guesses), and (iii) more accurate (because it corrects for truncation errors). Furthermore, we achieve all of that with zero additional computational cost because the asymptotic analysis is semi-analytical and the correction terms in the Pareto extrapolation are introduced only at one grid point.

To illustrate the usefulness of our approach, we present two exercises. First, using a simple heterogeneous-agent model that admits a closed-form solution (and a Pareto wealth distribution) as a laboratory, we show that the error with existing methods can be substantial, while it is minimal with our approach. Specifically, we find that the error with our method (with an exponentially-spaced grid with 100 points) ranges from 0.002 to 0.6% depending on the choice of the largest grid point, whereas it is 3.3–21% with the usual (“Truncation”) method that does not introduce correction terms. Simulation-based methods with 10,000 agents also have about 11% of error. The errors in the existing methods are especially severe when the Pareto exponent is smaller than 2, which is typical for wealth (1.5) and firm size (close to 1, Zipf’s law).

Second, we develop a Merton-Bewley-Aiyagari (MBA) model that features persistent idiosyncratic endowment and investment risk, borrowing constraint, portfolio decision, recursive utility, and endogenous Pareto-tailed wealth distribution. To the best of our knowledge, our paper is the first to solve such a complicated model without relying on (restrictive) closed-form solutions, and we obtain tractability through the asymptotic analysis of the individual optimization problem and the Pareto exponent formula. We provide a step-by-step approach to solving the model and use it as a laboratory to conduct a counterfactual experiment. We show that relaxing the borrowing limit from 25% of annual income to 250% increases inequality by reducing the bottom 50% wealth share from 11% to 6.7%, and decreases welfare by 8.2% in consumption equivalent due to increased poverty.

## 1.1 Related literature

Our paper is related to a large literature that spans across many disciplines, including quantitative macroeconomics, economic theory on consumption and portfolio choices, mathematical and statistical results on Pareto tails, and numerical analysis.

It is well-known in the quantitative macroeconomics literature that idiosyncratic unemployment risk and incomplete financial markets alone are insufficient to generate a sufficiently dispersed wealth distribution (Krueger, Mitman, and Perri, 2016). Recently, Stachurski and Toda (2018) have theoretically proved that in standard Bewley-Huggett-Aiyagari models, the wealth distribution necessarily inherits the tail property of the income distribution, and therefore standard heterogeneous-agent models cannot explain the wealth distribution. They also argue that introducing other ingredients such as random discount factors (Krusell and Smith, 1998), idiosyncratic investment risk (Quadrini, 2000; Cagetti and De Nardi, 2006), and random birth/death (Carroll, Slacalek, Tokuoka, and White, 2017; McKay, 2017) can generate fat tails. However, because these papers are all numerical, it is not clear how to build and solve general heterogeneous-agent models that feature fat-tailed wealth distributions. Our paper contributes to the quantitative macroeconomics literature by showing the usefulness of the theoretical analysis of the asymptotic problem and providing a general solution algorithm for such models.

As mentioned in the introduction, since numerical methods are in general not well-suited for studying the tail behavior of the wealth distribution, most papers that study the power law behavior in the wealth distribution use analytical solutions. Nirei and Souma (2007) and Benhabib, Bisin, and Zhu (2011) solve growth models with idiosyncratic investment risk and use the properties of Kesten (1973) processes to obtain a Pareto wealth distribution. Moll (2014), Toda (2014), Arkolakis (2016), Benhabib, Bisin, and Zhu (2016), and Nirei and Aoki (2016) consider stochastic birth/death and obtain the double Pareto wealth distribution based on the mechanism of Reed (2001).<sup>1</sup> For reviews of generative mechanisms of Pareto tails used in these papers, see Gabaix (2009). Our paper bridges this literature on power law in economics and quantitative macroeconomics by showing that the theoretical insight carries over to rich quantitative models.

The usefulness of the asymptotic problem for computing the Pareto exponent in general models that admit no closed-form solutions was pointed out by Toda (2018b).<sup>2</sup> However, Toda (2018b) does not consider the solution algorithm for general equilibrium models with fat-tailed wealth distributions. The asymptotic linearity of consumption policies has been known for a long time since at least Huggett (1993) and Krusell and Smith (1998), among others. Benhabib, Bisin, and Zhu (2015) show the asymptotic linearity in an i.i.d. environment and obtain a Pareto lower bound for the wealth distribution. In Appendix A, we argue that similar results should hold in richer models. To analytically characterize the Pareto exponent of the wealth

---

<sup>1</sup>Other recent applications include firm dynamics (Acemoglu and Cao, 2015), asset pricing (Toda and Walsh, 2015, 2017), dynamics of inequality (Gabaix, Lasry, Lions, and Moll, 2016; Aoki and Nirei, 2017; Cao and Luo, 2017; Kasa and Lei, 2018), bequests (Zhu, 2018), and entrepreneurship (Jones and Kim, 2018).

<sup>2</sup>The asymptotic problem is related to the “method of moderation” in Carroll, Tokuoka, and Wu (2012), who bound the consumption policy function from above and below by closed-form solutions to improve accuracy and stability.

distribution in a general Markovian environment, we apply the recent results from [Beare and Toda \(2017\)](#).

Our paper is also related to the literature on solution methods for heterogeneous-agent models such as [Krusell and Smith \(2006\)](#), [Algan, Allais, and Den Haan \(2008\)](#), [Reiter \(2009, 2010\)](#), [Den Haan \(2010a,b\)](#), and [Algan, Allais, Den Haan, and Rendahl \(2014\)](#), among others. In particular, we use the insight from [Algan, Allais, and Den Haan \(2008\)](#) and [Winberry \(2018\)](#), who approximate cross-sectional distributions using finite-dimensional parametric families. In our case, because economic theory suggests that the upper tail of the wealth distribution is Pareto and it is possible to compute the Pareto exponent from the solution to the asymptotic problem, we use this Pareto distribution to approximate the upper tail. Although we do not take a stance on how to deal with the rest of the distribution, we use [Young \(2010\)](#)'s non-stochastic simulation to compute the wealth distribution from the transition probability matrix arising from the law of motion.

The closest paper to ours in spirit is [Achdou, Han, Lasry, Lions, and Moll \(2017\)](#). They recast the Bewley-Huggett-Aiyagari model in continuous-time, which allows them to obtain a number of novel characterizations and results, including closed-form expressions for the stationary wealth distribution (in a special case) and the marginal propensity to consume of agents close to the borrowing constraint. They also prove that a stationary equilibrium exists and is unique when the intertemporal elasticity of substitution is weakly above one. Finally, they leverage finite-difference methods and propose a fast solution algorithm that can be applied to much more general heterogeneous-agent models in continuous time. While our paper is different—we focus on the complications arising with fat-tailed wealth distributions—we share the same goal of bridging the gap between theoretical and quantitative work in macroeconomics.

## 2 Solving heterogeneous-agent models with Pareto tails

In this section we propose a new solution algorithm for heterogeneous-agent models with fat-tailed wealth distributions based on Pareto extrapolation. We first point out the issues with existing solution algorithms, and then outline our new method.

### 2.1 Issues with existing algorithms

Suppose that we want to solve a [Bewley \(1977, 1983\)](#)-[Huggett \(1993\)](#)-[Aiyagari \(1994\)](#) model numerically when the wealth distribution could be fat-tailed. The conventional solution algorithm (which we refer to as the “Truncation” method throughout the paper) would be roughly as follows.

1. The researcher sets up a finite grid for wealth denoted by  $\mathcal{W}_N = \{w_n\}_{n=1}^N$ , where  $N$  is the number of grid points and  $w_1 < \dots < w_N$ . Suppose there are also other exogenous state variables (e.g., income, return on wealth, etc.), which can take  $S$  possible values indexed by  $s = 1, \dots, S$ . Given the guess of the equilibrium object (e.g., interest rate, wage, etc.), we can solve the individual optimization problem on the  $S \times N$  grid using dynamic programming.

2. Having solved the individual optimization problem and obtained the law of motion for individual wealth, the researcher constructs the  $SN \times SN$  transition probability matrix  $P$  of all state variables. The stationary distribution  $\pi \in \mathbb{R}_+^{SN}$  is obtained by solving  $P'\pi = \pi$  (so  $\pi$  is an eigenvector of  $P'$  corresponding to the eigenvalue 1).
3. Finally, the researcher imposes the market clearing condition by integrating the individual decision rules (capital, labor, etc.) over the grid using the stationary distribution  $\pi$  to find the equilibrium objects (interest rate, wage, etc.).

There are two potential issues with this truncation method when the stationary wealth distribution is fat-tailed, both of which are related. First, consider the largest grid point  $w_N$ . This grid point in principle does not represent just the point  $w = w_N$ , but the entire interval  $w \in [w_N, \infty)$ . Therefore when we construct the transition probability from  $w_N$  to other grid points, instead of assuming that the current wealth state  $w$  is concentrated at  $w_N$ , we need to take into account that  $w$  is really distributed over the interval  $[w_N, \infty)$  according to the (true) stationary distribution. Since the interval  $[w_N, \infty)$  contains substantial probability mass when the wealth distribution is fat-tailed, failing to account for this will overestimate the transition probability to lower wealth states, and hence underestimate the top tail probability.

Second, suppose that we use the stationary distribution  $\pi = (\pi_{sn})$  to compute aggregate quantities used in market clearing conditions. For concreteness, consider the aggregate wealth

$$W = \sum_{s=1}^S \sum_{n=1}^N \pi_{sn} w_n. \quad (2.1)$$

The right-hand side of (2.1) essentially supposes that the top tail is concentrated on the grid point  $w_N$ , whereas in fact it is distributed over the interval  $[w_N, \infty)$ . Thus failing to account for this will underestimate the aggregate wealth, which affects the computation of equilibrium through market clearing conditions.

Of course, one may choose a very large truncation point  $w_N$  (say, one million times the aggregate wealth) to reduce the truncation error, but that is computationally inefficient because it will either increase the number of grid points (making the solution algorithm slower) or decrease the grid density (making the solution less accurate). As we show in Section 3, the solution accuracy of the usual truncation method is poor when the wealth distribution is fat-tailed, even if we take a very large truncation point.

One may also argue that the above two issues are specific to the particular algorithm that involves truncation, and other methods such as simulation (Aiyagari, 1994; Krusell and Smith, 1998) may not be subject to those issues. As we see below, however, the situation is equally problematic. Simulation-based methods essentially use the law of large numbers to approximate the market clearing condition. Suppose we simulate  $I$  agents and compute the sample mean of wealth  $\frac{1}{I} \sum_{i=1}^I w_i$ . The question is how fast the sample mean converges to the population mean. If the Pareto exponent  $\zeta$  exceeds 2, then wealth has finite variance and we can apply the Central Limit Theorem. In this case the sample mean converges at rate  $I^{1/2}$ . If  $\zeta < 2$  on the other hand, it is well-known that the rate of convergence to the stable law is only  $I^{1-1/\zeta}$ .<sup>3</sup>

<sup>3</sup>See, for example, Durrett (2010, Theorem 3.7.2) for an accessible proof. Based on this insight, Gabaix (2011)

Therefore solving a model accurately may require an impractically large number of agents.

As an illustration, Table 1 shows the order of error  $I^{\max\{-1/2, 1/\zeta-1\}}$  in the sample mean for various sample size  $I$  and Pareto exponent  $\zeta$ .<sup>4</sup> If  $\zeta \geq 2$  and we use 10,000 agents (the number used in Aiyagari (1994)), then the order of the error in the sample mean is  $10000^{-1/2} = 1/100 = 1\%$ . However, the error order is much larger if the Pareto exponent is smaller. With  $\zeta = 1.5$  (a typical number for the wealth distribution according to Pareto (1897), Klass, Biham, Levy, Malcai, and Solomon (2006), and Vermeulen (2018)), the error order with 10,000 agents is 4.6%, which is substantial. If the Pareto exponent is 1.1 (a typical number for the firm size distribution, which obeys Zipf’s law (Axtell, 2001)), then even with ten billion agents ( $I = 10^{10}$ ), which is about the same order of magnitude as the world population, the error order is still 12.3%. To drive the error down to 1%, quite a modest number, the required sample size for  $\zeta = 1.1$  is  $I = 100^{\frac{\zeta}{\zeta-1}} = 10^{22}$  (ten sextillion), which is about the same order of magnitude as the number of stars in the universe or sand grains on earth.<sup>5</sup> Therefore we cannot expect to solve such models accurately using simulation.

Table 1: Order of error  $I^{\max\{-1/2, 1/\zeta-1\}}$  in sample mean.

Sample size $I$	Pareto exponent $\zeta$			
	$\geq 2$	1.5	1.3	1.1
$10^0 = 1$	1.00000	1.00000	1.00000	1.00000
$10^2$	0.10000	0.21544	0.34551	0.65793
$10^4$	0.01000	0.04642	0.11938	0.43288
$10^6$	0.00100	0.01000	0.04125	0.28480
$10^8$	0.00010	0.00215	0.01425	0.18738
$10^{10}$	0.00001	0.00046	0.00492	0.12328

## 2.2 The Pareto extrapolation algorithm

Our new solution algorithm for heterogeneous-agent models, which we call the “Pareto extrapolation” method, differs from the usual “truncation” method only when computing the stationary distribution and aggregating individual behavior. Therefore we focus the description of the algorithm only at this aggregation step.

The Pareto extrapolation method consists of three main sub-steps that correct the truncation errors in the standard algorithm:

### The Pareto extrapolation algorithm.

1. Solve an “asymptotic” individual optimization problem semi-analytically and compute the theoretical Pareto exponent  $\zeta$ .

argues that a substantial fraction of aggregate fluctuations is due to idiosyncratic shocks to large firms.

<sup>4</sup>In Table 12 in Appendix D, we assess the accuracy of the Aiyagari model in Section 3 using simulation and obtain similar results to Table 1.

<sup>5</sup><http://www.abc.net.au/science/articles/2015/08/19/4293562.htm>



2. Construct the  $SN \times SN$  transition probability matrix by approximating the wealth distribution for  $w \geq w_N$  by a Pareto distribution with exponent  $\zeta$ .
3. Compute the aggregate wealth by approximating the wealth distribution for  $w \geq w_N$  by a Pareto distribution with exponent  $\zeta$ .

Below, we explain each step in more detail.

### 2.2.1 Computing the theoretical Pareto exponent

Our method uses the theoretical Pareto exponent  $\zeta$  to correct the truncation errors. Thus the first step is to compute the theoretical Pareto exponent of the wealth distribution implied by individual behavior.

For this purpose we can use the insight from [Toda \(2018b\)](#). Since the tail property of the wealth distribution depends on the behavior of wealthy agents, and for those agents labor income is negligible compared to capital income (because labor income enters additively to the budget constraint, whereas capital income is proportional to wealth), we can consider a simplified problem where the labor income is zero. Assuming that agents have homothetic preferences (e.g., additive CRRA, Epstein-Zin, etc.), which is almost always the case in applications, this simplified problem becomes a homogeneous problem in the sense that all control variables scale with wealth. We refer to this problem as the “asymptotic” problem. Such problems can be solved semi-analytically even in a Markovian (non-i.i.d.) environment as shown by [Toda \(2014, Theorem 5\)](#), and the decision rules become linear in wealth. ([Appendix A](#) formally defines the asymptotic problem and discusses the asymptotic linearity of policy functions in an abstract dynamic programming setting.)

For concreteness, suppose that agents solve an optimal consumption-savings problem of the form

$$\text{maximize} \quad E_0 \sum_{t=0}^{\infty} [\beta(1-p)]^t \frac{c_t^{1-\gamma}}{1-\gamma} \quad (2.2a)$$

$$\text{subject to} \quad w_{t+1} = R_{s_t}(w_t - c_t + y_{s_t}), \quad (2.2b)$$

$$w_t \geq \underline{w}. \quad (2.2c)$$

Here  $\beta > 0$  is the discount factor,  $p \in [0, 1)$  is the birth/death probability (infinitely-lived case corresponds to  $p = 0$ ),  $\gamma > 0$  is the relative risk aversion,  $c_t$  is consumption,  $w_t$  is wealth at the beginning of period  $t$  excluding current labor income,  $s_t$  is some Markov state,  $y_s$  is income in state  $s$ ,  $R_s > 0$  is the gross return on wealth in state  $s$ , and  $\underline{w}$  is minimum wealth. By definition, the asymptotic problem studies the limiting case when  $w \rightarrow \infty$ . In this case, income  $y$  and minimum wealth are negligible, so we replace the budget constraint (2.2b) and

borrowing constraint (2.2c) by

$$w_{t+1} = R_{s_t}(w_t - c_t), \quad (2.3a)$$

$$w_t \geq 0, \quad (2.3b)$$

respectively. Note that the problem is now homogeneous because the utility function is homothetic: an agent twice as rich will consume twice as much, state-by-state. We can maximize a homothetic function subject to homogeneous constraints of the form (2.3) semi-analytically quite efficiently, as explained in Toda (2014) in detail. After solving this problem, the law of motion for wealth becomes linear,  $w' = G_s w$  for some gross growth rate  $G_s > 0$ .

Now let us go back to the general case and write the law of motion of the asymptotic problem as

$$w_{i,t+1} = G_{i,t+1} w_{it},$$

where  $G_{i,t+1} > 0$  is the gross growth rate of wealth between time  $t$  and  $t + 1$  for individual  $i$  and  $w_{it}$  is wealth. Thus, in the asymptotic problem, the law of motion for wealth necessarily satisfies **Gibrat (1931)**'s law of proportional growth. Assuming that agents enter/exit the economy at constant probability  $p > 0$ , **Beare and Toda (2017)** show that under mild conditions the stationary wealth distribution has a Pareto upper tail and characterize the Pareto exponent  $\zeta$ , as follows. Suppose that there are finitely many idiosyncratic states (other than wealth) denoted by  $s = 1, \dots, S$ , and let  $P = (p_{ss'})$  be the transition probability matrix, which we assume to be irreducible. For  $z \in \mathbb{R}$ , let

$$D(z) = \text{diag} \left( \mathbb{E} \left[ e^{z \log G_{i,t+1}} \mid s_{it} = 1 \right], \dots, \mathbb{E} \left[ e^{z \log G_{i,t+1}} \mid s_{it} = S \right] \right) \quad (2.4)$$

be the diagonal matrix consisting of the conditional moment generating functions of the log growth rate  $\log G_{i,t+1}$ . For a square matrix  $A$ , let  $\rho(A)$  denote its spectral radius (the maximum modulus of all eigenvalues of  $A$ ). Then under mild conditions **Beare and Toda (2017)** show that the equation

$$\rho(PD(z)) = \frac{1}{1-p} \quad (2.5)$$

has a unique positive solution  $z = \zeta > 0$ , and that the stationary wealth distribution has a Pareto upper tail with exponent  $\zeta$ . **Toda (2018b)** argues that if agents are infinitely lived but there exists a stationary distribution due to other mechanisms than random entry/exit (e.g., borrowing constraint), then we can just set  $p = 0$  in (2.5) to compute the theoretical Pareto exponent.

We can summarize this step as follows.

#### Computing the theoretical Pareto exponent.

1. Verify that the utility function is asymptotically homothetic (e.g., CRRA, HARA, Epstein-Zin) in consumption.
2. Define the asymptotic problem (e.g., no labor income, no borrowing constraint).

3. Solve the asymptotic problem semi-analytically and derive the law of motion  $w' = G_s w$ .
4. Define the diagonal matrix (2.4) of conditional moment generating functions of log growth rates.
5. The theoretical Pareto exponent  $\zeta > 0$  is the solution to (2.5).

See [Toda \(2018b\)](#) for more details about the actual implementation of this step.

## 2.2.2 Constructing the transition probability matrix

Having characterized the theoretical Pareto exponent  $\zeta > 0$ , the next step is to construct the transition probability matrix for all state variables.

Let  $w' = g(w, s)$  be the law of motion for wealth for the original (non-asymptotic) problem, which can be obtained by numerically solving the individual optimization problem using dynamic programming on the grid  $S \times \mathcal{W}_N$ , where  $\mathcal{W}_N = \{w_n\}_{n=1}^N$  is the grid for wealth. Let

$$I_n = [w_n, w_{n+1}), \quad n = 1, \dots, N-1$$

be the half-open interval with endpoints  $w_n$  and  $w_{n+1}$ . Let  $I_N = [w_N, \infty)$ . For  $n = 1, \dots, N$ , let us construct the transition probability as follows.

