

The Double Power Law in Consumption and Implications for Estimating Asset Pricing Models

Alexis Akira Toda* Kieran Walsh*

December 31, 2012

Abstract

Contrary to conventional wisdom, we find that the cross-sectional distributions of U.S. consumption and consumption growth are better described by the *double Pareto-lognormal* (dPIN) distribution, which obeys the power law in both tails, than by the lognormal distribution. Since the power law exponent is about 4 in both cases, the population moments of order > 4 or < -4 are not likely to exist. In light of this finding, we reevaluate the models of a number of consumption-based asset pricing papers that use the Consumer Expenditure Survey to build ‘Law of Large Numbers’ stochastic discount factors (SDFs). We draw three main conclusions. First, many estimators in the literature may be inconsistent. Second, the power law in consumption appears to have the ability to generate spurious acceptance of asset pricing models in explaining the equity premium. While dividing the sample into age cohorts allows us to mitigate the problem, the data do not support the SDFs that are robust to the power law. Third, estimation using simulated data from a general equilibrium model supports our theory that the power law interferes with testing and estimating heterogeneous agent asset pricing models. Our results are important for any econometric analysis with potentially fat-tailed data, not just for analyses of consumption data and asset prices.

1 Introduction

The famous “equity premium puzzle” of Mehra and Prescott (1985) is the failure of the representative agent, consumption-based asset pricing model of Lucas (1978) to reconcile, for a reasonable value of relative risk aversion, the U.S. aggregate consumption time series with the equity premium time series. Many authors have thus explored the ability of heterogeneous agent and incomplete markets models to resolve the equity premium puzzle and other inconsistencies between the representative agent approach and financial/macro data. While this literature consists mostly of theoretical papers,¹ some authors use household-level consumption data (Consumption Expenditure Survey, CEX) to empirically

*Department of Economics, Yale University. We benefited from comments by Tony Smith. acknowledges the financial support from the Cowles Foundation, the Nakajima Foundation, and Yale University. KW acknowledges financial support from the Cowles Foundation and Yale University.

¹See Constantinides and Duffie (1996), Heaton and Lucas (1996), Krusell and Smith (1997, 1998), Guvenen (2009), Geanakoplos (2010), among many others.

analyze the heterogeneous agent, incomplete market approach.² The CEX papers have so far yielded mixed results regarding the theory. This is puzzling because the approaches these papers take do not differ much except with respect to seemingly minor details such as the length of the sample period, data cleaning methods, and approximation methods such as log-linearization.³

This paper has two main contributions. First, contrary to conventional wisdom, we find that the cross-sectional distributions of consumption and consumption growth are better described by the *double Pareto-lognormal* (dPIN) distribution, which obeys the power law in both tails, than by the lognormal distribution. More specifically, we find that dPIN is not rejected by goodness-of-fit tests in about 80% of our sample, whereas the lognormal distribution is almost always rejected. The estimated power law exponent is around 4 for both consumption and consumption growth. Our finding does not contradict Battistin et al. (2009), who support the lognormal distribution *within age cohorts*, because we find the power law in the entire cross-section. That the power law emerges in the entire cross-section but not necessarily within age cohorts is theoretically supported by one of the authors' incomplete market general equilibrium model (Toda, 2012c).

The second contribution is an analysis of the implications of the power law for estimating and testing consumption-based asset pricing models. Since the power law exponent is about 4, the population moments of consumption and consumption growth of order > 4 or < -4 are not likely to exist. If this is the case, the estimation of relative risk aversion γ using some 'Law of Large Numbers' stochastic discount factors (SDFs) proposed in the literature may be inconsistent when the GMM criterion does not have a limit for γ larger than 4. But what is the effect of this on γ estimation and model fit? To address this issue we repeatedly sample from the data to generate bootstrap samples for γ and the pricing error (unexplained equity premium). We find that the histogram of the pricing errors is bimodal, with one peak at zero and the other at a significantly nonzero value. The zero pricing errors tend to correspond to γ estimates larger than 4, in the moment nonexistence range. Therefore, it seems that forming the sample analogs of nonexistent moments may mechanically aid in explaining the equity premium.

To see whether this odd behavior is due to the power law, we perform two robustness checks. First, we drop the top and bottom 100 data points, which correspond to less than 0.05% of the entire sample. Second, we split the entire sample into age cohorts and estimate an overidentified model, which is less susceptible to the power law issue. In both cases the zero pricing error peak disappears, suggesting that this peak is an artifact of the power law. After reevaluating each stochastic discount factor model using the age cohort robust estimation, we find that none of the models is supported by data. For example, overidentifying tests reject all of the models we consider.

We also perform Monte Carlo studies of simulated consumption data using the incomplete market general equilibrium model of Toda (2012c), which

²Prominent examples are Brav et al. (2002), Cogley (2002), Vissing-Jørgensen (2002), Balduzzi and Yao (2007), and Kocherlakota and Pistaferri (2009). See Ludvigson (2012) for a survey of this literature.

³One exception is Kocherlakota and Pistaferri (2009), who test the standard Euler equation as well as the inverse Euler equation, which arises when there is private information but agents are able to use insurance to achieve constrained Pareto efficient allocations.

is analytically solvable, and again find that the γ estimates zeroing the pricing error are often in the nonexistence range. Furthermore, in many Monte Carlo runs, the GMM criterion has multiple troughs, one near γ and one in the nonexistence range. Sometimes, the latter trough is the global minimum of the criterion. The age cohort robust method, however, almost never exhibits multiple troughs. Since in this case the model is correctly specified and there is a power law, the fat tails appear to be causing the estimation failure.

We draw three main conclusions from our study. First, the distributions of observed consumption and consumption growth have much heavier tails than economists typically believe. Moments outside the range of $[-4, 4]$ are not likely to exist. Second, in the presence of fat tails the estimation of model parameters by standard econometric techniques may fail due to the nonexistence of moments. However, since the calculation of sample moments is always possible (even if the population moments do not exist) and the estimation results may appear reasonable, the researcher faces the danger of falsely accepting the model. Also, we establish the bootstrap as a simple but effective method of finding spurious results. In our case, the histogram of the bootstrapped variable is bimodal, exhibiting a sharp, spurious peak. Third, the odd behavior we observe using CEX data also occurs when we use simulated data, which has fat tails by construction. Broadly, we interpret our results as a rejection of models with identical constant relative risk aversion utility and interior solutions, not of heterogeneous agent, consumption-based theories of asset prices in general.

Finally, note that our results are not specific to analyses of consumption data and the equity premium. Our warnings could apply to any econometric study with potentially fat-tailed data.

The rest of the paper is organized as follows. Section 2 summarizes the literature of testing consumption-based asset pricing models using CEX data and discusses potential inconsistencies under nonexistence of moments. Section 3 defines the power law, discusses an incomplete market general equilibrium model that generates power law in the cross-sectional consumption distribution, and documents the evidence for the power law in consumption and consumption growth. Section 4 evaluates all of stochastic discount factors proposed in the CEX literature, introduces our robust, cohort-based method, and presents bootstrap results. Section 5 performs Monte Carlo studies with simulated data.

2 Summary of asset pricing models

2.1 Literature

Consider an economy with a continuum of agents. Assuming every agent has an identical, additively separable, constant relative risk aversion (CRRA) utility function $E_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\gamma}}{1-\gamma}$, the Euler equation

$$c_{it}^{-\gamma} = E_t[\beta c_{i,t+1}^{-\gamma} R_{t+1}] \quad (2.1)$$

holds for any asset return R_{t+1} and interior agent i . If markets are complete, since individual consumption c_{it} is proportional to aggregate consumption C_t , by the Euler equation (2.1) we obtain

$$C_t^{-\gamma} = E_t[\beta C_{t+1}^{-\gamma} R_{t+1}].$$

Therefore ignoring the discount factor β ,

$$m_{t+1}^{\text{RA}} = \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma}$$

is a valid stochastic discount factor (SDF). This representative agent model has been extensively studied and repeatedly rejected (Hansen and Singleton, 1983).

If markets are incomplete, we cannot substitute aggregate consumption for individual consumption in the Euler equation because the marginal rate of substitution is not equalized across agents. Since individual consumption data are short and contain substantial measurement error, however, it is often undesirable to estimate the individual Euler equation (2.1) directly. To deal with the issue of measurement error, the literature has considered ‘Law of Large Numbers’ stochastic discount factors. For example, rewriting the Euler equation using the intertemporal marginal rate of substitution (IMRS) and averaging across agents, Brav et al. (2002) and Cogley (2002) derive

$$1 = \text{E}_t [\beta G_{-\gamma, t+1} R_{t+1}],$$

where $G_{\eta, t+1} = \text{E}_i[(c_{i, t+1}/c_{it})^\eta]$ is the η -th cross-sectional moment of consumption growth between time t and $t + 1$. Thus ignoring the discount factor β ,

$$m_{t+1}^{\text{IMRS}} = G_{-\gamma, t+1}$$

is a valid stochastic discount factor. Of course, with a finite sample we cannot compute the population moment $G_{-\gamma, t}$ and hence the exact SDF. However, we still can use the sample analog

$$\hat{m}_{t+1}^{\text{IMRS}} = \hat{G}_{-\gamma, t+1} = \frac{1}{I} \sum_{i=1}^I \left(\frac{c_{i, t+1}}{c_{it}} \right)^{-\gamma}$$

because averaging the Euler equations (2.1) is valid in the population (with a continuum of agents) as well as in the sample (with a finite number of agents). Brav et al. (2002) and Cogley (2002) employ linearized versions of the sample analog of the IMRS SDF. While the representative agent approach considers just the cross-sectional mean of the consumption distribution, these papers try the cross-sectional mean, variance, and skewness of the consumption *growth* distribution. Although the analysis of Cogley (2002) is unable to eradicate the pricing error for $\gamma < 15$, Brav et al. (2002) find that the IMRS SDF explains the equity premium for $\gamma \approx 3.5$. These two papers clean the data and approximate the SDF differently, and the literature does not provide a clear explanation for the conflicting results. We conjecture that the discrepancy could be due to the heaviness of the consumption growth distribution tail: the authors deal with outliers differently, which could greatly impact skewness calculations, even when the 3rd moment exists. Vissing-Jørgensen (2002) follows a similar approach, but her focus is the estimation of γ and not the equity premium.

Averaging the Euler equation (2.1) directly, Balduzzi and Yao (2007) derive

$$C_{-\gamma, t} = \text{E}_t [C_{-\gamma, t+1} R_{t+1}],$$

where $C_{\eta, t} = \text{E}_i[c_{it}^\eta]$ is the η -th cross-sectional moment of consumption at time t . Therefore

$$m_{t+1}^{\text{MU}} = \frac{C_{-\gamma, t+1}}{C_{-\gamma, t}}$$

is also a valid stochastic discount factor. They replicate the result of Brav et al. (2002) at the quarterly frequency but show that the IMRS SDF is not valid for monthly consumption growth. The main point of Balduzzi and Yao (2007) is that the MU SDF, which they argue is less affected by measurement error than is the IMRS one, zeroes the pricing error at $\gamma \approx 10$ when they consider only households with at least \$2,000 of financial assets. Also, assuming the consumption distribution is lognormal (a premise that we reject), Balduzzi and Yao (2007) show MU SDF is a closed-form function of the change in mean and variance of the consumption distribution. This “BY” SDF performs similarly to MU.⁴

Kocherlakota and Pistaferri (2009) take a somewhat different approach. Instead of the Euler equation (2.1), they start from the inverse Euler equation, which holds in a private information setting when agents use insurance companies to achieve constrained Pareto optimal allocation. They derive the SDF

$$m_{t+1}^{\text{PIPO}} = \frac{C_{\gamma,t}}{C_{\gamma,t+1}}.$$

While the above four papers study CEX data from the early 1980s through the mid 1990s, Kocherlakota and Pistaferri further the literature by analyzing the longer sample from 1980 to 2004 and incorporating data from U.K. and Italy to perform overidentifying tests. In this longer sample, they reject the MU SDF even when restricting analysis to households that meet various asset thresholds, which is in conflict with the findings of Balduzzi and Yao (2007). The primary result of Kocherlakota and Pistaferri (2009) is that the PIPO model zeroes the pricing error at the GMM estimate $\gamma \approx 5$. Also, imposing a common γ across U.S., U.K., and Italy, overidentifying tests reject MU and RA but not PIPO.

Finally, Krueger et al. (2008) test the heterogeneous agent, *complete market* model with limited enforcement and find that the equity premium cannot be explained unless the relative risk aversion γ is as large as 60.

Table 1 below summarizes the literature of testing the heterogeneous agent, incomplete market models. In summary, the literature had (i) rejected the representative agent (RA) model, (ii) generated mixed support for IMRS, (iii) confirmed MU (with less data) and then rejected it (with more recent data), and (iv) provided seemingly strong evidence for PIPO. However, Kocherlakota and Pistaferri (2009) do not consider IMRS and BY with their more expansive dataset. Moreover, none of the above authors explicitly discusses the presence or implications of fat tails in the cross-sectional distribution of consumption or consumption growth.⁵ In what follows we reevaluate these SDFs, and after accounting for the fat tails we find that none of the SDFs is supported.

2.2 ‘Law of Large Numbers’ stochastic discount factors and nonexistent moments

Because individual consumption data contain substantial measurement error and households participate in surveys for only short horizons, the above-mentioned literature considers ‘Law of Large Numbers’ SDFs. By averaging across agents in

⁴Kocherlakota and Pistaferri (2009) suggest (on page 560 of their paper) that Balduzzi and Yao (2007) use log-linearized SDFs, but this does not appear to be the case.

⁵Kocherlakota (1997) discusses the possibility of fat tails in aggregate consumption growth in the context of a representative agent model.

Table 1. Estimation of relative risk aversion γ and tests of stochastic discount factors. \checkmark (X) indicates support for (rejection of) an SDF. The number next to \checkmark is the estimate of γ when not rejected. Q (M) indicates quarterly (monthly) consumption growth.

