RISK, UNCERTAINTY, AND THE DYNAMICS OF INEQUALITY

KENNETH KASA AND XIAOWEN LEI

Abstract. This paper studies the dynamics of wealth inequality in a continuous-time Blanchard/Yaari model. Its key innovation is to assume that idiosyncratic investment returns are subject to (Knightian) uncertainty. In response, agents formulate robust portfolio policies (Hansen and Sargent, 2008). These policies are nonhomothetic; wealthy agents invest a higher fraction of their wealth in uncertain assets yielding higher mean returns. This produces an endogenous feedback mechanism that amplifies inequality. It also produces an accelerated rate of convergence, which helps resolve a puzzle recently identified by Gabaix, Lasry, Lions, and Moll (2016). We ask the following question - Suppose the US was in a stationary distribution in 1980, and the world suddenly became more 'uncertain'. Could this uncertainty explain both the magnitude and pace of recent US wealth inequality? Using detection error probabilities to discipline the degree of uncertainty, we conclude that an empirically plausible increase in uncertainty can account for about half of the recent increase in top wealth shares.

JEL Classification Numbers: D31, D81

1. Introduction

It is well known that models of idiosyncratic labor income risk, in the tradition of Aiyagari (1994), cannot explain observed inequality. Although these models shed some light on the lower end of the wealth distribution, they cannot generate sufficient concentrations of wealth in the right-tail (Huggett, 1996).[1] In response, a more recent literature considers models of idiosyncratic investment risk. These so-called 'random growth' models *can* generate the sort of power laws that characterize observed wealth distributions.[2]

Although investment risk models are successful in generating empirically plausible wealth distributions, they suffer from two drawbacks. First, existing applications focus on *stationary* distributions. However, what's notable about recent US wealth inequality is that it has increased. This suggests that some parameter characterizing the stationary distribution must have changed. It's not yet clear what changed. Second, Gabaix, Lasry, Lions, and Moll (2016) have recently shown that standard investment risk models

Date: August, 2017.
Kasa: Department of Economics, Simon Fraser University, `kkasa@sfu.ca`. Lei: Department of Economics, Simon Fraser University, `leixiaowen04@gmail.com`. We would like to thank Evan Anderson, Jess Benhabib, Ben Moll and Jun Nie for helpful comments. We are especially grateful to an anonymous referee for many useful suggestions.

[1]Benhabib, Bisin, and Luo (2017) note that models based on idiosyncratic labor income risk cannot generate wealth distributions with fatter tails than the distribution of labor income.

[2]The original idea dates back to Champernowne (1953) and Simon (1955). Recent examples include Benhabib, Bisin, and Zhu (2011) and Toda (2014). Gabaix (2009) provides a wide ranging survey of power laws in economics and finance. Benhabib and Bisin (2017) survey their application to the distribution of wealth.

based on Gibrat's Law cannot account for the *rate* at which inequality has increased. Top wealth shares have approximately doubled over the past 35-40 years. Standard model parameterizations suggest that this increase should have taken at least twice as long.

Our paper addresses both of these drawbacks. The key idea is to assume that agents confront (Knightian) uncertainty when investing. Following Hansen and Sargent (2008), we assume agents have a benchmark model of investment returns. In standard random growth models, agents fully trust their benchmark model. That is, they confront risk, not uncertainty. In contrast, here agents distrust their model, in a way that cannot be captured by a conventional Bayesian prior. Rather than commit to single model/prior, agents entertain a *set* of alternative models, and then optimize against the worst-case model. Since the worst-case model depends on an agent's own actions, agents view themselves as being immersed in a dynamic zero-sum game. Solutions of this game produce 'robust' portfolio policies. To prevent agents from being unduly pessimistic, in the sense that they attempt to hedge against empirically implausible alternatives, the hypothetical 'evil agent' who selects the worst-case model is required to pay a penalty that is proportional to the relative entropy between the benchmark model and the worst-case model.

This is not the first paper to study robust portfolio policies. Maenhout (2004) applied Hansen-Sargent robust control methods to a standard Merton-style consumption/portfolio problem. He showed that when the entropy penalty parameter is constant, robust portfolio policies are nonhomothetic, i.e., portfolio shares depend on wealth levels. He went on to show that homotheticity can be preserved if the penalty parameter is scaled by an appropriate function of wealth. Subsequent work has followed Maenhout (2004) by scaling the entropy penalty, and thereby confining attention to homothetic portfolio policies.

Here we do *not* scale the entropy penalty parameter. The problematic long-run implications of nonhomotheticity are not an issue for us, since we study an overlapping generations economy. We show that if the coefficient of relative risk aversion exceeds one, robustness concerns dissipate with wealth. As a result, wealthier agents choose to invest a higher fraction of their wealth in higher yielding assets.[3] This produces a powerful inequality amplification effect. It also provides a novel answer to the question of why inequality began increasing around 1980, not just in the US, but in many other countries as well. Many have argued that the world became more 'turbulent' around 1980. Some point to increased globalization. Others point to technology. Whatever the source, micro evidence supports the notion that individuals began to face greater idiosyncratic risk around 1980.[4] Given this, it seems plausible that idiosyncratic uncertainty increased as well.[5]

Idiosyncratic uncertainty also helps resolve the transition rate puzzle of Gabaix, Lasry, Lions, and Moll (2016). They show that models featuring 'scale dependence', in which shocks to growth rates depend on the level of income or wealth, produce faster transition

[3]Although there is widespread agreement that wealthier individuals earn higher average returns, it is not clear whether this reflects portfolio composition effects, as here, or whether it reflects higher returns *within* asset categories. See below for more discussion.

[4]See, e.g., Gottschalk and Moffitt (1994), Ljungqvist and Sargent (1998), Kambourov and Manovskii (2009).

[5]Note, in our model it is sufficient that agents *perceive* an increase in risk. The increase itself might not actually occur, but fears of its existence would still be relevant if they are statistically difficult to reject.

rates than traditional random growth models based on Gibrat's Law. Robust portfolio policies induced by uncertainty produce a form of scale dependence. We analytically characterize inequality dynamics using the Laplace transform methods popularized by Moll and his co-authors. Although the model itself is nonlinear, this nonlinearity only arises when the inverse of the entropy penalty parameter is nonzero. For small degrees of uncertainty the parameter is close to zero. This allows us to employ classical perturbation methods to obtain approximate analytical solutions of the Laplace transform of the Kolmogorov-Fokker-Planck (KFP) equation, which then yield approximations of the transition rates.

To illustrate the quantitative significance of uncertainty induced inequality, we suppose the US economy was in a stationary distribution without uncertainty in 1980. Even without uncertainty wealth is concentrated at the top due to a combination of investment luck and longevity luck. Assuming agents live/work on average about 40 years, we find that the wealth share of the top 1% is 24.3%, roughly equal to the observed 1980 share. We then inject uncertainty into the economy by setting the (inverse) entropy penalty parameter to a small nonzero value, while keeping all other parameters the same. This increases the top 1% wealth share to 36.9%, close to its current value of about 40%. If this increased inequality had been generated by a change in some other parameter, the transition rate at the mean level of wealth would be only 1.14%, implying a *half-life* of more than 60 years. Thus, assuming we are currently at least 90% of the way to a new stationary distribution, it should have taken 200 years to get here, rather than the observed 35-40 years! However, if increased inequality was instead generated by increased uncertainty, the transition rate at the mean more than triples, to 3.85%. This reduces the model implied transition time from 200 years to about 60 years; still longer than observed, but significantly closer.

Besides Gabaix, Lasry, Lions, and Moll (2016), the only other paper we are aware of that studies wealth inequality dynamics is Aoki and Nirei (2017). Like us, they study a version of the Blanchard-Yaari OLG model. Also like us, a portfolio composition effect is the key force behind increased inequality. However, there are several important differences. First, they focus on income inequality rather than wealth inequality. Second, their model lacks a natural pertubation parameter, so they resort to numerical solutions of the KFP equation. They find that if the variance of idiosyncratic productivity shocks is calibrated to those of publicly traded firms, the model produces transition rates that are comparable to those in the data. However, if privately held firms are included, which is more consistent with the model, transition rates are too slow. Third, and most importantly, the underlying mechanism in their paper is different. They argue that reductions in top marginal income tax rates were the trigger that produced increased inequality. In support, they cite Piketty, Saez, and Stantcheva (2014), who report evidence on top income shares and tax rates from 18 OECD countries for the period 1960-2010. They show that countries experiencing the largest reductions in top marginal income tax rates also experienced the largest increases in top income inequality. We do not dispute the role that tax policy likely played in growing inequality. However, an interesting additional result in Piketty, Saez, and Stantcheva (2014) is that if you split the sample in 1980, the link between taxes and inequality increases markedly. This is exactly what you would expect to find if increased uncertainty coincided with tax reductions in 1980.

The remainder of the paper is organized as follows. Section 2 quickly reviews evidence on US wealth inequality. Section 3 outlines the model. Section 4 compares the stationary distribution of wealth with and without uncertainty. Section 5 provides an approximate analytical characterization of the transition rate between these two distributions and compares it to a numerical solution. Section 6 shows how detection error probabilities can be used to calibrate the entropy penalty parameter (Anderson, Hansen, and Sargent, 2003). Section 7 offers a few concluding remarks, while an Appendix provides proofs and outlines an extension to recursive preferences.

2. Motivation

Our paper focuses on wealth inequality. While there are no doubt important interactions between labor income inequality and wealth inequality (especially in the US), we in some sense are 'tying our hands' by considering a model where labor income heterogeneity plays no role. Hence, here we only present evidence on wealth inequality.

Arguably the best current estimates of US wealth inequality come from Saez and Zucman (2016). They combine data from the Fed's Survey of Consumer Finances with IRS data on capital income. Wealth estimates are computed by capitalizing reported income data. Attempts are made to include the value of assets that do not generate capital income (e.g., pensions and life insurance), but Social Security is excluded. The value of owner occupied housing is computed from data on property taxes and mortgage interest payments. The major omission is human capital. One potential problem with this data is that the methodology assumes rates of return within asset categories are identical across households. This can produce biased estimates if returns are correlated with wealth within asset categories (Fagereng, Guiso, Malacrino, and Pistaferri, 2016b).[6]

In principle, we could look at the entire cross-sectional distribution, but since current interest (and our model) are focused on the right-tail, Figure 1 simply reports the top 1% wealth share.

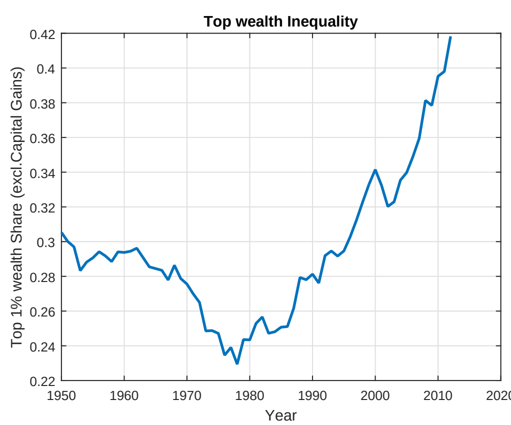

Figure 1. Top 1% Wealth Share in the USA

[6]Kopczuk (2015) discusses the pros and cons of the capitalization method. He notes that other estimation methods (e.g., those based on estate taxes) show a much more gradual increase in inequality.

The most striking feature of this plot is the U-shaped pattern of top wealth shares. Wealth has always been concentrated at the top, but top wealth shares actually declined from 1950-80. This was part of much longer process that began during WWI, which has been discussed by Piketty (2014) and others. Here we focus on the increase that began around 1980. From its minimum of 23% in 1978, the top 1% wealth share has increased steadily ever since. By 2012 it had roughly doubled, to 42%.

Who are these top 1%? In 2012, the top 1% consisted of the wealthiest 1.6 million families, with a minimum wealth of about $4 million. Average wealth in this group is about $13.8 million. As noted by Saez and Zucman (2016), wealth has become even more concentrated than indicated by the conventional focus on the top 1%. The top 0.1% now own 22% of US wealth, more than triple its value in 1978. To belong to this exclusive club, one needs a *minimum* wealth of $20.6 million. Although we focus on the top 1%, it should become clear below that our model would likely do even better at accounting for the top 0.1%.

It is of course debatable whether an economy is ever in a stationary equilibrium, but our basic modeling strategy here is to suppose that the US economy was in such a state around 1980. We then ask whether an empirically plausible increase in (Knightian) uncertainty could have produced the sort of rapid increase in inequality observed in Figure 1. To address this question we need a model.

3. The Model

The basic model here is the workhorse continuous-time Blanchard-Yaari OLG model. Benhabib, Bisin, and Zhu (2016) have recently studied the implications of this model for wealth inequality in the absence of uncertainty (i.e., when agents fully trust their models of asset returns). The key simplifying assumption, which makes the model so tractable, is that lifetimes are exponentially distributed. This eliminates life-cycle effects. Unlike Benhabib, Bisin, and Zhu (2016), we ignore bequests. Unintentional bequests are not an issue, because another key simplifying assumption of the Blanchard-Yaari model is the existence of perfectly competitive annuity markets. This allows agents to borrow and lend at a constant rate despite their mortality risk. Of course, ignoring bequests is not innocuous when studying wealth inequality. Dynastic wealth accumulation undoubtedly plays a role in observed wealth inequality. However, it is less clear whether the role of bequests has *changed*. If anything, it has likely decreased (Saez and Zucman, 2016). Since we are primarily interested in explaining the rapid rise in inequality, it seems safe to ignore bequests.