**Case 1:  $n < N$ .** Take the lower grid point of  $I_n$ , which is  $w_n$ . If  $g(w_n, s) \in I_k$  for some  $k < N$ , then we can take  $\theta \in [0, 1)$  such that

$$g(w_n, s) = (1 - \theta)w_k + \theta w_{k+1} \iff \theta = \theta_{nk} = \frac{g(w_n, s) - w_k}{w_{k+1} - w_k}. \quad (2.6)$$

We can then assign probabilities  $1 - \theta, \theta$  to the grid points  $w_k, w_{k+1}$  (i.e., states  $k$  and  $k + 1$ ), respectively (Figure 1). If  $g(w_n, s) < w_1$  or  $g(w_n, s) \geq w_N$ , then just assign probability 1 to state 1 or  $N$ . (Assigning probabilities to neighboring grid points to match the law of motion this way is essentially the same as what [Young \(2010\)](#) calls “non-stochastic simulation”.)

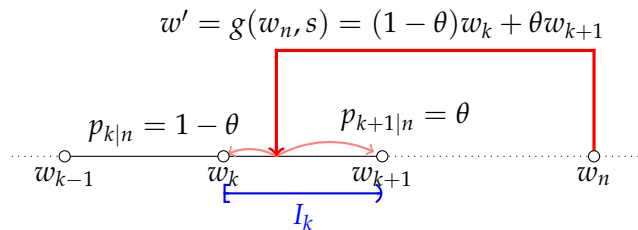


Figure 1: Construction of transition probabilities from a grid point.

**Case 2:  $n = N$ .** Suppose for the moment that there is an untruncated grid  $\mathcal{W}_\infty = \{w_n\}_{n=1}^\infty$ , and for  $n \geq N$  we know the probability of  $w = w_n$  conditional on  $w \in I_N \cap \mathcal{W}_\infty$ . Let this

probability be denoted by  $q_n$ . By definition, we have  $\sum_{n=N}^{\infty} q_n = 1$ . Now for each  $n \geq N$ , we can do precisely as in the previous case, and add probabilities  $(1 - \theta_{nk})q_n$  and  $\theta_{nk}q_n$  (where  $\theta_{nk}$  is defined by (2.6)) to the grid points  $w_k, w_{k+1}$  whenever  $w' = g(w_n, s) \in I_k$  for  $k < N$  (Figure 2). If  $g(w_n, s) < w_1$  or  $g(w_n, s) \geq w_N$ , then just add probability  $q_n$  to the transition to state 1 or  $N$ . The nice thing is that for large enough  $n$ , the next period's state  $w' = g(w_n, s)$  is likely large (contained in  $I_N$ ), so we only need to compute  $\theta_{nk}$  for finitely many  $n$  (say  $n = N, \dots, N'$ ).

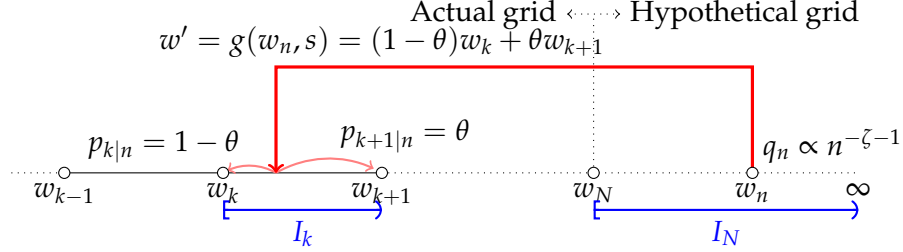


Figure 2: Construction of transition probabilities from a hypothetical grid point.

Now it remains to compute the conditional probability  $q_n$ . Assume that for  $n \geq N$ , the grid spacing  $w_{n+1} - w_n$  is some constant  $h > 0$ . Assuming that the stationary distribution has a Pareto upper tail with exponent  $\zeta > 1$ , we can set  $q_n \propto n^{-\zeta-1}$  for  $n \geq N$ . (If  $\zeta \leq 1$ , then the mean is infinite, which is impossible in equilibrium. In this case we exit from the loop and use a different guess for the equilibrium object.) As mentioned before, for  $n \geq N'$  the next state will always be  $N$  ( $w' = g(w_n, s) \in I_N$ ), so there is no need to compute  $q_n$  individually. Approximating the infinite sum by an integral, we can compute

$$\sum_{n=N'}^{\infty} q_n \propto \sum_{n=N'}^{\infty} (nh)^{-\zeta-1} \approx h^{-\zeta-1} \int_{N'}^{\infty} x^{-\zeta-1} dx = h^{-\zeta-1} \frac{1}{\zeta} (N')^{-\zeta}.$$

Imposing the condition  $\sum_{n=N}^{\infty} q_n = 1$ , we can obtain

$$\begin{cases} q_n = Cn^{-\zeta-1}, & (N \leq n < N') \\ \sum_{n=N'}^{\infty} q_n = \frac{C}{\zeta} (N')^{-\zeta}, \end{cases}$$

where the constant of proportionality  $C$  is given by

$$\frac{1}{C} = \frac{1}{\zeta} (N')^{-\zeta} + \sum_{n=N}^{N'-1} n^{-\zeta-1}.$$

We can summarize this step as follows.

**Constructing the transition probability matrix.**

Let  $\zeta > 1$  be the theoretical Pareto exponent,  $\mathcal{W}_N = \{w_n\}_{n=1}^N$  be the grid, and  $\{g(w, s)\}_{s=1}^S$  be the law of motion.

1. Choose a spacing parameter  $h > 0$ .

2. Imagine a hypothetical infinite grid  $\mathcal{W}_\infty = \{w_n\}_{n=1}^\infty$ , where  $w_n = w_N + (n - N)h$  for  $n > N$ , and  $\mathcal{W}_\infty$  agrees with  $\mathcal{W}_N$  for  $n \leq N$ .
3. Linearly extrapolate the law of motion  $g(w, s)$  for  $w > w_N$  (using the slope between the last two grid points, or better yet, the theoretical slope from the solution to the asymptotic problem). Take  $N' \geq N$  such that  $g(w_{N'}, s) > w_N$  for all  $s$ :

$$N' = \min \{n \geq N \mid \forall s, g(w_N + (n - N)h, s) > w_N\}.$$

Thus, if  $w' = G_s w$  is the asymptotic law of motion, after some algebra we obtain

$$N' = N + \max_s \left\lceil \frac{w_N - g(w_N, s)}{G_s h} \right\rceil, \quad (2.7)$$

where  $\lceil x \rceil$  denotes the smallest integer exceeding  $x$ .

4. For each  $(s, n), (s', n') \in \{1, \dots, S\} \times \{1, \dots, N\}$ , compute the transition probability from state  $(s, n)$  to  $(s', n')$  using the non-stochastic simulation described above: use Case 1 for  $n < N$  and Case 2 for  $n = N$ . Collect the probabilities into an  $SN \times SN$  transition probability matrix  $P$ .

A few remarks are in order. First, the above algorithm has essentially zero additional computational cost, despite its complicated appearance. The reason is that extrapolation from the Pareto distribution is used *only* at the largest grid point  $w_N$ . Thus, although we are computing transition probabilities from  $SN$  points, which the usual truncation algorithm needs to compute anyway, the Pareto extrapolation algorithm requires only  $S \times 1 = S$  additional operations, which is negligible. In our numerical implementation in Section 3, we find that the computing time of this step is trivial, and therefore we do not report it.

Second, the  $SN \times SN$  transition probability matrix  $P$  is sparse. To see this, let us evaluate the number of nonzero elements of  $P$ . For each  $s$  and  $n < N$ , there are at most two states the next wealth can take. For  $n = N$ , in principle the next wealth state can be anything. Therefore the number of nonzero elements of  $P$  is at most

$$2S^2(N - 1) + S^2N = S^2(3N - 2).$$

Thus the fraction of nonzero elements of  $P$  is

$$\frac{S^2(3N - 2)}{(SN)^2} = \frac{3N - 2}{N^2} \rightarrow 0$$

as  $N \rightarrow \infty$ , so  $P$  is sparse. [Achdou, Han, Lasry, Lions, and Moll \(2017\)](#) mention that “[c]ontinuous time imparts a number of computational advantages relative to discrete time [... , which] relate to [...] the fact that continuous-time problems with discretized state space are, by construction, very sparse.” While it is true that continuous-time problems have some advantages over discrete-time problems (e.g., partial differential equations versus nonlinear difference equa-

tions), discrete-time problems also do possess sparsity if appropriately solved.

Third, although we have implicitly assumed that  $N'$  in (2.7) is larger than  $N$ , for particular models it may be  $N' \leq N$ , which is true if and only if  $g(w_N, s) \leq w_N$  for all  $s$ . In that case we do not need to consider any extrapolation since the true distribution is not fat-tailed, and the algorithm becomes identical to the truncation method.

Finally, the Pareto extrapolation method requires the spacing parameter  $h > 0$ . Since the Pareto extrapolation algorithm uses a hypothetical evenly-spaced grid (with grid spacing  $h$ ) beyond the largest grid point  $w_N$ , the most natural choice for  $h$  is  $w_N - w_{N-1}$ , the distance between the two largest actual grid points. Conducting numerical experiments similar to those in Section 3, we have found that this choice is numerically optimal.

### 2.2.3 Computing the aggregate wealth

When computing the equilibrium, we need to impose the market clearing condition in some way or another. In Aiyagari models the relevant market clearing condition is for capital, and the demand side is often trivial. To compute the supply of capital (wealth), we can do as follows. First let  $P$  be the  $SN \times SN$  transition probability matrix computed above. Let  $\pi = (\pi_{sn})$  be its stationary distribution, where  $\pi_{sn}$  is the probability of being in state  $(s, n)$ . Note that since  $\pi$  is the (unique) eigenvector of  $P'$  corresponding to the eigenvalue 1 and the matrix  $P$  is sparse by construction, computing  $\pi$  is not an issue. The aggregate wealth is (in principle) then  $\sum_{s,n} \pi_{sn} w_n$ , as in (2.1). The only caveat is that for state  $N$ , the probability is not concentrated on the grid point  $w_N$  but it is a Pareto distribution with exponent  $\zeta$  and minimum size  $w_N$ . Since its density conditional on  $x \geq w_N$  is

$$f(x) = \zeta w_N^\zeta x^{-\zeta-1},$$

the conditional mean of wealth is

$$\int_{w_N}^{\infty} x \zeta w_N^\zeta x^{-\zeta-1} dx = \frac{\zeta}{\zeta-1} w_N.$$

Thus we can compute the aggregate wealth as

$$\mathbb{E}[w] \approx \sum_{s=1}^S \left( \sum_{n=1}^{N-1} \pi_{sn} w_n + \pi_{sN} \frac{\zeta}{\zeta-1} w_N \right). \quad (2.8)$$

Comparing (2.8) to (2.1), we can see that the usual truncation method introduces an error because the last term is  $w_N$  instead of  $\frac{\zeta}{\zeta-1} w_N$ . Since  $w_N$  is typically large, if  $\zeta$  is close to 1 (Zipf's law), then failing to account for the term  $\frac{\zeta}{\zeta-1}$  will introduce significant error.

The same correction applies to computing the integral of more general functions. Suppose we would like to compute the expectation of the power function  $w^\nu$  for some power  $\nu$ . For example,  $\nu = 1$  corresponds to aggregate wealth,  $\nu = 2$  the variance of wealth, and  $\nu = 1 - \gamma$  with  $\gamma > 0$  appears in calculating the welfare for CRRA preferences with relative risk aversion  $\gamma > 0$ . Assuming that wealth has a Pareto upper tail with exponent  $\zeta > \nu$ , then the conditional

expectation of the upper tail is

$$\mathbb{E}[w^\nu | w \geq w_N] = \int_{w_N}^{\infty} \zeta w_N^\zeta x^{\nu-\zeta-1} dx = \frac{\zeta}{\zeta-\nu} w_N^\nu.$$

Therefore the analog of (2.8) is

$$\mathbb{E}[w^\nu] \approx \sum_{s=1}^S \left( \sum_{n=1}^{N-1} \pi_{sn} w_n^\nu + \pi_{sN} \frac{\zeta}{\zeta-\nu} w_N^\nu \right), \quad (2.9)$$

which implies that we need to multiply the value for the largest grid point by the factor  $\frac{\zeta}{\zeta-\nu}$ . Similarly, noting that

$$\mathbb{E}[w^\nu \log w | w \geq w_N] = \mathbb{E} \left[ \frac{d}{d\nu} w^\nu \Big| w \geq w_N \right] = \frac{\zeta}{(\zeta-\nu)^2} w_N^\nu + \frac{\zeta}{\zeta-\nu} w_N^\nu \log w,$$

setting  $\nu = 0$  we obtain

$$\mathbb{E}[\log w] \approx \sum_{s=1}^S \left( \sum_{n=1}^{N-1} \pi_{sn} \log w_n + \pi_{sN} \left( \log w_N + \frac{1}{\zeta} \right) \right). \quad (2.10)$$

Therefore we need to add  $1/\zeta$  to the value for the largest grid point when computing the expectation of log wealth.

### 3 Evaluating solution accuracy

As in any new numerical method, the first order of business is to evaluate the solution accuracy. In this regard, [Den Haan, Judd, and Juillard \(2010\)](#) “find it troublesome that [...] the accuracy of numerical solutions obtains so little attention by so many authors these days.” One reason why accuracy gets little attention may be due to the lack of benchmark closed-form solutions for heterogeneous-agent models.<sup>6</sup> For this purpose, we present a simple (minimal) heterogeneous-agent model with idiosyncratic investment risk that admits a semi-analytical solution, which we use as a benchmark for evaluating numerical solutions.

#### 3.1 Model

We consider a standard [Aiyagari \(1994\)](#) model, except that the model features no idiosyncratic labor income risk (to make the model analytically tractable) but only investment risk (to generate a fat-tailed wealth distribution). The production side is completely standard: there is a representative firm with Cobb-Douglas production function  $F(K, L) = AK^\alpha L^{1-\alpha}$ , where  $A > 0$  is productivity and  $\alpha \in (0, 1)$  is the capital share. Capital depreciates at rate  $\delta$  each period. There are two types of agents, capitalists and workers, of whom there is a mass 1 continuum each. Workers are identical, supply one unit of labor inelastically, and consume the entire wage

<sup>6</sup>In the context of representative-agent asset pricing models, several authors such as [Collard and Juillard \(2001\)](#), [Schmitt-Grohé and Uribe \(2004\)](#), and [Farmer and Toda \(2017\)](#) use the closed-form solution of [Burnside \(1998\)](#) to evaluate solution accuracy.

(hand-to-mouth).<sup>7</sup>

For reasons that will become clear shortly, capitalists are born and go bankrupt with probability  $p$  each period (Yaari (1965)–Blanchard (1985) perpetual youth model). Newborn agents are exogenously endowed with initial wealth  $w_0 > 0$ , and capital is destroyed after bankruptcy. Capitalists have constant relative risk aversion (CRRA) utility as in (2.2a) and supply capital to the firm. Importantly, the gross return on capital is *not* risk-free as

$$R_f = F_K(K, 1) + 1 - \delta = A\alpha K^{\alpha-1} + 1 - \delta, \quad (3.1)$$

but rather  $z_s R_f$ , where  $s = 1, \dots, S$  denotes the exogenous Markov state and  $z_s > 0$  is the gross return on capital *relative* to the risk-free rate (essentially the excess return). Let  $P = (p_{ss'})$  be the transition probability matrix, which we assume to be irreducible. We assume that  $E[z_s] = 1$ , so capital income is just a zero-sum redistribution of aggregate capital income across capitalists. An interpretation is that capitalists earn persistent heterogeneous returns (Fagereng, Guiso, Malacrino, and Pistaferri, 2016a; Cao and Luo, 2017) because some are more skillful in using capital (or just lucky) than others. The initial state of a newborn capitalist is drawn from the stationary distribution  $\pi = (\pi_1, \dots, \pi_S)'$  of the transition probability matrix  $P$ .

The timing is as follows. A capitalist enters period  $t$  with some resource (units of consumption good)  $w_t$ . He decides how much to consume  $c_t$ , and the remaining amount  $k_{t+1} := w_t - c_t$  is installed as capital. At the beginning of period  $t + 1$ , production takes place by pooling all capital, and the capitalist receives the proceed  $w_{t+1} = z_{s_t} R_f k_{t+1}$ , where  $R_f$  is the gross risk-free rate in (3.1) and  $z_{s_t}$  is the predetermined gross excess return.<sup>8</sup> Thus the budget constraint of a capitalist is

$$w' = z_s R_f (w - c). \quad (3.2)$$

A stationary equilibrium consists of aggregate capital  $K$ , gross risk-free rate  $R_f$ , optimal consumption rule  $\{c_s(w)\}_{s=1}^S$ , and a stationary distribution  $\Gamma(w, s)$  such that (i) given  $R_f$ , the optimal consumption rule maximizes the utility (2.2a) subject to the budget constraint (3.2), (ii) firms maximize profits, so (3.1) holds, (iii) the capital market clears, so

$$K = \int (w - c_s(w)) d\Gamma(w, s), \quad (3.3)$$

and (iv)  $\Gamma(w, s)$  is the stationary distribution of the law of motion

$$(w, s) \mapsto \begin{cases} (z_s R_f (w - c_s(w)), s'), & \text{with probability } (1 - p)p_{ss'}, \\ (w_0, s'), & \text{with probability } p\pi_{s'}. \end{cases}$$

By exploiting homotheticity, we can solve the model semi-analytically as discussed in Appendix B. We also prove that a stationary equilibrium exists and the wealth distribution has a Pareto upper tail.

<sup>7</sup>The hand-to-mouth assumption is only for simplicity. Although we can also assume that workers behave optimally, it is inessential for our purpose of discussing numerical algorithms and evaluating the solution accuracy.

<sup>8</sup>We can also allow for the possibility that the gross excess returns are risky by using  $z_{s_{t+1}}$  instead of  $z_{s_t}$ .



### 3.2 Solution accuracy with various grids

We use a numerical example to evaluate the solution accuracy with various grids. We consider the parameter values in Table 2. By solving the equilibrium conditions discussed in Appendix B, we obtain the gross risk-free rate  $R_f = 1.0972$ , aggregate capital  $K = 3.4231$ , and Pareto exponent  $\zeta = 1.2826$ .

Table 2: Parameter values of the Aiyagari model.

Parameter	Symbol	Value
Discount factor	$\beta$	0.96
Relative risk aversion	$\gamma$	2
Bankruptcy probability	$p$	0.025
Gross excess return	$z$	(0.95,1.05)
Transition probability matrix	$P$	$\begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$
Productivity	$A$	1
Capital share	$\alpha$	0.38
Capital depreciation rate	$\delta$	0.08
Initial wealth	$w_0$	1

For the numerical solution, we consider both the conventional truncation method as well as the proposed Pareto extrapolation method with various wealth grid, truncation point, and number of grid points. For the Pareto extrapolation spacing parameter  $h$ , we always take  $h = w_N - w_{N-1}$ , the distance between the two largest grid points.

#### 3.2.1 Evenly-spaced grid

We first consider an  $N$ -point evenly-spaced grid on  $(0, \bar{w}]$ , where we set the truncation point to  $\bar{w} = 10, 20, 40$  and the number of points to  $N = 100, 200, 400$ . Therefore the wealth grid is  $\{nd\}_{n=1}^N$ , where  $d = \bar{w}/N$  is the distance between grid points.<sup>9</sup>

We evaluate the solution accuracy as follows. To ensure that all the differences of the numerical solutions from the analytical one are entirely due to the construction of the transition probability matrix, instead of solving for the equilibrium numerically for each method, we use the equilibrium risk-free rate and consumption policies from the semi-analytical solution to compute the stationary distribution on the wealth grid, and then compute the implied aggregate capital using (2.1) and (2.8) for the Truncation and Pareto extrapolation methods, respectively. For this exercise, our primary interest is the relative error  $|\hat{K}/K - 1|$ , where  $K, \hat{K}$  are the aggregate capital from the semi-analytical and numerical solutions, respectively. Table 3 shows the results.