Paper	Sample	IMRS	MU	PIPO
Cogley (2002)	1980–1994	X		
BCG (2002)	1982–1996	\checkmark 3.5		
BY (2007)	1982–1995	Q : \checkmark 5 M : X	\checkmark 10	
KP (2009)	1980–2004		X	\checkmark 5
This paper	1980–2004	X	X	X

an incomplete markets model, one could form SDFs consisting of cross-sectional moments of consumption and consumption growth. Assuming measurement error “averages out,” using the time series of CEX sample analogs of these moments, one can both estimate the model and evaluate its fit, even when individual household time series are short and poorly measured. If the population moments exist, these Law of Large Numbers SDFs are justified: the sample moments converge in probability as the sample size tends to infinity, and hence the GMM estimator of the relative risk aversion coefficient γ is consistent. For example, the IMRS, MU, and PIPO SDFs can be justified if $E[(c_{i,t+1}/c_{it})^{-\gamma}]$, $E[c_{it}^{-\gamma}]$, and $E[c_{it}^\gamma]$ exist, respectively, where the expectation is for i , fixing t . The minimized value of the GMM criterion then tells us about the fit of the model.

But what if these population moments do not exist? The following theorem gives a negative answer.

Theorem 2.1 (Feller (1946)). *Let X_1, X_2, \dots be i.i.d. with $E[|X_1|] = \infty$ and let $S_n = X_1 + \dots + X_n$. Let a_n be a sequence of positive numbers with a_n/n weakly increasing. Then $\limsup_{n \rightarrow \infty} |S_n|/a_n = 0$ or ∞ almost surely according as $\sum_{n=1}^{\infty} P(|X_1| \geq a_n) < \infty$ or $= \infty$.*

Letting $X_i = (c_{i,t+1}/c_{it})^{-\gamma}$, $c_{it}^{-\gamma}$, and c_{it}^γ for the case of IMRS, MU, PIPO, respectively, it follows that the existence of the relevant moment is also *necessary* for the consistency of the GMM estimator. This is because no matter what normalizing constant a_n ($a_n = n^\alpha$ with $\alpha \geq 1$, say) we choose, the quantity S_n/a_n (and hence the sample analog of each SDF above) either converges to zero or diverges. In other words, if an SDF estimate contains sample analogs of nonexistent moments, its asymptotic behavior will be model independent!

We establish in Section 3 that for both consumption and consumption growth, moments of order beyond 4 are unlikely to exist. Which papers in the literature would be affected by this degree of nonexistence? The analyses of Brav et al. (2002) and Cogley (2002) are likely not. First, both papers only approximate m_t^{IMRS} and never calculate a sample moment beyond the third. Second, Brav et al. (2002) trim the tails of the consumption growth distribution, which would help mitigate erratic behavior stemming from fat tails. However, trimming the tails of consumption growth is problematic because consumption growth is an endogenous variable.⁶

⁶To see this, observe that the theory predicts the (unconditional) Euler equation

$$1 = E[\beta g_{t+1}^{-\gamma} R_{t+1}] \tag{2.2}$$

where $g_{t+1} = c_{t+1}/c_t$ is consumption growth. Suppose, for example, that the researcher trims

Indeed, these authors were probably aware of the odd behavior of higher moments in the CEX. Cogley (2002) attributes this to measurement error, writing,

I stop at a third-order approximation because measurement error makes it difficult to estimate high-order moments. This is also the reason why I work with an approximate rather than exact discount factor (p. 314, footnote 8).

Similarly, Brav et al. (2002) comment,

The SDF is expressed in terms of the cross-sectional mean, variance, and skewness of the household consumption growth rate. The motivation for this procedure is that the estimation of the cross-sectional moments may be less susceptible to outliers than the estimation of the [exact SDF]: the estimates of the cross-sectional moments are independent of $[\gamma]$, whereas the [exact SDF] is very sensitive to outliers in the household consumption growth when $[\gamma]$ is large (p. 811).

Balduzzi and Yao (2007), in contrast, study exact SDFs. In their successful IMRS run, they estimate $\gamma \approx 5$, encroaching on the nonexistence range. However, they trim the tails in accordance with Brav et al. (2002), so we do not suspect moment nonexistence drives this result (although trimming is problematic). On the other hand, their MU findings rest on sample estimates of the -10 th moment of the consumption distribution, which do not exist according to our characterization of the lower tail. In other words, the support for MU in this case may be a “false positive.” While the successful BY SDF requires just the second moment to exist, it assumes the lognormality of the consumption distribution, which we reject in favor of the power law.

Similarly, the pro-PIPO evidence in Kocherlakota and Pistaferri (2009) relies on sample estimates of the 5th moment of the consumption distribution, which we suggest does not exist.

3 Double power law in consumption

In this section we introduce the notion of the double power law and show both theoretically and empirically that the cross-sectional distribution of consumption obeys the double power law.

3.1 Definitions and notations

Power law A nonnegative random variable X obeys the *power law* (in the upper tail) with exponent $\alpha > 0$ if

$$\lim_{x \rightarrow \infty} x^\alpha P(X > x) > 0$$

the tails of consumption growth by dropping observations outside the range $[\underline{g}, \bar{g}]$. Then the researcher is in fact testing the conditional moment restriction

$$1 = E \left[\beta g_{t+1}^{-\gamma} R_{t+1} \mid \underline{g} \leq g_{t+1} \leq \bar{g} \right], \quad (2.3)$$

which is different from (2.2). Note that even if the model is correct (*i.e.*, (2.2) holds), the conditional moment restriction (2.3) is almost always false for generic thresholds (\underline{g}, \bar{g}) .

exists (Pareto, 1896, 1897; Mandelbrot, 1960, 1961, 1963).⁷ Recently, many economic variables has been shown to obey the power law not only in the upper tail but also in the lower tail,⁸ meaning that

$$\lim_{x \rightarrow 0} x^{-\beta} P(X < x) > 0$$

exists for some exponent $\beta > 0$. Toda (2012a) introduces the concept of the *double power law*, which means that the power law holds in both the upper and the lower tails. If X obeys the double power law with exponents (α, β) , then X^η obeys the double power law with exponents $(\alpha/\eta, \beta/\eta)$ if $\eta > 0$ and $(-\beta/\eta, -\alpha/\eta)$ if $\eta < 0$. To see this, for example if $\eta > 0$ we have

$$P(X^\eta > x) = P(X > x^{\frac{1}{\eta}}) \sim x^{-\frac{\alpha}{\eta}}$$

as $x \rightarrow \infty$, and other cases are similar. An important implication of this fact is that the η th moment $E[X^\eta]$ exists if and only if $-\beta < \eta < \alpha$. Since many econometric techniques rely on the existence of some moments, recognizing a power law is important. For instance, Kocherlakota (1997) tests the representative agent, consumption-based capital asset pricing model (CAPM) by considering the possibility that aggregate consumption growth may have fat tails.

Double Pareto and double Pareto-lognormal distributions A canonical distribution that obeys the double power law is the *double Pareto* distribution (Reed, 2001), which has the probability density function (PDF)

$$f_{\text{dP}}(x) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta} \frac{1}{M} \left(\frac{x}{M}\right)^{-\alpha-1}, & (x \geq M) \\ \frac{\alpha\beta}{\alpha+\beta} \frac{1}{M} \left(\frac{x}{M}\right)^{\beta-1}, & (0 \leq x < M) \end{cases} \quad (3.1)$$

where $M > 0$ is a scale parameter (the mode if $\beta > 1$), and $\alpha, \beta > 0$ are shape parameters (power law exponents). The classical Pareto distribution with minimum size M is a special case of the double Pareto distribution by letting $\beta \rightarrow \infty$ in (3.1).

The double Pareto distribution is rarely observed in reality because its PDF has a cusp at M . An example of a distribution with a smooth density that obeys the double power law is the *double Pareto-lognormal* distribution (Reed, 2003), abbreviated as dPIN throughout the paper. A dPIN random variable is defined as the product of independent double Pareto and lognormal random variables. Its density is

$$f_{\text{dPIN}}(x) = \frac{\alpha\beta}{\alpha+\beta} \left[e^{\frac{\alpha^2\sigma^2}{2} + \alpha\mu} \Phi\left(\frac{\log x - \mu}{\sigma} - \alpha\sigma\right) x^{-\alpha-1} + e^{\frac{\beta^2\sigma^2}{2} - \beta\mu} \Phi\left(-\frac{\log x - \mu}{\sigma} - \beta\sigma\right) x^{\beta-1} \right],$$

where α, β are the power law exponents of the double Pareto variable (with $M = 1$), μ, σ are the mean and the standard deviation of the logarithm of the lognormal variable, and $\Phi(\cdot)$ denotes the cumulative distribution function

⁷See Gabaix (1999, 2009) for reviews of mechanisms generating the power law.

⁸Examples are income (Reed and Wu, 2008; Toda, 2011, 2012a) and city size (Reed, 2002; Giesen et al., 2010).

(CDF) of the standard normal distribution. As is clear from the above density, the double Pareto-lognormal distribution obeys the double power law with exponents α, β . The double Pareto distribution and the lognormal distribution are special cases of the double Pareto-lognormal distribution by letting $\sigma \rightarrow 0$ and $\alpha = \beta \rightarrow \infty$, respectively. This is an important point because it means the lognormal distribution, which is nested within dPIN, can be tested against dPIN by the likelihood ratio test.

Laplace and normal-Laplace distributions Instead of working with double Pareto and dPIN random variables, it is often easier to work with their logarithms. The logarithm of a double Pareto random variable is said to be *Laplace*.⁹ The logarithm of a dPIN random variable is called *normal-Laplace*, which is the convolution of independent normal and Laplace random variables. See Appendix A for more details on these distributions.

3.2 Theory

The conventional wisdom is that the cross-sectional distribution of consumption is lognormal. Using the consumption Euler equation, Battistin et al. (2009) show that consumption is approximately lognormal *within age cohorts*. If households “die” with constant probability and are reborn, by Theorem A.1 the double power law emerges in the entire cross-section.

There is an alternative model generating the double power law in consumption based on the heterogeneous agent, incomplete markets general equilibrium model studied by Toda (2012c). Because the model is highly tractable, we can create an artificial economy with a consumption distribution that obeys the double power law and then use it as a laboratory for studying the properties of various stochastic discount factors in asset pricing models.

Here is a simplified version of the model. Agents are indexed by $i \in I = \{1, \dots, I\}$ ¹⁰ and have a standard additive CRRA utility function

$$E_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\gamma}}{1-\gamma}.$$

Agent i starts with initial wealth (capital) $w_{i0} > 0$ but has no future endowment. There are J technologies (investment projects) indexed by $j \in J = \{1, \dots, J\}$. Production takes one period and is subject to aggregate and idiosyncratic shocks: when agent i allocates k units of goods at the end of time t to technology j , he receives $A_{i,t+1}^j k$ units of good at the beginning of period $t + 1$, where $A_{i,t+1}^j > 0$ is the productivity of technology j for agent i . Think of these technologies as private equity, human capital, or agriculture in private land. We can interpret a technology *without* idiosyncratic risk (which is a special case) as a firm whose shares are publicly traded. Letting θ_{it}^j be the fraction

⁹Hence the Laplace and the double Pareto distributions have the same relation as the normal and the lognormal distributions. As an interesting historical remark, Laplace discovered the Laplace distribution in 1774, which predates by a quarter of century the discovery of the normal distribution by himself and Gauss in the early 1800s. For more historical background, see Kotz et al. (2001).

¹⁰In Toda (2012c), there are a continuum of agents with mass 1, but the result goes through with finite agents.

of agent i 's wealth invested in project j at time t , where $\theta_{it}^j \geq 0$ and $\sum_j \theta_{it}^j = 1$, the gross return on portfolio $\theta_{it} = (\theta_{it}^1, \dots, \theta_{it}^J)$ is denoted by

$$R_{i,t+1}(\theta_{it}) = \sum_{j=1}^J A_{i,t+1}^j \theta_{it}^j.$$

The sequential budget constraint of agent i is therefore

$$(\forall t) w_{i,t+1} = R_{i,t+1}(\theta_{it})(w_{it} - c_{it}) \geq 0. \quad (3.2)$$

Assume that the productivity $A_{i,t+1}^j$ decomposes into aggregate and idiosyncratic components as $A_i^j = a_i^j A^j$, where we have dropped the time subscripts, and assume that the idiosyncratic shock a_i^j is i.i.d. across agents conditional on the history of aggregate shocks. The vector of the aggregate components is denoted by $\mathbf{A} = (A^1, \dots, A^J)$. We impose the normalization $\mathbb{E} \left[a_i^j \mid \mathbf{A} \right] = 1$. We assume that there is no insurance markets for the idiosyncratic shocks $\mathbf{a}_i = (a_i^1, \dots, a_i^J)$ (incomplete markets). We can think of the stock market as a technology with no idiosyncratic shocks, *i.e.*, $a_i^j = 1$ for all i . For simplicity assume that the aggregate productivities $\{\mathbf{A}_{t+1}\}_{t=0}^\infty$ are i.i.d. over time¹¹ and the distribution of $a_{i,t+1}^j$ conditional on public and private information depends only on \mathbf{A}_{t+1} .

In addition to these technologies, agents can trade an arbitrary set of assets in zero net supply, such as the risk-free asset or derivatives. A sequential equilibrium is defined by a sequence of individually optimal consumption and portfolio choices and asset prices such that the asset markets clear. Toda (2012c) proves that zero net supply assets are not traded in equilibrium (although they are priced) and that the optimal consumption and portfolio rule and the gross risk-free rate satisfy

$$\theta^* \in \arg \max_{\theta \in \Delta^{J-1}} \frac{1}{1-\gamma} \mathbb{E}[R(\theta)^{1-\gamma}], \quad (3.3a)$$

$$c(w) = (1 - (\beta \mathbb{E}[R(\theta^*)^{1-\gamma}])^{\frac{1}{\gamma}}) w, \quad (3.3b)$$

$$R_f = \frac{\mathbb{E}[R(\theta)^{1-\gamma}]}{\mathbb{E}[R(\theta)^{-\gamma}]}. \quad (3.3c)$$

That is, the portfolio choice and the propensity to consume out of wealth are common across all agents and constant over time, and the gross risk-free rate is constant over time. By the budget constraint (3.2) and the optimal consumption rule (3.3b), we obtain

$$c_{i,t+1} = (\beta \mathbb{E}[R(\theta^*)^{1-\gamma}])^{\frac{1}{\gamma}} R_{i,t+1}(\theta^*) c_{it}. \quad (3.4)$$

This ‘‘equation of motion’’ has the same structure as that in Appendix B. Consequently, if households ‘‘die’’ with constant probability and are reborn, then the cross-sectional consumption distribution obeys the double power law. The power law exponents can be computed by decomposing the growth rate

¹¹Toda (2012c) develops this model in a Markov setting.