The economy is comprised of a measure 1 continuum of finitely-lived agents. Death occurs at Poisson rate, δ. Hence, mean lifetimes are δ^{-1}. When an agent dies he is replaced by a new agent with wealth w_0. This initial wealth can either be interpreted as the capitalized value of expected (riskless) labor income or, following Benhabib, Bisin, and Zhu (2016), as transfers funded by capital income taxes. In the latter case, rates of return should be interpreted as net of taxes. The important point is that w_0 is identical across all newborn agents.

Agents can invest in three assets: (1) A risk-free technology.[7] The value of the risk-free asset follows the deterministic process

$$dQ = \tilde{r}Qdt$$

(2) Annuities issued by competitive insurance companies, with rate of return r. By no arbitrage and zero profits, $r = \tilde{r} + \delta$. Hence, agents devote all their risk-free investments to annuities. (3) A private/idiosyncratic risky technology. Agents are assumed to share the following benchmark model for the returns on risky capital

$$dS = \mu S dt + \sigma S dB \tag{3.1}$$

where μ is the common mean return, and σ is the common idiosyncratic volatility. The noise, dB, is an increment to a standard Brownian motion, and is assumed to be uncorrelated across individuals.

Again, the key departure point of this paper is to suppose that agents have doubts about their idiosyncratic investment opportunities. In other words, they confront 'uncertainty', not risk. By assumption, these doubts cannot be captured by specifying a finite-dimensional Bayesian prior over alternative models. Instead, agents fear a broad spectrum of nonparametric alternatives. These alternatives might reflect omitted variables or complicated nonlinearities. As emphasized by Hansen and Sargent (2008), we want agents to be prudent, not paranoid, so that they only hedge against models that could have plausibly generated the historically observed data. To operationalize this, let q_t^0 be the probability measure defined by the Brownian motion process in the benchmark model (3.1), and let q_t be some alternative probability measure, defined by some competing model. The (discounted) relative entropy between q_t and q_t^0 is then defined as follows:[8]

$$\mathcal{R}(q) = \int_0^\infty e^{-\rho t} \left[\int \log\left(\frac{dq_t}{dq_t^0}\right) dq_t \right] dt \tag{3.2}$$

Evidently, $\mathcal{R}(q)$ is just an expected log-likelihood ratio statistic, with expectations computed using the distorted probability measure. It can also be interpreted as the Kullback-Leibler 'distance' between q_t and q_t^0. From Girsanov's Theorem we have

$$\int \log\left(\frac{dq_t}{dq_t^0}\right) dq_t = \frac{1}{2}\tilde{E}\int_0^t |h_s|^2 ds$$

where $\tilde{E}$ denotes expectations with respect to the distorted measure q_t, and h_s represents a square-integrable process that is progressively measureable with respect to the filtration generated by q_t. Again from Girsanov's Theorem, we can view q_t as being induced by the following drift distorted Brownian motion[9]

$$\tilde{B}(t) = B(t) - \int_0^t h_s ds$$

[7]This can be interpreted as a small open-economy assumption.

[8]See Hansen, Sargent, Turmuhambetova, and Williams (2006) for a detailed discussion of robust control in continuous-time models, and in particular, on the role of discounting in the definition of relative entropy.

[9]There are some subtleties here arising from the possibility that $\tilde{B}$ and B generate different filtrations. See Hansen, Sargent, Turmuhambetova, and Williams (2006) for details.

which then defines the following conveniently parameterized set of alternative models

$$dS = (\mu + \sigma h)Sdt + \sigma S d\tilde{B}$$

Agents are assumed to have time-additive CRRA preferences. Each agent wants to construct robust consumption and portfolio policies that perform adequately for all h's that lie within some convex set centered around the benchmark model in (3.1). To do this, agents enlist the services of a hypothetical 'evil agent', who is imagined to select models that *minimize* their utility. That is, agents view themselves as being engaged in the following dynamic zero-sum game:

$$V(w_0) = \max_{c,\alpha} \min_h E \int_0^\infty \left(\frac{c^{1-\gamma}}{1-\gamma} + \frac{1}{2\varepsilon} h^2 \right) e^{-(\rho+\delta)t} dt \tag{3.3}$$

subject to

$$dw = [(r + \alpha(\mu - r))w - c + \alpha\sigma w h]dt + \alpha w \sigma \cdot dB \tag{3.4}$$

The parameters have their usual interpretations: w is the agent's wealth; c is his rate of consumption; α is the share of wealth invested in the risky technology; γ is the coefficient of relative risk aversion, and ρ is the rate of time preference. The key new parameter here is ε. It penalizes the evil agent's drift distortions. If ε is small, the evil agent pays a big cost when distorting the benchmark model. In the limit, as $\varepsilon \to 0$, the evil agent sets $h = 0$, and agents no longer doubt their risky technologies. Decision rules and the resulting equilibrium converge to those in Benhabib, Bisin, and Zhu (2016). Conversely, as ε increases, agents exhibit a greater preference for robustness.

We can solve the robustness game using dynamic programming. Ito's Lemma implies the Hamilton-Jacobi-Bellman equation can be written

$$(\rho + \delta)V(w) = \max_{c,\alpha} \min_h \left\{ \frac{c^{1-\gamma}}{1-\gamma} + \frac{1}{2\varepsilon} h^2 + [(r + \alpha(\mu - r))w - c + \alpha\sigma w h]V'(w) + \frac{1}{2}\alpha^2\sigma^2 w^2 V''(w) \right\} \tag{3.5}$$

The first-order conditions deliver the following policy functions in terms of the unknown value funcion, $V(w)$:

$$h = -\varepsilon\sigma\alpha w V'(w) \tag{3.6}$$

$$\alpha = -\frac{(\mu - r)V'(w)}{[V''(w) - \varepsilon(V'(w))^2]w\sigma^2} \tag{3.7}$$

$$c = [V'(w)]^{-1/\gamma} \tag{3.8}$$

Even before solving the model, these policy functions reveal a lot about the equilibrium. First, notice that robustness takes the form of *pessimism*, i.e., by (3.6) the drift distortion h is negative since $\alpha > 0$ and $V'(w) > 0$. Second, notice that the distortion increases with volatility ($\sigma\alpha w$). When volatility increases, it becomes statistically more difficult to rule out alternative models. Third, notice that it is not *a priori* obvious how the magnitude of the distortion depends on wealth. If $V(w)$ is concave, so that wealthier agents have a lower marginal utility of wealth, then (ceteris paribus) weathier agents will be less pessimistic. Having money in the bank allows you to relax. However, the volatility term offsets this. For a given portfolio share, α, wealthier agents have more at stake. This makes them more worried, and triggers a greater preference for robustness and a more pessimistic drift distortion.[10] This volatility effect will be reinforced if the first effect leads wealthier agents to invest a higher fraction of their wealth in the risky technology.[11]

Turning to the portfolio policy in (3.7), we can see that robustness in some sense makes the investor more risk averse, in that it subtracts a positive term from $V''(w)$. However, as noted by Maenhout (2004), unless the utility function is quadratic ($\gamma = -1$) or logarithmic ($\gamma = 1$), the resulting portfolio policy is nonhomothetic, and α will depend on w. Observe that the implicit degree of risk aversion falls as w increases if $V(w)$ is concave, so that $V'(w)$ decreases with wealth. In the limit, as $V'(w) \to 0$, the preference for robustness completely dissipates, and the limiting portfolio is the same as when $\varepsilon = 0$. Notice that this portfolio is just the classic Merton portfolio

$$\alpha_0 = \frac{\mu - r}{\gamma\sigma^2} \tag{3.9}$$

since one can readily verify that $V(w) \sim w^{1-\gamma}$ in this case. Finally, notice that the consumption function in (3.8) is the usual one. However, that does not mean uncertainty is irrelevant to saving decisions, since in general $V(w)$ will depend in complicated ways on the value of ε.

This analysis of the policy functions was premised on the properties of the unknown value function, $V(w)$. To verify and quantify these effects we must solve for $V(w)$. To do so we first substitute the policy functions in (3.6)-(3.8) into the HJB equation in (3.5), and then solve the resulting second-order ODE. Unfortunately, the resulting equation is nonlinear, and does not possess a known analytical solution. However, the fact that it *does* have a known analytical solution when $\varepsilon = 0$ opens the door to a classical perturbation approximation. That is, we look for a solution of the form[12]

$$V(w) = V_0(w) + \varepsilon V_1(w) + O(\varepsilon^2)$$

We know that $V_0(w)$ will just be the standard Merton solution, $V_0(w) = A_0 \frac{w^{1-\gamma}}{1-\gamma}$, where A_0 is a complicated but well known function of the underlying parameters. The $V_1(w)$ function turns out to satisfy a *linear* ODE with a nonhomogeneous forcing term given by $V_0(w)$. The results are given by

Proposition 3.1. *A first-order pertubation approximation of the value function for the HJB equation in* (3.5) *is given by*

$$V(w) = A_0 \frac{w^{1-\gamma}}{1-\gamma} + \varepsilon A_1 \frac{w^{2(1-\gamma)}}{2(1-\gamma)} + O(\varepsilon^2) \tag{3.10}$$

[10] To quote Janis Joplin (or actually, Kris Kristofferson): 'Freedom's just another word for nothing left to lose'.

[11] Below we will show that this is the case if and only if $\gamma > 1$.

[12] See Anderson, Hansen, and Sargent (2012) for a more sophisticated (and accurate) perturbation approximation to robust control problems.

with

$$A_0 = \left[\frac{1}{\gamma}\left(\rho+\delta-(1-\gamma)\left(r+\frac{(\mu-r)^2}{2\gamma\sigma^2}\right)\right)\right]^{-\gamma}$$

$$A_1 = \frac{1}{2}\left[\frac{\sigma^2(\alpha_0 A_0)^2}{r-A_0^{-1/\gamma}+\frac{1}{2}\gamma\sigma^2\alpha_0^2(1+\beta)-(\rho+\delta)/\beta}\right]$$

where $\beta = 2(1-\gamma)$ *and* α_0 *is given by the nonrobust portfolio share in equation* (3.9).

Proof. See Appendix A. □

This result immediately yields the following corollary, which characterizes the policy functions:

Corollary 3.2. *First-order perturbation approximations of the optimal portfolio policy,* $\alpha(w)$*, and saving rate,* $s(w) = 1 - c(w)/w$*, are given by*

$$\alpha(w) = \alpha_0 - \varepsilon\alpha_0\left(\frac{A_0^2+(\gamma-1)A_1}{\gamma A_0}\right)w^{1-\gamma} \tag{3.11}$$

$$s(w) = 1 - A_0^{-1/\gamma} + \varepsilon\frac{1}{\gamma}\left(A_0^{-1/\gamma-1}A_1\right)w^{1-\gamma} \tag{3.12}$$

where α_0 *is the nonrobust portfolio policy in* (3.9)*, and* (A_0, A_1) *are constants defined in Proposition 3.1.*

Proof. See Appendix B. □

Notice that there are two cases when these policy functions are homothetic (i.e., independent of wealth). First, and most obviously, is when $\varepsilon = 0$. In this case, the agent has no doubts, and the model degenerates to the standard case of risky investment, with no uncertainty. Second, when $\gamma = 1$ (i.e., preferences are logarithmic), the approximate decision rules become independent of wealth. In this case, the saving rate assumes its usual Permanent Income value, $s = 1 - (\rho + \delta)$. Although the portfolio policy is homothetic, uncertainty still matters, but its effect is isomorphic to enhanced (constant) relative risk aversion. It simply scales down the Merton portfolio share, $\alpha = \alpha_0(1 - \varepsilon A_0)$.

More generally, however, the portfolio share and saving rate depend on wealth. This dependence is the key mechanism in our paper. Whether our model can *quantitatively* account for observed wealth inequality dynamics depends on the quantitative properties of these functions for plausible values of the parameters.

3.1. **Comment on Units.** A convenient implication of homotheticity is that model parameterizations become independent of scale. Units do not matter. Here that is not the case, so we must be a careful when selecting parameters. Notice that the key source of nonhomotheticity is the term $\varepsilon w V'(w)$. (In principle, α matters too, but its value is bounded between $[0, 1]$). This determines the order of magnitude of the drift distortion. Later we discuss how detection error probabilities can be used to calibrate this distortion. For now, just note that for a given empirically plausible drift distortion, the units of ε become linked to those of w. For a given distortion, larger values of w can either increase or decrease $wV'(w)$, depending on the value of γ. Whichever way it goes, ε must move inversely so as to maintain the given drift distortion. Hence, the units of w are still irrelevant as long as we are careful to adjust the magnitude of the penalty parameter appropriately. The only caveat is that our perturbation approximation presumes a 'small' value of ε, so for given values of the remaining parameters, this puts limits on the units of w.

3.2. **Benchmark Parameterization.** To illustrate the quantitative effects of model uncertainty on saving and portfolio choice we use the following benchmark parameter values

TABLE 1
Benchmark Parameter Values

w_0	μ	r	δ	σ^2	ρ	γ	ε
100	.0586	.044	.026	.0085	.021	1.31	.045

We discuss these parameters in more detail in the following section. For now it is sufficient to focus on γ, since it determines the nature of the saving and portfolio policies. Empirical evidence yields a wide range of estimates, depending on data and model specifics. However, most estimates suggest $\gamma > 1$. We set $\gamma = 1.31$. Figure 2 plots the resulting portfolio policy:

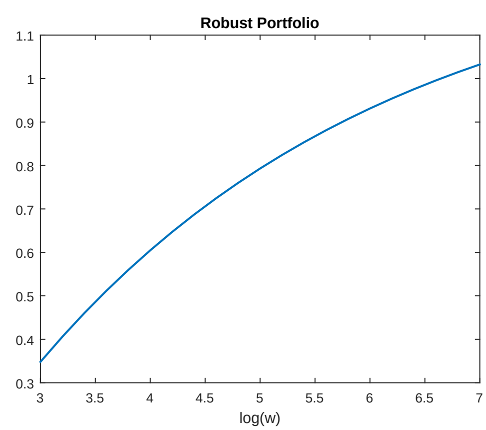

Figure 2. Robust Portfolio Share

Later we shall see that it is more convenient when computing the cross-sectional wealth distribution to work in terms of the natural log of wealth, so the horizontal axis plots $\log(w)$. Note that log wealth at birth is $\log(w_0) = \log(100) \approx 4.6$. Hence, newborns invest about 65% of their wealth in the risky technology. If they are lucky, and their wealth grows, so does their investment share in the risky technology. If log wealth increases to 7, implying a ten-fold increase in the level of wealth, the investment share reaches 100%.