We can make a few observations from Table 3. First, the conventional truncation method is extremely poor in calculating the aggregate capital: the relative error is about 27–40% depending on the specification. On the other hand, the Pareto extrapolation method is astonishingly more accurate. Second, for the truncation method, choosing a larger truncation point

<sup>9</sup>Note that we exclude 0 from the wealth grid because in equilibrium agents never hit 0 due to the Inada condition.

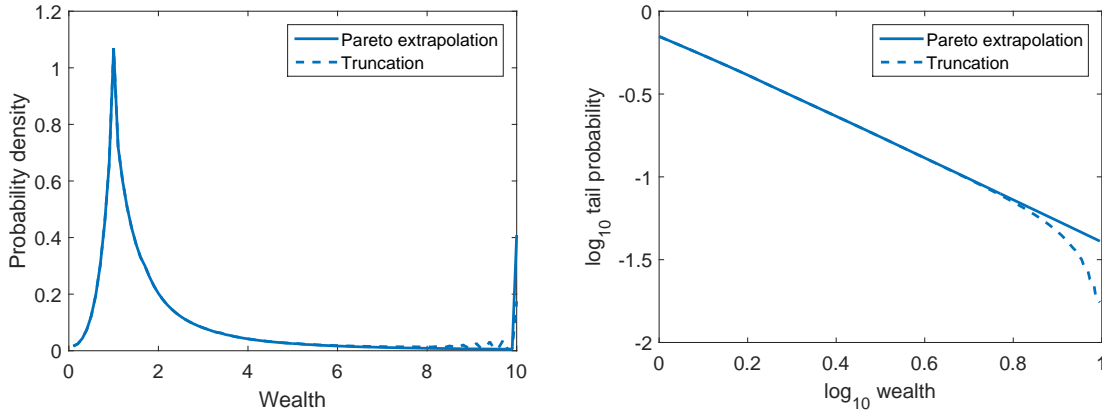
Table 3: Relative error (%) in aggregate capital for the truncation and Pareto extrapolation methods with an evenly-spaced grid.

Method:	Truncation			Pareto extrapolation		
$\bar{w}$	$N = 100$	200	400	100	200	400
10	40.00	39.69	39.62	0.214	0.105	0.052
20	33.58	32.88	32.63	0.430	0.097	0.043
40	26.99	27.56	27.03	3.588	0.331	0.046

Note:  $N$ : number of grid points;  $\bar{w}$ : wealth truncation point.

$\bar{w}$  somewhat improves the accuracy, probably because it misses less of the upper tail. On the other hand, the accuracy in the Pareto extrapolation method is not necessarily monotonic in  $\bar{w}$ . There seems to be a trade-off between less truncation (larger  $\bar{w}$ ) and higher density of grid points (smaller  $d = \bar{w}/N$ ).

Figure 3a shows the stationary wealth distribution using  $\bar{w} = 10$  and  $N = 100$ . The two methods are indistinguishable except at the upper tail. To study the tail behavior, Figure 3b plots the tail probability in a log-log scale. As we can see, the graphs show a straight line pattern, which is consistent with the theoretical Pareto distribution. However, the graph for the truncation method becomes concave towards the upper tail, which implies that it underestimates the tail probability. On the other hand, the Pareto extrapolation method shows a straight line pattern including the very top tail.



(a) Stationary wealth distribution.

(b) Log-log plot of wealth distribution.

Figure 3: Wealth distribution in the Aiyagari model.

These seemingly small differences have an enormous impact on aggregate quantities, as we have seen in Table 3. To assess the robustness, Figure 4a shows the aggregate capital for the semi-analytical solution as well as the Pareto extrapolation and truncation solutions when we change the initial wealth in the range  $w_0 \in [0.2, 5]$ . For all cases we set  $\bar{w} = 10$  and  $N = 100$ . The horizontal axis shows the corresponding equilibrium Pareto exponent. The graph for the Pareto extrapolation method is indistinguishable from the semi-analytical solution except when the Pareto exponent is very close to 1 (Zipf's law), in which case the truncation method is especially poor.

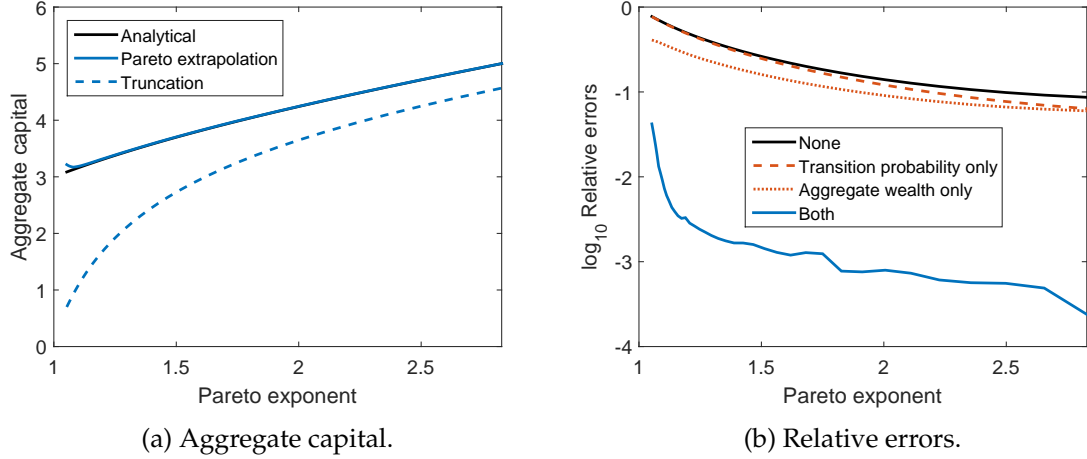


Figure 4: Solution accuracy of the Aiyagari model.

Note: “None”, “Transition probability only”, “Aggregate wealth only”, and “Both” refer to (i) not using Pareto extrapolation and using it to (ii) constructing the transition probability matrix only as in Section 2.2.2, (iii) calculating aggregate wealth only as in Section 2.2.3, and (iv) both, respectively.

Figure 4b shows the relative errors  $\left| \widehat{K}/K - 1 \right|$  in a semi log scale. For this exercise, we consider four solution methods that correspond to using/not using Pareto extrapolation when constructing the transition probability matrix and/or calculating aggregate capital. For example, “None” and “Both” in Figure 4b correspond to the truncation and Pareto extrapolation solutions, respectively. According to the figure, using Pareto extrapolation for only one step (either constructing the transition probability matrix or calculating aggregate capital) improves the accuracy only slightly, and correcting the aggregate capital matters more. However, combining both increases the solution accuracy dramatically.

The intuition for this (surprising) result is as follows. According to (2.8), the two sources of errors introduced by the truncation method (incorrect transition probability matrix *and* incorrect aggregate wealth held by agents at the top grid point) interact with each other. With the truncation method, the last term  $\pi_{sN} \frac{\zeta}{\zeta-1} w_N$  in (2.8) becomes  $\tilde{\pi}_{sN} w_N$ , with typically  $\tilde{\pi}_{sN} < \pi_{sN}$ . Therefore, errors in  $\pi_{sN}$  are inflated by a factor  $\frac{\zeta}{\zeta-1}$ , which is large if  $\zeta > 1$  is small.

### 3.2.2 Exponentially-spaced grid

One may argue that the poor performance of the truncation method in Table 3 is due to the fact that the truncation point  $\bar{w} = 10, 20, 40$  is relatively small compared to the aggregate capital  $K = 3.4231$ . What if we take  $\bar{w}$  much larger, say a million? Then we can no longer use evenly-spaced grids because there will be too few points to cover the bottom of the wealth distribution. Therefore we need to consider an exponentially-spaced grid.

In more general models, the state variable may become negative (e.g., asset holdings), which causes a problem for constructing an exponentially-spaced grid because we cannot take the logarithm of a negative number. One way to get around this issue is to shift the state space by some positive number (shift parameter), take an evenly-spaced grid in logs, take the exponential, and shift back the grid, as explained in Appendix C. We choose the shift parameter such that the median grid point corresponds to the capital in a representative-agent model,

which in our case is  $K_{RA} = ((1/(\beta(1-p)) - 1 + \delta)/(A\alpha))^{\frac{1}{\alpha-1}} = 4.5577$ . Since an exponential grid allows us to use far fewer points than an evenly-spaced grid, we consider  $N = 25, 50, 100$ . (We have checked that increasing  $N$  further also increases the accuracy.) For the truncation point, we consider  $\bar{w} = 10^2, 10^3, 10^4, 10^5, 10^6$ . Table 4 shows the results.

Table 4: Relative error (%) in aggregate capital for the truncation and Pareto extrapolation methods with an exponentially-spaced grid.

Method:	Truncation			Pareto extrapolation			
	$\bar{w}$	$N = 25$	50	100	25	50	100
	$10^2$	27.54	23.36	21.08	1.697	0.186	0.588
	$10^3$	20.85	15.27	12.32	2.423	0.785	0.002
	$10^4$	17.12	10.78	7.62	2.553	0.905	0.184
	$10^5$	14.93	8.06	4.93	2.546	0.879	0.222
	$10^6$	13.19	6.27	3.32	2.445	0.807	0.210

Note:  $N$ : number of grid points;  $\bar{w}$ : wealth truncation point.

The accuracy of the truncation method somewhat improves by using the exponential grid with a large truncation point  $\bar{w}$  and many points. However, even with a large truncation like a million and  $N = 100$  grid points, the error still exceeds 3%. (Increasing  $N$  further decreases the error, but only slightly for the truncation method.) Again, the Pareto extrapolation method is overwhelmingly more accurate.

### 3.3 Equilibrium implications of solution accuracy

So far we have evaluated the accuracy of each solution method by comparing the implied aggregate capital to the one from the semi-analytical solution. However, aggregate capital is usually not a quantity of interest. Therefore we now evaluate the solution accuracy by solving for the entire equilibrium.

We consider four possibilities, truncation or Pareto extrapolation with evenly- or exponentially-spaced grid. Since according to Table 4 the truncation method seems to work best when there are many points and the truncation point is large, to give it the best chance, we use  $N = 100$  and  $\bar{w} = 10^6$  for the exponential grid. For the evenly-spaced grid, we also take  $N = 100$  for a fair comparison, and based on Table 3, we choose a small truncation point  $\bar{w} = 10$  to give the Pareto extrapolation method the best chance. To ensure that all the differences of the numerical solutions from the analytical one are entirely due to the construction of the transition probability matrix, for each guess of the equilibrium risk-free rate, we use the semi-analytical solution to the optimal consumption-savings problem to compute the law of motion for wealth for all numerical methods. The first three rows of Table 5 show the equilibrium risk-free rate, aggregate capital, and Pareto exponent.

As expected, the truncation method with the evenly-spaced grid is poor, so we do not discuss it further. The truncation method with the exponentially-spaced grid performs reasonably well, with relative errors of 0.046% for the interest rate and 0.45% for aggregate capital. The relative errors for the Pareto extrapolation method are 0.0026%, 0.026%, and 0.050% for the interest rate, aggregate capital, and Pareto exponent, respectively, for both the evenly- and

Table 5: Numerical solutions to the Aiyagari model.

Method:	Truncation		Pareto		Analytical
	Even	Exp.	Even	Exp.	
Grid:					
$N$ :	100		100		
$\bar{w}$ :	10	$10^6$	10	$10^6$	
$R_f$	1.1184	1.0977	1.0972	1.0972	1.0972
$K$	2.8530	3.4073	3.4239	3.4222	3.4231
$\zeta$	NA	NA	1.2833	1.2820	1.2826
Top 0.01%	0.03	16.04	13.27	18.10	13.19
Top 0.1%	0.31	26.30	22.06	27.86	21.91
Top 1%	3.16	41.84	36.67	42.83	36.39
Top 10%	31.19	65.36	60.90	65.79	60.44

Note: the table presents equilibrium quantities in the Aiyagari model using both analytical and numerical methods.  $N$ : number of grid points;  $\bar{w}$ : wealth truncation point;  $R_f$ : gross risk-free rate;  $K$ : aggregate capital;  $\zeta$ : Pareto exponent; Top  $x\%$ : wealth share (%) of the wealthiest  $x\%$ . “Truncation” and “Pareto” refer to the truncation and Pareto extrapolation methods for solving the equilibrium, and “Analytical” shows results from the semi-analytical solution. “Even” and “Exp.” refer to the evenly- and exponentially-spaced grids. For the top wealth shares in the analytical solution, see Footnote 10.

exponentially-spaced grids. Given that the truncation method with an exponential grid is already reasonably accurate, however, one may wonder what is the point of improving it further.

One obvious reason to prefer the Pareto extrapolation method is that we can obtain 20 times more accurate results at no additional computational cost, so there is just no reason not to use it. However, a more compelling reason is that the interest rate, aggregate capital, and Pareto exponent are not the only quantities of interest (although the truncation method cannot even compute the Pareto exponent). One may be interested in other quantities, such as the top 1% wealth share.

We construct the top wealth shares as follows. For each grid point, we can compute the aggregate wealth held by agents at least as rich as that grid point (using either truncation or Pareto extrapolation, depending on the solution method). Dividing that number by aggregate wealth gives the top wealth share at that grid point. By interpolating between points, we can define the top wealth shares inside the grid. To compute the top wealth shares outside the grid, we do as follows. For the Pareto extrapolation method, we use the theoretical Pareto exponent  $\zeta$  to extrapolate the wealth share beyond the largest grid point. More precisely, let  $\pi_N = \sum_{s=1}^S \pi_{sN}$  be the probability mass on the largest grid point  $w_N$  in the Pareto extrapolation method. Then the density for  $x \geq w_N$  is given by  $f(x) = \pi_N \zeta w_N^\zeta x^{-\zeta-1}$ , where  $\zeta$  is the Pareto exponent. Using this, the tail probability  $\Pr(X \geq x)$  is proportional to  $x^{-\zeta}$ , whereas the total wealth held by wealthy agents  $E[X; X \geq x]$  is proportional to  $x^{-\zeta+1}$ . Therefore the wealth share  $s(p)$  of the wealthiest fraction  $p \in (0, 1)$  of agents is given by  $s(p) = Cp^{1-1/\zeta}$ , where the constant of proportionality  $C$  can be easily calculated from  $\pi_N$ ,  $w_N$ ,  $\zeta$ , and aggregate capital  $K$ . Taking the logarithm, we obtain

$$\log s = (1 - 1/\zeta) \log p + \log C,$$

so top wealth shares are linear in a log-log scale.

For the truncation method, since it is not obvious how to extrapolate the top wealth share beyond the largest grid point, we simply interpolate by a cubic spline using the point  $(0, 0)$  (by definition, the top 0% wealth share is 0) and all the grid points. The last four rows of Table 5 present some top wealth shares for representative groups.<sup>10</sup> Figure 5a plots the wealth shares of the richest 0.01–10%.

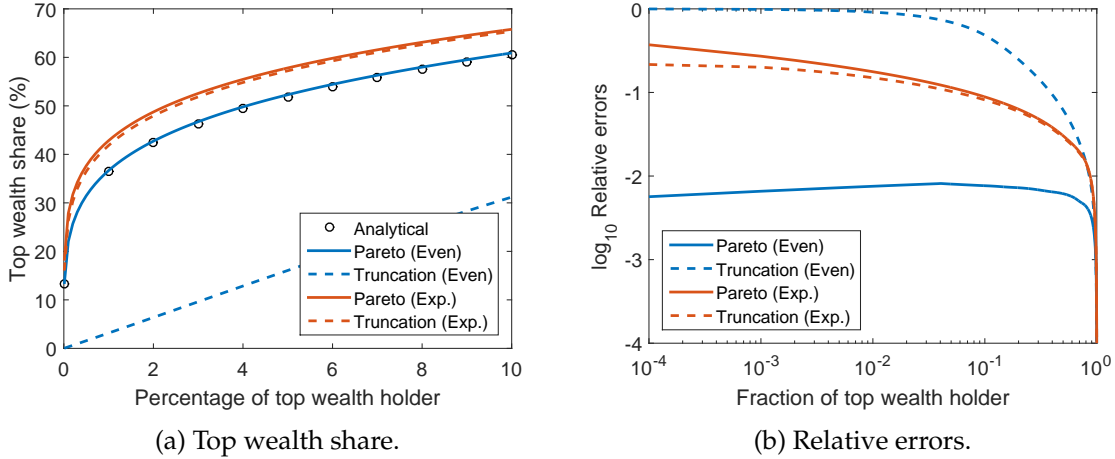


Figure 5: Top wealth shares in the Aiyagari model.

Note: “Analytical”, “Pareto”, and “Truncation” refer to the semi-analytical solution and the numerical solutions using the Pareto extrapolation and truncation methods, respectively. “Even” and “Exp.” refer to the evenly- and exponentially-spaced grid.

Interestingly, both the truncation and Pareto extrapolation methods with the exponentially-spaced grid *overestimate* the top wealth shares. On the other hand, the Pareto extrapolation method with the evenly-spaced grid essentially gives identical results to the semi-analytical solution. Since the  $\log_{10}$  relative errors in Figure 5b are uniformly less than  $-2$ , the top wealth shares from the Pareto extrapolation method with the evenly-spaced grid are accurate up to two significant digits.

Why is the Pareto extrapolation method with exponentially-spaced grid poor at computing top wealth shares, despite its accuracy in computing the equilibrium according to Table 5? The reason is probably because the exponentially-spaced grid (with a large truncation such as  $10^6$ ) places relatively few points in the *middle* of the distribution, by its very nature that it places many points in the bottom (because the exponential function is flat at the bottom) and the top (because the truncation point is large) of the distribution. In fact, the exponential grid places only 10 (16) points out of 100 between the 5 and 95 percentile (1 and 99 percentile) of the wealth distribution, whereas the evenly-spaced grid places 77 (95) points. However, the Pareto tail result of [Beare and Toda \(2017\)](#) does not say anything about the middle of the distribution. Since the top wealth shares depend on the entire shape of the distribution, placing few points in the middle (where it can be non-Pareto) makes the solution less accurate.

<sup>10</sup>Technically, for the semi-analytical solution we cannot compute the exact top wealth shares since the functional form of the wealth distribution is unknown (we only know the tail behavior characterized by the Pareto exponent  $\zeta$ ). For this case, to compute the stationary distribution, we use the Pareto extrapolation method with a highly accurate 3,000-point evenly- or exponentially-spaced grid with truncation  $\bar{w} = 100, 10^6$ , respectively. The results with the two methods were identical up to three significant digits, so we take it as the truth.

### 3.4 What grid should we use?

From the above observations, one might be tempted to use the evenly-spaced grid all the time because it dominates the exponential grid in every aspect according to Table 5. However, we do not think that is the way to go. The reason is that in more general models, the policy functions are nonlinear (only asymptotically linear), unlike in the model of Section 3.1, which is exactly linear. To deal with potential nonlinearity, researchers usually need to use a large enough grid, such as the exponential. However, we would like to assign enough points to the middle of the distribution to approximate the non-Pareto part. Is there a simple grid that achieves both objectives?

An easy way is to first construct an exponential grid as discussed in Appendix C, and then replace the bottom half points by an evenly-spaced grid. We refer to this grid as the hybrid (affine-exponential) grid. Since the Pareto extrapolation method does not require a large truncation point (in fact, in Table 3 smaller  $\bar{w}$  is generally more accurate) because it can recover the tail behavior of the wealth distribution using a single parameter, the Pareto exponent  $\zeta$ , when using the hybrid grid, we can also choose a relatively small truncation point such as  $\bar{w} = 1000$ . Redoing the above exercise with this specification, the relative errors in the interest rate, aggregate capital, and Pareto exponent are now  $2.7 \times 10^{-7}$ ,  $2.7 \times 10^{-6}$ , and  $5.3 \times 10^{-6}$ , respectively, which are minuscule even comparing to the evenly- or exponentially-spaced grids. The top 10, 1, 0.1, 0.01% wealth shares are, respectively, 61.02, 37.70, 23.28, 14.35%, which are off from the true values only by about one percentage point (compared to about five percentage points error for the exponential grid). Therefore the hybrid grid seems to achieve the right balance between high accuracy and large coverage. In fact, there are now many grid points in the middle of the distribution: 47 (61) points out of 100 between the 5 and 95 percentile (1 and 99 percentile) of the wealth distribution.