$G_{i,t+1} = (\beta \mathbb{E}[R(\theta^*)^{1-\gamma}])^{\frac{1}{\gamma}} R_{i,t+1}(\theta^*)$ as in (B.3). For the benchmark “symmetric” case explained in Appendix B, the power law exponents are given by

$$\alpha = \beta = \frac{\sqrt{2\delta}}{\sigma_c}, \quad (3.5)$$

where $0 < \delta < 1$ is the death probability between periods and $\sigma_c > 0$ is the volatility of the idiosyncratic component of log consumption.

3.3 Evidence for consumption

Data construction We use the same data as the real, seasonally adjusted, quarterly household consumption data used in Kocherlakota and Pistaferri (2009) constructed from the Consumption Expenditure Survey (CEX).¹² By visual inspection of histograms, we find that the cross-sectional distribution of log consumption is bell-shaped but has tails heavier than the normal distribution, so we model log consumption as normal-Laplace. Note that this modeling choice is theoretically justified: if in the previous model newborn agents inherit wealth and if the initial wealth distribution is lognormal with the same mean as the cross-sectional distribution with constant standard deviation σ_0 , then the cross-sectional consumption distribution becomes dPIN with parameter $(\mu_t, \sigma_0, \alpha, \beta)$, where the power law exponents $\alpha = \beta$ are given by (3.5). Here the location parameter μ_t is time-dependent but other parameters are not. This is because under the i.i.d. assumption, by the law of large numbers the coefficients of the quadratic equation (B.5), which determines the power law exponents, are almost surely constant over time.

It is well known that the consumption data in the CEX are subject to substantial measurement error, but the double power law survives if the tails of the error distribution of log consumption are thinner than exponential: the distribution of observed log consumption will still have exponential tails in that case.¹³ For example, if observed consumption is the product of actual consumption and an independent lognormally distributed measurement error (with log standard deviation σ_ϵ), then the observed consumption is still dPIN with parameters $(\mu_t, \sigma, \alpha, \beta)$, where $\sigma^2 = \sigma_0^2 + \sigma_\epsilon^2$ (observed log consumption is normal-Laplace with the same parameters).

Now suppose that observed log consumption is normal-Laplace with parameters $(\mu_t, \sigma, \alpha, \beta)$. Normalizing log consumption by subtracting the population mean, normalized log consumption is normal-Laplace with parameters $(\mu, \sigma, \alpha, \beta)$ (where μ is such that the mean is zero), which do not depend on the sample. Since the CEX samples the same households once in a quarter, by the above reasoning *monthly* data of normalized consumption has the same distribution as the *quarterly* data of normalized consumption obtained by pooling three consecutive monthly data, which contains no overlapping households. Making the sample size approximately three times larger in this way, we are taking a conservative position since it is easier to reject a particular parametric model with more data.¹⁴

¹²The data is available at <http://www.jstor.org/stable/10.1086/599761>.

¹³On the other hand, if the error distribution has fatter tails than exponential, then the (observed) consumption data will be of little use because most of it will consist of measurement error.

¹⁴Of course, one may object pooling different data sets (although here it is theoretically

Parameter estimation Having constructed quarterly normalized log consumption data, we estimate the normal-Laplace parameters $(\mu, \sigma, \alpha, \beta)$ for each quarter by maximum likelihood. Since the two power law exponents α, β are almost the same, as predicted by theory (3.5), we estimate the parameters of the symmetric normal-Laplace distribution by maximum likelihood. The likelihood ratio test failed to reject symmetry ($\alpha = \beta$) in 77 out of 98 quarters at significance level 0.05. Therefore we choose the symmetric normal-Laplace distribution as the benchmark model for normalized log consumption.

Figure 1 shows the histograms of log consumption for 1985:Q1, 1990:Q2, 1995:Q3, and 2000:Q4, together with the fitted symmetric normal-Laplace density plotted in the range between the minimum and the maximum log consumption of each quarter (other quarters look similar). According to Figure 1, the symmetric normal-Laplace distribution fits the log consumption data very well.

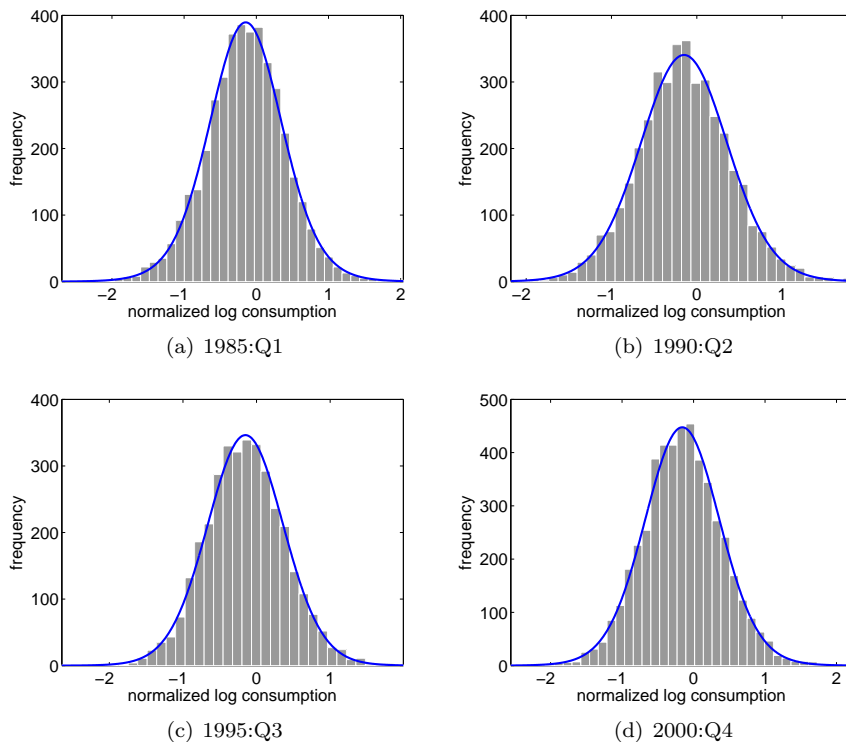


Figure 1. Histogram and normal-Laplace density fitted to normalized log consumption data.

We can also see that there are large positive and negative values that would be very unlikely if the distribution were normal. To see this visually, Figure 2 shows the quantile-quantile plot (QQ plot) of log consumption against the normal distribution. Here the actual quantiles are smaller (larger) than the case of the normal distribution in the lower (upper) tail beyond two standard errors, suggesting that the log consumption distribution has fatter tails than the normal distribution.

justified). As a robustness check, we also perform all subsequent analysis with the original monthly data.

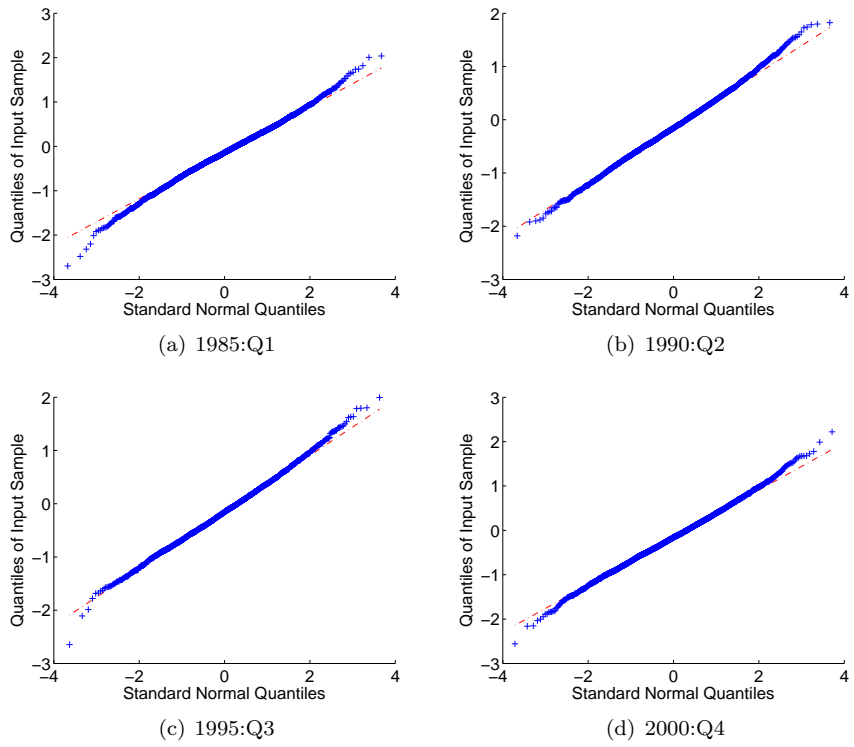


Figure 2. Quantile-quantile plot of normalized log consumption against the normal distribution.

Figure 3 shows the estimate of the power law exponent α for each quarter. According to Figure 3, the power law exponent α is around 4 (the average across all quarters was 4.06) and in the range of [3, 5.5]. The standard error computed by bootstrapping 500 times was smaller than 0.15 in every quarter (the average across all quarters was 0.073).

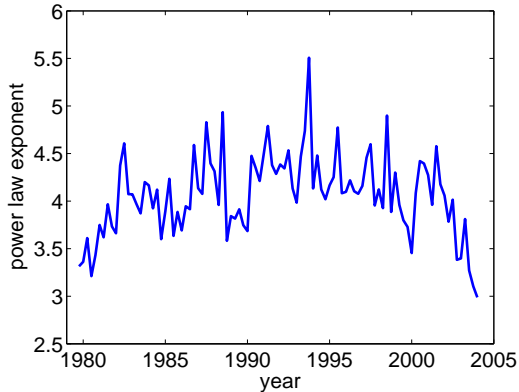


Figure 3. Power law exponent obtained by fitting a symmetric normal-Laplace distribution to quarterly U.S. normalized log consumption data.

Goodness-of-fit and model selection We evaluate the goodness-of-fit of the double Pareto-lognormal distribution (with $\alpha = \beta$) and the lognormal distribution. We perform both the Kolmogorov-Smirnov test (Massey, 1951) and the Anderson-Darling test (Anderson and Darling, 1952) and compute the P value by bootstrapping 500 times for each quarter. (See Appendix C for details. Matlab codes are available upon request.) Letting $F(x)$ be the theoretical distribution and $F_N(x)$ the empirical cumulative distribution function of data $\mathbf{x} = (x_1, \dots, x_N)$, the Kolmogorov-Smirnov and the Anderson-Darling test statistics are based on the sup norm and the L^2 norm

$$\sup_x |F_N(x) - F(x)|, \\ \int_{-\infty}^{\infty} \frac{(F_N(x) - F(x))^2}{F(x)(1 - F(x))} dF(x),$$

respectively. Note that the Kolmogorov-Smirnov test has a low power for detecting deviations from the theoretical distribution in the tails because $F_N(x) - F(x)$ tends to zero as $x \rightarrow \pm\infty$. On the other hand, the Anderson-Darling test can detect deviations in the tails because the weighting function $[F(x)(1 - F(x))]^{-1}$ tends to infinity as $x \rightarrow \pm\infty$. Hence, with the Anderson-Darling test the deviations in the tails are much more penalized. Since the existence of moments of the consumption distribution depends only on the tail behavior, for our purpose clearly the Anderson-Darling test is more appropriate. However, we also perform the Kolmogorov-Smirnov test because it is widely used.

Table 2 shows the P value of the Kolmogorov-Smirnov test for the double Pareto-lognormal specification with $\alpha = \beta$ and the lognormal specification. The double Pareto-lognormal distribution is not rejected at significance level 0.05 in

79 quarters out of 98. On the other hand, the lognormal distribution is rejected in 73 quarters. Table 3 shows the P value of the Anderson-Darling test, which is more relevant because we are interested in the tail behavior. While dPIN is not rejected in 64 quarters out of 98, lognormal is rejected in 92 quarters. The fact that the lognormal distribution is rejected much more often by the Anderson-Darling test suggests that the lognormal distribution fails to fit the tails of the data. These findings are strong evidence for the double power law conjecture of consumption and against its lognormality.

Table 2. P values of the Kolmogorov-Smirnov test for fitting a distribution to quarterly U.S. normalized consumption data. P value larger than 0.05 shown in **boldface**.

Model	double Pareto-lognormal, $\alpha = \beta$				lognormal				
	Year	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1979	-	-	-	0.15	-	-	-	-	0.002
1980	0.50	0.13	0.096	0.018	0.001	0.001	0.001	0.001	0.001
1981	0.28	0.38	0.43	0.02	0.001	0.001	0.03	0.001	0.001
1982	0.64	0.058	0.074	0.60	0.001	0.002	0.008	0.082	0.082
1983	0.78	0.84	0.59	0.91	0.014	0.007	0.077	0.032	0.032
1984	0.71	0.38	0.14	0.008	0.017	0.004	0.13	0.001	0.001
1985	0.34	0.012	0.022	0.022	0.001	0.001	0.001	0.001	0.001
1986	0.41	0.068	0.78	0.58	0.001	0.001	0.027	0.28	0.28
1987	0.28	0.47	0.78	0.49	0.02	0.035	0.14	0.08	0.08
1988	0.97	0.88	0.01	0.62	0.073	0.24	0.007	0.001	0.001
1989	0.044	0.11	0.21	0.85	0.001	0.047	0.11	0.001	0.001
1990	0.32	0.16	0.056	0.53	0.001	0.025	0.001	0.024	0.024
1991	0.62	0.23	0.90	0.32	0.11	0.11	0.12	0.02	0.02
1992	0.26	0.32	0.21	0.042	0.001	0.051	0.27	0.001	0.001
1993	0.70	0.21	0.17	0.63	0.002	0.043	0.014	0.16	0.16
1994	0.29	0.062	0.31	0.78	0.001	0.004	0.077	0.052	0.052
1995	0.59	0.36	0.12	0.004	0.044	0.002	0.16	0.018	0.018
1996	0.038	0.068	0.092	0.20	0.001	0.012	0.003	0.006	0.006
1997	0.41	0.81	0.24	0.26	0.016	0.43	0.079	0.13	0.13
1998	0	0.066	0.17	0.51	0.001	0.023	0.037	0.016	0.016
1999	0.026	0.034	0.16	0	0.02	0.001	0.052	0.001	0.001
2000	0.018	0.004	0.17	0.45	0.001	0.001	0.008	0.05	0.05
2001	0.42	0.43	0.93	0.94	0.023	0.004	0.17	0.13	0.13
2002	0.29	0.88	0.53	0.66	0.001	0.01	0.006	0.001	0.001
2003	0.34	0.018	0.47	0.038	0.001	0.001	0.001	0.001	0.001
2004	0.90	-	-	-	0.001	-	-	-	-

We also compare the performance of dPIN to other parametric distributions using the Bayesian Information Criterion, BIC (Schwarz, 1978). The parametric distributions that we consider are lognormal, gamma, and generalized beta II (GB2) (McDonald, 1984). GB2 has four parameters a, b, p, q with density

$$f_{\text{GB2}}(x) = \frac{ax^{ap-1}}{b^{ap}B(p, q)(1 + (x/b)^a)^{p+q}},$$

where $b > 0$ is a scale parameter, $a, p, q > 0$ are shape parameters, and $B(p, q)$ denotes the beta function. The generalized beta II distribution contains the

Table 3. P values of the Anderson-Darling test for fitting a distribution to quarterly U.S. normalized consumption data. P value larger than 0.05 shown in **boldface**.