Figure 2 reveals the key feedback mechanism in our model. Wealth begets wealth because it leads agents to invest in higher yielding technologies.[13] Notice, however, that inequality is not only a result of wealthy agents investing more than average in higher yielding assets. The effect is symmetric, in that bad luck and low wealth makes agents pessimistic, which leads them to hold most of their wealth in the safe asset. The very poorest agents invest only about 30% in the risky technology. Our model therefore provides a novel explanation for why poor households hold most, if not all, of their wealth in safe low-yielding assets. It is also broadly consistent with the empirical evidence in Carroll (2002), who examines portfolio data from the Survey of Consumer Finances for the period 1962-95. He finds that the top 1% hold about 80% of their wealth in risky assets, while the bottom 99% hold (on average) about 40%. (See his Table 4).

The other force driving inequality is saving. Empirical evidence suggests wealthy households have higher saving rates (Saez and Zucman, 2016). Wealth increasing saving rates also produce deviations from Gibrat's Law, which amplifies inequality. Figure 3 plots individual saving rates as a function of log wealth:

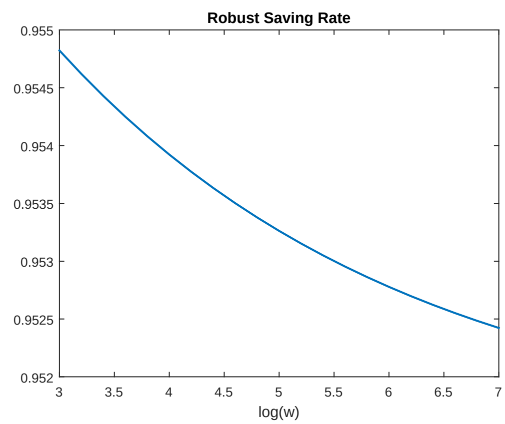

FIGURE 3. Robust Saving Rate

Evidently, the model's predictions along this dimension are counterfactual. Saez and Zucman (2016) report 'synthetic' saving rate estimates that increase with wealth. Dynan, Skinner, and Zeldes (2004) also find increasing saving rates in a variety of US data sets. Note that a saving rate that decreases with wealth actually *reduces* inequality, since poorer households save a larger fraction of their wealth. What's happening here is that, for a given portfolio allocation, wealthy agents are less pessimistic and expect higher mean returns. Because the intertemporal elasticity of substitution $(1/\gamma)$ is less than one, higher expected returns reduce the savings rate, since the income effect dominates the substitution effect. This suggests that we could avoid this problem by employing recursive preferences, which

[13]From eq. (3.11) and the expressions for (A_0, A_1) in Proposition 3.1, we can see that $A_0 > 0$ whenever $\gamma > 1$. Hence, a sufficient condition for $\alpha'(w) > 0$ is that $A_1 > 0$. A necessary and sufficient condition for $A_1 > 0$ is that $\gamma^{-1}[r - (\rho + \delta)] + \frac{1}{2}\frac{\rho+\delta}{\gamma-1} + \frac{1}{2}\sigma^2\alpha_0^2[3(1-\gamma)+1] > 0$. A sufficient condition for this is that $\gamma < 1.33$. Alternatively, if $\gamma > 1.33$, then the condition will be satisfied if $r \geq \rho + \delta$ and σ^2 is sufficiently small.

delivers a separation between risk aversion and intertemporal substitution. Appendix C outlines such an extension using a Duffie-Epstein (1992) aggregator with an intertemporal elasticity of substitution of one. It shows that for very similar parameter values to those in Table 1, saving rates actually increase with wealth.[14]

Together, Figures 2 and 3 suggest that whether uncertainty amplifies inequality depends on the relative strength of the portfolio effect versus the saving effect. In the following section we shall see that, at least for these parameter values, the portfolio effect dominates, and uncertainty amplifies inequality. This result is perhaps not too surprising if you look closely at Figures 2 and 3. Notice that the decline in saving is very mild. The poorest households save only a few tenths of a percent more than the wealthiest. In contrast, the portfolio allocation effect is quite strong.

The mechanism driving these portfolio and saving dynamics is a pessimistic drift distortion. Earlier it was noted that it is not at all clear whether wealthy agents are more or less pessimistic in equilibrium. On the one hand, wealth provides financial security, so wealthy agents can afford to be less robust. This effect operates via the decline in $V'(w)$ in the evil agent's policy function. On the other hand, greater financial security enables you to take on more risk. As a result, wealthier agents have more to lose from model misspecification. This effect operates via the scaling term, αw, in the evil agent's policy function. The following result shows that for small values of ε the first effect dominates if $\gamma > 1$.

Corollary 3.3. *To a first-order approximation, the equilibrium distortion function, $h(w)$, is increasing if and only if $\gamma > 1$.*

Proof. Note that the portfolio effect on $h(w)$ is 2nd-order in ε. To a first-order approximation we can set $\alpha = \alpha_0$. Therefore, we have (ignoring inessential constants)

$$h(w) \sim -\varepsilon w V'(w)$$

Again, to a first-order approximation, we can ignore the $\varepsilon V_1(w)$ component of $V(w)$. Hence,

$$h(w) \sim -\varepsilon w V_0'(w) = -\varepsilon A_0 w^{1-\gamma} \qquad \Rightarrow \qquad h'(w) \sim -\varepsilon(1-\gamma)w^{-\gamma}$$

□

To interpret the $h(w)$ function it is convenient to map it into an implied drift distortion. Notice that if $h(w)$ is substituted into the HJB equation in (3.5) the agent's problem becomes equivalent to the standard one, but with an endogenous, wealth-dependent, drift equal to $\mu + \sigma h(w)$. Figure 4 plots this implied drift distortion using the benchmark parameter values given in Table 1:

Perhaps not surprisingly, the drift distortion mirrors the portfolio policy depicted in Figure 2. Agents begin life with a small degree of pessimism, equivalent to about a 60 basis point negative drift distortion. However, if they are unlucky, and they move into the left tail of the wealth distribution, their pessimism grows to more than a 1.0 percentage point negative drift distortion. As seen in Figure 2, this is enough to nearly keep them

[14]Borovicka (2016) makes a similar point. He notes that long-run survival in models featuring heterogeneous beliefs depends on a delicate interplay between risk aversion, which governs portfolio choice, and intertemporal substitution, which governs saving.

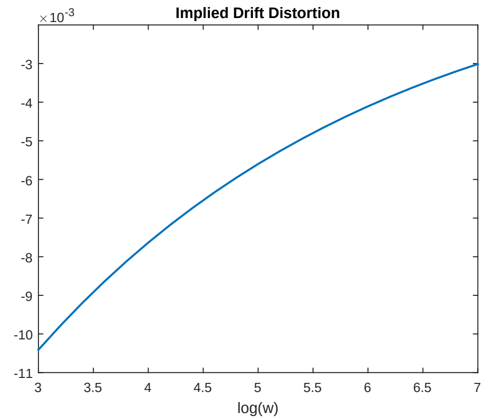

FIGURE 4. Drift Distortion

out of the market. Hence, robustness generates something like a 'poverty trap', in that pessimism and low wealth become self-fulfilling.[15]

3.3. Comment on Preferences vs. Environment. Before turning to the aggregate implications of the model, we briefly comment on an identification issue. As noted by Hansen, Sargent, Turmuhambetova, and Williams (2006), from a mathematical perspective continuous-time robust control is just a special (nonhomothetic) case of Duffie and Epstein's (1992) Stochastic Differential Utility (SDU) preferences. For a *constant* value of ε, the two are observationally equivalent. Following Hansen and Sargent (2008), we prefer the robust control interpretation because it views ε as an aspect of the environment, which is subject to change. It is *changes* in the distribution of wealth we are attempting to explain. In contrast, SDU views ε as an attribute of preferences, and for the usual reasons, we prefer to think of preferences as being time invariant.

4. Stationary Distributions

So far we have focused on the problem and decision rules of a single agent. Our primary interest, however, is on wealth inequality. To address this we must aggregate the individual wealth dynamics and compute the cross-sectional distribution of wealth. The first step is to substitute the optimal policy functions into the individual's budget constraint in equation (3.4). When doing this it is important to observe that, as in Hansen and Sargent (2008), we assume fears of model misspecification are solely in the mind of the agent. The agent's benchmark model is the true data-generating process, the agent just doesn't know it. Uncertainty still matters, however, because it influences behavior.

It turns out to be mathematically convenient to work in terms of log wealth, $x = \log(w)$. Using Ito's Lemma to translate from w to x gives us:

[15]Note that if the aggregate economy exhibited growth, represented here as a steadily rising w_0, then for a given ε pessimism would disappear from the economy, and so would the robustness-induced poverty trap. Hence, our model is more suited to explain the temporary effects of *change* in ε.

Proposition 4.1. *To a first-order approximation in* ε*, individual log wealth follows the diffusion process*

$$dx = (a_0 + \varepsilon a_1 e^{(1-\gamma)x})dt + (b_0 + \varepsilon b_1 e^{(1-\gamma)x})dB \tag{4.13}$$

with

$$\begin{aligned} a_0 &= r - A_0^{-1/\gamma} + \gamma\sigma^2\alpha_0^2 - \frac{1}{2}b_0^2 \\ a_1 &= \frac{\gamma A_1 A_0^{-1/\gamma} - (\sigma\gamma\alpha_0)^2(A_0^2 + (\gamma-1)A_1)}{A_0\gamma^2} - b_0 b_1 \\ b_0 &= \sigma\alpha_0 \\ b_1 &= -\sigma\gamma\alpha_0\left[\frac{A_0^2 + (\gamma-1)A_1}{A_0\gamma^2}\right] \end{aligned}$$

where (A_0, A_1) *are constants defined in Proposition 3.1 and* α_0 *is the nonrobust portfolio share defined in equation* (3.9).

Proof. See Appendix D. □

Notice that the drift and diffusion coefficients are constant when either $\varepsilon = 0$ or $\gamma = 1$. Also notice that the exponential terms in the coefficients damp to zero as wealth grows when $\gamma > 1$. However, when $(a_1, b_1) < 0$ this diminishing effect implies that the mean growth rate and volatility of wealth are increasing in the level of wealth.

Next, let $f(t, x)$ be the time-t cross-sectional distribution of wealth. It obeys the following Kolmogorov-Fokker-Planck (KFP) equation:

$$\frac{\partial f}{\partial t} = -\frac{\partial[(a_0 + \varepsilon a_1 e^{(1-\gamma)x})f]}{\partial x} + \frac{1}{2}\frac{\partial^2[(b_0 + \varepsilon b_1 e^{(1-\gamma)x})^2 f]}{\partial x^2} - \delta f + \delta\zeta_0 \tag{4.14}$$

where ζ_0 is a Dirac delta function at $x_0 = \log(w_0)$. This is a *linear* PDE, which can be solved using (two-sided) Laplace transforms.

$$\mathcal{L}\{f(x)\} \equiv F(t, s) \equiv \int_{-\infty}^{\infty} f(t, x)e^{-sx}dx$$

Applying the Laplace transform to equation (4.14) and dropping $O(\varepsilon^2)$ terms yields the following approximation of the cross-sectional wealth dynamics:[16]

$$\frac{\partial F}{\partial t} = \Lambda(s)F(t, s) + \varepsilon\Phi(s-\beta)F(t, s-\beta) + \delta e^{-sx_0} \tag{4.15}$$

where $\beta \equiv 1 - \gamma$ and

$$\Lambda(s) = \frac{1}{2}b_0^2 s^2 - a_0 s - \delta \tag{4.16}$$

$$\Phi(s) = b_0 b_1 s^2 + (2b_0 b_1 \beta - a_1)s + \beta(b_0 b_1 \beta - a_1) \tag{4.17}$$

The $\Lambda(s)$ and $\Phi(s)$ functions pick-up the derivatives of the KFP equation, since $\mathcal{L}\{f'(x)\} = sF(s)$ and $\mathcal{L}\{f''(x)\} = s^2F(s)$. The fact that derivatives are converted to a simple multiplication is what makes Laplace transforms so useful. This method is less commonly

[16]Remember, as with moment-generating functions, which the Laplace transform generalizes, all the information about the cross-sectional distribution is contained in the transform.

applied to problems with variable coefficients, as is the case here, since the Laplace transform of the product of two functions is not the product of their Laplace transforms.[17] However, the particular form of variable coefficients here, which involves multiplication by an exponential function, is one case that works out nicely. In particular, the so-called 'shift theorem' implies $\mathcal{L}\{e^{\beta x}f(x)\} = F(s-\beta)$. This result follows rather obviously from the defintion of a Laplace transform after a simple change of variables.