Based on these observations, we recommend the following strategy for solving heterogeneous-agent models with fat-tailed wealth distribution.

#### **Strategy for solving fat-tailed heterogeneous-agent models.**

1. Before solving the model, find out a typical scale for the state variable (wealth), perhaps by solving a representative-agent model without any shock.
2. Solve the heterogeneous-agent model using the Pareto extrapolation method with the hybrid affine-exponential grid. More concretely,
  - (a) Construct the exponentially-spaced grid with a truncation point about 1000 times the typical scale for the state variable.
  - (b) Replace the bottom half grid points by an evenly-spaced grid.
3. After solving the model, recompute the wealth distribution from the already computed equilibrium law of motion using a finer grid if necessary.

Note that there are many other possibilities. In the value function iteration step, since we do not need the wealth distribution, we can just use an exponentially-spaced grid with relatively few points to increase the speed. When computing the market clearing condition, we can interpolate the policy functions on an evenly-spaced grid and then use Pareto extrapolation method for accuracy.

The beauty of the Pareto extrapolation method is that we do not need to worry about the truncation error, because it is designed to correct for it. One can thus choose the grid by focusing on the nonlinearity in the policy functions and placing enough points in the middle (non-Pareto) part of the wealth distribution. With the truncation method, on the other hand, we are forced to use the exponential grid with a huge truncation point to reduce the truncation error. However, doing so *necessarily* places too many points in the upper tail of the wealth distribution (where in fact we do not need any points because the behavior is governed by a single parameter, the Pareto exponent) and too few points in the middle (non-Pareto) part, leading to inaccuracy.

## 4 Merton-Bewley-Aiyagari model

Having now established that the solution method we propose is accurate, we apply it to study wealth inequality in an incomplete market general equilibrium model in the spirit of [Aiyagari \(1994\)](#). Agents face an income fluctuation problem as in [Bewley \(1977, 1983\)](#) and those who choose to invest face uninsurable investment risk that leads to an investor's problem similar to [Merton \(1969\)](#) and [Samuelson \(1969\)](#), although in a Markovian setting as studied in [Krebs \(2006\)](#) and [Toda \(2014\)](#). The model (which we refer to as the Merton-Bewley-Aiyagari, or MBA model) generates a fat-tailed wealth distribution, where the Pareto exponent is shaped by rich general equilibrium effects. We provide a step-by-step approach to solving the model and analyzing its quantitative implications.

We use the model to showcase how the asymptotic analysis of the individual problem can be used to solve for a general equilibrium, and how solution accuracy affects the quantitative implications of the model.

### 4.1 Model

Time is discrete and denoted by  $t = 0, 1, \dots$

**Agents** The economy is populated by a unit measure of infinitely-lived agents with Epstein-Zin preferences

$$U_t = \left( (1 - \beta)c_t^{1-1/\varepsilon} + \beta E_t[U_{t+1}^{1-\gamma}]^{\frac{1-1/\varepsilon}{1-\gamma}} \right)^{\frac{1}{1-1/\varepsilon}}, \quad (4.1)$$

where  $c_t > 0$  is consumption,  $U_t > 0$  is continuation utility,  $\beta \in (0, 1)$  is the discount factor,  $\gamma > 0$  is the coefficient of relative risk aversion, and  $\varepsilon > 0$  is the elasticity of intertemporal substitution.<sup>11</sup> Agents differ in productivity/ability states denoted by  $s \in S = \{1, \dots, S\}$ , which

<sup>11</sup>By considering the limit  $\gamma \rightarrow 1$ , we interpret  $E[U^{1-\gamma}]^{\frac{1}{1-\gamma}}$  as  $\exp(E[\log U])$  if  $\gamma = 1$ . Similarly, we interpret  $((1 - \beta)c^{1-1/\varepsilon} + \beta v^{1-1/\varepsilon})^{\frac{1}{1-1/\varepsilon}}$  as  $c^{1-\beta} v^\beta$  if  $\varepsilon = 1$ .



evolve over time according to a Markov chain with irreducible transition probability matrix  $P = (p_{ss'})$ . The idiosyncratic productivity states are independent and identically distributed across agents and we assume the law of large numbers for the continuum as in Sun (2006). Therefore if  $\pi = (\pi_1, \dots, \pi_S)'$  denotes the (unique) stationary distribution of the transition probability matrix  $P$ , at any point in time exactly fraction  $\pi_s > 0$  of agents are in state  $s$ .

**Labor and financial markets** A type  $s$  agent has labor productivity  $h_s \geq 0$  and earns labor income  $(1 - \tau_h)\omega h_s$ , where  $\omega > 0$  is the “piece-rate” wage determined in equilibrium and  $\tau_h \in [0, 1)$  is a labor income tax. We assume that tax proceeds are wasted. A type  $s$  agent has investment ability  $z_s > 0$  and earns excess returns in the financial market, as described below. Without loss of generality we assume that  $z_1 \leq \dots \leq z_S$ , so the Markov state  $s$  indexes the agent’s investment ability. There are two types of assets, risk-free and risky. The risky asset is subject to an investment income tax  $\tau_I \in [0, 1)$ . Let  $R_f$  be the gross risk-free rate determined in equilibrium. We assume that the ex post gross return on risky investment for an investor in state  $s$  is

$$R_{sj} = (1 + (1 - \tau_I)(z_s + \epsilon_j - 1))R_f, \quad (4.2)$$

where  $z_s$  is the investment ability and  $\epsilon_j$  is a zero-mean i.i.d. random variable that can take  $J$  possible values  $\epsilon_1 < \dots < \epsilon_J$ . Let  $p_j > 0$  be the probability of state  $j$ . Thus high-skilled investors earn higher returns on average ( $z_s$ ), but there is some element of luck ( $\epsilon_j$ ). We can interpret this as lack of diversification. We assume that  $z_1 + \epsilon_1 > 0$ , so agents have limited liability even in the worst possible state. To prevent arbitrage, we also assume that  $z_S + \epsilon_1 < 1$ , so even the most skilled investor underperforms the risk-free asset with positive probability. For notational simplicity let

$$z_{sj} = \tau_I + (1 - \tau_I)(z_s + \epsilon_j) \quad (4.3)$$

be the after-tax ex post excess return in state  $(s, j)$ . Then (4.2) can be compactly written as  $R_{sj} = z_{sj}R_f$ . From the above assumptions, note that (i)  $z_{sj}$  is increasing in both  $s$  and  $j$ , (ii)  $z_{sj} > 0$  for all  $s, j$ , and (iii)  $z_{S1} < 1$ .

**Technology** Technology is represented by a representative firm with a constant-returns-to-scale production function  $F(K, L)$ . Capital depreciates at rate  $\delta \in [0, 1]$ . Therefore the firm’s problem is

$$\max_{K, L \geq 0} \left[ -K + \frac{1}{R_f}(F(K, L) - \omega L + (1 - \delta)K) \right]. \quad (4.4)$$

That is, the firm buys capital  $K$  at end of time  $t$ , hires labor to produce, and pays the profit and depreciated capital to the shareholders (who discount using the risk-free rate since there is no aggregate risk).

**Budget constraint** Letting  $w$  be the financial wealth at the beginning of the period, the budget constraint of an agent is

$$w' = R_f(w + (1 - \tau_h)\omega h_s - I - c) + z_{sj}R_f I \geq w, \quad (4.5)$$

where  $\underline{w}$  is an exogenous minimum wealth constraint and  $I \geq 0$  is the investment in the risky asset.

**Equilibrium** Our equilibrium concept is the stationary equilibrium defined as follows.

**Definition 4.1** (Stationary equilibrium). A stationary equilibrium consists of a gross risk-free rate  $R_f$ , a piece-rate wage  $\omega$ , aggregate capital  $K$ , aggregate labor  $L$ , optimal decision rules  $\{c_s(w), I_s(w)\}_{s=1}^S$ , value functions  $\{v_s(w)\}_{s=1}^S$ , and a stationary distribution  $\Gamma(w, s)$  such that

1. given  $R_f$  and  $\omega$ , aggregate capital  $K$  and aggregate labor  $L$  solves the profit maximization problem (4.4),
2. given  $R_f$  and  $\omega$ , for each  $s$  the optimal decision rule  $(c_s(w), I_s(w))$  maximizes the recursive utility (4.1) subject to the budget and borrowing constraint (4.5), i.e.,

$$v_s(w) = \max_{c, I \geq 0} \left( (1 - \beta)c^{1-1/\varepsilon} + \beta \mathbb{E} \left[ v_{s'}(w')^{1-\gamma} \mid s \right]^{\frac{1-1/\varepsilon}{1-\gamma}} \right)^{\frac{1}{1-1/\varepsilon}}, \quad (4.6)$$

3. the capital market clears, so

$$K = \int (w + (1 - \tau_h)\omega h_s - I_s(w) - c_s(w)) d\Gamma(w, s) + \int z_s I_s(w) d\Gamma(w, s), \quad (4.7)$$

4. the labor market clears, so

$$L = \sum_{s=1}^S \pi_s h_s, \quad (4.8)$$

5.  $\Gamma(w, s)$  is the stationary distribution of the law of motion for  $(w, s) \in [\underline{w}, \infty) \times S$  defined by

$$(w, s) \mapsto (R_f(w + (1 - \tau_h)\omega h_s - I_s(w) - c_s(w)) + z_{sj}R_f I_s(w), s') \quad (4.9)$$

with probability  $p_j p_{ss'}$ .

Note that an investor who invests  $I$  units of wealth supplies  $z_s I$  units of capital to the firm (see (4.7)), so investors not only supply funds but also “expertise”. This assumption (together with  $\mathbb{E}[\varepsilon_j] = 0$ ) makes the (pre-tax) gross return to investment  $(z_s + \varepsilon_j)R_f$  consistent with the aggregate resource constraint.

## 4.2 Solving the model

### 4.2.1 Asymptotic problem

Since the solution to the asymptotic problem plays an important role in the Pareto extrapolation algorithm (especially for computing the theoretical Pareto exponent), in this section we discuss the properties of the asymptotic problem.

We first derive the asymptotic problem and convert it in a more convenient form. Since there is no labor income in the asymptotic problem, the budget constraint (4.5) becomes

$$w' = R_f(w - I - c) + z_{sj}R_f I \geq 0, \quad (4.10)$$

where  $z_{sj}$  is as in (4.3). Letting  $\theta = \frac{I}{w-c} \geq 0$ , (4.10) becomes

$$w' = R(\theta)(w - c), \quad (4.11)$$

where  $R(\theta) = R_f(z_{sj}\theta + 1 - \theta)$  is the gross portfolio return. The following proposition characterizes the solution to the asymptotic problem.

**Proposition 4.2** (Asymptotic problem). *Suppose  $\gamma \neq 1$ . For  $s = 1, \dots, S$ , define*

$$\rho_s = \max_{0 \leq \theta \leq \bar{\theta}_s} \mathbb{E} \left[ (z_{sj}\theta + 1 - \theta)^{1-\gamma} \mid s \right]^{\frac{1}{1-\gamma}}, \quad (4.12)$$

where  $\bar{\theta}_s := \frac{1}{1-z_{s1}} > 0$ . Then  $\rho_s$  is well-defined and there exists a unique maximizer  $\theta_s^*$ . Letting  $D = \text{diag}(\rho_1^{1-\gamma}, \dots, \rho_S^{1-\gamma})$ , the asymptotic problem has a solution if and only if

$$\beta R_f^{1-1/\varepsilon} \rho(DP)^{\frac{1-1/\varepsilon}{1-\gamma}} < 1, \quad (4.13)$$

where  $\rho(DP)$  is the spectral radius of  $DP$ . Under this condition, the value function of the asymptotic problem is given by  $v_s(w) = b_s w$ , where  $b = (b_1, \dots, b_S) \gg 0$  is the unique positive solution to the system of nonlinear equations

$$b_s = \begin{cases} \left( (1-\beta)^\varepsilon + \beta^\varepsilon \left( R_f \rho_s \mathbb{E} \left[ b_{s'}^{1-\gamma} \mid s \right]^{\frac{1}{1-\gamma}} \right)^{\varepsilon-1} \right)^{\frac{1}{\varepsilon-1}} & (\varepsilon \neq 1) \\ (1-\beta)^{1-\beta} \beta^\beta \left( R_f \rho_s \mathbb{E} \left[ b_{s'}^{1-\gamma} \mid s \right]^{\frac{1}{1-\gamma}} \right)^\beta & (\varepsilon = 1) \end{cases} \quad (4.14)$$

for  $s = 1, \dots, S$ . The optimal consumption-investment rules of the asymptotic problem are

$$c_s(w) = \bar{c}_s w := (1-\beta)^\varepsilon b_s^{1-\varepsilon} w, \quad (4.15a)$$

$$I_s(w) = \bar{I}_s w := \theta_s^* (1 - (1-\beta)^\varepsilon b_s^{1-\varepsilon}) w. \quad (4.15b)$$

**Remark.** Clearly  $\theta_s^*$  is independent of  $R_f$ . By (4.14),  $b_s$  is increasing in  $R_f$ . Therefore consumption  $c_s(w)$  is increasing and investment  $I$  is decreasing in  $R_f$  if  $\varepsilon < 1$ , and the opposite is true if  $\varepsilon > 1$ .

## 4.2.2 General equilibrium

Since the diagonal matrix  $D = \text{diag}(\rho_1^{1-\gamma}, \dots, \rho_S^{1-\gamma})$  depends only on exogenous parameters, the spectral condition (4.13) puts a restriction on  $R_f$ . The following lemma puts further restrictions on  $R_f$  based on equilibrium considerations.

**Lemma 4.3** (Finite aggregate wealth). *Let everything be as in Proposition 4.2 and suppose that (4.13) holds. Define*

$$G_s := (1 - (1-\beta)^\varepsilon b_s^{1-\varepsilon}) R_f (\mathbb{E} [z_{sj} \mid s] \theta_s^* + 1 - \theta_s^*) \quad (4.16)$$

and  $G = (G_1, \dots, G_S)'$ . Then in equilibrium it must be

$$\rho(P \text{ diag } G) < 1. \quad (4.17)$$

The intuition for Lemma 4.3 is as follows. Using the budget constraint and the optimal consumption and portfolio rules for asymptotic agents established in Proposition 4.2, the expected growth rate of wealth in state  $s$  becomes  $G_s$  in (4.16). The spectral condition (4.17) ensures that wealth of rich agents does not grow on average and makes the aggregate wealth finite, which must be the case in stationary equilibrium.

Under the assumptions of Lemma 4.3, the following lemma provides an explicit algorithm for computing the Pareto exponent  $\zeta$ .

**Lemma 4.4** (Pareto exponent). *Let everything be as in Lemma 4.3 and*

$$G_{sj} = (1 - (1 - \beta)^\varepsilon b_s^{1-\varepsilon}) R_f (z_{sj} \theta_s^* + 1 - \theta_s^*) > 0$$

be the ex post gross growth rate of wealth for asymptotic agents in state  $(s, j)$ . Define the conditional moment generating function of log growth rate by

$$M_s(z) = \mathbb{E} \left[ e^{z \log G_{sj}} \mid s \right] = \sum_{j=1}^J p_j G_{sj}^z$$

and the diagonal matrix  $D(z) = \text{diag}(M_1(z), \dots, M_S(z))$ . Suppose that  $p_{ss} > 0$  for all  $s$  and  $G_{sJ} > 1$  for some  $s$ . Then there exists a unique solution  $z = \zeta > 1$  to

$$\rho(PD(z)) = 1.$$

The Pareto exponent of the wealth distribution is  $\zeta > 1$ . If  $G_{sJ} \leq 1$  for all  $s$ , then the wealth distribution does not have a Pareto tail.

Finally, it is trivial to show that the equilibrium risk-free rate must satisfy

$$R_f > 1 - \delta, \quad (4.18)$$

otherwise the demand for capital would be infinite. Based on the above results, we can solve for the general equilibrium as follows.

1. For each  $s$ , solve the optimal portfolio problem of the asymptotic agents (4.12). Compute the bounds on the gross risk-free rate  $R_f$  implied by (4.13) and (4.18).
2. For a range of  $R_f$  within the bounds, solve the system of nonlinear equations (4.14), compute  $G_s$  by (4.16), and update the bound on  $R_f$  implied by (4.17).
3. Given a guess of  $R_f$  within the bound, compute the capital demand  $K^d$  implied by

profit maximization<sup>12</sup>

$$R_f = F_K(K^d, L) + 1 - \delta.$$

Then, compute the wage as  $\omega = F_L(K^d, L)$ .

4. Given  $R_f$  and  $\omega$ , solve the individual optimization problem using dynamic programming (see Appendix E for details), compute the stationary distribution from the law of motion (4.9) using the Pareto extrapolation algorithm, and compute the aggregate capital supply  $K^s$  (4.7) (with the correction term as in (2.8)) and labor (4.8).
5. If excess demand  $K^d - K^s$  is within error tolerance, stop. Otherwise, update  $R_f$  (steps 3 and 4 only).

### 4.3 Calibration

A time period represents a year and we calibrate the model to the U.S. economy.

**Preferences, technology, and taxes** We assume a Cobb-Douglas production function  $F(K, L) = K^\alpha L^{1-\alpha}$ , where  $\alpha \in (0, 1)$  represents the capital share. We set the technological and preference parameters  $(\beta, \gamma, \alpha)$  to standard values (see Table 6). For the elasticity of intertemporal substitution, most macro papers assume that it is less than 1, while most finance papers assume that it is greater than 1. To be neutral, we set  $\varepsilon = 1$ , which is also supported by studies using disaggregated data to estimate the elasticity (Mankiw and Zeldes, 1991; Attanasio and Weber, 1993; Beaudry and van Wincoop, 1996; Vissing-Jørgensen, 2002). We set the investment tax rate to  $\tau_I = 0.389$  to match the 2017 combined corporate income tax rate. We set the labor income tax rate to  $\tau_h = 0.260$  to match the 2017 average income tax rate for a single filer without children.<sup>13</sup> Finally, we set the borrowing limit to 1/4 of average annual labor income as in Kaplan, Moll, and Violante (2018). Our calibration implies that average value of  $h_s$  in a stationary equilibrium is 1, so  $\underline{w} = -\frac{1}{4}\omega$ .

Table 6: Parameter values.

Parameter	Symbol	Value
Discount factor	$\beta$	0.96
Relative risk aversion	$\gamma$	2
Elasticity of intertemporal substitution	$\varepsilon$	1
Capital share	$\alpha$	0.38
Investment income tax rate	$\tau_I$	0.389
Labor income tax rate	$\tau_h$	0.260
Borrowing limit	$\underline{w}$	$-\frac{1}{4}\omega$

<sup>13</sup>We obtain the data on corporate and labor income tax rates from the OECD tax database tables II.1 and I.6, respectively.

**Exogenous individual states** We assume that labor productivity  $\{h_t\}_{t=0}^{\infty}$  and investment ability  $\{z_t\}_{t=0}^{\infty}$  depend contemporaneously on two Markov state variables  $s^\pi$  and  $s^\tau$ , with associated transition probability matrices  $P^\pi$  and  $P^\tau$ . We interpret the first state  $s^\pi$  as a “permanent component”, which affect both labor productivity and investment ability and takes three states: low, high, and high-entrepreneur. The second state  $s^\tau$ , which we call the “transitory component” affects only labor productivity and takes three values: low, average and high. The index  $s = s^\pi \times s^\tau$  can thus take 9 states.