Model	double Pareto-lognormal, $\alpha = \beta$				lognormal			
Year	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1979	-	-	-	0.034	-	-	-	0
1980	0.77	0.074	0.002	0.03	0	0	0	0
1981	0.076	0.20	0.47	0.012	0	0	0	0
1982	0.29	0.044	0.002	0.60	0	0	0	0.004
1983	0.57	0.55	0.75	0.91	0	0.002	0.004	0.016
1984	0.61	0.39	0.17	0.004	0.004	0	0.018	0
1985	0.076	0.014	0.004	0	0	0	0	0
1986	0.19	0.058	0.78	0.50	0	0	0.002	0.27
1987	0.42	0.36	0.65	0.07	0	0.004	0.13	0.004
1988	0.70	0.58	0.01	0.59	0.008	0.006	0.002	0
1989	0.084	0.18	0.36	0.30	0	0	0.004	0
1990	0.16	0.24	0.30	0.43	0	0.002	0.004	0.002
1991	0.56	0.21	0.27	0.022	0.02	0.008	0.022	0
1992	0.002	0.32	0.24	0.02	0	0.028	0.11	0
1993	0.15	0.068	0.38	0.35	0	0	0.03	0.14
1994	0.026	0.006	0.15	0.27	0	0	0	0.006
1995	0.48	0.004	0.04	0.018	0.004	0	0.008	0
1996	0.002	0.014	0.006	0.018	0	0	0	0
1997	0.25	0.67	0.062	0.032	0.002	0.092	0.048	0
1998	0	0.12	0.002	0.48	0	0	0.002	0.002
1999	0.006	0	0.17	0	0.002	0	0.002	0
2000	0	0	0.012	0.23	0	0	0	0.01
2001	0.05	0.12	0.90	0.92	0	0	0.058	0.004
2002	0.12	0.50	0.56	0.41	0	0	0	0
2003	0.036	0.038	0.49	0.056	0	0	0	0
2004	0.94	-	-	-	0	-	-	-

exponential, gamma, lognormal, Weibull, Singh-Maddala (Singh and Maddala, 1976), and Dagum (Dagum, 1977) distributions as special or limiting cases, but not the double Pareto or the dPIN.¹⁵ Out of 98 quarters, dPIN performed best in 78 quarters, GB2 in 15 quarters, and lognormal in 5 quarters. Therefore, among a large class of parametric distributions, dPIN provides the best fit to the consumption distribution.

Testing the existence of moments directly Although characterizing the cross-sectional consumption distribution is interesting in its own right, for comparing various stochastic discount factors in incomplete markets asset pricing models proposed in the literature, whether a moment exists or not is more important than which parametric distribution fits to data. Fortunately, there is a simple bootstrap test for testing the existence of moments directly (Fedotenkov, 2011), explained in Appendix D.

Figure 4 shows the upper and lower bounds of the order of moments for which the existence is not rejected at significance level 0.05. The average of the upper and lower bounds across all quarters is 5.76 and -6.24 , which are close to the estimated power law exponent (around 4).

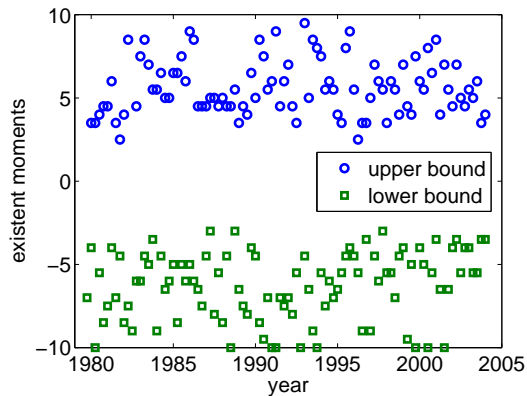


Figure 4. Range of existent moments for quarterly U.S. consumption data.

Testing the lognormal distribution against dPIN Since the lognormal distribution is nested within the double Pareto-lognormal distribution (by letting the power law exponents α, β to infinity), we can test the lognormal distribution against dPIN by the likelihood ratio test. The test rejects the lognormal distribution at significance level 0.05 in every quarter except 1993:Q4, with P value 0.08 (which gives the largest power law exponent 5.5 in Figure 3). Therefore the consumption distribution is better described by dPIN than lognormal when we look at the entire sample.

This finding does not contradict to those of Battistin et al. (2009) because they look at the consumption distribution *within age cohorts*, not the entire

¹⁵However, GB2 obeys the double power law with exponent $\alpha = aq$ and $\beta = ap$. Because power law exponents can be estimated easily, the maximum likelihood estimation of GB2 parameters is quicker and more stable by maximizing over (a, b, α, β) instead of (a, b, p, q) .

cross-section. Since the double power law emerges from the geometric age distribution (Theorem A.1), we would expect that the cross-sectional consumption distribution is more lognormal within age cohorts than in the entire cross-section. To evaluate this conjecture, we perform the likelihood ratio test and goodness-of-fit tests for the lognormal distribution for each age cohort. The groups are household head age 30 or less, 31 to 40, 41 to 50, 51 to 60, and 60 or more. The likelihood ratio test fails to reject lognormality in 46, 38, 37, 56, and 32 quarters out of 98 for each age group, respectively. The Kolmogorov-Smirnov test fails to reject lognormality in 67, 69, 63, 72, and 43 quarters for each age group, and the Anderson-Darling test fails to reject in 51, 51, 50, 64, and 23 quarters, respectively. Therefore the lognormal distribution fits reasonably well to the cross-sectional distribution of consumption for each age group, in agreement with Battistin et al. (2009). Our finding that the double power law emerges only in the entire cross-section and not within age cohorts strongly suggests that it is the age distribution driving the power law (as in Appendix B) and not measurement error: we have no reason to believe that measurement error is correlated with age.

The range of moment existence is $[-6.6, 8.1]$ for 30 or less, $[-7.8, 7.5]$ for 31 to 40, $[-8.4, 7.0]$ for 41 to 50, $[-8.5, 8.0]$ for 51 to 60, and $[-8.5, 6.5]$ for 61 or more. These ranges are wider than for the entire cross section. Since we only checked moments of order between -10 and 10 , the actual range of existence is likely to be even wider.

Robustness with monthly data So far, we have normalized and pooled consecutive months to make the sample size larger, and the model of Section 3.2 justifies this procedure. In case the model is false, as a robustness check, we perform all previous exercises with the original monthly data.

Figure 5(a) shows the estimated power law exponents. The power law exponent α is around 4 (the average across all months was 4.31), which is qualitatively the same as with the pooled data (Figure 3). However, occasionally there are large exponents (*i.e.*, tails are thin) of around 10 because with smaller sample sizes there are fewer extreme observations, which determine the tail behavior.

The likelihood ratio test rejects the lognormal distribution against dPIN in 240 months out of 291 (82% of the time). The Kolmogorov-Smirnov test failed to reject dPIN in 263 months (90% of the time) and rejected the lognormal distribution in 104 months (36% of the time), whereas the Anderson-Darling test failed to reject dPIN in 250 months (86% of the time) and rejected the lognormal distribution in 165 months (57% of the time). Considering the low power of the Kolmogorov-Smirnov test, the result of the Anderson-Darling test is more relevant. Again dPIN fits much better than the lognormal in monthly data.

Figure 5(b) shows the upper and lower bounds of the order of moments for which the existence is not rejected at significance level 0.05. The averages of the upper and lower bounds across all months are 6.73 and -7.16 , which are slightly larger than the estimated power law exponent (around 4.3), but the results are qualitatively the same as with quarterly data.

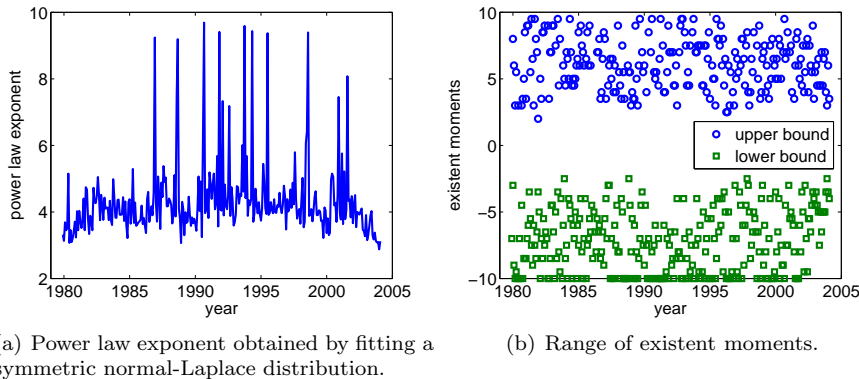


Figure 5. Estimation results with monthly consumption.

3.4 Evidence for consumption growth

Next we turn to the distribution of consumption growth. By the equation of motion (3.4) for consumption, we obtain

$$\frac{c_{i,t+1}}{c_{it}} = (\beta \mathbb{E}[R(\theta^*)^{1-\gamma}])^{\frac{1}{\gamma}} R_{i,t+1}(\theta^*),$$

so the distribution of consumption growth is the same as that of portfolio return $R_{i,t+1}(\theta^*)$, except for a multiplicative constant. Since asset returns are known to have fat tails, we conjecture that idiosyncratic productivity shocks also have fat tails, meaning the high order moments of consumption growth may not exist.

Given the robustness of the Laplace distribution and the double Pareto distribution (Theorem A.1) and considering the possibility of measurement error, the most natural parametric distribution to fit consumption growth is again the double Pareto-lognormal distribution. Therefore we estimate the normal-Laplace parameters $(\mu, \sigma, \alpha, \beta)$ for each *month* (pooling months is not justified for consumption growth) by maximum likelihood. Since the two power law exponents α, β are almost the same, we further estimate the parameters of the symmetric normal-Laplace distribution by maximum likelihood. The likelihood ratio test rejects symmetry ($\alpha \neq \beta$) in only 6 months out of 291 at significance level 0.05. Therefore we choose the symmetric normal-Laplace distribution as the benchmark model for log consumption growth.

Figure 6 shows the histograms of log consumption growth for March 1985, June 1990, September 1995, and December 2000, together with the fitted symmetric normal-Laplace density plotted in the range between the minimum and the maximum log consumption growth of each month (other months look similar). Again the normal-Laplace distribution fits very well. Compared to the log consumption distribution, the peak of log consumption growth distribution is sharper. Therefore, by visual inspection alone the normal distribution seems inappropriate. In fact, the likelihood ratio test rejects the lognormality of consumption growth against dPIN in every single month.

Figures 7(a) and 7(b) show the estimated power law exponent and the upper and lower bounds of the order of moments for which the existence is not rejected at significance level 0.05. The average of the exponent and the upper and lower bounds across all months was 4.05, 6.80, and -6.83 , respectively.

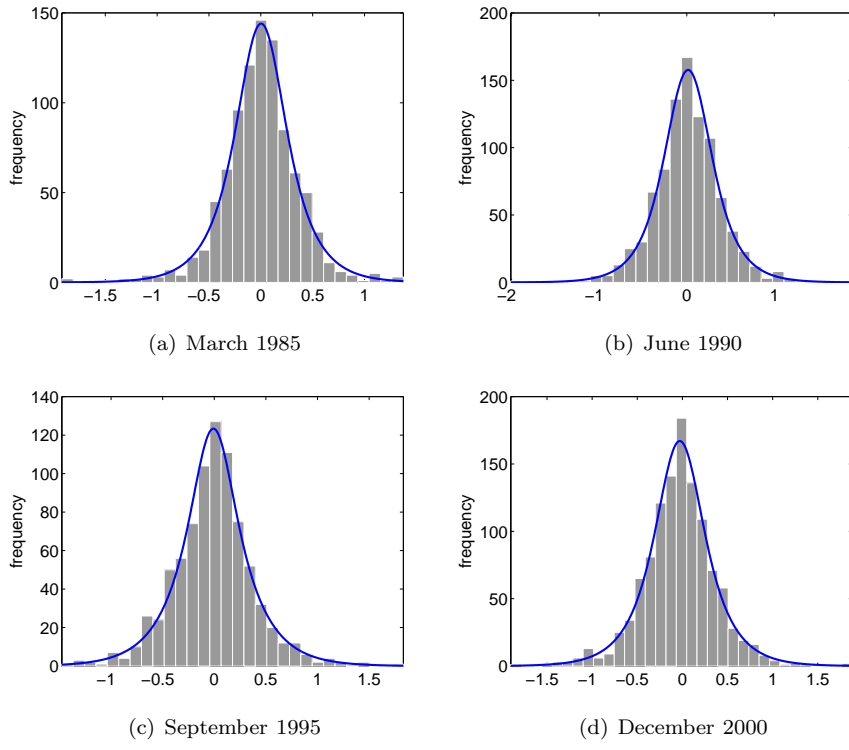


Figure 6. Histogram and normal-Laplace density fitted to log consumption growth.

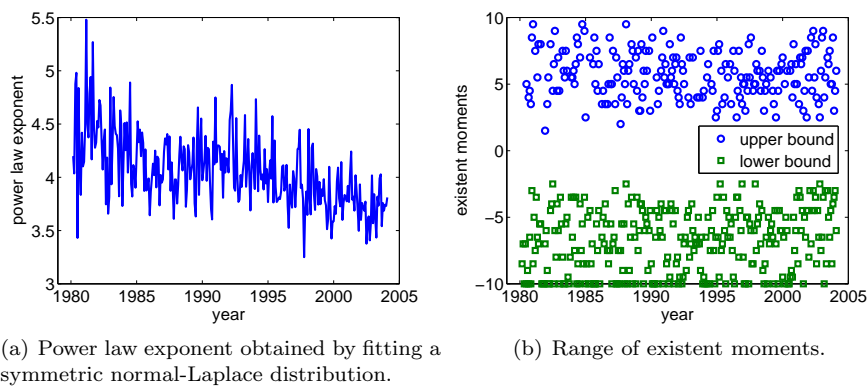


Figure 7. Estimation results with consumption growth.