The $F(t, s-\beta)$ term on the right-hand side of (4.15) makes the KFP equation nonstandard. Later we shall obtain an approximate solution by approximating the discrete shift with a derivative. For now, notice that in the nonrobust case, when $\varepsilon = 0$, the problem becomes entirely standard, and we get:

Proposition 4.2. *The nonrobust stationary distribution of log wealth is double exponential, with a mode at* $x_0 = \log(w_0)$,

$$f(x) = c_1 h(x_0 - x)e^{\phi_1 x} + c_2 h(x - x_0)e^{\phi_2 x} \tag{4.18}$$

where $h(\cdot)$ *is the Heaviside (unit-step) function, and* (c_1, c_2) *are constants of integration chosen to ensure continuity at the switch point,* x_0*, and satisfaction of the adding-up constraint* $\int_{-\infty}^{\infty} f(x) = 1$. *The exponents* (ϕ_1, ϕ_2) *are the positive and negative roots, respectively, of the quadratic* $\Lambda(s) = 0$*, where* $\Lambda(s)$ *is given by equation* (4.16).

Proof. Since this result is standard, we merely sketch a proof. First, the Laplace transform of the stationary distribution can be obtained by setting $\partial F/\partial t = 0$ in equation (4.15), and then solving the resulting algebraic equation for $F(s)$. When $\varepsilon = 0$, this just gives $F(s) = -\delta e^{-sx_0}/\Lambda(s)$. Next, we invert $F(s)$ to get $f(x)$. This can either be accomplished by expanding $\Lambda(s)^{-1}$ into partial fractions, and using the result that $\mathcal{L}^{-1}\{(s-\phi)^{-1}\} = e^{\phi x}$, or by using contour integration and the residue calculus, noting that the singularity in the left-half plane corresponds to $x > 0$, while the singularity in the right-half plane corresponds to $x < 0$. □

In applications we are really more interested in the distribution of the *level* of wealth. However, this distribution follows immediately from the previous result:

Corollary 4.3. *In the nonrobust* ($\varepsilon = 0$) *case, the stationary distribution of wealth is double Pareto*

$$f(w) = c_1 h(w_0 - w)w^{\phi_1 - 1} + c_2 h(w - w_0)w^{\phi_2 - 1} \tag{4.19}$$

Proof. The Jacobian of the transformation from x to w is w^{-1}. □

The fact that the conventional (nonrobust) distribution of wealth is Pareto is no accident. This is a well documented feature of the data, and the main attraction of constant parameter idiosyncratic investment risk models is that they quite naturally generate a

[17] Instead, the Laplace transform of the *convolution* of two functions is the product of their Laplace transforms.

steady-state Pareto distribution.[18] The key parameter here is ϕ_2,

$$\phi_2 = \frac{a_0 - \sqrt{a_0^2 + 2\delta b_0^2}}{b_0^2} < 0$$

As $|\phi_2|$ decreases, top wealth shares increase. Although values of $\phi_2 \in [-1, 0)$ generate a stationary distribution, mean wealth becomes infinite if $|\phi_2| \leq 1$, so we restrict parameters to satisfy the constraint $|\phi_2| > 1$. The comparative statics of this parameter are intuitive. For example, $|\phi_2|$ decreases as δ decreases. When people live longer, saving and portfolio decisions become more important, and there is more time to accumulate wealth. $|\phi_2|$ also decreases when σ^2 increases, since luck plays a role in generating inequality.

The basic strategy for computing the robust stationary distribution is the same as above, but now the equation characterizing this distribution is the more complicated functional equation

$$\Lambda(s)F(s) + \varepsilon\Phi(s-\beta)F(s-\beta) + \delta = 0 \tag{4.20}$$

Although a perturbation approximation with respect to ε seems natural, the distribution changes quite quickly, even for $\varepsilon \approx 0$, making a first-order approximation unreliable. This will become apparent later. So instead we approximate (4.20) by converting it to an ODE in s, using β as a step-size. This presumes β is 'small', so that $\gamma \approx 1$. Solving this ODE yields the following result,

Proposition 4.4. *To a first-order approximation in β, the robust stationary distribution of log wealth is double exponential, with a mode at $x_0 = \log(w_0)$,*

$$f_r(x) = c_{1r}h(x_0 - x)e^{\psi_1 x} + c_{2r}h(x - x_0)e^{\psi_2 x} \tag{4.21}$$

where $h(\cdot)$ is the Heaviside (unit-step) function, and (c_{1r}, c_{2r}) are constants of integration chosen to ensure continuity at the switch point, x_0, and satisfaction of the adding-up constraint $\int_{-\infty}^{\infty} f(x) = 1$. The exponents (ψ_1, ψ_2) are the positive and negative roots, respectively, of the quadratic

$$\varepsilon + \Lambda(s) + \beta\Phi'(s-\beta) = 0$$

where $\Phi'(s-\beta)$ is the derivative of the $\Phi(s)$ function in (4.17) evaluated at $(s-\beta)$.

Proof. See Appendix E. □

Hence, we have distilled the question of how uncertainty influences wealth inequality into the question of how ψ_2 compares to ϕ_2. If $|\psi_2| < |\phi_2|$, then uncertainty leads to greater inequality. In general, there are two competing forces. On the one hand, doubts create pessimism, which discourages investment, which in turn reduces the mean growth rate of wealth for *everyone*. Lower mean growth reduces inequality. On the other hand, since pessimism depends on the level of of wealth, this creates heterogeneity in the growth rate of wealth (or 'scale dependence' in the language of Gabaix, Lasry, Lions, and Moll (2016)),

[18]Reed (2001) appears to have been the first to note that the combination of exponentially distributed lifetimes and idiosyncratic investment returns following geometric Brownian motions generates a double Pareto wealth distribution. Toda (2014) shows that this result is robust to a number of extensions, e.g., aggregate shocks, recursive preferences, endogenously determined interest rates, and heterogeneous initial wealth. However, he assumes homotheticity and wealth independent portfolio shares.

and this heterogeneity *increases* inequality. We shall see that for reasonable parameter values the heterogeneity effect dominates.

Although it is not obvious from the statement of Proposition 4.4, one can also show that the robust stationary distribution satisfies the following correspondence principle,

Corollary 4.5. *As $\varepsilon \to 0$, the robust stationary distribution in eq.* (4.21) *converges pointwise to the nonrobust stationary distribution in eq.* (4.18).

Proof. See Appendix D. □

4.1. **Top Wealth Shares.** It should be emphasized that Propositions 4.2 and 4.4 characterize the *entire* cross-sectional distribution of wealth. They could be used to compute Gini coefficients, Theil/entropy indices, or any other inequality measure of interest. However, in line with recent debate, we focus here on top wealth shares, in particular, on the top 1% wealth share. Given the stationary distribution $f_\infty(w)$, this is defined as follows

$$\text{top 1\% share} = \frac{\int_{\hat{w}}^{\infty} w f_\infty(w)dw}{\int_0^{\infty} w f_\infty(w)dw}$$

where $\hat{w}$ solves $\int_0^{\hat{w}} f_\infty(w)dw = .99$. With a double Pareto distribution, these integrals can be computed analytically. This yields,

Lemma 4.6. *Given a double Pareto distribution as in eq.* (4.19)*, the top 1% wealth share is given by*

$$\textit{top 1\% share} = \frac{c_2 \frac{\hat{w}^{1+\phi_2}}{1+\phi_2}}{c_2 \frac{w_0^{1+\phi_2}}{1+\phi_2} - c_1 \frac{w_0^{1+\phi_1}}{1+\phi_1}} \tag{4.22}$$

where $\hat{w} = w_0 \left[1 + \phi_2(.99/\bar{c} - 1/\phi_1)\right]^{1/\phi_2}$ *and* $\bar{c} = \phi_1\phi_2/(\phi_2 - \phi_1)$.

Standard (one-sided) Pareto distributions have a well known fractal nature, meaning that any given wealth share is a scalar multiple of any other. For example, if we ignore the left-half of the above distribution, the top 0.1% share would simply be $10^{-\phi_2^{-1}-1}$ as large as the top 1% share, while the top 10% share would be $\left(\frac{1}{10}\right)^{-\phi_2^{-1}-1}$ larger. Evidently, wealth shares implied by double Pareto distributions bear a (slightly) more complicated relationship to each other.

4.2. **Benchmark Parameterization.** Here we plot and compare the stationary distributions for two economies. In the first, agents only confront idiosyncratic risk, meaning $\varepsilon = 0$. We assume this distribution described the US economy in 1980. In the second, agents confront idiosyncratic 'uncertainty', meaning ε is set to a small positive value. We assume this is the distribution toward which the US economy has been gravitating. All the remaining parameters are set to the benchmark values displayed in Table 1.

As discussed earlier, initial wealth, w_0, is arbitrary, as long as we are careful to adjust ε to maintain a 'reasonable' drift distortion. (Later we discuss how we define reasonable, but Figure 4 shows that it is only around 1%, even for the poorest agents). We simply normalize initial wealth to $w_0 = 100$. Moskowitz and Vissing-Jorgensen (2002) provide evidence from the 1989 SCF on the distribution of private equity returns, and report a value of 6.9% for the median return, so we set $\mu = 5.86\%$, to be conservative. The more important parameters are δ, σ^2, and γ. It is natural to calibrate δ to the mean work life, keeping in mind that this has been changing over time and that the model abstracts from important life-cycle effects. We simply set $\delta = 0.026$, which implies a mean work life of about 38 years.[19] Calibrating σ^2 is more difficult. The model is based on idiosyncratic investment risk, so using values from the stock market would be misleading. Benhabib and Bisin (2017) cite evidence on the returns to owner-occupied housing and private equity which suggest a standard deviation in the range of 10-20%, so to be conservative we set $\sigma = 9.2\%$. Finally, the basic mechanism in our paper hinges on people becoming less pessimistic as they become wealthier. For this we need $\gamma > 1$. Fortunately, this is consistent with a wide range of evidence. However, there is less agreement on the magnitude of γ. Asset market data imply large values. However, as noted by Barillas, Hansen, and Sargent (2009), these large values are based on models that abstract from model uncertainty. They show that a modest amount of uncertainty can substitute for extreme degrees of risk aversion, and their argument applies here as well. In addition, our approximate solution strategy presumes that $1 - \gamma$ is small, so we set $\gamma = 1.31$.

Figure 5 displays the two distributions, using the results in Propositions 4.2 and 4.4,

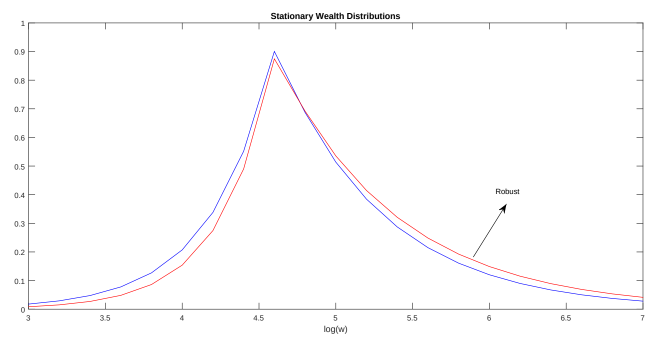

FIGURE 5. Stationary Distributions

The main thing to notice is that the robust distribution has a fatter right tail, reflecting greater wealth inequality. The nonrobust distribution has a right-tail exponent of $\phi_2 = -1.45$, whereas the robust distribution has a right-tail exponent of $\psi_2 = -1.28$. Using Lemma 4.6 we find that the top 1% share in the nonrobust economy is 24.3%, roughly equal to the actual value in 1980 depicted in Figure 1. In contrast, the top 1% share in the robust economy is 36.9%, which represents a significant increase, but is still 5 percentage points lower than the data. We could of course increase robust inequality to more closely match recent data simply by increasing ε. We argue later that higher values of ε would still be plausible in terms of drift distortions and detection error probabilities. However,

[19]It should be noted that setting $\delta = .026$ implies an implausibly large number of very old agents. For example, 7.4% of our agents are over 100 years old. Even worse, 0.55% are over 200! The natural way to avoid this counterfactual prediction is to introduce age-dependent death probabilities. Unfortunately, this would make the model *much* less tractable analytically. An alternative strategy would be to increase δ to reduce (though not eliminate) the number of Methuselahs, and then to offset the effects of this on inequality by incorporating bequests.

higher values of ε start to produce implausible portfolio policies. We should also note that the robust distribution comes somewhat closer to matching the growth in inequality in the extreme right-tail which, according to Saez and Zucman (2016), is where most of the action has been. They find that the share of the top 0.1% increased from 7% in 1978 to 22% in 2012. Our model generates an increase from 11.9% to 22.3%.

It is easier to visualize the extreme right tail using a log-log scale, as depicted in Figure 6. Since the distribution of log wealth is exponential, these are simply linear.

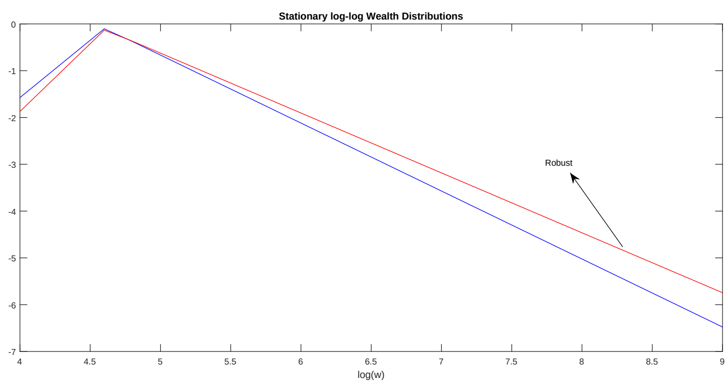

Figure 6. Stationary Distributions: Log/Log Scale

These plots extend to the top 0.1% in the nonrobust case (i.e., $e^{-7} \approx .001$), and to the top 0.2% in the robust case (i.e., $e^{-6} \approx .002$). Since median wealth is about 111 in the nonrobust case, and about 123 in the robust case, these plots suggest that the threshold top 0.2% individual is about 66 times wealthier than the median (i.e., $e^9 = 8,103 \approx 66 \times 123$). For comparison, Saez and Zucman's (2016) data show that the threshold 0.1% individual is more than 400 times wealthier than the median, whose wealth is about \$50,000 (i.e., $20.6\text{mill}/50,000 = 412$). Although a 0.2% threshold would produce a smaller number, it seems safe to conclude that the model does not fully capture the extreme wealth levels that occur in the extreme right tail of the US wealth distribution. This is not too surprising given that the model abstracts from labor income inequality and bequests.