**Labor productivity** Labor productivity in state  $s$  is the product of a permanent component and a transitory component:

$$h_s = h_s^\pi h_s^\tau.$$

The permanent component  $h_s^\pi$  takes two values: 0.3980 (low) and 1.6020 (high and high-entrepreneur). The high state workers (high and high-entrepreneur) thus earn a wage rate 4.03 times higher than low state workers, in line with the ratio of the mean annual income of the top half to the bottom half of full-time workers in the U.S.<sup>14</sup> We interpret agents as dynasties with perfect altruism, and assume that the permanent component of labor productivity is very persistent and changes on average every 40 years. Moreover, we choose transition probabilities for the permanent component as to imply that, in a stationary equilibrium, 50% of the agents are in the low state and 3.7% are in the high-entrepreneur state. We chose the value of 3.7% to match the fraction of households that invest at least half of their net worth in a business (see footnote 15). The transition probability matrix  $P_{h^\pi}$  is thus given by

$$P_{h^\pi} = \begin{bmatrix} 0.9875 & 0.0116 & 0.0009 \\ 0.0125 & 0.9866 & 0.0009 \\ 0.0125 & 0.0116 & 0.9759 \end{bmatrix}. \quad (4.19)$$

We model the process for the transitory component of labor productivity  $h^\tau$  as an AR(1) in logarithm

$$\log h_t^\tau = \rho \log h_{t-1}^\tau + \sigma \eta_t, \quad (4.20)$$

where  $\eta_t \sim N(0, 1)$ . [Guvenen \(2009\)](#) estimates values of  $\rho = 0.821$  and  $\sigma = 0.170$  by estimating a model that allows for heterogeneous lifetime earning profiles. We discretize the process (4.20) over a three-point grid using the method proposed by [Farmer and Toda \(2017\)](#), while imposing that the unconditional mean of  $h^\tau$  is equal to one. We obtain values for  $h^\tau$  of 0.6584 (low state), 0.9150 (average state), and 1.5115 (high state). The resulting transition probability matrix  $P_{h^\tau}$  is given by

$$P_{h^\tau} = \begin{bmatrix} 0.8290 & 0.1630 & 0.0080 \\ 0.0815 & 0.8370 & 0.0815 \\ 0.0080 & 0.1630 & 0.8290 \end{bmatrix}. \quad (4.21)$$

<sup>14</sup>We use data from the American Community Survey and restrict the sample to employed individuals aged between 20 and 60. For every year, we truncate the sample at the 5th and 95th percentile and then compute the ratio of the mean annual income in the bottom and top half of the sample. We then average the ratio over the 2000-2016 period and obtain an average ratio of 4.03.

Since  $h^\pi$  and  $h^\tau$  are independent and both have an unconditional mean of one, we have that the average labor productivity  $h_s$  is one in equilibrium.

**Investment returns** We assume that investment ability  $z$  is fully determined by the permanent component  $s^\pi$ . For the low and high states, we set  $z = 1$ , which implies that those agents do not earn an excess return on their investments (i.e., they earn the risk-free rate  $R_f$ ). Since investment returns are risky, those agents will never invest. For the high-entrepreneur state, we set  $z = 1.029$ , which implies an annual excess return of 2.9%. This is roughly the difference between the average return on risky financial assets of households at the 80th and 60th percentiles of the financial wealth distribution in Norway as reported in [Fagereng, Guiso, Malacrino, and Pistaferri \(2016b\)](#), Figure 2).

To calibrate the distribution of idiosyncratic investment return shocks  $\epsilon$ , we use microdata from the Survey of Consumer Finances and construct a measure of the rate of return on business investment (business income over the market value of the business) for each household.<sup>15</sup> Using the nonparametric discretization method proposed by [Toda \(2018a\)](#) on the de-meaned data, we obtain a discrete distribution for  $\epsilon$ , which takes values  $(-0.0836, 0.0761, 0.3795)$  with probability  $(0.6345, 0.2822, 0.0833)$ .

Putting the pieces together, we have  $S = 9$  exogenous individual states and  $J = 3$  idiosyncratic investment return shock states. The transition probability matrix for the exogenous individual states is given by  $P = P_{h^\tau} \otimes P_{h^\pi}$ , where  $\otimes$  is the Kronecker product and  $P_{h^\pi}, P_{h^\tau}$  are defined in (4.19) and (4.21), respectively. Table 7 summarizes the dependence of labor productivity and investment ability on the exogenous individual states  $s$ .

Table 7: Exogenous individual states.

$s$	Component		Productivity/ability	
	permanent	transitory	labor ( $h_s$ )	investment ( $z_s$ )
1	low	low	0.2620	1.00
2	low	average	0.3642	1.00
3	low	high	0.6016	1.00
4	high	low	1.0547	1.00
5	high	average	1.4659	1.00
6	high	high	2.4215	1.00
7	high-entrepreneur	low	1.0547	1.029
8	high-entrepreneur	average	1.4659	1.029
9	high-entrepreneur	high	2.4215	1.029

#### 4.4 Quantitative results

Figure 6a shows the aggregate capital supply and demand curves for a range of values of  $R_f$ . The demand curve is determined by profit maximization of the representative firm, while the

<sup>15</sup>We use data from the Survey of Consumer Finances for the years 2001, 2004, 2007, 2010, 2013, and 2016 and keep only households who have at least 50% of their net worth invested in businesses. Business income `busseincfarm` is defined as “Income from business, sole proprietorship, and farm” while the market value of the businesses `bus` is defined as “Total value of business(es) in which the household has either an active or nonactive interest”. We winsorize the rate of return at the 5% to extreme values.

supply curve is obtained by aggregating net capital supply from households, either supplied directly to the firm or intermediated by investors (see (4.7)). The intersection of the curves pins down the risk-free rate that clears the capital market. We obtain a value of 1.0199 (Table 8).

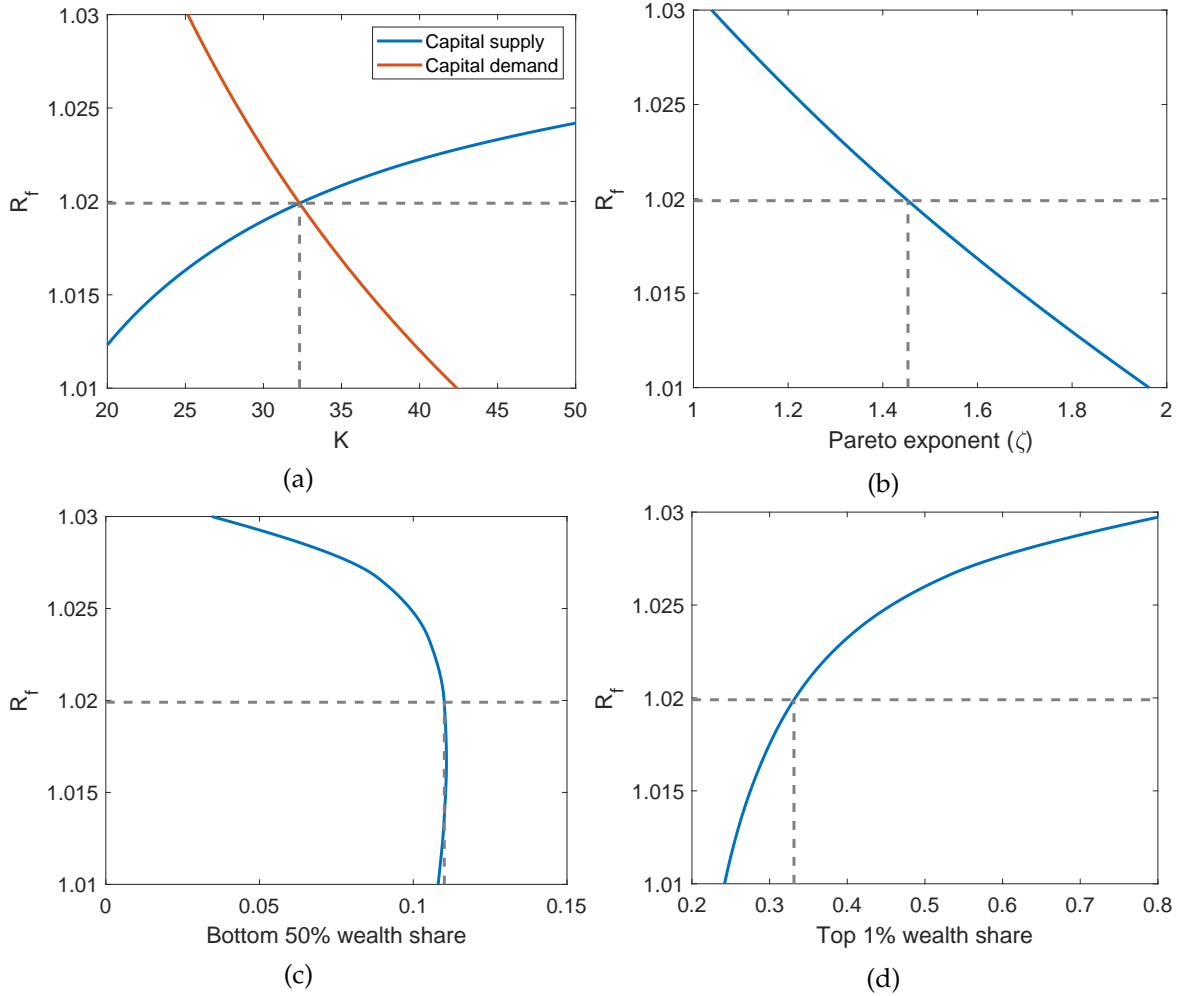


Figure 6: Equilibrium.

The equilibrium interest rate in turn determines the Pareto exponent  $\zeta$  of the wealth distribution (Figure 6b). Higher interest rates  $R_f$  are associated with lower Pareto exponents  $\zeta$  (higher inequality). The reason is that the asymptotic growth rate of wealth increases linearly with  $R_f$  (see Lemma 4.4 for the case  $\varepsilon = 1$ ). Intuitively, the “pace” at which rich agents get richer is increasing in the interest rate, so there is more concentration of wealth at the top of the distribution in economies with higher interest rates. We obtain an equilibrium value of  $\zeta = 1.4532$  (Table 8), close but somewhat lower than what is estimated for the U.S. (1.52 according to Table 8 of Vermeulen, 2018).

While the wealth distribution exhibits a Pareto upper tail, the Pareto exponent  $\zeta$  does not fully summarize the extent of wealth inequality. For example, the borrowing constraint is an important determinant of the wealth share of the poorest 50% of agents, yet  $\zeta$  depends only on the behavior of rich agents, who are not affected by the borrowing constraint.<sup>16</sup> Figures 6c and

<sup>16</sup>To be precise,  $\zeta$  can be computed using only the solution to the asymptotic problem (see Section 4.2.1), which does not depend on the borrowing constraint.



Table 8: Equilibrium objects.

Object	Symbol	Value
Aggregate capital	$K$	32.3127
Risk-free rate	$R_f$	1.0199
Wage rate	$\omega$	2.3225
Pareto exponent	$\zeta$	1.4532

6d highlight the dependence of the wealth share of the poorest 50% of agents as well as the top 1% on  $R_f$ . Higher interest rates are associated with a lower wealth share held by the bottom 50% and a higher share held by the top 1%, but notice that the dependence of the bottom 50% wealth share on  $R_f$  is much weaker. Table 9 presents the wealth shares in the model and in the data (Survey of Consumer Finances).<sup>17</sup> The model qualitatively replicates two important features of the data: the poorest 50% of households holds little wealth (11%) while the top 1% accounts for a large share of (33%).

Table 9: Wealth shares (%).

Groups	Model	Data	Pareto
[0, 50)	11.01	1.79	19.44
[50, 90)	35.83	25.09	31.79
[90, 99)	20.01	38.17	24.98
[99, 100]	33.14	34.95	23.79

The last column of Table 9 presents the wealth shares associated with a Pareto distribution with the same Pareto exponent as in the model.<sup>18</sup> The Pareto distribution generates a bottom 50% wealth share nearly twice as high as in the model (19.44% versus 11.01%). One drawback of analytical models that imply a “pure-Pareto” wealth distribution is that they do not allow for negative wealth levels, and therefore cannot match the low wealth share of the bottom 50% of agents. Figure 7a presents the CDF for the lower range of wealth levels. Notice that the borrowing constraint binds for 6.9% of households. Figure 7b plots the complementary CDF of the wealth distribution on exponential axis. As predicted by theory, the upper tail of the distribution appears to converge to the theoretical Pareto slope.

#### 4.5 Equilibrium effects of relaxing the borrowing constraint

We now revisit a well-known experiment which consists in relaxing the borrowing constraint.<sup>19</sup> As noted earlier, the borrowing constraint binds for many agents, including investors. Presumably, relaxing the borrowing constraint would increase aggregate investment. We increase the borrowing limit tenfold, from one quarter of average earnings ( $w = -0.25\omega$ , see Table 6) to 2.5 years of average annual earnings ( $w = -2.5\omega$ ). Figures 8a and 8b plot the policy functions

<sup>17</sup>We compute the wealth shares in the data using the Survey of Consumer Finances and average over the survey years 2001, 2004, 2007, 2010, 2013, and 2016.

<sup>18</sup>The Pareto distribution with exponent  $\zeta$  has a CDF given by  $F(x) = 1 - (x/\bar{x})^{-\zeta}$  over  $[\bar{x}, \infty)$ . The top wealth shares associated with the pure Pareto distribution do not depend on the minimum size  $\bar{x} > 0$ .

<sup>19</sup>Cagetti and De Nardi (2006) is the first paper, to our knowledge, to conduct the experiment in an heterogeneous-agent model with entrepreneurial income.

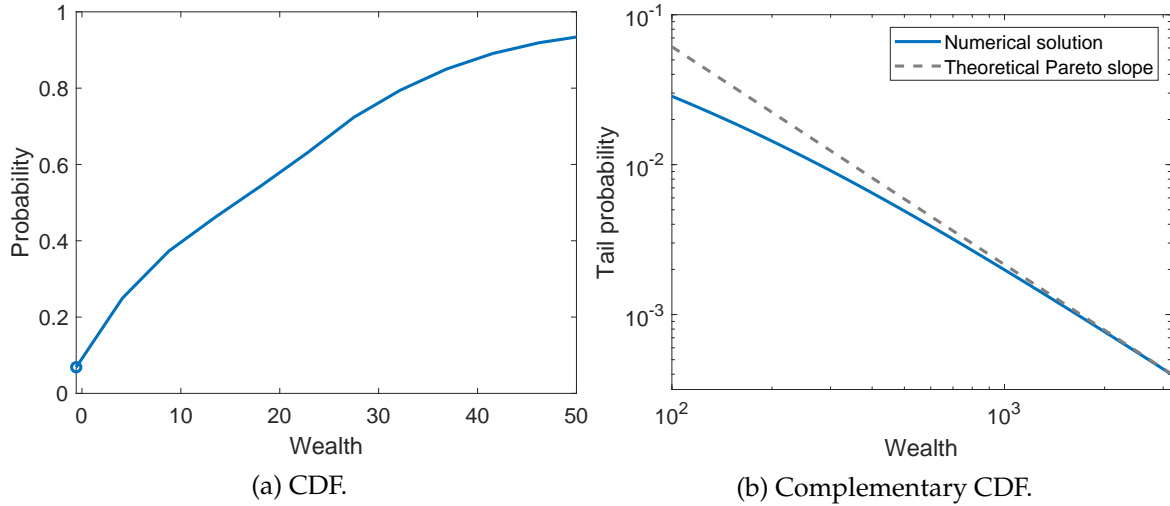


Figure 7: Wealth distribution.

in the benchmark model and the counterfactual model. The direct effect of relaxing the borrowing constraint is that it increases the levels of investment and consumption of poor agents (those near the borrowing limit). However, the policy also has the perverse effect creating more poverty. Figure 8c highlights the fact that, in a stationary equilibrium, more agents are debtors. Accounting for the increasing share of debtors, the bottom 50% wealth share decreases starkly, from 11.01% to 6.72% (see Table 10).

Table 10: Wealth shares (%).

Groups	Benchmark	Counterfactual
[0, 50)	11.01	6.72
[50, 90)	35.83	36.08
[90, 99)	20.01	21.35
[99, 100]	33.14	35.84

Another indirect effect of the policy is that precautionary savings decrease. While it is true that the amount of capital supplied by investors increases, the decline in precautionary savings dominates and translates in a leftward shift of the capital supply curve. As a result, the interest rate increases and the aggregate capital stock decreases (see Table 11).

Table 11: Equilibrium objects.

Object	Symbol	Benchmark	Counterfactual	Change (%)
Aggregate capital	$K$	32.3127	31.7990	-1.59
Risk-free rate	$R_f$	1.0199	1.0215	0.15
Wage rate	$\omega$	2.3225	2.3084	-0.61
Pareto exponent	$\zeta$	1.4532	1.4249	-1.95
Social welfare	$\mathcal{W}$	1.1452	1.0514	-8.19

The lower capital stock leads to lower wages, while the higher interest further increases

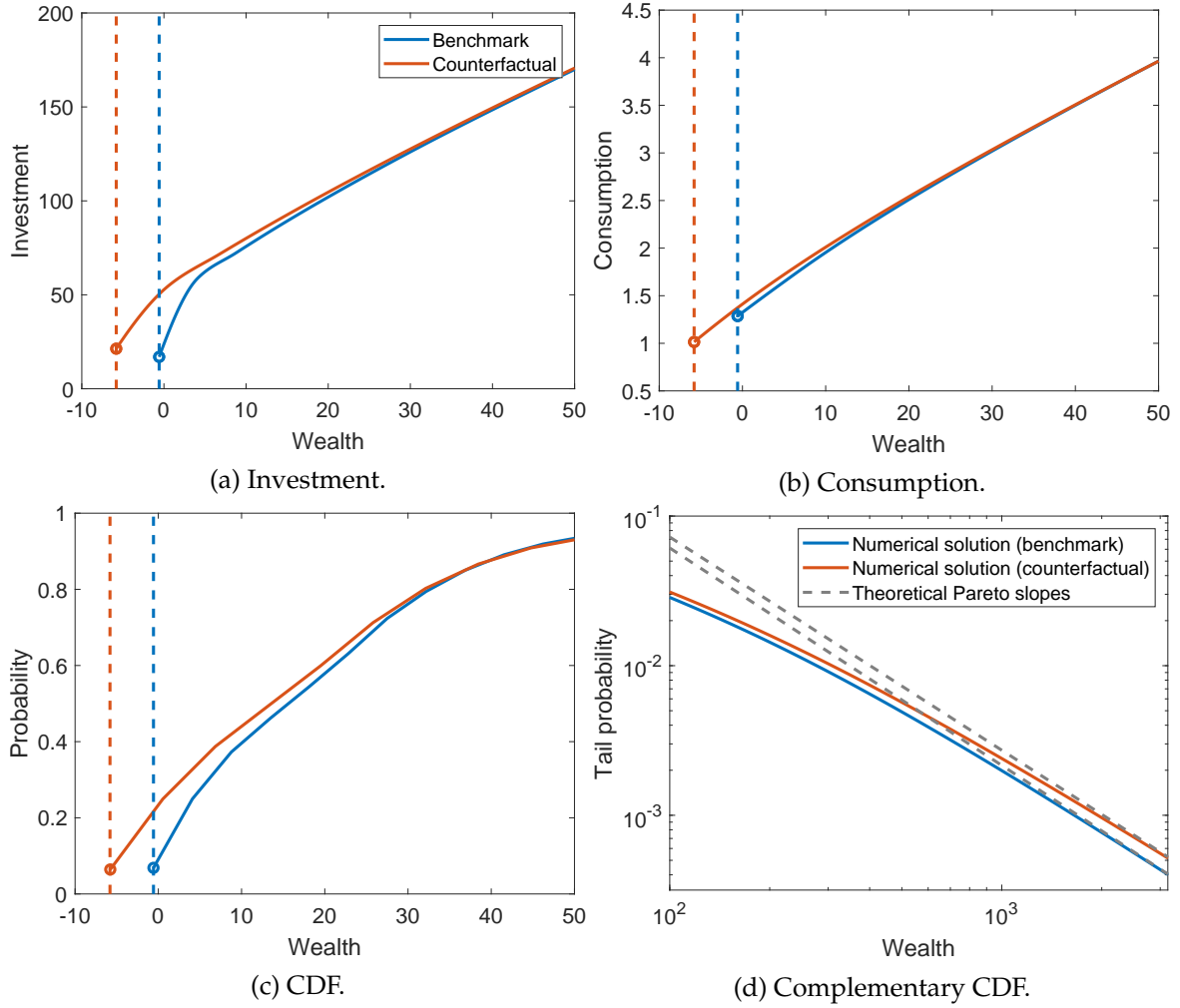


Figure 8: Policy functions (for entrepreneurs) and the wealth distribution.