Finally, we evaluate the goodness-of-fit of the double Pareto-lognormal distribution (with $\alpha = \beta$) and the lognormal distribution. While the Kolmogorov-Smirnov test and the Anderson-Darling test fail to reject dPIN in 265 and 260 months, respectively, the Kolmogorov-Smirnov test rejects the lognormal distribution in all but 3 months and the Anderson-Darling test rejects the lognormal distribution in every single month. Therefore dPIN fits consumption growth even better than it does the consumption level. Using BIC, dPIN provides the best fit in all but 7 months. GB2 performs best in 6 months, and gamma performs best in one month. Overall, the performance of dPIN is outstanding.

4 Estimation and robustness

In this section we estimate the relative risk aversion coefficient γ using various asset pricing models and study the robustness of the performance of each model.

4.1 Data

As in Section 3, we use the real, seasonally adjusted consumption data in Kocherlakota and Pistaferri (2009) constructed from the CEX. Their dataset has monthly observations from December 1979 to February 2004, but each number corresponds to a household’s consumption over the previous 3 months. So, while there are households for each month, no household appears in consecutive months. Therefore, despite the fact that we have an SDF and excess return realization for each month, the data for each month reflect a quarter of information, and the return series are 3 month moving averages. For example, the sample analog of the PIPO SDF is defined by

$$\widehat{m}_t^{\text{PIPO}}(\gamma) = \frac{\frac{1}{I_{t-3}} \sum_{i=1}^{I_{t-3}} c_{i,t-3}^\gamma}{\frac{1}{I_t} \sum_{i=1}^{I_t} c_{it}^\gamma},$$

where I_t is the number of households at time t and c_{it} is the consumption of household i at time t . This gives 288 SDF observations for RA, MU, and PIPO and 287 for IMRS (we lose one quarter for IMRS because household IDs were reset in 1986). In total, we have 410,788 consumption data points and 270,428 consumption growth data points (there are fewer consumption growth data points because many households participate the survey for only one quarter). See Kocherlakota and Pistaferri (2009) for further details on the construction of real consumption and the U.S. equity premium.

In one of our exercises described below, we split households into age groups, where we define the age of a household by the age of the oldest head of household. We get these ages from the NBER Consumer Expenditure Survey Family-Level Extracts webpage.¹⁶ However, since the NBER data cover fewer households, the merge leaves us with 373,785 data points for consumption levels and 253,549 for growth. (Thus we lose 9% of consumption level and 6% of consumption growth, respectively.) Also, for consumption growth the number of SDF observations falls to 286 due to changes in the CEX in 1996.

¹⁶http://www.nber.org/data/ces_cbo.html

4.2 Estimation

For any SDF $j \in \{\text{RA}, \text{IMRS}, \text{MU}, \text{PIPO}\}$, define

$$f_t^j(\gamma) = \left(\widehat{m}_t^j(\gamma) \right) (R_t^s - R_t^b),$$

$$g_T^j(\gamma) = \frac{1}{T} \sum_{t=1}^T f_t^j(\gamma),$$

where T is the number of observation for SDF j , R_t^s is the stock market return, and R_t^b is the Treasury bill rate. The GMM estimator of γ and the pricing error are

$$\widehat{\gamma}^j = \arg \min_{\gamma} T \left(g_T^j(\gamma) \right)^2,$$

$$e^j = g_T^j(\widehat{\gamma}^j) = \frac{1}{T} \sum_{t=1}^T \left(\widehat{m}_t^j(\widehat{\gamma}^j) \right) (R_t^s - R_t^b).$$

Following Kocherlakota and Pistaferri (2009), we report the Newey and West (1987) standard errors (with truncation parameter equal to 4), which account for the sampling error in the time series but abstract from uncertainty regarding cross-sectional moments of consumption.

In addition, we report standard errors from a bootstrap procedure based on Politis and Romano (1994) that also account for the sampling error in the cross-section. We sample with replacement from the original data to generate B bootstrap samples, indexed by $b = 1, \dots, B$. Each is of length T and has statistical properties like the original sample. Each bootstrap sample yields risk aversion estimate $\widehat{\gamma}_b^j$ and pricing error e_b^j . The bootstrap standard error is the sample standard error of $\left\{ \widehat{\gamma}_b^j \right\}_{b=1}^B$. The explicit procedure for generating each sample b is as follows:

1. For each $t \in \mathbf{T} = \{1, \dots, T\}$, draw with replacement I_t observations from $\{c_{it}\}_{i=1}^{I_t}$, yielding $\{c_{it}^b\}_{i=1}^{I_t}$.
2. Let M be the average block length and set $p = 1/M$. (We choose $M = \sqrt{T}$.) Draw τ_1^b uniformly from \mathbf{T} . For $s = 2, \dots, T$, with probability $1 - p$ set $\tau_s^b = \tau_{s-1}^b + 1$ modulo T (hence $\tau_s^b = 1$ if $\tau_{s-1}^b = T$), and with probability p draw τ_s^b uniformly from \mathbf{T} .
3. The bootstrap sample b consists of all \widetilde{c}_{is}^b , $s = 1, \dots, T$, where we define $\widetilde{c}_{is}^b = c_{i, \tau_s^b}^b$ for $i = 1, \dots, I_{\tau_s^b}$.

The process for bootstrapping consumption growth and asset returns is analogous. The one caveat concerns the calculation of SDF $j \in \{\text{RA}, \text{MU}, \text{PIPO}\}$. Consider PIPO for example. We use

$$\widehat{m}_s^{\text{PIPO}, b}(\gamma) = \frac{\frac{1}{I_{\tau_s^b-3}} \sum_{i=1}^{I_{\tau_s^b-3}} \left(c_{i, \tau_s^b-3}^b \right)^\gamma}{\frac{1}{I_{\tau_s^b}} \sum_{i=1}^{I_{\tau_s^b}} \left(\widetilde{c}_{is}^b \right)^\gamma}.$$

That is, the bootstrap time s SDF is formed from actual time τ_s^b and $\tau_s^b - 3$ data, not actual time τ_s^b and τ_{s-3}^b data in order to preserve the statistical properties of the SDF.

Below, we use $B = 500$ bootstrap replications.

4.3 Results and robustness to outliers

The first two columns of the Table 4 below replicate Table 2 on p. 581 of Kocherlakota and Pistaferri (2009) and display the estimation results for IMRS. The IMRS SDF, like MU, fails to explain the equity premium.

Table 4. GMM estimation of relative risk aversion (RRA) γ and pricing errors e^j . The first and second numbers in parentheses are the Newey-West and bootstrap standard errors.

Model	Full KP sample		Without Outliers	
	RRA (γ)	Pricing error	RRA (γ)	Pricing error
RA	53.26 (29.41) (20.19)	0.000	53.10 (30.85) (21.24)	-0.000
MU	1.52 (5698) (0.90)	0.019	2.51 (9960) (1.86)	0.019
PIPO	5.33 (1.42) (1.98)	0.000	2.23 (8010) (1.68)	0.019
IMRS	0.03 (1035) (0.08)	0.019	0.03 (1297) (0.21)	0.019

The first and second numbers in parentheses are, respectively, the Newey-West and bootstrap standard errors. Our Newey-West standard errors are slightly larger than those in Kocherlakota and Pistaferri, probably because they use the Hansen and Hodrick (1980) serial correlation correction. For RA and PIPO, the bootstrap standard errors (which account for cross-sectional sampling error) are similar to the Newey-West ones. This is not true for MU and IMRS. However, it is not clear how to interpret these Newey-West numbers anyway, because in each case γ is exactly identified but the pricing error is away from 0 (as returns are quarterly, 0.02 is essentially the entire equity premium). For this reason, Kocherlakota and Pistaferri do not report a standard error for MU. Another reason for the large standard errors in MU and IMRS might be again the fat tails. For example, even if the pricing error $f_t^j(\gamma) = \hat{m}_t^j(\gamma)(R_t^s - R_t^b)$ has a finite first moment, it may not have a finite second moment, in which case we cannot apply the standard asymptotic theory of the GMM estimator.

Oddly, $\hat{\gamma}^{\text{IMRS}}$ is very low. One explanation for this result is that since the exact IMRS SDF is the weighted average of the $-\gamma$ th power of each household's consumption growth, whenever γ is large the SDF will be huge because there are always households with consumption growth much smaller than 1. Therefore the GMM criterion may be a huge number when γ is large. In this way, small consumption growth observations may drive γ toward zero.

The last two columns of Table 4 are what we get when we drop a small

number of outliers relative to the total number of data points. Specifically, we drop the top 100 and bottom 100 consumption observations from the entire sample. For IMRS, we also drop the top 100 and bottom 100 consumption growth observations. As there are 410,788 data points, for consumption levels the points we drop account for less than 0.05% of the entire sample. Note that these outliers are spread roughly uniformly across the quarters, so on average we are dropping less than 1 observation point per quarter since there are 288 quarters.

We see that the results for the RA SDF, which should not be affected by the nonexistence of higher moments, barely change. We continue to reject MU and IMRS. PIPO, however, is now unable to explain the equity premium. Figure 8 below shows the GMM criterion for PIPO as a function of γ , with and without the outliers. Just a few outliers generate the trough at 5.33. Perhaps not coincidentally, $\gamma = 5$ is near the beginning of the moment nonexistence region. Without the outliers, the γ estimate falls to 2.23, well within the existence region.

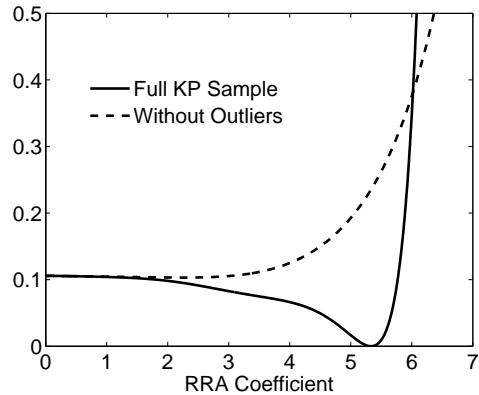


Figure 8. PIPO GMM criterion with and without largest and smallest 100 consumption outliers out of 410,788.

Another way to see that the positive PIPO evidence may be the result of the power law tails is to analyze the bootstrap distributions for the pricing error and γ estimate. Figure 9 displays histograms of e_b^{PIPO} , with and without outliers. We see that when we bootstrap with all data, there is an odd mass of pricing errors at 0. Without outliers, the pricing error bootstrap distribution is centered around e^{PIPO} , as it should be, with much less mass at zero.

Finally, Figure 10 is a scatter plot of the bootstrap estimates $\hat{\gamma}_b^j$ and pricing errors e_b^j . There is a clear inverse relationship between the pricing error and the γ estimate. Indeed, *most* of the zero pricing errors correspond to γ estimates in the moment nonexistence range (> 4); when the pricing error is greater than 0.01, the corresponding γ estimate tends to be less than 3.

We take this collection of observations as strong evidence that the heavy tail of the consumption distribution aids *mechanically* in zeroing the pricing error. At least, such CEX-based asset pricing exercises seem quite sensitive to outliers. This, in conjunction with measurement error and the limited cross-sectional sample size, leads one to question the usefulness of CEX ‘Law of Large

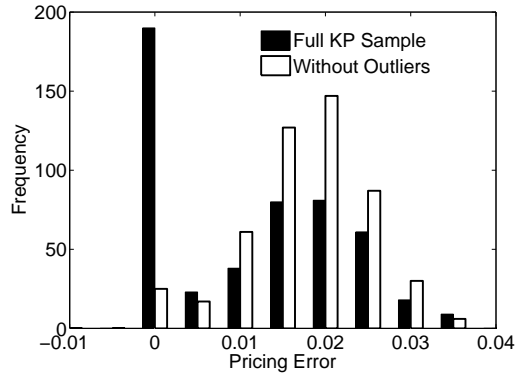


Figure 9. Histogram of bootstrapped PIPO pricing errors with and without largest and smallest 100 consumption outliers out of 410,788.

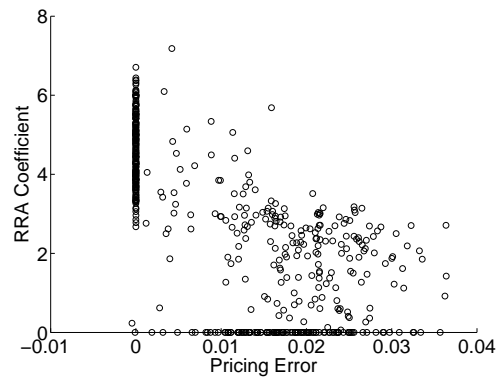


Figure 10. Scatter plot of bootstrapped PIPO RRA estimates and pricing errors.

4.4 Robust estimation using age data

What can we do to circumvent the fat tail issue in the estimation of γ ? One solution is to find an exogenous variable such that the conditional consumption distribution does not have fat tails. When we tested the double power law conjecture of consumption in Section 3.3, for each quarter t we divided the cross-section into five age cohorts, 30 years or younger, 31 to 40, 41 to 50, 51 to 60, and older than 60. Call these $H_{t,1}, \dots, H_{t,5}$. We found that at each t , within cohort the consumption distribution is approximately lognormal. At least, more moments exist within cohort than for the entire cross-section. With this in mind, we perform an overidentified GMM exercise that (i) is less susceptible to the nonexistent moment issue¹⁸ and (ii) allows for overidentifying tests of the different models.

Specifically, we exploit the fact that the Euler equation aggregation in Section 2 that gave us the SDFs also works within a particular age cohort because age is an exogenous variable. That is, instead of averaging across all agents, we can average across a particular age group. For example, we can form the $H_{t,5}$ (> 60) MU SDF by

$$\widehat{m}_t^{\text{MU},5}(\gamma) = \frac{\frac{1}{I_{t,5}} \sum_{i \in H_{t,5}} c_{it}^{-\gamma}}{\frac{1}{I_{t-3,5}} \sum_{i \in H_{t-3,5}} c_{i,t-3}^{-\gamma}}.$$

For any $j \in \{\text{RA}, \text{IMRS}, \text{MU}, \text{PIPO}\}$, define

$$F_t^j(\gamma) = \begin{bmatrix} \widehat{m}_t^{j,1}(\gamma) \\ \vdots \\ \widehat{m}_t^{j,5}(\gamma) \end{bmatrix} (R_t^s - R_t^b),$$

$$G_T^j(\gamma) = \frac{1}{T} \sum_{t=1}^T F_t^j(\gamma).$$

The overidentified GMM estimator of γ is

$$\widehat{\gamma}^j = \arg \min_{\gamma} T G_T^j(\gamma)' W G_T^j(\gamma),$$

where W is the weighting matrix. (We always use the identity matrix as the weighting matrix.)