Finally, it should also be observed that while the model sheds light on the right tail of the wealth distribution, it does less well on the left tail. Actual wealth distributions are strongly skewed to the right, with a mode that is significantly smaller than the median. In contrast, the above distributions are nearly symmetric. For example, in the robust case the median is only 23% larger than the mode. Again, this is not too surprising, since the model abstracts from borrowing constraints and other frictions that bind at low wealth levels. Still, it is noteworthy that the robust distribution is more skewed than the nonrobust distribution.

5. Transition Dynamics

As emphasized by Gabaix, Lasry, Lions, and Moll (2016) (GLLM), rising inequality is not really a puzzle. A number of possible mechanisms have been identified, for example, taxes, technology, globalization, and so on. Each of these can be interpreted as altering one or more of the parameters of a standard random growth model, in way that generates empirically plausible increases in inequality. However, GLLM show that if you just perturb the parameters of a standard random growth model, transition rates are *far* too slow. This is especially true in the tails of the distribution, which converge at a slower rate. Hence, the real puzzle is why inequality has increased so *rapidly*.

GLLM argue that the key to resolving this puzzle is to consider deviations from Gibrat's Law, which produce heterogeneity in mean growth rates or volatility. They outline two possibilities. The first features 'type dependence', in which there are both high growth types and low growth types, with individuals experiencing (exogenous) stochastic switches between types. Although type dependence seems like a plausible way to introduce heterogeneity into labor income dynamics, it seems less persuasive as a description of heterogeneity in wealth dynamics. The second model they outline features 'scale dependence', in which growth rates depend on the level of wealth or income. This seems better suited to wealth dynamics. They show that scale dependence can be captured by simply adding an aggregate *multiplicative* shock to the Brownian motion describing individual log income or wealth. Interestingly, they show that scale dependence modeled in this way produces an *infinitely* fast transition, since it produces an immediate jump in the Pareto exponent.

Although we think increased uncertainty (as opposed to increased risk) provides an interesting and plausible explanation of the rise in steady state inequality, its real advantage is to generate more rapid transition rates. It does this by producing a form of scale dependence. In our model, wealth dependent pessimism generates wealth dependent portfolio policies, which in turn generates wealth dependent growth rates. This form of scale dependence is more complicated than the simple multiplicative aggregate shock considered in GLLM, and so we do not obtain their simple and dramatic increase in transition rates. Nevertheless, we now show that the scale dependence produced by nonhomothetic portfolio policies generates transition rates that are much closer to the data than those of traditional random growth models.

To characterize the transition rate we must solve the differential equation in (4.15). Although this describes the evolution of the Laplace transform, remember the Laplace transform embodies all the information in the distribution itself. In fact, it does so in an especially revealing way. In particular, the convergence rate of $F(s)$ at a particular value of $(-s)$ tells us the convergence rate of the $(-s)th$ moment of $f(x)$.[20] Without uncertainty ($\varepsilon = 0$), the solution of (4.15) is straightforward, and is the same as in GLLM

$$F(s,t) = \left[F^0_\infty(s) - F^1_\infty(s)\right] e^{\Lambda(s)t} + F^1_\infty(s)$$

where $F^0_\infty(s)$ is the Laplace transform of the initial stationary distribution and $F^1_\infty(s)$ is the Laplace transform of the new stationary distribution. In our model we assume the only parameter that changes is ε, so here some other parameter would have to change so that $F^0_\infty(s) \neq F^1_\infty(s)$. The convergence rate of the $(-s)$th moment is just $\Lambda(s)$. Since our parameter values are similar to those in GLLM, it is no surprise we get similar results

[20]The minus sign derives from the fact that we defined the Laplace transform as $\int_{-\infty}^{\infty} f(x)e^{-sx}dx$, while a moment-generating function is defined as $\int_{-\infty}^{\infty} f(x)e^{sx}dx$.

when $\varepsilon = 0$. For example, the rate of convergence of mean wealth is

$$\Lambda(-1) = -.0114$$

implying a *half-life* of 60 years! Convergence in the tails is even slower.[21]

Unfortunately, solution of equation (4.15) when $\varepsilon \neq 0$ is more complicated. The exponential functions in the drift and diffusion coefficients of (4.13) produce a discrete shift in the Laplace transform, so that both $F(s,t)$ and $F(s-\beta,t)$ appear on the right-hand side of the differential equation. While a first-order perturbation approximation of the stationary distribution is defensible, a first-order perturbation approximation of $F(s,t)$ along the entire convergence path is more problematic, given the discrete shift, which essentially creates a 'singular perturbation' problem.

Our strategy for solving (4.15) is to approximate the discrete shift with a differential equation. This yields,

$$\frac{\partial F}{\partial t} = \Lambda(s)F(t,s) + \varepsilon\Phi(s-\beta)\left[F(t,s) - \beta\frac{\partial F}{\partial s}\right] + \delta e^{-sx_0} \tag{5.23}$$

where $\beta \equiv 1-\gamma$. Note this is a *partial* differential equation. However, it is amenable to a standard separation of variables approach. In particular, we posit a solution of the form,

$$F(s,t) = H(t)G(s) + F_\infty(s)$$

and obtain separate ODEs for $H(t)$ and $F_\infty(s)$. Solving the ODE for $F_\infty(s)$ produces the stationary distribution, characterized in Propostion 4.4. Solving the ODE for $H(t)$ then gives us the convergence rate. The $G(s)$ function is determined by boundary conditions, and it turns out to be $G(s) = F^0_\infty(s) - F^1_\infty(s)$, where $F^0_\infty(s)$ is the Laplace transform of the initial stationary distribution (given in Propostion 4.2), and $F^1_\infty(s)$ is the Laplace transform of the new stationary distribution (given in Proposition 4.4). We relegate the details to the Appendix, and merely state the following result

Proposition 5.1. *Let $\tilde{F}(t,s) \equiv H(t)G(s)$ be the solution of the homogeneous part of the PDE in* (5.23)*, which describes the dynamics of the cross-sectional distribution of wealth. Then to a first-order approximation in β, $\tilde{F}(t,s)$ is given by*

$$\tilde{F}(t,s) = \tilde{F}(0,s)e^{[\Lambda(s)+\varepsilon\Phi(s-\beta)]t} \tag{5.24}$$

where $\tilde{F}(0,s) = F^0_\infty(s) - F^1_\infty(s)$.

Proof. See Appendix F. □

The convergence rate is determined by the function $\Lambda(s) + \varepsilon\Phi(s-\beta)$. Naturally, when $\varepsilon = 0$, the convergence rate is the same as in the nonrobust case. Note that the convergence rate depends on s, implying that different moments converge at different rates. Figure 7 plots this function. It extends from $s = 0$ to $s = -1.3$. Since the robust Pareto exponent is 1.28, the largest moment that exists for both distributions is 1.28.

The top line plots $\Lambda(s)$, the nonrobust convergence rate. As in GLLM, convergence is slower in the tails. At the mean ($s = -1$), the convergence rate is 1.14%. Slower tail convergence reflects the fact that parametric disturbances move along the distribution like

[21] Unless $|\phi_2| > 2$, the variance does not exist. For the benchmark parameter values the largest moment that exists is about 1.5.

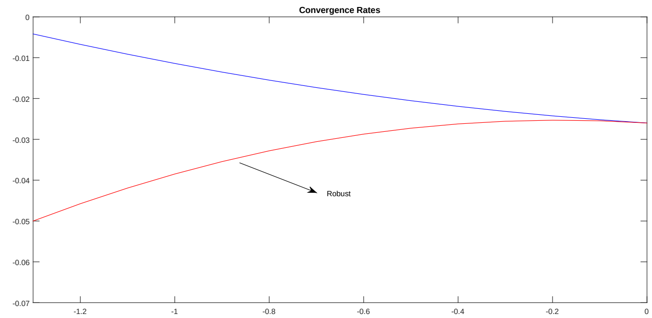

FIGURE 7. Convergence Rates

a 'traveling wave', first hitting young individuals with low wealth, then spreading to older, more wealthy individuals. With exponentially distributed lifetimes, most individuals are young, and possess little wealth.

The bottom line plots the robust convergence rate. Note that convergence is *faster* in the tails. Higher-order moments convergence faster due to scale dependence. When $\gamma > 1$ and $(a_1, b_1) < 0$, the drift and diffusion coefficients of wealth increase with the level of wealth. This produces faster convergence in the tails. Now convergence at the mean takes place at a rate of 3.85%, more than three times faster than in the nonrobust case.

As emphasized by GLLM, standard random growth models obeying Gibrat's Law converge far too slowly to explain the recent rise in inequality. At a 1.14% convergence rate it would take more than 100 years to approach reasonably close to a new stationary distribution, so unless we are moving to an economy with far more inequality than already exists, random growth by itself seems inadequate. We claim that increased uncertainty and the resulting scale dependent growth dynamics provides a plausible explanation not only for the increase itself, but more importantly, for its rapid pace. A 3.85% convergence rate implies that it would take only 60 years to move 90% of the way to a new stationary distribution. Without uncertainty, this same transition would require more than 200 years. The strength of our argument increases the farther out in the tails we go.

Finally, one should remember that robustness induced amplification and acceleration occurs despite the fact that with additive preferences a counterfactually declining saving rate acts as a stabilizing force. As noted earlier in Section 3, and discussed in more detail in Appendix C, we can avoid this by using recursive preferences. Appendix G presents plots comparing stationary distributions and convergence rates when preferences are assumed to be recursive, as in Duffie and Epstein (1992). It shows that recursive preferences reinforce the paper's main findings concerning steady state inequality and convergence rates.

5.1. **Numerical Solution.** Analytical solutions are attractive because they reveal the underlying mechanisms at work. Still, it is important to keep in mind that all the previous results are just approximations. If we are willing to turn on the computer, we can avoid these approximations by solving the model numerically. In principle, we could even solve the agent's optimization problem numerically as well, but here we focus on the accuracy of our key result, related to convergence rates. Proposition 5.1 is based on approximating a shifted Laplace transform with a differential equation, and it is not entirely clear how good this approximation is. So here we numerically solve an $O(\varepsilon)$ approximation of the KFP equation in (4.14) using a standard discretization algorithm, based on central-difference approximations of the derivatives.[22] Even this $O(\varepsilon)$ approximation is nonstandard, due to the dependence of the drift and diffusion coefficients on x.

The parameters are the same as those in Table 1. The algorithm is initialized at the known stationary distribution when $\varepsilon = 0$. The boundary conditions are $f(\infty) = f(-\infty) = 0$, approximated using large positive and negative values of x. Figure 8 depicts the evolution of the top 1% wealth share for the 32 year period from 1981 to 2012, along with the actual Saez-Zucman (2016) data. The dotted line near the top is the share implied by the robust stationary distribution when $\varepsilon = .045$. For comparison, we also plot the paths generated by recursive preferences and a standard random growth model.

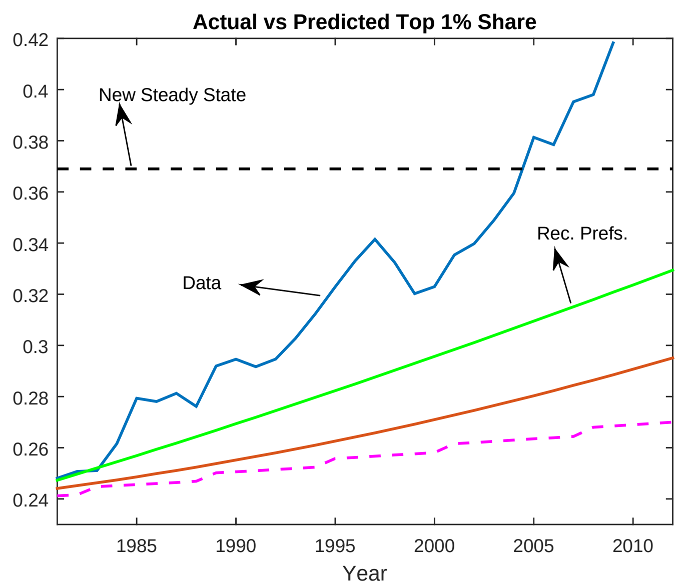

FIGURE 8. Actual vs Predicted Top Wealth Shares: Numerical Solution

The path of the benchmark time-additive model lies in the middle, showing an increase in in the top 1% share from about 24% to 30%. Although significant, this clearly understates the observed increase. Part of the apparent discrepancy simply reflects the fact that the new stationary distribution understates long-run inequality by about 5 percentage points. As noted earlier, although we could increase steady state inequality by increasing ε, this would produce counterfactual portfolio policies. However, the path discrepancy reflects more than a discrepancy in steady states. The convergence rate appears to be too slow as well, both in terms of the data and the model's prediction. Proposition 5.1 suggests the top 1% share should be converging at a rate of at least 3.85%, since that is the predicted mean convergence rate, and our model predicts the tails should converge

[22]Our matlab code is available upon request.

faster than the mean. In contrast, Figure 8 suggests the top 1% share is converging at about a 1.9% (average) rate. Hence, we are off by at least a factor of 2. This suggests that $\gamma - 1 = .31$ might be too large for the ODE to accurately approximate the shifted Laplace transform.