Note: The policy functions are presented only for the entrepreneur type (permanent component  $s^\pi = \text{high-entrepreneur}$ ) with average earnings (transitory component state  $s^\tau = \text{average}$ ).

wealth inequality (lower  $\zeta$ , see Table 11). We compute social welfare as

$$\mathcal{W} = \left( \int v_s(w)^{1-\gamma} d\Gamma(s, w) \right)^{\frac{1}{1-\gamma}}, \quad (4.22)$$

which is the certainty equivalent of the value function in the stationary equilibrium. Note that  $\mathcal{W}$  is in units of consumption because the Epstein-Zin utility (4.1) is also in units of consumption. To compute (4.22) numerically, we use the correction term from (2.9) with  $\nu = 1 - \gamma$  to extrapolate the term  $v_s(w)^{1-\gamma}$  off the grid. With lower wages and more inequality, welfare declines. In fact, agents would be willing to forgo 8.19% of consumption at every date and state to live in a world with a tight borrowing constraint!

## 5 Concluding remarks

This paper proposes a simple, systematic approach—Pareto extrapolation—to analyze and solve heterogeneous-agent models with fat-tailed wealth distributions. Pareto noticed that

wealth data displayed a striking empirical regularity.

Nous sommes tout de suite frappé du fait que les points ainsi déterminés, ont une tendance très marqué à se disposer en ligne droite.

(We are instantly struck by the fact that the points determined this way have a very marked tendency to be disposed in straight line.)

—Pareto (1897, pp. 304–305)

Our proposed method puts Pareto’s insight to work to tackle *models* of wealth inequality. Our approach makes the solution algorithm more transparent, efficient, and accurate with zero additional computational cost.

## References

- Daron Acemoglu and Dan Cao. Innovation by entrants and incumbents. *Journal of Economic Theory*, 157:255–294, May 2015. doi:[10.1016/j.jet.2015.01.001](https://doi.org/10.1016/j.jet.2015.01.001).
- Yves Achdou, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. Income and wealth distribution in macroeconomics: A continuous-time approach. NBER Working Paper 23732, 2017. URL <http://www.nber.org/papers/w23732>.
- S. Rao Aiyagari. Uninsured idiosyncratic risk and aggregate saving. *Quarterly Journal of Economics*, 109(3):659–684, August 1994. doi:[10.2307/2118417](https://doi.org/10.2307/2118417).
- Yann Algan, Olivier Allais, and Wouter J. Den Haan. Solving heterogeneous-agent models with parameterized cross-sectional distributions. *Journal of Economic Dynamics and Control*, 32(3):875–908, March 2008. doi:[10.1016/j.jedc.2007.03.007](https://doi.org/10.1016/j.jedc.2007.03.007).
- Yann Algan, Olivier Allais, Wouter J. Den Haan, and Pontus Rendahl. Solving and simulating models with heterogeneous agents and aggregate uncertainty. In Karl Schmedders and Kenneth L. Judd, editors, *Handbook of Computational Economics*, volume 3, chapter 6, pages 277–324. Elsevier, 2014. doi:[10.1016/B978-0-444-52980-0.00006-2](https://doi.org/10.1016/B978-0-444-52980-0.00006-2).
- Shuhei Aoki and Makoto Nirei. Zipf’s law, Pareto’s law, and the evolution of top incomes in the United States. *American Economic Journal: Macroeconomics*, 9(3):36–71, July 2017. doi:[10.1257/mac.20150051](https://doi.org/10.1257/mac.20150051).
- Costas Arkolakis. A unified theory of firm selection and growth. *Quarterly Journal of Economics*, 131(1):89–155, February 2016. doi:[10.1093/qje/qjv039](https://doi.org/10.1093/qje/qjv039).
- Orazio P. Attanasio and Guglielmo Weber. Consumption growth, the interest rate and aggregation. *Review of Economic Studies*, 60(3):631–649, July 1993. doi:[10.2307/2298128](https://doi.org/10.2307/2298128).
- Robert L. Axtell. Zipf distribution of U.S. firm sizes. *Science*, 293(5536):1818–1820, September 2001. doi:[10.1126/science.1062081](https://doi.org/10.1126/science.1062081).

- Brendan K. Beare and Alexis Akira Toda. Geometrically stopped Markovian random growth processes and Pareto tails. 2017. URL <https://arxiv.org/abs/1712.01431>.
- Paul Beaudry and Eric van Wincoop. The intertemporal elasticity of substitution: An exploration using a US panel of state data. *Economica*, 63(251):495–512, August 1996. doi:[10.2307/2555019](https://doi.org/10.2307/2555019).
- Jess Benhabib, Alberto Bisin, and Shenghao Zhu. The distribution of wealth and fiscal policy in economies with finitely lived agents. *Econometrica*, 79(1):123–157, January 2011. doi:[10.3982/ECTA8416](https://doi.org/10.3982/ECTA8416).
- Jess Benhabib, Alberto Bisin, and Shenghao Zhu. The wealth distribution in Bewley economies with capital income risk. *Journal of Economic Theory*, 159(A):489–515, September 2015. doi:[10.1016/j.jet.2015.07.013](https://doi.org/10.1016/j.jet.2015.07.013).
- Jess Benhabib, Alberto Bisin, and Shenghao Zhu. The distribution of wealth in the Blanchard-Yaari model. *Macroeconomic Dynamics*, 20:466–481, March 2016. doi:[10.1017/S1365100514000066](https://doi.org/10.1017/S1365100514000066).
- Truman F. Bewley. The permanent income hypothesis: A theoretical formulation. *Journal of Economic Theory*, 16(2):252–292, December 1977. doi:[10.1016/0022-0531\(77\)90009-6](https://doi.org/10.1016/0022-0531(77)90009-6).
- Truman F. Bewley. A difficulty with the optimum quantity of money. *Econometrica*, 51(5): 1485–1504, September 1983. doi:[10.2307/1912286](https://doi.org/10.2307/1912286).
- Olivier J. Blanchard. Debt, deficits, and finite horizons. *Journal of Political Economy*, 93(2): 223–247, April 1985. doi:[10.1086/261297](https://doi.org/10.1086/261297).
- Jaroslav Borovička and John Stachurski. Necessary and sufficient conditions for existence and uniqueness of recursive utilities. 2017. URL <https://arxiv.org/abs/1710.06526>.
- Craig Burnside. Solving asset pricing models with Gaussian shocks. *Journal of Economic Dynamics and Control*, 22(3):329–340, March 1998. doi:[10.1016/S0165-1889\(97\)00075-4](https://doi.org/10.1016/S0165-1889(97)00075-4).
- Marco Cagetti and Mariacristina De Nardi. Entrepreneurship, frictions, and wealth. *Journal of Political Economy*, 114(5):835–870, October 2006. doi:[10.1086/508032](https://doi.org/10.1086/508032).
- Dan Cao and Wenlan Luo. Persistent heterogeneous returns and top end wealth inequality. *Review of Economic Dynamics*, 26:301–326, October 2017. doi:[10.1016/j.red.2017.10.001](https://doi.org/10.1016/j.red.2017.10.001).
- Christopher D. Carroll, Kiichi Tokuoka, and Weifeng Wu. The method of moderation. 2012.
- Christopher D. Carroll, Jiri Slacalek, Kiichi Tokuoka, and Matthew N. White. The distribution of wealth and the marginal propensity to consume. *Quantitative Economics*, 8(3):977–1020, November 2017. doi:[10.3982/QE694](https://doi.org/10.3982/QE694).
- Fabrice Collard and Michel Juillard. Accuracy of stochastic perturbation methods: The case of asset pricing models. *Journal of Economic Dynamics and Control*, 25(6-7):979–999, June 2001. doi:[10.1016/S0165-1889\(00\)00064-6](https://doi.org/10.1016/S0165-1889(00)00064-6).

- Wouter J. Den Haan. Comparison of solutions to the incomplete markets model with aggregate uncertainty. *Journal of Economic Dynamics and Control*, 34(1):4–27, January 2010a. doi:[10.1016/j.jedc.2008.12.010](https://doi.org/10.1016/j.jedc.2008.12.010).
- Wouter J. Den Haan. Assessing the accuracy of the aggregate law of motion in models with heterogeneous agents. *Journal of Economic Dynamics and Control*, 34(1):79–99, January 2010b. doi:[10.1016/j.jedc.2008.12.009](https://doi.org/10.1016/j.jedc.2008.12.009).
- Wouter J. Den Haan, Kenneth L. Judd, and Michel Juillard. Computational suite of models with heterogeneous agents: Incomplete markets and aggregate uncertainty. *Journal of Economic Dynamics and Control*, 34(1):1–3, January 2010. doi:[10.1016/j.jedc.2009.07.001](https://doi.org/10.1016/j.jedc.2009.07.001).
- Richard Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York, fourth edition, 2010.
- Andreas Fagereng, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. Heterogeneity in returns to wealth and the measurement of wealth inequality. *American Economic Review: Papers and Proceedings*, 106(5):651–655, May 2016a. doi:[10.1257/aer.p20161022](https://doi.org/10.1257/aer.p20161022).
- Andreas Fagereng, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. Heterogeneity and persistence in returns to wealth. NBER Working Paper 22822, 2016b. URL <http://www.nber.org/papers/w22822>.
- Leland E. Farmer and Alexis Akira Toda. Discretizing nonlinear, non-Gaussian Markov processes with exact conditional moments. *Quantitative Economics*, 8(2):651–683, July 2017. doi:[10.3982/QE737](https://doi.org/10.3982/QE737).
- Xavier Gabaix. Power laws in economics and finance. *Annual Review of Economics*, 1:255–293, 2009. doi:[10.1146/annurev.economics.050708.142940](https://doi.org/10.1146/annurev.economics.050708.142940).
- Xavier Gabaix. The granular origins of aggregate fluctuations. *Econometrica*, 79(3):733–772, May 2011. doi:[10.3982/ECTA8769](https://doi.org/10.3982/ECTA8769).
- Xavier Gabaix, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. The dynamics of inequality. *Econometrica*, 84(6):2071–2111, November 2016. doi:[10.3982/ECTA13569](https://doi.org/10.3982/ECTA13569).
- Robert Gibrat. *Les Inégalités Économiques*. Librairie du Recueil Sirey, Paris, 1931.
- Fatih Guvenen. An empirical investigation of labor income processes. *Review of Economic Dynamics*, 12(1):58–79, January 2009. doi:[10.1016/j.red.2008.06.004](https://doi.org/10.1016/j.red.2008.06.004).
- Mark Huggett. The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control*, 17(5-6):953–969, September–November 1993. doi:[10.1016/0165-1889\(93\)90024-M](https://doi.org/10.1016/0165-1889(93)90024-M).
- Charles I. Jones and Jihee Kim. A Schumpeterian model of top income inequality. *Journal of Political Economy*, 2018. doi:[10.1086/699190](https://doi.org/10.1086/699190). Forthcoming.
- Greg Kaplan, Benjamin Moll, and Giovanni L. Violante. Monetary policy according to HANK. *American Economic Review*, 108(3):697–743, March 2018. doi:[10.1257/aer.20160042](https://doi.org/10.1257/aer.20160042).

- Kenneth Kasa and Xiaowen Lei. Risk, uncertainty, and the dynamics of inequality. *Journal of Monetary Economics*, 94:60–78, April 2018. doi:[10.1016/j.jmoneco.2017.11.008](https://doi.org/10.1016/j.jmoneco.2017.11.008).
- Harry Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131(1):207–248, 1973. doi:[10.1007/BF02392040](https://doi.org/10.1007/BF02392040).
- Oren S. Klass, Ofer Biham, Moshe Levy, Ofer Malcai, and Sorin Solomon. The Forbes 400 and the Pareto wealth distribution. *Economics Letters*, 90(2):290–295, February 2006. doi:[10.1016/j.econlet.2005.08.020](https://doi.org/10.1016/j.econlet.2005.08.020).
- Tom Krebs. Recursive equilibrium in endogenous growth models with incomplete markets. *Economic Theory*, 29(3):505–523, 2006. doi:[10.1016/S0165-1889\(03\)00062-9](https://doi.org/10.1016/S0165-1889(03)00062-9).
- Dirk Krueger, Kurt Mitman, and Fabrizio Perri. Macroeconomics and household heterogeneity. In John B. Taylor and Harald Uhlig, editors, *Handbook of Macroeconomics*, volume 2, chapter 11, pages 843–921. Elsevier, 2016. doi:[10.1016/bs.hesmac.2016.04.003](https://doi.org/10.1016/bs.hesmac.2016.04.003).
- Per Krusell and Anthony A. Smith, Jr. Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896, October 1998. doi:[10.1086/250034](https://doi.org/10.1086/250034).
- Per Krusell and Anthony A. Smith, Jr. Quantitative macroeconomic models with heterogeneous agents. In Richard Blundell, Whitney K. Newey, and Torsten Persson, editors, *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, volume 1 of *Econometric Society Monograph*, chapter 8, pages 298–340. Cambridge University Press, New York, 2006.
- Qingyin Ma and John Stachurski. Dynamic programming deconstructed. 2018. URL [http://johnstachurski.net/\\_downloads/dpd6.pdf](http://johnstachurski.net/_downloads/dpd6.pdf).
- N. Gregory Mankiw and Stephen P. Zeldes. The consumption of stockholders and non-stockholders. *Journal of Financial Economics*, 29(1):97–112, March 1991. doi:[10.1016/0304-405X\(91\)90015-C](https://doi.org/10.1016/0304-405X(91)90015-C).
- Alisdair McKay. Time-varying idiosyncratic risk and aggregate consumption dynamics. *Journal of Monetary Economics*, 88:1–14, June 2017. doi:[10.1016/j.jmoneco.2017.05.002](https://doi.org/10.1016/j.jmoneco.2017.05.002).
- Robert C. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics*, 51(3):247–257, August 1969. doi:[10.2307/1926560](https://doi.org/10.2307/1926560).
- Benjamin Moll. Productivity losses from financial frictions: Can self-financing undo capital misallocation? *American Economic Review*, 104(10):3186–3221, October 2014. doi:[10.1257/aer.104.10.3186](https://doi.org/10.1257/aer.104.10.3186).
- Makoto Nirei and Shuhei Aoki. Pareto distribution of income in neoclassical growth models. *Review of Economic Dynamics*, 20:25–42, April 2016. doi:[10.1016/j.red.2015.11.002](https://doi.org/10.1016/j.red.2015.11.002).
- Makoto Nirei and Wataru Souma. A two factor model of income distribution dynamics. *Review of Income and Wealth*, 53(3):440–459, September 2007. doi:[10.1111/j.1475-4991.2007.00242.x](https://doi.org/10.1111/j.1475-4991.2007.00242.x).

- Vilfredo Pareto. La legge della domanda. *Giornale degli Economisti*, 10:59–68, January 1895.
- Vilfredo Pareto. *La Courbe de la Répartition de la Richesse*. Imprimerie Ch. Viret-Genton, Lausanne, 1896.
- Vilfredo Pareto. *Cours d'Économie Politique*, volume 2. F. Rouge, Lausanne, 1897.
- Vincenzo Quadrini. Entrepreneurship, saving, and social mobility. *Review of Economic Dynamics*, 3(1):1–40, January 2000. doi:[10.1006/redy.1999.0077](https://doi.org/10.1006/redy.1999.0077).
- William J. Reed. The Pareto, Zipf and other power laws. *Economics Letters*, 74(1):15–19, December 2001. doi:[10.1016/S0165-1765\(01\)00524-9](https://doi.org/10.1016/S0165-1765(01)00524-9).
- Michael Reiter. Solving heterogeneous-agent models by projection and perturbation. *Journal of Economic Dynamics and Control*, 33(3):649–665, March 2009. doi:[10.1016/j.jedc.2008.08.010](https://doi.org/10.1016/j.jedc.2008.08.010).
- Michael Reiter. Solving the incomplete markets model with aggregate uncertainty by backward induction. *Journal of Economic Dynamics and Control*, 34(1):28–35, January 2010. doi:[10.1016/j.jedc.2008.11.009](https://doi.org/10.1016/j.jedc.2008.11.009).
- Paul A. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics*, 51(3):239–246, August 1969. doi:[10.2307/1926559](https://doi.org/10.2307/1926559).
- Stephanie Schmitt-Grohé and Martín Uribe. Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics and Control*, 28(4):755–775, January 2004. doi:[10.1016/S0165-1889\(03\)00043-5](https://doi.org/10.1016/S0165-1889(03)00043-5).
- John Stachurski and Alexis Akira Toda. An impossibility theorem for wealth in heterogeneous-agent models with limited heterogeneity. 2018. URL <https://arxiv.org/abs/1807.08404>.
- Yeneng Sun. The exact law of large numbers via Fubini extension and characterization of insurable risks. *Journal of Economic Theory*, 126(1):31–69, January 2006. doi:[10.1016/j.jet.2004.10.005](https://doi.org/10.1016/j.jet.2004.10.005).
- Alexis Akira Toda. Incomplete market dynamics and cross-sectional distributions. *Journal of Economic Theory*, 154:310–348, November 2014. doi:[10.1016/j.jet.2014.09.015](https://doi.org/10.1016/j.jet.2014.09.015).
- Alexis Akira Toda. Data-based automatic discretization of nonparametric distributions. 2018a. URL <http://arxiv.org/abs/1805.00896>.
- Alexis Akira Toda. Wealth distribution with random discount factors. *Journal of Monetary Economics*, 2018b. doi:[10.1016/j.jmoneco.2018.09.006](https://doi.org/10.1016/j.jmoneco.2018.09.006).
- Alexis Akira Toda and Kieran Walsh. The double power law in consumption and implications for testing Euler equations. *Journal of Political Economy*, 123(5):1177–1200, October 2015. doi:[10.1086/682729](https://doi.org/10.1086/682729).
- Alexis Akira Toda and Kieran James Walsh. Fat tails and spurious estimation of consumption-based asset pricing models. *Journal of Applied Econometrics*, 32(6):1156–1177, September/October 2017. doi:[10.1002/jae.2564](https://doi.org/10.1002/jae.2564).



- Philip Vermeulen. How fat is the top tail of the wealth distribution? *Review of Income and Wealth*, 64(2):357–387, June 2018. doi:[10.1111/roiw.12279](https://doi.org/10.1111/roiw.12279).
- Annette Vissing-Jørgensen. Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy*, 110(4):825–853, August 2002. doi:[10.1086/340782](https://doi.org/10.1086/340782).
- Thomas Winberry. A method for solving and estimating heterogeneous agent macro models. *Quantitative Economics*, 2018. URL <http://qeconomics.org/ojs/forth/740/740-2.pdf>. Forthcoming.
- Menahem E. Yaari. Uncertain lifetime, life insurance, and the theory of the consumer. *Review of Economic Studies*, 32(2):137–150, April 1965. doi:[10.2307/2296058](https://doi.org/10.2307/2296058).
- Eric R. Young. 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, January 2010. doi:[10.1016/j.jedc.2008.11.010](https://doi.org/10.1016/j.jedc.2008.11.010).
- Shenghao Zhu. A Becker-Tomes model with investment risk. *Economic Theory*, 2018. doi:[10.1007/s00199-018-1103-2](https://doi.org/10.1007/s00199-018-1103-2). Forthcoming.

## A Asymptotic homogeneous problem

In this appendix we describe how to derive the asymptotic homogeneous problem in an abstract dynamic programming setting. For the notation, we follow [Ma and Stachurski \(2018\)](#). Let

- $X$  be a set called the *state space*;
- $A$  be a set called the *action space*;
- $\Gamma : X \rightarrow A$  be a nonempty correspondence called the *feasible correspondence*;
- $g : X \times A \rightarrow X$  be a function called the *law of motion*;
- $\mathcal{V}$  be a subset of all functions from  $X$  to  $\mathbb{R} \cup \{-\infty\}$  called the set of *candidate value functions*;
- $Q : X \times A \times \mathcal{V} \rightarrow \mathbb{R} \cup \{-\infty\}$  be a map called the *state-action aggregator*.