We calculate standard errors via the above bootstrap procedure because the Newey-West standard errors may be misleading according to the results of Table 4. Furthermore, for each SDF we bootstrap a P value for the null hypothesis that the pricing error is 0 (that is, that the model is correct). The following is a description of the calculation of these P values:

¹⁷In their online appendix, Kocherlakota and Pistaferri (2009) find that the success of PIPO is robust to (i) dropping large SDF values and (ii) dropping observations with particularly high consumption-wealth ratios. Beyond their initial data-cleaning, however, they do not try dropping the largest and smallest observations.

¹⁸Technically, if the η th moment does not exist for the entire cross-section, it must not exist for at least one age group because the moment of the entire cross-section is a weighted average of the moments of the age groups.

1. Dropping the SDF superscript, let $G_{T,b}(\hat{\gamma}_b)$ be the vector of pricing errors corresponding to bootstrap sample b .
2. For each bootstrap sample b , define

$$J_{T,b} = T(G_{T,b}(\hat{\gamma}_b) - G_T(\hat{\gamma}))' W(G_{T,b}(\hat{\gamma}_b) - G_T(\hat{\gamma})).$$

Also define the minimized sample criterion

$$J_T = T G_T(\hat{\gamma})' W G_T(\hat{\gamma}).$$

3. Calculate the P value by

$$p = \frac{1}{B} \sum_{b=1}^B \mathbb{1}\{J_{T,b} \geq J_T\}.$$

Why should this work? The idea of the bootstrap is that the empirical distribution of $G_{T,b}$ around G_T approximates the distribution of G_T around G_∞ , which is 0 under the null. It follows that under the null the empirical distribution of $J_{T,b}$ approximates the distribution of J_T . Finally, if the null fails and G_T converges to something different from 0, then J_T is not properly centered and will diverge as $T \rightarrow \infty$.

Table 5 presents the robust GMM γ estimates and the bootstrapped P values.

Table 5. Robust GMM estimation of relative risk aversion (RRA) γ and P value of overidentifying tests. Numbers in parentheses are bootstrapped standard errors.

Model	RRA (γ)	P value
RA	2.62 (1.68)	0.00
MU	1.22 (0.63)	0.00
PIPO	1.88 (0.88)	0.00
IMRS	0.04 (0.10)	0.00

We see that with the robust method, the RA, MU, and PIPO γ s are all between 1 and 3, well within the moment existence range. The IMRS estimate, as before, is around 0. This is because unlike with the consumption *level*, the double power law in consumption *growth* holds also within age groups, so splitting households into age cohorts does not mitigate the fat tail issue for the IMRS SDF. The standard errors for the former three SDFs are, respectively, 1.68, 0.63, and 0.88, meaning the risk aversion estimates of these models are statistically close. Moreover, the overidentifying tests reject all of the models at the 1% significance level.

This result corroborates the rejection of RA and MU by Kocherlakota and Pistaferri (2009). Recall, though, that their exercise employs data from the U.K. and Italy and not from different age cohorts within the U.S. Our estimation and rejection of the IMRS SDF is a new but unsurprising result, given the performance of MU, which comes from the same model. With respect to PIPO, the model rejection

and low γ estimate are further evidence that the power law was a factor in the earlier success of PIPO. This is because, to reiterate, the approximate within-cohort lognormality means the robust GMM procedure dampens the impact of nonexistent moments. One more piece of evidence is Figure 11, which is a histogram of the average pricing error, $(\mathbf{1}'G_T(\hat{\gamma}))/5$, across bootstrap samples. As when we drop outliers (compare to Figure 9), there is no spike at 0.

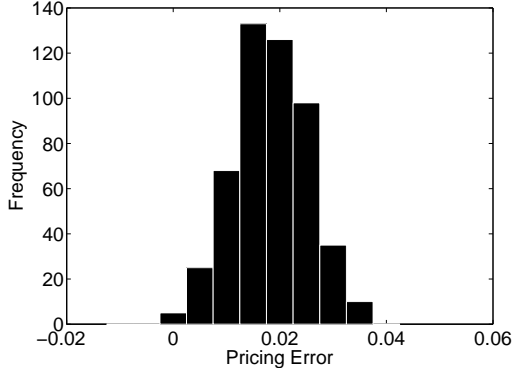


Figure 11. Histogram of bootstrapped PIPO pricing errors with robust GMM estimation.

5 Estimation with simulated data

So far, by both dropping outliers and bootstrapping the actual data, we have provided support for our conjecture that the power law interferes with model selection and the estimation of parameters. In this section we present more direct evidence of this interference. Specifically, we construct an artificial economy in which the consumption distribution obeys the double power law with a known exponent and then perform some of the above exercises using simulated data.

5.1 Simulating a GEI economy

To simulate a sufficiently rich example of a general equilibrium model with incomplete markets (GEI) described in Section 3.2, we need a stock market (to study the equity premium) and idiosyncratic shocks (to have a nondegenerate wealth distribution), so at least two technologies are necessary. Let technologies 1 and 2 be the “stock market” and “private equity”, with productivities $(A^1, A_i^2) = (A_s, a_i A_p)$. Assume that all shocks are lognormally distributed, with joint distribution

$$\begin{bmatrix} \log A_s \\ \log A_p \\ \log a_i \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu_s \\ \mu_p \\ -\sigma_i^2/2 \end{bmatrix}, \begin{bmatrix} \sigma_s^2 & \rho\sigma_s\sigma_p & 0 \\ \rho\sigma_s\sigma_p & \sigma_p^2 & 0 \\ 0 & 0 & \sigma_i^2 \end{bmatrix} \right).$$

($E[\log a_i] = -\sigma_i^2/2$ ensures that $E[a_i] = 1$.)

To numerically solve the optimal portfolio problem (3.3a), we write the joint distribution of the log returns of the two technologies as

$$\begin{bmatrix} \log A_s \\ \log(a_i A_p) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu_s \\ \mu_p - \sigma_i^2/2 \end{bmatrix}, \begin{bmatrix} \sigma_s^2 & \rho\sigma_s\sigma_p \\ \rho\sigma_s\sigma_p & \sigma_p^2 + \sigma_i^2 \end{bmatrix} \right) =: \mathcal{N}(\mu, \Sigma).$$

We discretize this distribution as follows. First we find a matrix C such that $\Sigma = CC'$, say using the Cholesky decomposition. Second, we approximate a standard normal variable using the technique described in Tanaka and Toda (2013) with $K = 81$ equally spaced grid points on the interval $[-4, 4]$, denoted by $\{x_k\}_{k=1}^K$, and obtain the probabilities $\{p_k\}$. Finally, we assign the probability $q_{kl} = p_k p_l$ to the point $(y_k, y_l) = \mu + C(x_k, x_l) \in \mathbb{R}^2$. This way, the multinomial distribution on the points $\{(y_k, y_l)\}_{1 \leq k, l \leq K}$ with probability $\{q_{kl}\}$ approximates $\mathcal{N}(\mu, \Sigma)$. Then we numerically solve the optimal portfolio problem (3.3a) and compute the optimal consumption rule (3.3b) and gross risk-free rate (3.3c).

We simulate an economy with N agents and T time periods as follows. First, to create the panel of ages, we generate $N \times T$ Bernoulli variables with death probability $0 < \delta < 1$. The initial wealth of each household at time 0 is $w_{i0} = 1$. We assume that a newborn household inherits the cross-sectional average wealth times a lognormal perturbation w_{ini} , where $\log w_{\text{ini}} \sim \mathcal{N}(-\sigma_0^2/2, \sigma_0^2)$, i.i.d. across agents and time. Second, we generate T aggregate shocks $\{(A_{s,t+1}, A_{p,t+1})\}_{t=0}^{T-1}$ and $N \times T$ idiosyncratic shocks $\{(a_{i,t+1})_{i \in I}\}_{t=0}^{T-1}$ and compute the consumption path of each household, denoted by $\{c_{it}\}$. Finally, we multiply c_{it} by the ‘‘measurement error’’ e^ϵ , where $\epsilon \sim \mathcal{N}(-\sigma_\epsilon^2/2, \sigma_\epsilon^2)$, again i.i.d. across agents and time. In this way we obtain a sequence of stock market returns $\{A_{s,t+1}\}_{t=0}^{T-1}$ and an $N \times T$ panel of observed consumption and age.

Because the initial wealth and the measurement error are both lognormal, by Theorem A.1 the cross-sectional consumption distribution for large enough time periods becomes approximately double Pareto-lognormal. One may calculate the power law exponent α either theoretically using (3.5) or numerically by fitting the normal-Laplace distribution to the log observed consumption distribution by maximum likelihood. (In our simulation they are almost the same number, as they should be.) We find that the economy typically converges to a stochastic steady state after $4/\delta$ periods (4 times the average age of households). In practice, we generate a much longer sample (say $5/\delta$) than T and use only the last T .

5.2 Calibration and estimation

We simulate a GEI economy at the quarterly frequency. We set the one period discount factor $\beta = 0.99$ (0.96 annually), relative risk aversion $\gamma = 4$, death probability $\delta = (1/75)/4$ (an average lifespan of 75 years), expected stock market and private equity returns $\mu_s = \mu_p = 0.07/4$ (7% annually), volatilities $(\sigma_s, \sigma_p, \sigma_i) = (0.15, 0.1, 0.15)/\sqrt{4}$, correlation between the stock market and private equity $\rho = 0.5$, and standard deviation of initial log wealth and measurement error $(\sigma_0, \sigma_\epsilon) = (0.5, 0.25)$. Some households live for much longer than 75 years, which is clearly unrealistic if we interpret lifespan literally as the age of a household member. However, we interpret households as dynasties, and we allow the possibility of very long lived households in order to generate the cross-sectional power law.¹⁹ We simulate the economy 200 times, each run consisting of 300 quarters and 4000 households at any given time (when one dies, it is immediately replaced by a new one).

¹⁹Note, by (3.5), that the power law exponent α is proportional to $\sqrt{\delta}$, so a longer average lifespan $1/\delta$ yields a lower power law exponent.

With these parameters the implied quarterly equity premium is 1.47% (5.86% annually), the risk-free rate computed by (3.3c) is 1.14% annually, and the power law exponent computed by (3.5) is 3.24 for consumption. In this case, we know that there is a power law in consumption, and we know that if not for this reason, the MU SDF would give us consistent estimates of γ , using simulated data. The question then is, how does MU behave in the presence of the power law? Since the power law exponent is close to but below the true RRA coefficient ($\gamma = 4$), we expect that the MU SDF will perform poorly.

We uncover three interesting observations. First, the PIPO and MU SDFs are identical in the context of this model for the following reason. Under the model of Section 3.2, the cross-sectional consumption distribution at time t is dPIN with parameter $(\mu_t, \sigma_t, \alpha_t, \beta_t)$, where $\sigma_t = \sigma = \sqrt{\sigma_0^2 + \sigma_\epsilon^2}$ is time independent and $\alpha_t = \beta_t$. Now suppose that X is dPIN with parameter $(\mu, \sigma, \alpha, \beta)$. Then the η th moment of X is

$$E[X^\eta] = \frac{\alpha\beta}{(\alpha - \eta)(\beta + \eta)} e^{\mu\eta + \frac{1}{2}\sigma^2\eta^2}.$$

(To derive this, use the fact that dPIN is the product of the lognormal and the double Pareto distributions, compute the moments for each, and take the product.) Hence if c_t is dPIN with parameter $(\mu_t, \sigma_t, \alpha_t, \beta_t)$ and $\alpha_t = \beta_t$, then after some algebra we obtain

$$\frac{m_{t+1}^{\text{PIPO}}}{m_{t+1}^{\text{MU}}} = \frac{E[c_t^\gamma] E[c_t^{-\gamma}]}{E[c_{t+1}^\gamma] E[c_{t+1}^{-\gamma}]} = e^{-\gamma^2(\sigma_{t+1}^2 - \sigma_t^2)},$$

so the ratio of PIPO and MU SDFs depends only on the difference of σ^2 and not on other parameters. But for our model we know that $\sigma_t = \sigma$ is time independent, so PIPO and MU SDFs are identical. This might explain why the MU and PIPO estimation results in Table 4 without outliers and in Table 5 are similar. We conjecture that, in general, the PIPO and MU SDFs will behave similarly when higher moments of the consumption distribution do not vary much over time, even when the mean fluctuates.

Second, across simulations there is an inverse relationship between the γ estimate and the pricing error. As with the bootstrap PIPO exercise, when the MU model is able to almost exactly zero the pricing error, the γ estimate is often well above the start of the nonexistence range, > 3.24 (Figure 12(a)). However, splitting households into age groups²⁰ and performing the robust GMM, we no longer see this pattern: the large γ estimates corresponding to the zero pricing error in Figure 12(a) has disappeared in the robust GMM of Figure 12(b).

Third, in some runs (54 out of 200 simulations), the GMM criterion has multiple troughs, one near the true γ and one in the moment nonexistence range. Indeed, for these 54 second troughs, the 10th and 90th percentiles are at γ equal to, respectively, 6.70 and 19.42. It seems nonexistent moments may introduce spurious troughs, and in some instances, only the spurious one is close to zero. This occurs in 12 out of 200 simulations. Figure 13 illustrates this scenario. If sample versions of nonexistence moments do indeed cause GMM to behave this way, they have the ability to cause researchers to fail to reject incorrect models.

²⁰The age group cutoffs are at the 20th, 40th, 60th, and 80th percentiles, pooling all simulated age observations.