Given the discrepancy in the benchmark model, it is useful to examine the convergence properties of the model with recursive preferences. As discussed in the Appendix, our analytical approximations suggest that recursive preferences should generate faster convergence, since an increasing savings rate now reinforces the portfolio channel. The top (green) line in Figure 8 depicts the numerical convergence path using recursive preferences. As predicted, it converges more rapidly, reaching a top 1% share of about 33% by 2012. Hence, with recursive preferences we can account for about half of the observed 18 pecentage point increase in the top 1% share.

Although these numerical results might seem disappointing, remember that traditional random growth models do even worse. Following GLLM, suppose we instead attribute increased inequality to increased risk, as opposed to increased uncertainty. In doing this, we must be careful to shut down the portfolio channel, since in our model increased risk reduces investment in the higher yielding asset. Ceteris paribus, this *reduces* inequality. So in an effort to stay as close as possible to GLLM, we simply fix $\alpha = 1$, and assume the equity premium is 1%. Initial volatility is calibrated to match the 1981 top 1% wealth share (of about 24%). This yields an initial standard deviation of 12.2%. We then assume σ increases so as to match the uncertainty model's steady state share of about 37%. This implies a standard deviation of 13.4%. We then numerically calculate the transition path to the new long run steady state. The bottom dotted line displays the results. As discovered by GLLM, the risk model generates a very slow transition. The top 1% share increases by only a few percentage points over the entire 32 year period. Again, it does worse because it lacks any sort of feedback induced scale effects.

Finally, as noted in Section 2, there are reasons to believe the Saez-Zucman data overstate the actual increase in top wealth shares. If this is indeed the case, then our results might be much closer than they appear. For example, a 9 percentage point increase in the top 1% share would more than fully account for the increase registered by SCF or estate tax data.

6. Detection Error Probabilities

We have asserted that increased (Knightian) uncertainty provides a plausible explanation for the timing, magnitude, and rate of increase of US wealth inequality. Here we explain the basis of this assertion. The key mechanism in our model is wealth dependent pessimism, which produces wealth dependent portfolio allocations. Pessimism is formalized using a max-min objective function, in which agents optimize against an endogenous worst-case scenario. A criticism of this approach is to ask why agents should expect the worst. Following Hansen and Sargent (2008), we answer this critique by constraining the set of alternative models the agent considers. In particular, we suppose agents only hedge against models that could have plausibly generated the observed data.

To quantify plausiblity we use 'detection error probabilities'. The idea is to think of agents as statisticians who attempt to discriminate among models using likelihood ratio statistics. When likelihood ratio statistics are large, detection error probabilities are small, and models are easy to distinguish. Detection error probabilities will be small when models are very different, or when there is a lot of data. Classical statistical inference is based on adopting a null hypothesis, and fixing the probability of falsely rejecting this hypothesis. Detection error probabilities treat the null and alternative symmetrically, and average between Type I and Type II errors. In particular,

$$DEP = \frac{1}{2}\text{Prob}(H_A|H_0) + \frac{1}{2}\text{Prob}(H_0|H_A)$$

Hence, a DEP is analogous to a p-value. Our results would therefore be implausible if the DEP is too small. Small DEPs imply agents are hedging against models that could easily be rejected based on observed data.

Models in our economy are continuous-time stochastic processes. Likelihood ratio statistics are integrals of the conditional relative entropies between models, and as discussed previously, these entropies are determined by the evil agent's policy function, $h(x)$, where x is the logarithm of wealth. In particular, Anderson, Hansen, and Sargent (2003) provide the following bound

$$\text{avg} \quad DEP \leq \frac{1}{2} E \exp\left\{-\frac{1}{8}\int_0^T h^2(x)dx\right\}$$

where T is the sample length. This bound is difficult to evaluate since x is stochastic. As a result, we compute the following (state-dependent) approximation, avg $\quad DEP(x) \leq \frac{1}{2}\exp\left\{-\frac{1}{8}Th^2(x)\right\}$. By fixing x, this overstates the bound at values of x where $h(x)$ is small, while understating it at values of x where $h(x)$ is large. Figure 9 plots this approximation using the benchmark parameter values and assuming $T = 200$.

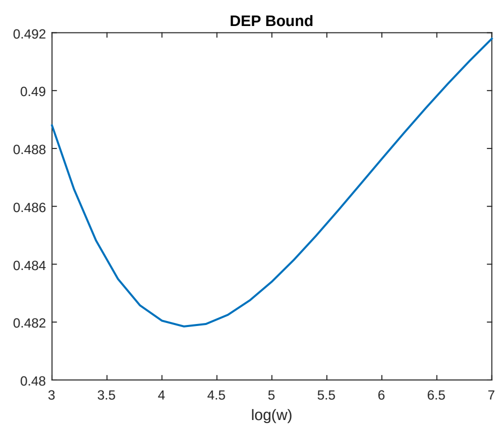

Figure 9. Detection Error Probability Bound

Notice the DEP bound is nonmonotonic. This reflects the nonmonotonicity of $h(x)$, which itself reflects the endogeneous nature of pessimism in the model. The agents who are least pessimistic in equilibrium are the very poor, because they choose to avoid the risky technology, and the very rich, because they have a low marginal utility of wealth. Agents in the middle have the most to worry about. When interpreting this plot one should keep in mind that a fixed T might not be appropriate given the model's OLG structure. In particular, one could argue that T should be age-specific. This would be the case if agents weight their own experience more heavily than the entire past history of data, as evidence suggests they do.[23] If this were the case, the DEP would be smaller than Figure 9 portrays for large values of x, while being larger for small values of x.

The main point to notice, however, is that the DEPs are quite large, for all values of x. This is true despite the fact that setting $T = 200$ overstates the data that is actually available to agents. (Remember, our time unit here is a year). Although AHS suggest the bound is not always tight, the numbers in Figure 9 are so large that we suspect more refined estimates would still preserve the basic conclusion that the degree of pessimism here is plausible. In fact, another way to see this is to simply suppose investment outcomes are i.i.d coin flips. With a 10% standard deviation and 200 years of data, a 95% confidence interval for the mean would be $\pm 2 \times \frac{.10}{\sqrt{200}} \approx 1.4\%$. The implied drift distortions reported in Figure 4 are well within this interval.

7. Conclusion

The recent increase in US wealth inequality raises three important questions: (1) What caused it?, (2) Why did it start around 1980?, and (3) Why has it been so rapid? There are already many answers to the first two questions. Although we think our proposed explanation - increased uncertainty - provides an interesting and novel answer to these two questions, we view it more as a complement than a substitute to existing explanations. However, as recently emphasized by Gabaix, Lasry, Lions, and Moll (2016), answering the third question is more challenging. Standard random growth models cannot explain why top wealth shares more than doubled in little more than a single generation. To account for the rate of increase in inequality, they argue it is necessary to extend random growth models by incorporating either 'type dependence' or 'scale dependence', which produce heterogeneity in growth rates. Using the tools of robust control theory, we show that increased uncertainty produces a natural form of scale dependence, which generates significantly more rapid convergence rates.

More work needs to be done to make increased uncertainty a fully convincing explanation of recent US wealth inequality. First, the notion that the world became more uncertain around 1980 seems plausible, but should be more carefully documented. There is abundant evidence that idiosyncratic labor income risk increased around this time (Ljungqvist and Sargent, 1998), but risk is not the same as uncertainty, and wealth is not the same as labor income. Robust control provides a link between objective risk and subjective uncertainty by viewing distortions as 'hiding behind' objective risk. Hence, it would be interesting to consider an extension to stochastic volatility, which would open the door to stochastic uncertainty as well. In fact, there is no reason to view risk and uncertainty as competing explanations. Using panel data on the Forbes 400, Gomez (2017) finds that inequality during 1980-95 was primarily driven by increased risk, while return heterogeneity (possibly driven by increased uncertainty) was primarily responsible for increased top wealth shares during the period 1995-2015. Combining both increased risk and increased

[23] See, e.g., Malmendier and Nagel (2016).

uncertainty might help to close the gap in Figure 8. Second, the key source of scale dependence in our model is that wealthy individuals earn higher mean returns. Again, there is abundant evidence in support of this (although the evidence is less clear using after-tax returns). However, it is not clear whether this return differential reflects portfolio allocation decisions, as in our model, or whether wealthy individuals receive higher returns within asset categories as well. Using tax records from Norway, Fagereng, Guiso, Malacrino, and Pistaferri (2016a) argue that wealthy investors receive higher returns within asset categories. Hence, it would be interesting to consider a model where both forces are at work. Finally, we have argued that wealth dependent pessimism is empirically plausible because the drift distortions it produces could not be statistically rejected based on historical data. However, we suspect our detection error probability bound is not very tight, and one could easily obtain tighter bounds using simulations.

References

Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109, 659–84.

Anderson, E. W., L. P. Hansen, and T. J. Sargent (2003): "A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection," *Journal of the European Economic Association*, 1, 68–123.

——— (2012): "Small Noise Methods for Risk-Sensitive/Robust Economies," *Journal of Economic Dynamics and Control*, 36, 468–500.

Aoki, S., and M. Nirei (2017): "Zipf's Law, Pareto's Law, and the Evolution of Top Incomes in the U.S.," *American Economic Journal: Macroeconomics*, 9, 36–71.

Barillas, F., L. P. Hansen, and T. J. Sargent (2009): "Doubts or Variability?," *Journal of Economic Theory*, 144, 2388–2418.

Benhabib, J., and A. Bisin (2017): "Skewed Wealth Distributions: Theory and Empirics," NBER Working Paper No. 21924, forthcoming in *Journal of Economic Literature*.

Benhabib, J., A. Bisin, and M. Luo (2017): "Earnings Inequality and Other Determinannts of Wealth Inequality," *American Economic Review: Papers & Proceedings*, 107, 593–97.

Benhabib, J., A. Bisin, and S. Zhu (2011): "The Distribution of Wealth and Fiscal Policy in Economies with Finitely-Lived Agents," *Econometrica*, 79, 123–57.

——— (2016): "The Distribution of Wealth in the Blanchard-Yaari Model," *Macroeconomic Dynamics*, 20, 466–81.

Borovicka, J. (2016): "Survival and Long-Run Dynamics with Heterogeneous Beliefs Under Recursive Preferences," working paper, New York University.

Carroll, C. D. (2002): "Porfolios of the Rich," in *Household Portfolios*, ed. by L. Guiso, M. Haliassos, and T. Jappelli, pp. 389–430. MIT Press.

Champernowne, D. (1953): "A Model of Income Distribution," *Economic Journal*, 63, 318–51.

Duffie, D., and L. G. Epstein (1992): "Stochastic Differential Utility," *Econometrica*, 60, 353–94.

Dynan, K. E., J. Skinner, and S. P. Zeldes (2004): "Do the Rich Save More?," *Journal of Political Economy*, 112, 397–444.

Epstein, L. G. (1999): "A Definition of Uncertainty Aversion," *Review of Economic Studies*, 66, 579–608.

Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri (2016a): "Heterogeneity and Persistence in Returns to Wealth," NBER Working Paper No. 22822.

——— (2016b): "Heterogeneity in Returns to Wealth and the Measurement Of Wealth Inequality," *American Economic Review: Papers & Proceedings*, 106, 651–55.

Gabaix, X. (2009): "Power Laws in Economics and Finance," *Annual Review of Economics*, 1, 255–93.

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

Ghirardato, P., and M. Marinacci (2002): "Ambiguity Made Precise: A Comparative Foundation," *Journal of Economic Theory*, 102, 251–89.

Gomez, M. (2017): "What Drives the Recent Rise in Top Wealth Shares?," Princeton University Working Paper.

Gottschalk, P., and R. Moffitt (1994): "The Growth of Earnings Instability in the US Labor Market," *Brookings Papers on Economic Activity*, 2, 217–54.

Hansen, L. P., and T. J. Sargent (2008): *Robustness.* Princeton University Press, Princeton.

Hansen, L. P., and T. J. Sargent (2011): "Robustness and Ambiguity in Continuous Time," *Journal of Economic Theory*, 146, 1195–223.

Hansen, L. P., T. J. Sargent, G. A. Turmuhambetova, and N. Williams (2006): "Robust Control and Model Misspecification," *Journal of Economic Theory*, 128, 45–90.

Hayashi, T., and J. Miao (2011): "Intertemporal Substitution and Recursive Smooth Ambiguity Preferences," *Theoretical Economics*, 6, 423–72.

Huggett, M. (1996): "Wealth Distribution in Life-Cycle Economies," *Journal of Monetary Economics*, 38, 469–94.

Kambourov, G., and I. Manovskii (2009): "Occupational Mobility and Wage Inequality," *Review of Economic Studies*, 76, 731–59.

Kaplan, W. (1962): *Operational Methods for Linear Systems*. Addison-Wesley.

Klibanoff, P., M. Marinacci, and S. Mukerji (2005): "A Smooth Model of Decision Making Under Ambiguity," *Econometrica*, 73, 1849–92.

——— (2009): "Recursive Smooth Ambiguity Preferences," *Journal of Economic Theory*, 144, 930–76.

Kopczuk, W. (2015): "What Do We Know about the Evolution of Top Wealth Shares in the United States?," *Journal of Economic Perspectives*, 29, 47–66.

Ljungqvist, L., and T. J. Sargent (1998): "The European Unemployment Dilemma," *Journal of Political Economy*, 106, 514–550.

Maenhout, P. (2004): "Robust Portfolio Rules and Asset Pricing," *Rev. of Financial Studies*, 17, 951–83.

Malmendier, U., and S. Nagel (2016): "Learning from Inflation Experiences," *Quarterly Journal of Economics*, 131, 53–87.