Then we say that the value function  $v \in \mathcal{V}$  satisfies the Bellman equation if

$$v(x) = \max_{a \in \Gamma(x)} Q(x, a, v(g(x, a))) \quad (\text{A.1})$$

for all  $x \in X$ .

**Definition A.1** (Asymptotic homogeneity). We say that the dynamic programming problem is *asymptotically homogeneous* if has the following properties:

- $X = X_1 \times X_2$ , where  $\mathbb{R}_+ \subset X_1 \subset \mathbb{R}$ ;
- $\Gamma(x) = \Gamma_1(x_1, x_2) \times \Gamma_2(x_2)$ , where  $x = (x_1, x_2) \in X_1 \times X_2$  and  $\mathbb{R}_+^d \subset \Gamma_1(x_1, x_2) \subset \mathbb{R}^d$  for some  $d$ ;
- $g(x, a) = g_1(x_1, x_2, a_1, a_2) \times g_2(x_2, a_2)$ , where  $x = (x_1, x_2) \in X_1 \times X_2$  and  $(a_1, a_2) \in \Gamma_1(x_1, x_2) \times \Gamma_2(x_2)$ ;
- $\lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} \Gamma_1(\lambda x_1, x_2) = \tilde{\Gamma}_1(x_1, x_2)$  exists for  $(x_1, x_2) \in X_1 \times X_2$ ;
- $\lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} g_1(\lambda x_1, x_2, \lambda a_1, a_2) = \tilde{g}_1(x_1, x_2, a_1, a_2)$  exists for  $(x_1, x_2) \in X_1 \times X_2$  and  $(a_1, a_2) \in \Gamma_1(x_1, x_2) \times \Gamma_2(x_2)$ ;
- $\lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} Q(\lambda x_1, x_2, \lambda a_1, a_2, \lambda v) = \tilde{Q}(x_1, x_2, a_1, a_2, v)$  exists.

**Lemma A.2.** *Suppose that the dynamic programming problem is asymptotically homogeneous. Then*

(i)  $\tilde{\Gamma}_1$  is homogeneous of degree 1 in  $x_1$ : for any  $\lambda > 0$  we have

$$\tilde{\Gamma}_1(\lambda x_1, x_2) = \lambda \tilde{\Gamma}_1(x_1, x_2).$$

(ii)  $\tilde{g}_1$  is homogeneous of degree 1 in  $(x_1, a_1)$ : for any  $\lambda > 0$  we have

$$\tilde{g}_1(\lambda x_1, x_2, \lambda a_1, a_2) = \lambda \tilde{g}_1(x_1, x_2, a_1, a_2).$$

(iii)  $\tilde{Q}$  is homogeneous of degree 1 in  $(x_1, a_1, v)$ : for any  $\lambda > 0$  we have

$$\tilde{Q}(\lambda x_1, x_2, \lambda a_1, a_2, \lambda v) = \lambda \tilde{Q}(x_1, x_2, a_1, a_2, v).$$

*Proof.* By the definition of  $\tilde{\Gamma}_1$ , for any  $\lambda > 0$  we have

$$\begin{aligned} \tilde{\Gamma}_1(\lambda x_1, x_2) &= \lim_{\lambda' \rightarrow \infty} \frac{1}{\lambda'} \Gamma_1(\lambda' \lambda x_1, x_2) \\ &= \lambda \lim_{\lambda' \rightarrow \infty} \frac{1}{\lambda' \lambda} \Gamma_1(\lambda' \lambda x_1, x_2) = \lambda \tilde{\Gamma}_1(x_1, x_2). \end{aligned}$$

The proofs of the other claims are similar. □

When the dynamic programming problem is asymptotically homogeneous, we define the asymptotic problem as follows.

**Definition A.3.** Suppose that the dynamic programming problem is asymptotically homogeneous. Then the Bellman equation of the asymptotic problem corresponding to (A.1) is defined by

$$v(x_1, x_2) = \max_{(a_1, a_2) \in \Gamma_1(x_1, x_2) \times \Gamma_2(x_2)} \tilde{Q}(x_1, x_2, a_1, a_2, v(\tilde{g}_1(x_1, x_2, a_1, a_2), g_2(x_2, a_2))). \quad (\text{A.2})$$

The following lemma shows that we can reduce the dimension of the asymptotic problem by 1.

**Lemma A.4.** *Suppose that the dynamic programming problem is asymptotically homogeneous. Consider the following “normalized” Bellman equation:*

$$\tilde{v}(x_2) = \max_{(a_1, a_2) \in \tilde{\Gamma}_1(1, x_2) \times \Gamma_2(x_2)} \tilde{Q}(1, x_2, a_1, a_2, \tilde{g}_1(1, x_2, a_1, a_2) \tilde{v}(g_2(x_2, a_2))). \quad (\text{A.3})$$

If (A.3) has a solution  $\tilde{v}(x_2)$ , then  $v(x_1, x_2) = x_1 \tilde{v}(x_2)$  is a solution to the asymptotic Bellman equation (A.2). Furthermore, letting  $\tilde{a} = (\tilde{a}_1, \tilde{a}_2)$  be the policy function of the normalized Bellman equation (A.3), the policy function  $a = (a_1, a_2)$  of the asymptotic Bellman equation (A.2) is given by  $a_1(x_1, x_2) = x_1 \tilde{a}_1(x_2)$  and  $a_2(x_1, x_2) = \tilde{a}_2(x_2)$ .

*Proof.* Immediate by multiplying both sides of (A.3) by  $x_1 > 0$  and using the homogeneity of  $\tilde{\Gamma}_1, \tilde{g}_1, \tilde{Q}$  established in Lemma A.2.  $\square$

The following proposition shows that if a dynamic programming problem is asymptotically homogeneous, then the value function and policy functions are asymptotically linear.

**Proposition A.5.** *Suppose that the dynamic programming problem is asymptotically homogeneous. Suppose that the Bellman equation (A.1) has a solution  $v(x)$ , and it can be computed by value function iteration starting from  $v(x) \equiv 0$ . Then under some regularity conditions, the value function and policy functions are asymptotically linear: we have*

$$\begin{aligned} v(x_1, x_2) &= x_1 \tilde{v}(x_2) + o(x_1), \\ a_1(x_1, x_2) &= x_1 \tilde{a}_1(x_2) + o(x_1), \\ a_2(x_1, x_2) &= \tilde{a}_2(x_2) + o(x_1) \end{aligned}$$

as  $x_1 \rightarrow \infty$ , where  $\tilde{v}(x_2)$ ,  $\tilde{a}_1(x_2)$ , and  $\tilde{a}_2(x_2)$  are defined as in the normalized Bellman equation (A.3).

*Proof.* Define the operator  $T : \mathcal{V} \rightarrow \mathcal{V}$  by the right-hand side of (A.1). Let  $v^{(0)} \equiv 0$  and  $v^{(k)} = T v^{(k-1)} = T^k 0$ . Let us show by induction that

$$\lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} v^{(k)}(\lambda x_1, x_2) = \tilde{v}^{(k)}(x_1, x_2)$$

exists. If  $k = 0$ , the claim is trivial since  $v^{(0)} \equiv 0$ . Suppose the claim holds for some  $k - 1$ . Then by Lemma A.2, we obtain

$$\begin{aligned} \frac{1}{\lambda} v^{(k)}(\lambda x_1, x_2) &= \frac{1}{\lambda} (T v^{(k-1)})(\lambda x_1, x_2) \\ &= \max_{(a_1, a_2) \in \frac{1}{\lambda} \Gamma_1(\lambda x_1, x_2) \times \Gamma_2(x_2)} Q \left( \lambda x_1, x_2, \lambda a_1, a_2, v^{(k-1)} \left( \lambda \frac{1}{\lambda} g_1(\lambda x_1, x_2, \lambda a_1, a_2), g_2(x_2, a_2) \right) \right). \end{aligned}$$

Using the asymptotic homogeneity of  $\Gamma_1, g_1, Q$  established in Lemma A.2, the asymptotic homogeneity of  $v^{(k-1)}$ , and assuming that we can interchange the limit and maximization (e.g., assuming enough conditions to apply the Maximum Theorem), it follows that  $v^{(k)}$  is asymptotically homogeneous. Since by assumption  $v^{(k)} \rightarrow v$  as  $k \rightarrow \infty$  point-wise, assuming that the limit of  $k \rightarrow \infty$  and  $\lambda \rightarrow \infty$  can be interchanged (which is the case if  $v^{(k)}$  converges to

$v$  monotonically, which is often the case in particular applications), then  $v$  is asymptotically homogeneous in the sense that  $\lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} v(\lambda x_1, x_2)$  exists.

Now that asymptotic homogeneity of  $v$  is established, from (A.1) we obtain

$$v(\lambda x_1, x_2) = \max_{a \in \Gamma(\lambda x_1, x_2)} Q(\lambda x_1, x_2, a, v(g(\lambda x_1, x_2, a))).$$

Dividing both sides by  $\lambda > 0$  and letting  $\lambda \rightarrow \infty$ , using the asymptotic homogeneity of  $\Gamma_1$ ,  $g_1$ ,  $Q$ , and  $v$ , we obtain the asymptotic Bellman equation (A.2). Thus if in particular (A.3) has a unique solution  $\bar{v}(x_2)$ , by Lemma A.4 it must be

$$\lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} v(\lambda x_1, x_2) = x_1 \bar{v}(x_2).$$

Consequently, setting  $x_1 = 1$  and  $\lambda = x_1$ , we obtain  $v(x_1, x_2) = x_1 \bar{v}(x_2) + o(x_1)$ . The proof for the policy functions is similar.  $\square$

## B Proofs

### B.1 Proof of results in Section 3

We use the following notations. Let  $\tilde{\beta} = \beta(1 - p)$  be the effective discount factor. For a vector  $v = (v_1, \dots, v_S)'$ , let  $v^{(\alpha)} = (v_1^\alpha, \dots, v_S^\alpha)'$  be the vector of  $\alpha$ -th powers and  $\text{diag}(v)$  the diagonal matrix whose  $s$ -th diagonal element is  $v_s$ .

The following proposition characterizes the solution to the capitalist's optimal consumption-savings problem.

**Proposition B.1.** *Let  $z = (z_1, \dots, z_S)'$  be the vector of gross excess returns. A solution to the optimal consumption-savings problem exists if and only if*

$$\tilde{\beta} R_f^{1-\gamma} \rho(\text{diag}(z^{(1-\gamma)})P) < 1. \quad (\text{B.1})$$

*Under this condition, the value function and optimal consumption rule are*

$$V_s(w) = b_s \frac{w^{1-\gamma}}{1-\gamma}, \quad (\text{B.2a})$$

$$c_s(w) = b_s^{-1/\gamma} w, \quad (\text{B.2b})$$

where  $b = (b_1, \dots, b_S)'$   $\gg 0$  is the smallest solution to the system of nonlinear equations

$$b_s = \left(1 + (\tilde{\beta}(z_s R_f)^{1-\gamma} \mathbb{E}[b_{s'} | s])^{1/\gamma}\right)^\gamma, \quad s = 1, \dots, S. \quad (\text{B.3})$$

*Proof.* Immediate from Toda (2018b, Proposition 1).  $\square$

Let us simplify the equilibrium condition (3.3) by exploiting the linearity in Proposition B.1. Using the budget constraint (3.2) and the optimal consumption rule (B.2b), the individual wealth dynamics is

$$w' = z_s R_f (1 - b_s^{-1/\gamma}) w =: G_s w.$$

Letting  $W_s$  be the aggregate wealth held by agents in state  $s$ , by accounting we obtain

$$W_{s'} = p\pi_{s'}w_0 + (1-p) \sum_{s=1}^S p_{ss'}G_s W_s.$$

Letting  $\pi = (\pi_1, \dots, \pi_S)'$ ,  $G = (G_1, \dots, G_S)'$ , and  $W = (W_1, \dots, W_S)'$ , in matrix form this becomes

$$W = pw_0\pi + (1-p)P'(\text{diag } G)W \iff W = pw_0(I - (1-p)P' \text{diag } G)^{-1}\pi.$$

Let  $m_s = b_s^{-1/\gamma} \in (0,1)$  be the marginal propensity to consume out of wealth in state  $s$  and  $m = (m_1, \dots, m_S)'$ . Then the vector of saving rates is given by  $1 - m$ , where  $1 = (1, \dots, 1)'$  is the vector of ones. Using this, the aggregate capital supply is given by

$$K = (1 - m)'W = pw_0(1 - m)'(I - (1-p)P' \text{diag } G)^{-1}\pi, \quad (\text{B.4})$$

assuming  $(1-p)\rho(P' \text{diag } G) < 1$ . (If this inequality is violated, we just set  $K = \infty$ .) On the other hand, by (3.1) the aggregate capital demand is

$$K = \left( \frac{R_f - 1 + \delta}{A\alpha} \right)^{\frac{1}{\alpha-1}}. \quad (\text{B.5})$$

Equating (B.4) and (B.5), the market clearing condition (3.3) becomes

$$0 = f(R_f) := pw_0(1 - m)'(I - (1-p)P' \text{diag } G)^{-1}\pi - \left( \frac{R_f - 1 + \delta}{A\alpha} \right)^{\frac{1}{\alpha-1}}. \quad (\text{B.6})$$

The following theorem shows that a stationary equilibrium exists and that the stationary wealth distribution has a Pareto upper tail.

**Theorem B.2.** *A stationary equilibrium exists if and only if there exists  $\underline{R} > 1 - \delta$  such that*

$$\tilde{\beta}\underline{R}^{1-\gamma}\rho(\text{diag}(z^{(1-\gamma)})P) < 1 \quad (\text{B.7})$$

and  $f(\underline{R}) < 0$ , where  $f$  is given by (B.6). The equilibrium is unique if  $\gamma < 1$ . If in addition  $p_{ss} > 0$  and  $G_s > 1$  for some  $s$ , then the stationary wealth distribution has a Pareto upper tail with exponent  $\zeta > 1$  that satisfies

$$\rho(P \text{diag } G^{(\zeta)}) = \frac{1}{1-p}. \quad (\text{B.8})$$

*Proof.* The existence of equilibrium follows from a continuity argument similar to Toda (2018b, Theorem 3), which we only sketch for space considerations. The condition (B.7) ensures that (B.1) holds for  $R_f > \underline{R}$  sufficiently close to  $\underline{R}$ . Then we can show that the individual optimization problem has a solution and the aggregate wealth is finite for some range  $R_f \in [\underline{R}, \bar{R})$ , and that the aggregate wealth (as well as supply of capital) diverges to  $\infty$  as  $R_f \uparrow \bar{R}$ . Since  $f(\underline{R}) < 0$  by assumption and  $f(\bar{R}) = \infty$ , by the intermediate value theorem there exists  $R_f \in (\underline{R}, \bar{R})$  that satisfies the market clearing condition (B.6). Uniqueness of equilibrium when  $\gamma < 1$  follows by the exact same argument as in Toda (2018b, Theorem 3).

The Pareto tail result follows from the general theorem in [Beare and Toda \(2017\)](#) and a similar argument to [Toda \(2018b, Theorem 4\)](#).  $\square$

Numerically solving for the equilibrium is straightforward. Given the guess of the interest rate  $R_f > \underline{R}$ , solve for the fixed point  $b = (b_s)$  using [\(B.3\)](#), and solve [\(B.6\)](#) to obtain the equilibrium risk-free rate.

## B.2 Proof of results in Section 4

**Proof of Proposition 4.2.** The maximization problem [\(4.12\)](#) is equivalent to

$$\max_{0 \leq \theta \leq \bar{\theta}_s} \frac{1}{1-\gamma} \mathbb{E} \left[ (z_{sj}\theta + 1 - \theta)^{1-\gamma} \mid s \right].$$

Let  $f(\theta)$  be the objective function of this problem. Since by assumption  $z_s + \epsilon_1 < 1$  and  $z_1 < \dots < z_s$ , we have  $z_s + \epsilon_1 < 1$  for all  $s$ . Therefore  $\bar{\theta}_s > 0$  and

$$f'(\theta) = \mathbb{E} \left[ (z_{sj}\theta + 1 - \theta)^{-\gamma} (z_{sj} - 1) \mid s \right] \rightarrow -\infty$$

as  $\theta \uparrow \bar{\theta}_s = \frac{1}{1-z_{s1}}$ . Furthermore,

$$f''(\theta) = -\gamma \mathbb{E} \left[ (z_{sj}\theta + 1 - \theta)^{-\gamma-1} (z_{sj} - 1)^2 \mid s \right] < 0$$

for  $\theta \in (0, \bar{\theta}_s)$ , so  $f$  is strictly concave. Therefore there exists a unique  $\theta_s^*$  that maximizes [\(4.12\)](#), and hence  $\rho_s$  is well-defined.

Assume  $\epsilon \neq 1$ . By the discussion in the text, the Bellman equation of the asymptotic problem is

$$v_s(w) = \max_{\substack{0 \leq \theta \leq \bar{\theta}_s \\ 0 \leq c \leq w}} \left( (1-\beta)c^{1-1/\epsilon} + \beta \mathbb{E} \left[ (v_{s'}(R(\theta)(w-c)))^{1-\gamma} \mid s \right]^{\frac{1-1/\epsilon}{1-\gamma}} \right)^{\frac{1}{1-1/\epsilon}},$$

where the upper bounds on  $c, \theta$  ensure that  $w' \geq 0$  and

$$R(\theta) = R_{sj}\theta + R_f(1-\theta) = R_f(z_{sj}\theta + 1 - \theta)$$

is the gross portfolio return. By homogeneity, the value function must be of the form  $v_s(w) = b_s w$ . Substituting into the Bellman equation, we obtain

$$b_s w = \max_{\substack{0 \leq \theta \leq \bar{\theta}_s \\ 0 \leq c \leq w}} \left( (1-\beta)c^{1-1/\epsilon} + \beta(w-c)^{1-1/\epsilon} \mathbb{E} \left[ (b_{s'} R(\theta))^{1-\gamma} \mid s \right]^{\frac{1-1/\epsilon}{1-\gamma}} \right)^{\frac{1}{1-1/\epsilon}}.$$

Noting that  $R(\theta)$  does not depend on  $s'$  and  $j$  is independent of  $s'$ , using the definition of  $\rho_s$  in [\(4.12\)](#), we can rewrite this as

$$b_s w = \max_{0 \leq c \leq w} \left( (1-\beta)c^{1-1/\epsilon} + \beta(R_f \rho_s)^{1-1/\epsilon} (w-c)^{1-1/\epsilon} \mathbb{E} \left[ b_{s'}^{1-\gamma} \mid s \right]^{\frac{1-1/\epsilon}{1-\gamma}} \right)^{\frac{1}{1-1/\epsilon}}.$$

For notational simplicity let  $\kappa_s = R_f \rho_s \mathbb{E} \left[ b_{s'}^{1-\gamma} \mid s \right]^{\frac{1}{1-\gamma}}$ . Then the above problem becomes equivalent to

$$\max_c \frac{1}{1-1/\varepsilon} \left( (1-\beta)c^{1-1/\varepsilon} + \beta \kappa_s^{1-1/\varepsilon} (w-c)^{1-1/\varepsilon} \right).$$

Clearly this is a strictly concave function in  $c$ . Taking the first-order condition and solving for  $c$ , we obtain

$$c = \frac{(1-\beta)^\varepsilon}{(1-\beta)^\varepsilon + \beta^\varepsilon \kappa_s^{\varepsilon-1}} w.$$

Substituting into the Bellman equation, after some algebra we obtain

$$b_s = \left( (1-\beta)^\varepsilon + \beta^\varepsilon (R_f \rho_s)^{\varepsilon-1} \mathbb{E} \left[ b_{s'}^{1-\gamma} \mid s \right]^{\frac{\varepsilon-1}{1-\gamma}} \right)^{\frac{1}{\varepsilon-1}},$$

which is (4.14). The optimal consumption rule then simplifies to  $c = (1-\beta)^\varepsilon b_s^{1-\varepsilon}$  and we obtain the optimal investment rule using  $\theta = \frac{I}{w-c}$ .