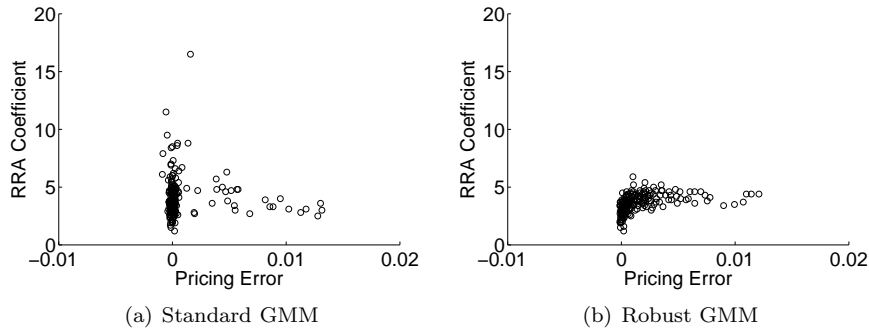


Figure 12. Scatter plot of simulated MU RRA estimates and pricing errors.

In contrast, the robust GMM criterion has multiple troughs in just 2 out of 200 simulations, and in neither case is the spurious one closer to zero.

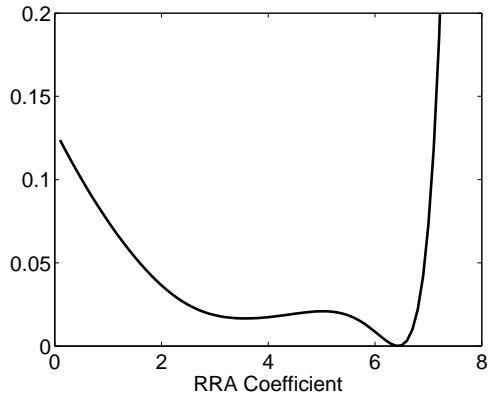


Figure 13. MU GMM criterion from simulation 9.

6 Discussion and Conclusion

In her recent consumption-based asset pricing survey (Ludvigson, 2012), Sydney Ludvigson writes,

It is difficult to draw general conclusions from the results [of the empirical heterogeneous consumer asset pricing literature]. The mixed results seem to depend sensitively on a number of factors, including the sample, the empirical design, on the method for handling and modeling measurement error, the form of cross-sectional aggregation of Euler equations across heterogeneous agents, and the implementation, if any, of linear approximation of the pricing kernel. A tedious but productive task for future work will be to carefully control for all of these factors in a single empirical study, so that researchers may better assess whether the household consumption heterogeneity we can measure in the data has the characteristics needed to explain asset return data.

In addition to establishing the excellent fit of the double Pareto-lognormal distribution to cross-sectional U.S. consumption and consumption growth data, we have addressed some of the issues presented in this quotation. Holding the estimation sample constant and accounting for the fat tails, we find that Consumer Expenditure Survey data do not seem to support any of the heterogeneous agent general equilibrium models we consider. Our bootstrap and simulation evidence show that it is at least plausible that the mixed results of this literature stem from the erratic behavior of sample analogs of nonexistent moments. However, even if all consumption moments exist, it is clear that tail observations impact these exercises. Thus, given the Consumer Expenditure Survey's measurement error and limited cross-sectional sample size, we question its usefulness in explaining asset prices.

The common elements in this literature are identical constant relative risk aversion utility, interior solutions (Euler equations), and use of the Consumer Expenditure Survey. Consequently, both authors of this paper are in the process of using different data sources to test the asset pricing implications of models with non-identical or non-constant relative risk aversion utility, with binding financial constraints, or with both. For example, Toda (2012b) builds a dynamic general equilibrium model with many agents that have heterogeneous preferences (except a common relative risk aversion coefficient γ) with very general recursive structures, proves that the $-\gamma$ th power of the market return is a valid stochastic discount factor, tests the asset pricing implications bypassing consumption data altogether, and fails to reject the model with $\gamma \approx 3$.

A Laplace and normal-Laplace distributions

Definitions A Laplace random variable is the logarithm of a double Pareto variable. By changing variables in (3.1) and setting $m = \log M$, the density of the Laplace distribution is given by

$$f_L(x) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta}e^{-\alpha|x-m|}, & (x \geq m) \\ \frac{\alpha\beta}{\alpha+\beta}e^{-\beta|x-m|}, & (x < m) \end{cases} \quad (\text{A.1})$$

where m is the mode and $\alpha, \beta > 0$ are scale parameters. A Laplace distribution is said to be *asymmetric* if $\alpha \neq \beta$. A comprehensive review of the Laplace distribution can be found in Kotz et al. (2001).

The logarithm of a double Pareto-lognormal variable is said to be *normal-Laplace* (Reed and Jorgensen, 2004; Reed and Wu, 2008), which is simply the convolution of independent normal and Laplace random variables. The normal-Laplace distribution has four parameters, a location parameter μ and three scale parameters $\sigma, \alpha, \beta > 0$, with probability density function

$$f_{NL}(x) = \frac{\alpha\beta}{\alpha+\beta} \left[e^{\frac{\alpha^2\sigma^2}{2}-\alpha(x-\mu)}\Phi\left(\frac{x-\mu}{\sigma}-\alpha\sigma\right) + e^{\frac{\beta^2\sigma^2}{2}+\beta(x-\mu)}\Phi\left(-\frac{x-\mu}{\sigma}-\beta\sigma\right) \right] \quad (\text{A.2})$$

and cumulative distribution function

$$F_{\text{NL}}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) - \frac{1}{\alpha+\beta} \left[\beta e^{\frac{\alpha^2\sigma^2}{2}-\alpha(x-\mu)} \Phi\left(\frac{x-\mu}{\sigma} - \alpha\sigma\right) - \alpha e^{\frac{\beta^2\sigma^2}{2}+\beta(x-\mu)} \Phi\left(-\frac{x-\mu}{\sigma} - \beta\sigma\right) \right]. \quad (\text{A.3})$$

Again the Laplace and the normal distributions are special cases of the normal-Laplace distribution by letting $\sigma \rightarrow 0$ and $\alpha = \beta \rightarrow \infty$, respectively, and therefore can be tested against the normal-Laplace distribution by the likelihood ratio test.

Characteristic function Using (A.1), the characteristic function of a Laplace random variable X is

$$\begin{aligned} \phi_X(t) &= \int_{-\infty}^m e^{itx} \frac{\alpha\beta}{\alpha+\beta} e^{-\beta|x-m|} dx + \int_m^{\infty} e^{itx} \frac{\alpha\beta}{\alpha+\beta} e^{-\alpha|x-m|} dx \\ &= \frac{e^{imt}}{1 - i\left(\frac{1}{\alpha} - \frac{1}{\beta}\right)t + \frac{t^2}{\alpha\beta}}, \end{aligned} \quad (\text{A.4})$$

from which we obtain the mean $m + \frac{1}{\alpha} - \frac{1}{\beta}$ and the variance $\frac{1}{\alpha^2} + \frac{1}{\beta^2}$. It is often useful to parameterize the Laplace distribution in terms of its characteristic function. Letting $a = \frac{1}{\alpha} - \frac{1}{\beta}$ be an asymmetry parameter and $\sigma = \sqrt{\frac{2}{\alpha\beta}}$ be a scale parameter in (A.4), we write $X \sim \mathcal{AL}(m, a, \sigma)$ if

$$\phi_X(t) = \frac{e^{imt}}{1 - iat + \frac{\sigma^2 t^2}{2}}. \quad (\text{A.5})$$

The mean, mode, and variance of $\mathcal{AL}(m, a, \sigma)$ are $m + a$, m , and $a^2 + \sigma^2$, respectively. Comparing (A.4) and (A.5), we obtain $1/\alpha - 1/\beta = a$ and $\alpha\beta = 2/\sigma^2$, so $-\alpha$ and β are the solutions to the quadratic equation

$$\frac{\sigma^2}{2}\zeta^2 - a\zeta - 1 = 0. \quad (\text{A.6})$$

Limit theorem Perhaps the most important property of the Laplace distribution is that it is the only limit distribution of geometric sums. Theorem A.1 below shows that it is a robust property that the limit of a geometric sum is a Laplace distribution.

Theorem A.1 (Toda (2012c)). *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of zero mean random variables such that the central limit theorem holds, $N^{-1/2} \sum_{n=1}^N X_n \xrightarrow{d} \mathcal{N}(0, \sigma^2)$; $\{a_n\}_{n=1}^{\infty}$ be a sequence such that $N^{-1} \sum_{n=1}^N a_n \rightarrow a$; and ν_p be a geometric random variable with mean $1/p$ independent from $\{X_n\}_{n=1}^{\infty}$. Then as $p \rightarrow 0$ we have*

$$p^{\frac{1}{2}} \sum_{n=1}^{\nu_p} (X_n + p^{\frac{1}{2}} a_n) \xrightarrow{d} \mathcal{AL}(0, a, \sigma).$$

B Generative model of double power law

This appendix, which draws heavily from Toda (2012c), describes a generative model of the double power law and explains its robustness.

Consider an economy consisting of many “units” (households, firms, cities, etc.). Let S_{it} be the “size” (income, wealth, consumption, revenue, population, etc.) of unit i at time t . Let us characterize the cross-sectional size distribution when units grow multiplicatively.

Equation of motion Suppose that the size grows multiplicatively (Gibrat (1931)’s law of proportionate effect) according to

$$S_{i,t+\Delta t} = G_{i,t+\Delta t} S_{it}, \quad (\text{B.1})$$

where Δt is the time step and $G_{i,t+\Delta t}$ is the gross growth rate of unit i between time t and $t + \Delta t$. Iterating (B.1), the log size of unit i is given by

$$\log S_{it} = \log S_{i0} + \sum_{n=1}^{N_{it}} \log G_{i,t+\Delta t-n\Delta t}, \quad (\text{B.2})$$

where S_{i0} is the initial size and N_{it} is the number of time periods unit i has been alive up to time t .

Distributional assumptions Assume that G_{it} decomposes into the aggregate and the purely idiosyncratic components such that $G_{it} = g_{it} G_t$. Since we are interested in the limiting case in which the time step Δt is small, define the “per unit of time” shocks by $X_{at} = \log G_t / \Delta t$ and $X_{it} = \log g_{it} / \sqrt{\Delta t}$. By construction the idiosyncratic shock X_{it} has zero mean conditional on the history of (the σ -algebra generated by) aggregate shocks $\mathcal{F}_t := \sigma(\{X_{as}\}_{s \leq t})$.²¹ Then we obtain

$$\log G_{it} = X_{at} \Delta t + X_{it} \sqrt{\Delta t}. \quad (\text{B.3})$$

Let us further assume that the idiosyncratic component $\{X_{it}\}$ is independent over time and that the central limit theorem holds.

If units are infinitely lived, by (B.2) with $N_{it} = t/\Delta t$, (B.3), and the Lindeberg-Feller central limit theorem, letting $\Delta t \rightarrow 0$ the cross-sectional size distribution (relative to initial size) becomes approximately lognormal.

Things dramatically change when units die. The simplest way to model death is to assume that units die at a constant Poisson rate $\delta > 0$. Think of “death” as breaking up of households, financially disastrous events, and so on. For now assume that all units start from (a common) initial size S_0 and when a unit dies, it is reborn with the same initial size S_0 . (We will relax this assumption later.) Hence, the population is constant overtime.

Emergence of double Pareto distribution Letting $p = \delta \Delta t$, the number of time periods unit i has been alive is $N_{it} = \min\{\nu_p, t/\Delta t\}$, where ν_p is distributed

²¹One can think of X_{at} as the “drift” and $E[X_{it}^2 | \{X_{as}\}_{s \leq t}]$ as the “volatility”.

as a geometric random variable with mean $1/p$. By (B.2) and (B.3), the log size of unit i at time t is given by

$$\log S_{it} = \log S_0 + p^{\frac{1}{2}} \sum_{n=1}^{N_{it}} \left(X_{i,t+\Delta t-n\Delta t} / \sqrt{\delta} + p^{\frac{1}{2}} X_{a,t+\Delta t-n\Delta t} / \delta \right). \quad (\text{B.4})$$

The following theorem shows that for large t the cross-sectional size distribution is approximately double Pareto with mode equal to the initial size.

Theorem B.1. *Let everything be as above, where $\{X_{at}\}$ and $\{X_{it}\}$ are the stochastic processes defined on a probability space (Ω, \mathcal{F}, P) describing the aggregate and idiosyncratic components as in (B.3) and $\mathcal{F}_t = \sigma(\{X_{as}\}_{s \leq t})$ is the σ -algebra generated by the aggregate shock. Assume that for the realization $\omega \in \Omega$ the time averages of the “drift” and “volatility” have limits:*

$$\begin{aligned} \mu_S(\omega) &:= \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t X_{as}(\omega) ds, \\ \sigma_S^2(\omega) &:= \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t E[X_{is}^2 | \mathcal{F}_s](\omega) ds. \end{aligned}$$

Then for $\omega \in \Omega$ the cross-sectional distribution of $\{S_{it}\}_{i \in I}$ converges in distribution to the double Pareto distribution with mode S_0 and power law exponents α, β as $t \rightarrow \infty$ and $\Delta t \rightarrow 0$, where $-\alpha$ and β are solutions to the quadratic equation

$$\frac{\sigma_S^2(\omega)}{2} \zeta^2 - \mu_S(\omega) \zeta - \delta = 0. \quad (\text{B.5})$$

Proof. See Toda (2012c). □

Robustness of the double power law So far we have assumed that units are ex ante identical, *i.e.*, they have the same initial size and obey the same stochastic process. However, the double power law holds under weaker assumptions.

First, instead of assuming a common initial size $S_{i0} = S_0$, assume that S_{i0} is random (i.i.d. across units). Since the double power law implies that the cross-sectional log size distribution has exponential tails (which follows from the argument relating the double Pareto distribution (3.1) and the Laplace distribution (A.1)), the double power law still holds as long as the distribution of the initial log size has tails thinner than exponential.

Second, instead of assuming a constant initial size S_0 , suppose that initial log size of a unit born at time t is the cross-sectional average of log size at time t . (Think about inheriting financial and human capital wealth when born.) Then (B.4) becomes

$$\log S_{it} = \log S_0 + \Delta t \sum_{n=1}^{t/\Delta t} X_{a,t+\Delta t-n\Delta t} + p^{\frac{1}{2}} \sum_{n=1}^{N_{it}} X_{i,t+\Delta t-n\Delta t} / \sqrt{\delta}. \quad (\text{B.6})$$

We can still apply Theorem A.1 to the third term in (B.6), use (B.5), and see that the log size distribution is approximately (symmetric) Laplace, with power law exponent

$$\alpha = \beta = \frac{\sqrt{2\delta}}{\sigma_S(\omega)}. \quad (\text{B.7})$$

The first and second terms of (B.6) simply determine the common mode.