Moskowitz, T., and A. Vissing-Jorgensen (2002): "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?," *American Economic Review*, 92, 745–78.

Piketty, T. (2014): *Capital in the Twenty-First Century*. Harvard University Press.

Piketty, T., E. Saez, and S. Stantcheva (2014): "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities," *American Economic Journal: Economic Policy*, 6, 230–71.

Reed, W. J. (2001): "The Pareto, Zipf and Other Power Laws," *Economics Letters*, 74, 15–19.

Saez, E., and G. Zucman (2016): "Wealth Inequality in the United States Since 1913: Evidence from Capitalized Income Tax Data," *Quarterly Journal of Economics*, 131, 519–78.

Simon, H. A. (1955): "On a Class of Skew Distribution Functions," *Biometrika*, 44, 425–40.

Skiadas, C. (2013): "Smooth Ambiguity Aversion toward Small Risks and Continuous-Time Recursive Utility," *Journal of Political Economy*, 121, 775–92.

Toda, A. A. (2014): "Incomplete Markets, Dynamics and Cross-Sectional Distributions," *Journal of Economic Theory*, 154, 310–48.

ONLINE APPENDIX

Appendix A. Proof of Proposition 3.1

After substituting the h policy function in eq. (3.6) into the HJB equation in eq. (3.5) we get

$$(\rho+\delta)V(w)=\max_{c,\alpha}\left\{\frac{c^{1-\gamma}}{1-\gamma}+[(r+\alpha(\mu-r))w-c]V'(w)+\frac{1}{2}\alpha^2\sigma^2w^2V''(w)-\frac{1}{2}\varepsilon\alpha^2\sigma^2w^2(V'(w))^2\right\} \tag{A.25}$$

Except for the last term multiplying ε, this is a standard Merton consumption/portfolio problem, with a well known solution. Hence, this suggests a perturbation approximation around ε. To obtain this, we posit

$$V(w)\approx V^0(w)+\varepsilon V^1(w)$$

Our goal is to solve for $V^0(w)$ and $V^1(w)$. With this approximation, a first-order approximation of the $c(w)$ policy function is

$$\begin{aligned} c(w) &= V_w^{-\frac{1}{\gamma}} \\ &\approx [V_w^0+\varepsilon V_w^1]^{-\frac{1}{\gamma}} \\ &\approx \left(V_w^0\right)^{-\frac{1}{\gamma}}-\varepsilon\frac{1}{\gamma}\left(V_w^0\right)^{-\frac{1}{\gamma}-1}V_w^1 \\ &\equiv c_0+\varepsilon c_1 \end{aligned} \tag{A.26}$$

and a first-order approximation of the $\alpha(w)$ policy is

$$\begin{aligned} \alpha(w) &= -\frac{(\mu-r)}{\sigma^2}\left[\frac{V_w}{w(V_{ww}-\varepsilon V_w^2)}\right] \\ &\approx -\frac{(\mu-r)}{\sigma^2}\left[\frac{wV_w^0}{w^2V_{ww}^0}+\varepsilon\frac{wV_w^1V_{ww}^0-wV_w^0(V_{ww}^1-(V_w^0)^2)}{w^2(V_{ww}^0)^2}\right] \\ &\equiv \alpha_0+\varepsilon\alpha_1 \end{aligned} \tag{A.27}$$

Substituting these approximations into HJB equation in (A.25) gives us

$$\begin{aligned} (\rho+\delta)[V^0+\varepsilon V^1] &= \frac{c_0^{1-\gamma}}{1-\gamma}+\varepsilon c_0^{-\gamma}c_1+(V_w^0+\varepsilon V_w^1)[(r+(\alpha_0+\varepsilon\alpha_1)(\mu-r))w-(c_0+\varepsilon c_1)]+\frac{1}{2}(\alpha_0+\varepsilon\alpha_1)^2\sigma^2w^2[V_{ww}^0+\varepsilon V_{ww}^1] \\ &\quad -\frac{1}{2}\varepsilon(\alpha_0+\varepsilon\alpha_1)^2\sigma^2w^2(V_w^0+\varepsilon V_w^1)^2 \end{aligned}$$

Matching terms of equal order and dropping terms of order $O(\varepsilon^2)$ gives us

$$\begin{aligned} \varepsilon^0: \quad (\rho+\delta)V^0 &= \frac{c_0^{1-\gamma}}{1-\gamma}+V_w^0[(r+\alpha_0(\mu-r))w-c_0]+\frac{1}{2}\alpha_0^2\sigma^2w^2V_{ww}^0 \\ \varepsilon^1: \quad (\rho+\delta)V^1 &= c_0^{-\gamma}c_1+V_w^1[(r+\alpha_0(\mu-r))w-c_0]+V_w^0[\alpha_1(\mu-r)w-c_1]+\sigma^2w^2[\alpha_0\alpha_1V_{ww}^0+\alpha_0^2V_{ww}^1]-\frac{1}{2}\alpha_0^2\sigma^2w^2(V_w^0)^2 \end{aligned}$$

Note this system is recursive. We can first solve the ε^0 equation for V^0, and then substitute this into the ε^1 equation. Solving for V^0 just gives the Merton solution. In particular, we conjecture $V^0(w)=\frac{A}{1-\gamma}w^{1-\gamma}$. Note, this implies $c_0=A^{-1/\gamma}w$ and $\alpha_0=\frac{(\mu-r)}{\gamma\sigma^2}$. After canceling the common $w^{1-\gamma}$ term we can solve for A. This produces the expression for A_0 stated in Proposition 3.1.

To solve the ε^1 equation we conjecture $V^1(w)=\frac{B}{\epsilon}w^{\epsilon}$, and try to solve for B and ϵ. From (A.26) and (A.27), this guess implies $c_1=-\gamma^{-1}A^{-1/\gamma-1}Bw^{2-\gamma}$ and $\alpha_1=-\left(\frac{\mu-r}{\sigma^2}\right)\left(\frac{A^2+(\gamma-1)B}{A\gamma^2}\right)w^{1-\gamma}$. Note, these now depend on w. Substituting these into the ε^1 equation, we find that if $\epsilon=2(1-\gamma)$, we can cancel out the terms in w and solve for B. Doing so produces the expression for A_1 stated in Proposition 3.1. □

Appendix B. Proof of Corollary 3.2

This follows immediately from the proof of Proposition 3.1. Here we fill in some of the details omitted in the above proof. Substituting the expressions for V^0 and V^1 into the expression for $\alpha(w)$ in eq. (A.27) gives

$$\begin{aligned}\alpha(w) &\approx \alpha_0 - \varepsilon\left(\frac{\mu - r}{\sigma^2}\right)\left[\frac{-\gamma A_0 A_1 w^{-3\gamma} - A_0 A_1(1-2\gamma)w^{-3\gamma} + A_0^3 w^{-3\gamma}}{\gamma^2 A_0^2 w^{-2\gamma-1}}\right] \\ &= \alpha_0 - \varepsilon\alpha_0\left[\frac{-\gamma A_1 - (1-2\gamma)A_1 + A_0^2}{\gamma A_0}\right] w^{1-\gamma}\end{aligned}$$

One can readily verify this is the same expression as stated in Corollary 3.2. Next, substituting the expressions for V^0 and V^1 into the expressions for $c(w)$ in eq. (A.26) gives

$$\frac{c(w)}{w} = A_0^{-1/\gamma} - \varepsilon\frac{1}{\gamma}\left(A_0^{-1/\gamma-1}A_1\right)w^{1-\gamma}$$

This then implies the expression for $s(w) = 1 - [c(w)/w]$ stated in Corollary 3.2. □

Appendix C. Recursive Preferences I

The main text considers a traditional robust control problem with observable states where an agent conditions on a given benchmark model, and then formulates policies that are robust to local unstructured perturbations around this model. As noted by Hansen, Sargent, Turmuhambetova, and Williams (2006), continuous-time versions of this problem are observationally equivalent to Duffie and Epstein's (1992) Stochastic Differential Utility (SDU) model of recursive preferences. Hence, risk aversion is not separately identified from ambiguity/uncertainty aversion.

This observational equivalence has sparked a more recent literature that attempts to distinguish risk aversion from both ambiguity aversion and intertemporal substitution. Hansen and Sargent (2008, chpts. 18 and 19) note that the key to separating risk aversion from ambiguity aversion is to introduce hidden state variables, which the agent attempts to learn about. Early robust control methods were criticized because they abstracted from learning. Ambiguity is then defined by distortions of the agent's estimates of the hidden states.[24]

Hidden states can be used to represent a wide range of unobservables. For example, time invariant hidden states can index alternative models. Here we assume the hidden state is an unobserved, potentially time-varying, mean investment return. In principle, we could allow the agent to be uncertain about both the dynamics conditional on a particular mean growth rate, as well as the mean itself. However, for our purposes it is sufficient to assume the agent is only uncertain about the mean.[25]

Distorted beliefs about the hidden state can be interpreted from the perspective of the Klibanoff, Marinacci, and Mukerji (2005) (KMM) model of smooth ambiguity aversion. In the KMM model an agent prefers act f to act g if and only if

$$\mathbb{E}_\mu\phi(\mathbb{E}_\pi u \circ f) \geq \mathbb{E}_\mu\phi(\mathbb{E}_\pi u \circ g)$$

where $\mathbb{E}$ is the expectation operator, π is a probability measure over outcomes conditional on a model, and μ is a probability measure over models. Ambiguity aversion is characterized by the properties of the ϕ function, while risk aversion is characterized by the properties of the u function. If ϕ is concave, the agent is ambiguity averse. KMM refer to $\mathbb{E}_\pi$ as 'first-order beliefs', while $\mathbb{E}_\mu$ is referred to as 'second-order beliefs'. Note that when ϕ is nonlinear, the implicit compound lottery defined by selecting a model with unknown parameters cannot be reduced to a single lottery over a 'hypermodel', as in Bayesian decision theory, so the distinction between models and parameters becomes important. Also note that from the

[24]This definition of ambiguity aversion is based on the axiomatization of Ghirardato and Marinacci (2002), which defines ambiguity aversion as deviations from subjective expected utility. Epstein (1999) proposes an alternative definition based on deviations from probabilistic sophistication.

[25]In the language of Hansen and Sargent (2008, chpt. 18), we activate the T^2-operator by setting $\theta_2 < \infty$, while deactivating the T^1-operator by setting $\theta_1 = \infty$. This is a subtle distinction, since at the end-of-the-day they both produce drift distortions. However, they do this in different ways.

perspective of smooth ambiguity aversion, evil agents and entropy penalized drift distortions are just a device used to produce a particular distortion in second-order beliefs about continuation values, i.e., where $\phi(V) \approx -\exp(-\varepsilon V)$.

The original KMM model was static. Klibanoff, Marinacci, and Mukerji (2009) extend it to a recursive, dynamic setting. However, their implicit aggregator is additive, so risk and intertemporal substitution cannot be distinguished. In response, Hayashi and Miao (2011) propose a model of generalized smooth ambiguity aversion by combining KMM with an Epstein-Zin aggregator. Unfortunately, as noted by Skiadas (2013), this model does not extend to continuous-time with Brownian information structures. Intuitively, first-order uncertainty (risk) is $O(dt)$, whereas second-order uncertainty (ambiguity) is $O(dt^2)$, and so it evaporates in the continuous-time limit. In response, Hansen and Sargent (2011) propose a trick to retain ambiguity aversion, even as the sampling interval shrinks to zero. In particular, they show that if the robustness/ambiguity-aversion parameter is scaled by the sampling interval, then ambiguity aversion will persist in the limit. Intuitively, even though second-order uncertainty becomes smaller and smaller as the sampling interval shrinks, because the agent effectively becomes more ambiguity averse at the same time, ambiguity continues to matter.

In what follows we outline a heuristic combination of the discrete-time generalized KMM preferences of Hayashi and Miao (2011) and the continuous-time scaling trick of Hansen and Sargent (2011). As far as we know, there are no formal decision-theoretic foundations for such a combination, at least not yet.

With recusive preferences, the agent's problem becomes

$$V_t = \max_{c,\alpha} \min_{h} E_t \int_t^\infty f(c_s, V_s) ds$$

where $f(c_s, V_s)$ is the (normalized) Duffie-Epstein aggregator,

$$f(c, V) = \varphi(1-\gamma)V \left[\log(c) - \frac{1}{1-\gamma} \log((1-\gamma)V)\right]$$

and where for simplicity we've assumed the the elasticity of intertemporal substitution is unity. The effective rate of time preference is $\varphi = \rho + \delta$, and the coefficient of relative risk aversion is $\gamma \neq 1$.[26] The budget constraint is the same as before

$$dw = \{[r + \alpha(\mu - r)]w - c + \alpha\sigma w h\}\, dt + \alpha\sigma w dB$$

The HJB equation is

$$0 = \max_{c,\alpha} \min_{h} \left\{ f(c, V) + \frac{1}{2\varepsilon} h^2 + ([r + \alpha(\mu - r)]w - c + \alpha\sigma w h)\, V'(w) + + \frac{1}{2}\alpha^2\sigma^2 w^2 V''(w) \right\}$$

Note that discounting is embodied in the properties of the aggregator. Also note that in contrast to Bayesian learning models, where the drift is regarded as an unknown parameter and its current estimate becomes a hedgeable state variable, here the drift is viewed as a *control* variable, which is selected by the agent to produce a robust portfolio.