To complete the proof it remains to show that the system of nonlinear equations (4.14) has a solution. For this purpose let us write  $\sigma = \frac{1-\gamma}{\varepsilon-1}$  and  $x_s = b_s^{1-\gamma}$ . Then we can rewrite (4.14) as

$$x_s = \left( (1-\beta)^\varepsilon + (\beta^{\varepsilon\sigma} (R_f \rho_s)^{1-\gamma} \mathbb{E} [x_{s'} \mid s])^{1/\sigma} \right)^\sigma,$$

which is equivalent to

$$x = ((1-\beta)^\varepsilon + (Kx)^{1/\sigma})^\sigma$$

for  $x = (x_1, \dots, x_S)'$  and  $K = \beta^{\varepsilon\sigma} R_f^{1-\gamma} \text{diag}(\rho_1^{1-\gamma}, \dots, \rho_S^{1-\gamma})P$ . Since this equation is identical to Equation (12) in [Borovička and Stachurski \(2017\)](#), by their Theorem 2.1, a necessary and sufficient condition for the existence of a unique fixed point is  $\rho(K)^{1/\sigma} < 1$ , which is equivalent to (4.13).

Finally we briefly comment on the case  $\varepsilon = 1$ . Although this case requires a separate treatment, it turns out that the equations are valid by taking the limit  $\varepsilon \rightarrow 1$ . To show (4.14), define  $g(\varepsilon) = \log((1-\beta)^\varepsilon + \beta^\varepsilon \kappa^{\varepsilon-1})$  for  $\kappa > 0$ . Then as  $\varepsilon \rightarrow 1$  we obtain

$$\log((1-\beta)^\varepsilon + \beta^\varepsilon \kappa^{\varepsilon-1})^{\frac{1}{\varepsilon-1}} = \frac{g(\varepsilon)}{\varepsilon-1} = \frac{g(\varepsilon) - g(1)}{\varepsilon-1} \rightarrow g'(1).$$

But since

$$g'(\varepsilon) = \frac{(1-\beta)^\varepsilon \log(1-\beta) + \beta^\varepsilon \kappa^{\varepsilon-1} \log(\beta \kappa)}{(1-\beta)^\varepsilon + \beta^\varepsilon \kappa^{\varepsilon-1}} \rightarrow (1-\beta) \log(1-\beta) + \beta \log(\beta \kappa)$$

as  $\varepsilon \rightarrow 1$ , it follows that

$$((1-\beta)^\varepsilon + \beta^\varepsilon \kappa^{\varepsilon-1})^{\frac{1}{\varepsilon-1}} \rightarrow (1-\beta)^{1-\beta} (\beta \kappa)^\beta,$$

which explains (4.14) for  $\varepsilon = 1$ . The existence and uniqueness of a positive solution can be proved by taking the logarithm of (4.14) and applying a contraction mapping argument to  $x = \log b$ .  $\square$

**Proof of Lemma 4.3.** By Proposition 4.2 and the budget constraint of the asymptotic problem, we obtain the the law of motion

$$w' = (1 - (1 - \beta)^\varepsilon b_s^{1-\varepsilon}) R_f(z_{s_j} \theta_s^* + 1 - \theta_s^*) w.$$

Taking the expectation conditional on  $s$  and using the definition of  $G_s$  in (4.16), we obtain  $E[w' | s] = G_s w$ . By the same derivation as (B.4), a necessary condition for aggregate wealth to be finite is  $\rho(P' \text{diag } G) < 1$ . Since  $\text{diag } G$  is diagonal, we obtain

$$\rho(P \text{diag } G) = \rho((\text{diag } G)' P') = \rho((\text{diag } G) P') = \rho(P' \text{diag } G) < 1. \quad \square$$

**Proof of Lemma 4.4.** Since by assumption  $G_{s_j} > 1$  for some  $s$ , we have  $M_s(z) \rightarrow \infty$  as  $z \rightarrow \infty$  for this  $s$ . Since by assumption  $p_{ss} > 0$  for all  $s$ , it follows that  $\rho(PD(z)) \rightarrow \infty$  as  $z \rightarrow \infty$ . Since  $D(1) = \text{diag } G$ , by (4.17) we obtain  $\rho(PD(1)) = \rho(P \text{diag } G) < 1$ . By the intermediate value theorem, there exists  $\zeta > 1$  such that  $\rho(PD(\zeta)) = 1$ . Uniqueness follows from the convexity of  $\rho(PD(z))$  established in Beare and Toda (2017). The Pareto tail result follows from (2.5) with  $p = 0$ .  $\square$

## C Constructing the exponential grid

Suppose we would like to construct an  $N$ -point exponential grid on a given interval  $(a, b)$ . This problem is nontrivial because  $a, b$  could be negative and we cannot take the logarithm of negative numbers. A natural idea to deal with such a case is as follows.

### Constructing the exponential grid.

1. Choose a shift parameter  $s > -a$ .
2. Construct an  $N$ -point evenly-spaced grid on  $(\log(a + s), \log(b + s))$ .
3. Take the exponential.
4. Subtract  $s$ .

The remaining question is how to choose the shift parameter  $s$ . Suppose we would like to specify the median grid point as  $c \in (a, b)$ . Since the median of the evenly-spaced grid on



$(\log(a + s), \log(b + s))$  is  $\frac{1}{2}(\log(a + s) + \log(b + s))$ , we need to take  $s > -a$  such that

$$\begin{aligned} c &= \exp\left(\frac{1}{2}(\log(a + s) + \log(b + s))\right) - s \\ \iff c + s &= \sqrt{(a + s)(b + s)} \\ \iff (c + s)^2 &= (a + s)(b + s) \\ \iff c^2 + 2cs + s^2 &= ab + (a + b)s + s^2 \\ \iff s &= \frac{c^2 - ab}{a + b - 2c}. \end{aligned}$$

Note that in this case

$$s + a = \frac{c^2 - ab}{a + b - 2c} + a = \frac{(c - a)^2}{a + b - 2c},$$

so  $s + a$  is positive if and only if  $c < \frac{a+b}{2}$ . Therefore, for any  $c \in \left(a, \frac{a+b}{2}\right)$ , it is possible to construct such a grid.

The remaining question is how to choose  $c$ , but we can use information from the problem we want to solve. Note that by construction, half the grid points will lie on the interval  $(a, c)$ . Therefore we should choose the number  $c$  such that  $c$  is a “typical” value for the state variable (e.g., initial capital, aggregate capital, etc.).

## D Simulating the Aiyagari model

In this appendix we assess the accuracy of the simulation method using the Aiyagari model in Section 3. We conduct simulations by recursively computing the wealth of  $I$  agents over  $T$  periods using the semi-analytical solution for the consumption policies and the risk-free rate. We initialize the wealth distribution at  $T = 0$  by setting  $w = 1$  for all agents. For every simulation, we compute the cross-sectional mean of the wealth distribution  $\widehat{K}$  in the terminal period  $T$ , which corresponds to the aggregate capital  $K$  in the model.<sup>20</sup> Given that  $\widehat{K}$  is a random variable, we compute the relative error for  $B = 10,000$  independent simulations and take the average to obtain the “mean relative error” defined by  $\frac{1}{B} \sum_{b=1}^B \left| \widehat{K}_b / K - 1 \right|$ , where  $K$  is the true aggregate capital and  $\widehat{K}_b$  is the numerical aggregate capital from simulation  $b$ . Finally, we compute a measure of the dispersion of results across simulations defined by  $\left| \widehat{K}_{p95} / \widehat{K}_{p5} - 1 \right|$ , where  $\widehat{K}_{p5}, \widehat{K}_{p95}$  denote the 5th and 95th percentiles of  $\widehat{K}$  across simulations. Table 12 shows the results for different combinations of sample size  $I$  and simulation length  $T$ .

A few remarks are in order. First, the simulation method performs poorly on average and the results are very dispersed across simulations. Second, increasing the number of agents  $I$  helps reduce both the mean relative error and the dispersion. The gains in accuracy associated with increasing  $I$  by an order of magnitude tend to be small, consistently with the fact that the sample mean of a fat-tailed distribution converges very slowly to the population mean. In fact, in our model the Pareto exponent is  $\zeta = 1.28$ , and the relative errors for sample sizes  $I = 10^4, 10^6$  in Table 12, which are about 11% and 7% respectively, are about the same order

<sup>20</sup>Since there is a unit continuum of agents in the model, the average wealth is equal to the aggregate wealth.

Table 12: Solution accuracy of the simulation method in the Aiyagari model.

$T$	$I$	Relative error (%)	Dispersion (%)	Time (sec)
1,000	$10^3$	21.93	92.97	0.07
	$10^4$	10.97	47.99	0.38
	$10^5$	6.70	27.53	3.13
	$10^6$	6.69	16.33	41.23
10,000	$10^3$	23.76	96.68	0.39
	$10^4$	20.42	50.07	3.76
	$10^5$	9.69	27.47	31.43
	$10^6$	6.64	16.07	414.30

Note:  $T$ : simulation length in model-years;  $I$ : number of simulated agents; Relative error:  $\frac{1}{B} \sum_{b=1}^B |\hat{K}_b/K - 1|$ , where  $\hat{K}_b$  is the aggregate capital in simulation  $b$  and  $K$  is the true value from the analytical solution; Dispersion:  $|\hat{K}_{p95}/\hat{K}_{p5} - 1|$ , where  $\hat{K}_{p5}, \hat{K}_{p95}$  are the 5th and 95th percentiles of aggregate capital across 1000 simulations; Time: computing time of 10,000 simulations in minutes.

of magnitude as the error order 11.9% and 4.1% in the column for  $\zeta = 1.3$  in Table 1. Third, increasing the simulation length beyond  $T = 1,000$  does not seem to improve accuracy, suggesting that the simulated wealth distribution has already converged after 1,000 periods.

The last column of Table 12 reports the computing time (without parallelization) associated with producing a *single* simulation using a machine equipped with an Intel Xeon E3-1245 3.5GHz processor and 16GB of memory. While increasing the sample size  $I$  by a factor of ten is associated with small accuracy gains, it implies a tenfold increase in the computing time. Given the inaccuracy and the large dispersion of results across simulations, we conclude that the simulation method is not a viable option for solving models with fat-tailed wealth distributions.

## E Algorithm for Merton-Bewley-Aiyagari model

### E.1 Euler and asset pricing equations

First, we derive the Euler and asset pricing equations. Noting that  $z_{sj}$  is strictly increasing in  $j$ , the borrowing constraint (4.5) can bind only in state  $j = 1$ . Therefore the Bellman equation (4.6) is equivalent to

$$\begin{aligned} \frac{1}{1 - 1/\varepsilon} v_s(w)^{1-1/\varepsilon} &= \max_{c, I \geq 0} \frac{1}{1 - 1/\varepsilon} \left( (1 - \beta)c^{1-1/\varepsilon} + \beta \mathbb{E} \left[ v_{s'}(w')^{1-\gamma} \mid s \right]^{\frac{1-1/\varepsilon}{1-\gamma}} \right), \\ R_f(w + (1 - \tau_h)\omega h_s + I(z_{s1} - 1) - c) &\geq \underline{w}, \\ I &\geq 0, \end{aligned}$$

where  $w'$  is as in (4.5). Let  $\mathcal{L}_s(c, I, w)$  be the Lagrangian and  $\lambda_s(w), \mu_s(w)$  be the corresponding Lagrange multipliers for the borrowing constraint and the nonnegativity constraint on invest-

ment:

$$\begin{aligned} \mathcal{L}_s(c, I, w) = & \frac{1}{1-1/\varepsilon} \left( (1-\beta)c^{1-1/\varepsilon} + \beta \mathbb{E} \left[ v_{s'}(w')^{1-\gamma} \mid s \right]^{\frac{1-1/\varepsilon}{1-\gamma}} \right) \\ & + \lambda_s(w) (R_f(w + (1-\tau_h)\omega h_s + I(z_{s1} - 1) - c) - \bar{w}) + \mu_s(w)I. \end{aligned}$$

The first-order condition for consumption is given by

$$(1-\beta)c_s(w)^{-1/\varepsilon} = \beta R_f \mathbb{E} \left[ v_{s'}(w')^{1-\gamma} \mid s \right]^{\frac{\gamma-1/\varepsilon}{1-\gamma}} \mathbb{E} [v_{s'}(w')^{-\gamma} v'_{s'}(w') \mid s] + \lambda_s(w)R_f. \quad (\text{E.1})$$

Differentiating both sides of the Bellman equation by  $w$ , it follows from the Envelope Theorem that

$$v_s(w)^{-1/\varepsilon} v'_s(w) = \beta R_f \mathbb{E} \left[ v_{s'}(w')^{1-\gamma} \mid s \right]^{\frac{\gamma-1/\varepsilon}{1-\gamma}} \mathbb{E} [v_{s'}(w')^{-\gamma} v'_{s'}(w') \mid s] + \lambda_s(w)R_f. \quad (\text{E.2})$$

By (E.1) and (E.2), we obtain the following expression for the derivative of the value function

$$v'_s(w) = (1-\beta) \left( \frac{c_s(w)}{v_s(w)} \right)^{-1/\varepsilon}. \quad (\text{E.3})$$

Substituting (E.3) into (E.1), we obtain the consumption Euler equation

$$c_s(w)^{-1/\varepsilon} = \beta R_f \mathbb{E} \left[ \left( \frac{v_{s'}(w')}{\mathbb{E} [v_{s'}(w')^{1-\gamma} \mid s]^{\frac{1}{1-\gamma}}} \right)^{1/\varepsilon-\gamma} c_{s'}(w')^{-1/\varepsilon} \mid s \right] + \frac{\lambda_s(w)R_f}{1-\beta}. \quad (\text{E.4})$$

The first-order condition for investment is given by

$$\begin{aligned} \beta R_f \mathbb{E} \left[ v_{s'}(w')^{1-\gamma} \mid s \right]^{\frac{\gamma-1/\varepsilon}{1-\gamma}} \mathbb{E} [v_{s'}(w')^{-\gamma} v'_{s'}(w')(z_{sj} - 1) \mid s] \\ + \lambda_s(w)R_f(z_{s1} - 1) + \mu_s(w) = 0. \end{aligned} \quad (\text{E.5})$$

Using the expression for  $v'_s(w)$  from (E.3) and rearranging, we obtain the following asset pricing equation

$$\begin{aligned} \mathbb{E} \left[ \left( \frac{v_{s'}(w')}{\mathbb{E} [v_{s'}(w')^{1-\gamma} \mid s]^{\frac{1}{1-\gamma}}} \right)^{1/\varepsilon-\gamma} c_{s'}(w')^{-1/\varepsilon} (z_{sj} - 1) \mid s \right] \\ = - \frac{\lambda_s(w)(z_{s1} - 1) + \mu_s(w)/R_f}{\beta(1-\beta)}. \end{aligned} \quad (\text{E.6})$$

## E.2 Policy function and value function iteration

First, we choose an exponential grid as described in Appendix C and replacing the bottom half grid points with an evenly-spaced grid. The grid has 50 points and we set the truncation point to 1000 times the typical scale of the model (steady-state capital stock in the representative-agent model with the same parametrization). The strategy is start with guesses for the value func-

tion  $v_s^{\text{old}}(w)$  and policy functions  $c_s^{\text{old}}(w)$ ,  $I_s^{\text{old}}(w)$  and update by doing the following steps for each individual state  $(s, w)$ . Equipped with the asymptotic solution to the individual problem  $\{\bar{c}_s, \bar{I}_s, \bar{v}_s\}_{s=1}^S$ , we choose the following initial guesses

$$c_s(w) = \bar{c}_s(w - \underline{w}), \quad I_s(w) = \bar{I}_s(w - \underline{w}), \quad v_s(w) = \bar{I}_s(w - \underline{w}).$$

Then, we iterate over steps (1), (2), and (3) until convergence.

1. **Consumption decision:** First, use the old decision rules  $c_s^{\text{old}}(w)$ ,  $I_s^{\text{old}}(w)$  to construct next period's wealth  $w'_{sj}(w)$  and the upper bound on consumption  $c_s^{\text{ub}}(w)$ :

$$\begin{aligned} w'_{sj}(w) &= R_f \left( w + (1 - \tau_h)\omega h_s + I_s^{\text{old}}(w)(z_{sj} - 1) - c_s^{\text{old}}(w) \right), \\ c_s^{\text{ub}}(w) &= w + (1 - \tau_h)\omega h_s - \underline{w}/R_f. \end{aligned}$$

Then, update the consumption using the Euler equation (E.4) as if the borrowing constraint was slack.

$$c_s^*(w) = \left( \beta R_f \mathbb{E} \left[ \left( \frac{v_{s'}^{\text{old}}(w'_{sj}(w))}{\mathbb{E} [v_{s'}^{\text{old}}(w'_{sj}(w))^{1-\gamma} | s]^{\frac{1}{1-\gamma}}} \right)^{1/\varepsilon-\gamma} c_{s'}^{\text{old}}(w'_{sj}(w))^{-1/\varepsilon} \middle| s \right] \right)^{-\varepsilon}.$$

Finally, impose the upper bound on consumption so that  $c_s^{\text{new}}(w) = \min \{c_s^*(w), c_s^{\text{ub}}(w)\}$  and compute the Lagrange multiplier

$$\lambda_s^{\text{new}}(w) = \frac{1 - \beta}{R_f} \left( c_s^{\text{new}}(w)^{-1/\varepsilon} - c_s^*(w)^{-1/\varepsilon} \right).$$

2. **Investment decision:** This step applies only to types  $s$  such that  $\mathbb{E} [z_{sj} | s] > 1$ , for other types set  $I_s(w) = 0$ .

Use the new consumption rule  $c_s^{\text{new}}(w)$  to construct an upper bound on investment

$$I_s^{\text{ub}}(w) = \frac{w + (1 - \tau_h)\omega h_s - c_s^{\text{new}}(w) - \underline{w}/R_f}{1 - z_{s1}}.$$

Construct a grid of feasible choices for investment over the range  $[0, I_s^{\text{ub}}(w)]$ . Using  $c_s^{\text{new}}(w)$ , compute next period's wealth for every choice of  $I$

$$w'_{sj}(w, I) = R_f \left( w + (1 - \tau_h)\omega h_s + I(z_{sj} - 1) - c_s^{\text{new}}(w) \right).$$

Finally, to solve for the optimal investment, conjecture that investment is interior and find the root  $I^*$  that solves the asset pricing equation (E.6) with  $\mu_s(w) = 0$ ,

$$\mathbb{E} \left[ \left( \frac{v_{s'}^{\text{old}}(w')}{\mathbb{E} [v_{s'}^{\text{old}}(w')^{1-\gamma} | s]^{\frac{1}{1-\gamma}}} \right)^{1/\varepsilon-\gamma} c_{s'}^{\text{new}}(w')^{-1/\varepsilon} (z_{sj} - 1) \middle| s \right] = -\frac{\lambda_s(w)^{\text{new}} (z_{s1} - 1)}{\beta(1 - \beta)}$$

and set  $I_s(w) = I^*$ . If a root does not exist and the left-hand side is everywhere positive, set  $I_s(w) = I_s^{\text{ub}}(w)$ . If a root does not exist and the left-hand side is everywhere negative, set  $I_s(w) = 0$ .

3. **Value function:** Construct next period's wealth using the new policy functions

$$w'_{sj}(w) = R_f (w + (1 - \tau_h)\omega h_s + I_s^{\text{new}}(w)(z_{sj} - 1) - c_s^{\text{new}}(w))$$

and update the value function

$$v_s^{\text{new}}(w) = \begin{cases} \left( (1 - \beta)c_s^{\text{new}}(w)^{1-1/\varepsilon} + \beta \mathbf{E} \left[ v_{s'}^{\text{old}}(w'_{sj}(w))^{1-\gamma} \mid s \right]^{\frac{1-1/\varepsilon}{1-\gamma}} \right)^{\frac{1}{1-1/\varepsilon}}, & (\varepsilon \neq 1) \\ c_s^{\text{new}}(w)^{1-\beta} \left( \mathbf{E} \left[ v_{s'}^{\text{old}}(w'_{sj}(w))^{1-\gamma} \mid s \right]^{\frac{1}{1-\gamma}} \right)^{\beta}. & (\varepsilon = 1) \end{cases}$$