Finally, suppose that there are finitely many types of units denoted by $h \in H = \{1, \dots, H\}$, each obeying a stochastic process for the growth rate $G_{i,t+\Delta t}^h$. Then the double power law holds for each type, with corresponding power law exponents $(\alpha_h, \beta_h)_{h \in H}$. Letting $\alpha = \min_h \alpha_h$ and $\beta = \min_h \beta_h$, the double power law with exponents α, β holds in the entire economy because the tail of the entire population is determined by the fattest tail among all subpopulations.

C Implementing goodness-of-fit tests

This Appendix explains how to implement the Kolmogorov-Smirnov test and the Anderson-Darling test by parametric bootstrap. Let $\mathbf{x} = (x_1, \dots, x_N)$ be the data and $F_N(x) = \frac{1}{N} \sum_{n=1}^N \mathbb{1}\{x \leq x_n\}$ be the empirical distribution function. Given a parametric model $\{f(x; \theta)\}_{\theta \in \Theta}$, the first step is to estimate θ by maximum likelihood. Let $\hat{\theta}$ be the maximum likelihood estimate, and compute the Kolmogorov-Smirnov and Anderson-Darling statistics defined by

$$\begin{aligned} \text{KS}^{\text{data}} &= \sup_x \left| F_N(x) - F(x; \hat{\theta}) \right|, \\ \text{AD}^{\text{data}} &= -N - \sum_{n=1}^N \frac{2n-1}{N} \left[\log F(x_n; \hat{\theta}) + \log(1 - F(x_{N+1-n}; \hat{\theta})) \right], \end{aligned}$$

where for the Anderson-Darling statistic the data must be sorted in ascending order: $x_1 \leq x_2 \leq \dots \leq x_N$. Second, we generate B bootstrap samples $\{\mathbf{x}^b\}_{b=1}^B$ (each of size N) drawn from $F(x; \hat{\theta})$, and for each bootstrap sample \mathbf{x}^b we compute the maximum likelihood estimate $\hat{\theta}^b$. Again compute the Kolmogorov-Smirnov and Anderson-Darling statistics defined by

$$\begin{aligned} \text{KS}^b &= \sup_x \left| F_N^b(x) - F(x; \hat{\theta}^b) \right|, \\ \text{AD}^b &= -N - \sum_{n=1}^N \frac{2n-1}{N} \left[\log F(x_n^b; \hat{\theta}^b) + \log(1 - F(x_{N+1-n}^b; \hat{\theta}^b)) \right], \end{aligned}$$

where $F_N^b(x)$ is the empirical distribution function of the synthetic data \mathbf{x}^b . Finally, compute the P value by $p = \frac{1}{B} \sum_{b=1}^B \mathbb{1}\{\text{KS}^b > \text{KS}^{\text{data}}\}$ for the Kolmogorov-Smirnov test and similarly for the Anderson-Darling test.

In order to implement the goodness-of-fit tests for the normal-Laplace distribution, we need to generate random numbers. Since the normal-Laplace distribution is the convolution of the normal and the Laplace distributions, it suffices to generate Laplace random numbers. Letting U, V be independent uniform random variables on $[0, 1]$,

$$X = -\frac{1}{\alpha} \log U + \frac{1}{\beta} \log V$$

has a Laplace distribution with mode 0 and exponents α, β . To see this, note that

$$\text{E}[\exp(it\lambda \log U)] = \text{E}[U^{it\lambda}] = \int_0^1 u^{it\lambda} du = \frac{1}{1 + i\lambda t},$$

so setting $\lambda = -1/\alpha, 1/\beta$, the characteristic function of X is

$$\phi_X(t) = \mathbb{E}[\exp(itX)] = \frac{1}{1 - it/\alpha} \frac{1}{1 + it/\beta} = \frac{1}{1 - i(\frac{1}{\alpha} - \frac{1}{\beta})t + \frac{t^2}{\alpha\beta}},$$

which is the characteristic function of the Laplace distribution with mode 0 and exponents α, β given by (A.4).²²

Because the maximum likelihood estimator of the parameters of the normal-Laplace distribution has no closed-form, implementing goodness-of-fit tests for the normal-Laplace distribution requires numerically maximizing the likelihood function for many times, which is time consuming.

D Testing moment existence

This Appendix explains how to test the existence of moments directly following Fedotenkov (2011). Suppose that the random variable X is nonnegative (consider $|X|$ if X can be negative) and $\{X_n\}_{n=1}^\infty$ are i.i.d. copies of X . If $\mathbb{E}[X^\eta] = \infty$, then the sample moment $\frac{1}{N} \sum_{n=1}^N X_n^\eta$ tends to infinity as $N \rightarrow \infty$. Therefore if we take a number $M(N)$ such that $M \rightarrow \infty$ and $M/N \rightarrow 0$ as $N \rightarrow \infty$, and $\{Y_m\}_{m=1}^\infty$ are independent and have the same distribution as X , then for $0 < \xi < 1$ the quantity

$$F = \mathbb{1} \left\{ \frac{1}{M} \sum_{m=1}^M Y_m^\eta \geq \xi \frac{1}{N} \sum_{n=1}^N X_n^\eta \right\}$$

tends to zero almost surely as $N \rightarrow \infty$, where $\mathbb{1}\{\cdot\}$ denotes the indicator function. This is because both $\xi \frac{1}{N} \sum_{n=1}^N X_n^\eta$ and $\frac{1}{M} \sum_{m=1}^M Y_m^\eta$ tend to infinity, but the former does so at a faster rate. On the other hand, if $\mathbb{E}[X^\eta]$ is finite, then by the law of large numbers F tends to 1 almost surely because both sample means converge to the same population mean, but since $0 < \xi < 1$ as N tends to infinity $\xi \frac{1}{N} \sum_{n=1}^N X_n^\eta$ is almost surely smaller than $\frac{1}{M} \sum_{m=1}^M Y_m^\eta$.

Given this result Fedotenkov (2011) constructs a bootstrap test of moment existence as follows. Let $\mathbf{x} = (x_1, \dots, x_N)$ be the data. First, we choose the bootstrap sample size $M(N)$, the parameter ξ , and bootstrap repetition B (Fedotenkov suggests taking $M(N) = \lfloor \log N \rfloor$, $\xi = 0.999$, and $B = 10,000$). Second, for each $b = 1, \dots, B$, we generate a bootstrap sample \mathbf{x}^b of size M drawn randomly with replacement from the data, and compute

$$F^b = \mathbb{1} \left\{ \frac{1}{M} \sum_{m=1}^M (x_m^b)^\eta \geq \xi \frac{1}{N} \sum_{n=1}^N x_n^\eta \right\}.$$

Finally, the P value is defined by $p = \frac{1}{B} \sum_{b=1}^B F^b$.

²²There is another way of generating Laplace random numbers: if Y, Z are independent exponential random variables with mean $1/\alpha, 1/\beta$, respectively, then $X = m + Y - Z$ is Laplace with mode m and exponents α, β .

References

- Theodore W. Anderson and Donald A. Darling. Asymptotic theory of certain ‘goodness of fit’ criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23(2):193–212, 1952. doi:10.1214/aoms/1177729437.
- Pierluigi Balduzzi and Tong Yao. Testing heterogeneous-agent models: An alternative aggregation approach. *Journal of Monetary Economics*, 54(2):369–412, March 2007. doi:10.1016/j.jmoneco.2005.08.021.
- Erich Battistin, Richard Blundell, and Arthur Lewbel. Why is consumption more log normal than income? Gibrat’s law revisited. *Journal of Political Economy*, 117(6):1140–1154, December 2009. doi:10.1086/648995.
- Alon Brav, George M. Constantinides, and Christopher C. Geczy. Asset pricing with heterogeneous consumers and limited participation: Empirical evidence. *Journal of Political Economy*, 110(4):793–824, August 2002. doi:10.1086/340776.
- Timothy Cogley. Idiosyncratic risk and the equity premium: Evidence from the consumer expenditure survey. *Journal of Monetary Economics*, 49(2):309–334, March 2002. doi:10.1016/S0304-3932(01)00106-4.
- George M. Constantinides and Darrell Duffie. Asset pricing with heterogeneous consumers. *Journal of Political Economy*, 104(2):219–240, April 1996.
- Camilo Dagum. A new model of personal income distribution: Specification and estimation. *Economie Appliquée*, 30:413–437, 1977.
- Igor Fedotenkov. Using a bootstrap method to test the existence of finite moments. 2011.
- William Feller. A limit theorem for random variables with infinite moments. *American Journal of Mathematics*, 68(2):257–262, April 1946.
- Xavier Gabaix. Zipf’s law for cities: An explanation. *Quarterly Journal of Economics*, 114(3):739–767, 1999. doi:10.1162/003355399556133.
- Xavier Gabaix. Power laws in economics and finance. *Annual Review of Economics*, 1:255–293, 2009. doi:10.1146/annurev.economics.050708.142940.
- John Geanakoplos. The leverage cycle. In Daron Acemoglu, Kenneth Rogoff, and Michael Woodford, editors, *NBER Macroeconomic Annual 2009*, volume 24, chapter 1, pages 1–65. University of Chicago Press, Chicago, 2010.
- Robert Gibrat. *Les Inégalités économiques*. Librairie du Recueil Sirey, Paris, 1931.
- Kristian Giesen, Arndt Zimmermann, and Jens Suedekum. The size distribution across all cities—double Pareto lognormal strikes. *Journal of Urban Economics*, 68(2):129–137, September 2010. doi:10.1016/j.jue.2010.03.007.
- Fatih Guvenen. A parsimonious macroeconomic model for asset pricing. *Econometrica*, 77(6):1711–1750, November 2009. doi:10.3982/ECTA6658.

- Lars Peter Hansen and Robert J. Hodrick. Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy*, 88(5):829–853, October 1980.
- Lars Peter Hansen and Kenneth J. Singleton. Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of Political Economy*, 91(2):249–265, April 1983.
- John Heaton and Deborah J. Lucas. Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy*, 104(3):443–487, June 1996.
- Narayana R. Kocherlakota. Testing the consumption CAPM with heavy-tailed pricing errors. *Macroeconomic Dynamics*, 1(3):551–567, September 1997. doi:10.1017/S136510059700401X.
- Narayana R. Kocherlakota and Luigi Pistaferri. Asset pricing implications of Pareto optimality with private information. *Journal of Political Economy*, 117(3):555–590, June 2009. doi:10.1086/599761.
- Samuel Kotz, Tomasz J. Kozubowski, and Krzysztof Podgórski. *The Laplace Distribution and Generalizations*. Birkhäuser, Boston, 2001.
- Dirk Krueger, Hanno Lustig, and Fabrizio Perri. Evaluating asset pricing models with limited commitment using household consumption data. *Journal of the European Economic Association*, 6(2-3):715–726, April-May 2008. doi:10.1162/JEEA.2008.6.2-3.715.
- Per Krusell and Anthony A. Smith, Jr. Income and wealth heterogeneity, portfolio choice, and equilibrium asset returns. *Macroeconomic Dynamics*, 1(2):387–422, June 1997. doi:10.1017/S1365100597003052.
- 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.
- Robert E. Lucas, Jr. Asset prices in an exchange economy. *Econometrica*, 46(6):1429–1445, November 1978.
- Sydney C. Ludvigson. Advances in consumption-based asset pricing: Empirical tests. In George M. Constantinides, Milton Harris, and René M. Stultz, editors, *Handbook of the Economics of Finance*, volume 2. Elsevier, Amsterdam, 2012.
- Benoît Mandelbrot. The Pareto-Lévy law and the distribution of income. *International Economic Review*, 1(2):79–106, May 1960.
- Benoît Mandelbrot. Stable Paretian random functions and the multiplicative variation of income. *Econometrica*, 29(4):517–543, October 1961.
- Benoît Mandelbrot. New methods in statistical economics. *Journal of Political Economy*, 71(5):421–440, October 1963.

- Frank J. Massey, Jr. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951. doi:10.1080/01621459.1951.10500769.
- James B. McDonald. Some generalized functions for the size distribution of income. *Econometrica*, 52(3):647–663, May 1984.
- Rajnish Mehra and Edward C. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2):145–161, March 1985. doi:10.1016/0304-3932(85)90061-3.
- Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, May 1987.
- Vilfredo Pareto. *Cours d'économie politique*. F. Rouge, Lausanne, 1896.
- Vilfredo Pareto. New theories of economics. *Journal of Political Economy*, 5(4):485–502, September 1897.
- Dimitris N. Politis and Joseph P. Romano. The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313, 1994. doi:10.1080/01621459.1994.10476870.
- 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.
- William J. Reed. On the rank-size distribution for human settlements. *Journal of Regional Science*, 42(1):1–17, February 2002. doi:10.1111/1467-9787.00247.
- William J. Reed. The Pareto law of incomes—an explanation and an extension. *Physica A*, 319(1):469–486, March 2003. doi:10.1016/S0378-4371(02)01507-8.
- William J. Reed and Murray Jorgensen. The double Pareto-lognormal distribution—a new parametric model for size distribution. *Communications in Statistics—Theory and Methods*, 33(8):1733–1753, 2004. doi:10.1081/STA-120037438.
- William J. Reed and Fan Wu. New four- and five-parameter models for income distributions. In Duangkamon Chotikapanich, editor, *Modeling Income Distributions and Lorenz Curves*, chapter 11, pages 211–223. Springer, New York, 2008. doi:10.1007/978-0-387-72796-7_11.
- Gideon Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978. doi:10.1214/aos/1176344136.
- S. K. Singh and G. S. Maddala. A function for size distribution of incomes. *Econometrica*, 44(5):963–970, September 1976.
- Ken'ichiro Tanaka and Alexis Akira Toda. Discrete approximations of continuous distributions by maximum entropy. *Economics Letters*, 2013. doi:10.1016/j.econlet.2012.12.020. Forthcoming.
- Alexis Akira Toda. Income dynamics with a stationary double Pareto distribution. *Physical Review E*, 83(4):046122, 2011. doi:10.1103/PhysRevE.83.046122.

Alexis Akira Toda. The double power law in income distribution: Explanations and evidence. *Journal of Economic Behavior and Organization*, 84(1):364–381, September 2012a. doi:10.1016/j.jebo.2012.04.012.

Alexis Akira Toda. Asset pricing with heterogeneous preferences. 2012b. URL <https://sites.google.com/site/aatoda111/papers>.

Alexis Akira Toda. Asset pricing and wealth distribution with heterogeneous investment returns. Submitted, 2012c. URL <https://sites.google.com/site/aatoda111/papers>.

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.