The first-order conditions for (α, h) are the same as before, while the first-order condition for c becomes:

$$c = \frac{\varphi(1-\gamma)V}{V'(w)}$$

If these are substituted into the HJB equation we get:

$$0 = f[c(V, V'), V] + (rw - c)V' - \frac{1}{2} \frac{(\mu - r)^2 (V')^2}{[V'' - \varepsilon(V')^2]\sigma^2}$$

Our goal is to compute the following first-order approximation

$$V(w) \approx V^0(w) + \varepsilon V^1(w)$$

[26] We could easily allow $\gamma = 1$ by slightly modifying the aggregator. However, this would be uninteresting, since preferences would then collapse to additive form given that we've already assumed the elasticity of intertemporal substitution is unity.

By inspection, it is clear that when $\varepsilon = 0$ it is natural to guess

$$V^0(w) = \frac{\hat{A}_0}{1-\gamma} w^{1-\gamma}$$

Note that this implies $c_0 = \varphi w$ and $\alpha_0 = (\mu - r)/\gamma\sigma^2$. Substituting into the HJB equation and cancelling out the common $w^{1-\gamma}$ term gives the following equation for $\hat{A}_0$

$$0 = \varphi\left[\log(\varphi) - \frac{1}{1-\gamma}\log(\hat{A}_0)\right] + (r-\varphi) - \frac{1}{2}\frac{(\mu-r)^2}{\gamma\sigma^2} \qquad \Rightarrow \qquad \hat{A}_0 = \exp\left\{(1-\gamma)\left[\log(\varphi) + \frac{r-\varphi}{\varphi} - \frac{1}{2}\frac{(\mu-r)^2}{\varphi\gamma\sigma^2}\right]\right\}$$

Next, matching the $O(\varepsilon)$ terms in the HJB equation yields the following ODE for $V^1(w)$

$$0 = \varphi(1-\gamma)\left\{V^1\left[\log(c_0) - \frac{1}{1-\gamma}\log((1-\gamma)V^0)\right] + V^0\left(\frac{c_1}{c_0} - \frac{1}{1-\gamma}\frac{V^1}{V^0}\right)\right\} + (r-\varphi)wV^1_w - \frac{1}{2}\alpha_0^2\sigma^2(wV^0_w)^2 + \frac{1}{2}\sigma^2 w^2[2\alpha_0\alpha_1 V^0_{ww} + \alpha_0^2 V^1_{ww}]$$

where

$$\begin{aligned} c_1 &= \varphi(1-\gamma)\frac{V^1 V^0_w - V^0 V^1_w}{(V^0_w)^2} \\ \alpha_1 &= -\gamma\alpha_0 \frac{wV^1_w V^0_{ww} - wV^0_w(V^1_{ww} - (V^0_w)^2)}{w^2(V^0_{ww})^2} \end{aligned}$$

Note that the expression for α_1 is the same as before. Also as before, note that the system is recursive, with the above solutions for (V^0, c_0, α_0) becoming inputs into the V^1 ODE. If you stare at this ODE long enough, you will see that a function of the following form will solve this equation

$$V^1(w) = \frac{\hat{A}_1}{\epsilon} w^{\epsilon}$$

where as before $\epsilon = 2(1-\gamma)$. Substituting in this guess, cancelling the common w^ϵ terms, and then solving for $\hat{A}_1$ gives

$$\hat{A}_1 = \frac{-\frac{1}{2}\alpha_0^2\sigma^2\hat{A}_0}{-\gamma\varphi\log(\varphi) + (1-\gamma)(r-\varphi) + \frac{\varphi\gamma}{2(\gamma-1)} - \varphi + \alpha_0^2\sigma^2[\frac{1}{2}(\gamma^2-1) + (\gamma-1)^2]}$$

From here, the analysis proceeds exactly as in the main text. We just need to replace (A_0, A_1) with $(\hat{A}_0, \hat{A}_1)$.The approximate saving rate now becomes

$$s(w) = 1 - \varphi + \varepsilon\frac{\varphi\hat{A}_1}{2\hat{A}_0} w^{1-\gamma}$$

If $\gamma > 1$, then this is increasing in w as long as $\hat{A}_1 < 0$, since $\hat{A}_0 > 0$.

To examine the quantitative properties of the model with recursive preferences, we use the same parameter values as those in Table 1, with three minor exceptions. First, since the model with recursive preferences seems to be somewhat less sensitive to the robustness parameter, we increased ε from 0.045 to 0.45. Second, we increased γ slightly from 1.31 to 1.5. Finally, we increased μ slightly, from 5.86% to 5.95%. These parameter values remain consistent with available empirical estimates. Figure 10 displays the resulting portfolio shares and savings rates.

The portfolio shares are similar to those in Figure 2, though wealth dependence is somewhat weaker. The key difference here is the saving rate. Now it is increasing in wealth. However, as in Figure 3, wealth dependence is very weak.

Appendix D. Proof of Proposition 4.1

Substituting the policy functions into the budget constraint gives

$$dw = [(r + \alpha(w)(\mu - r))w - c(w)]dt + \alpha(w)\sigma w dB$$

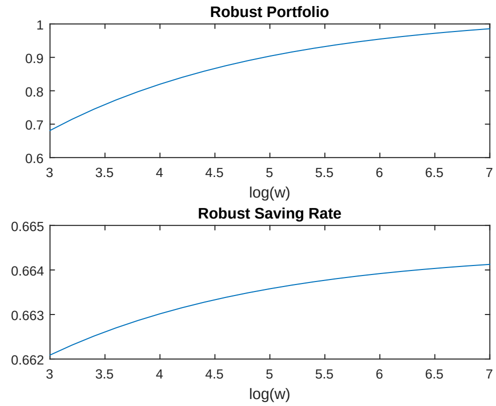

FIGURE 10. Recursive Preferences

where $\alpha(w)$ and $c(w)$ are the approximate policy functions given in Corollary 3.2. Collecting terms, individual wealth dynamics can be described

$$\frac{dw}{w} = a(w;\varepsilon)dt + b(w;\varepsilon)dB$$

where the drift coefficient is given by

$$\begin{aligned} a(w;\varepsilon) &= r + \frac{(\mu-r)^2}{\gamma\sigma^2} - A_0^{-1/\gamma} + \varepsilon\left\{\frac{-\gamma^2\sigma^2\alpha_0^2[A_0^2+(\gamma-1)A_1)] + \gamma A_1 A_0^{-1/\gamma}}{\gamma^2 A_0}\right\} w^{1-\gamma} \quad \text{(D.28)} \\ &\equiv \bar{a}_0 + \varepsilon\bar{a}_1 w^{1-\gamma} \end{aligned}$$

and the diffusion coefficient is given by

$$\begin{aligned} b(w;\varepsilon) &= \sigma\alpha_0 - \varepsilon\left\{\alpha_0\sigma\gamma\frac{A_0^2+(\gamma-1)A_1}{\gamma^2 A_0}\right\} w^{1-\gamma} \quad \text{(D.29)} \\ &\equiv b_0 + \varepsilon b_1 w^{1-\gamma} \end{aligned}$$

In a standard random growth model, these coefficients would be constant. Hence, the wealth dependence here reflects the 'scale dependence' of our model. Next, let $x = \log(w) \equiv g(w)$ denote log wealth. Ito's lemma implies

$$\begin{aligned} dx &= g'(w)dw + \frac{1}{2}w^2 b(w)^2 g''(w)dt \\ &= \frac{dw}{w} - \frac{1}{2}\left(b_0 + \varepsilon b_1 w^{1-\gamma}\right)^2 dt \end{aligned}$$

Substituting $e^{(1-\gamma)x} = w^{1-\gamma}$ into the right-hand side and dropping $O(\varepsilon^2)$ terms gives

$$dx = [\bar{a}_0 - \frac{1}{2}b_0^2 + \varepsilon(\bar{a}_1 - b_0 b_1)e^{(1-\gamma)x}]dt + (b_0 + \varepsilon b_1 e^{(1-\gamma)x})dB$$

Finally, defining $a_0 \equiv \bar{a}_0 - \frac{1}{2}b_0^2$ and $a_1 = \bar{a}_1 - b_0 b_1$ gives the result stated in Proposition 4.1. □

APPENDIX E. PROOF OF PROPOSITION 4.4

For convenience, we start by reproducing the KFP equation in (4.14)

$$\frac{\partial f}{\partial t} = -\frac{\partial[(a_0 + \varepsilon a_1 e^{(1-\gamma)x})f]}{\partial x} + \frac{1}{2}\frac{\partial^2[(b_0 + \varepsilon b_1 e^{(1-\gamma)x})^2 f]}{\partial x^2} - \delta f + \delta\zeta_0$$

Evaluating the derivatives gives

$$\frac{\partial f}{\partial t} = (-a_x + b_x^2 + bb_{xx})f + (2bb_x - a)\frac{\partial f}{\partial x} + \frac{1}{2}b^2\frac{\partial^2 f}{\partial x^2} - \delta f + \delta\zeta_0$$

where $a(x)$ and $b(x)$ are the drift and diffusion coefficients defined in equations (D.28) and (D.29) after the change of variables $w = e^x$. In general, these sorts of partial differential equations are not fun to solve. However, this PDE is *linear*, which opens the door to transform methods. The first step is to evaluate the derivatives of the $a(x)$ and $b(x)$ functions, and then drop the $O(\varepsilon^2)$ terms. Then we take the Laplace transform of both sides, with x as the transform variable. When doing this we use the following facts

$$\mathcal{L}\{\frac{\partial f}{\partial t}\} = \frac{\partial F(s)}{\partial t} \qquad \mathcal{L}\{\frac{\partial f}{\partial x}\} = sF(s) \qquad \mathcal{L}\{\frac{\partial^2 f}{\partial x^2}\} = s^2F(s) \qquad \mathcal{L}\{e^{\beta x}f\} = F(s-\beta) \qquad \mathcal{L}\{\zeta_0\} = 1$$

where $\mathcal{L}\{f(x)\} \equiv F(t,s) \equiv \int_{-\infty}^{\infty} f(t,x)e^{-sx}dx$ defines the (two-sided) Laplace transform. The first result follows from interchanging differentiation and integration, while the second and third results follow from integration by parts (using the boundary conditions $f(-\infty) = f(\infty) = 0$). The fourth result is called the 'shift theorem', and follows from the change of variable $s \to s - \beta$. The last result is more subtle. The fact that the Laplace transform of a delta function is just equal to 1 uses results from the theory of generalized functions.[27]

Following these steps produces equation (4.15) in the text, which we repeat here for convenience

$$\frac{\partial F}{\partial t} = \Lambda(s)F(t,s) + \varepsilon\Phi(s-\beta)F(t,s-\beta) + \delta \tag{E.30}$$

where $\beta \equiv 1 - \gamma$ and

$$\begin{aligned} \Lambda(s) &= \frac{1}{2}b_0^2s^2 - a_0s - \delta \\ \Phi(s) &= b_0b_1s^2 + (2b_0b_1\beta - a_1)s + \beta(b_0b_1\beta - a_1) \end{aligned}$$

To solve equation (E.30) we use the approximation $F(t,s-\beta) \approx F(t,s) - \beta\frac{\partial F}{\partial s}$. This gives us

$$F_t = \Lambda(s)F(t,s) + \varepsilon\Phi(s-\beta)\left[F(t,s) - \beta F_s\right] + \delta \tag{E.31}$$

where subscripts denote partial derivatives. Note that we are now back to solving a PDE. We solve this using a standard guess-and-verify/separation-of-variables strategy. In particular, we posit a solution of the following form

$$F(t,s) = H(t)G(s) + F_\infty(s)$$

Loosely speaking, $H(t)$ captures transition dynamics, $G(s)$ captures initial and boundary conditions, and $F_\infty(s)$ captures the new stationary distribution. Here we focus on this last component. Plugging this guess into (E.31) gives

$$H'G = \Lambda[HG] + \varepsilon\Phi[HG - \beta HG'] + \left\{[\Lambda + \varepsilon\Phi]F_\infty - \varepsilon\Phi\beta F'_\infty + \delta\right\} \tag{E.32}$$

where for convenience we have suppressed function arguments. The key point to notice here is that the last term in parentheses is independent of time, so we can solve it separately. Doing so gives us the robust stationary distribution. Another important observation here is that ε multiplies the derivative F'_∞. This 'singular perturbation' term makes conventional first-order perturbation approximation unreliable. To deal with this term we employ the change of variables $\hat{s} = s/\varepsilon$. With this change of variable we can write the ODE in parentheses as follows

$$F'_\infty = \frac{1}{\beta}\left(\varepsilon + \frac{\Lambda}{\Phi}\right)F_\infty + \frac{\delta}{\beta\Phi}$$

We can eliminate the nonhomogeneous term's dependence on $\hat{s}$ by defining $Q(\hat{s}) = \Phi(\hat{s}-\beta)F_\infty(\hat{s})$, which implies $F'_\infty = Q' - (\Phi'/\Phi)Q$. This delivers the following ODE

$$Q' = \left(\frac{\varepsilon + \Lambda + \beta\Phi'}{\beta\Phi}\right)Q + \frac{\delta}{\beta}$$

The general solution of this linear ODE is the sum of a particular solution to the nonhomogeneous equation and the solution of the homogeneous equation. However, we can ignore the homogeneous solution, since

[27] Kaplan (1962) provides a good discussion of Laplace transform methods.

we know that a stationary distribution does not exist when $\delta = 0$.[28] Stated in terms of F_∞, the particular solution is

$$F_\infty(\hat{s}) = \frac{-\delta}{\varepsilon + \Lambda(\hat{s}) + \beta\Phi'(\hat{s} - \beta)}$$

After expanding the denominator polynomial into partial fractions we obtain the result stated in Proposition 4.4. To prove the correspondence principle stated in Corollary 4.5 we can just reverse the change of variables back to s. Let (R_1, R_2) denote the two roots of $F(\hat{s})$, (ϕ_1, ϕ_2) denote the roots of $\Lambda(s)$, and (r_1, r_2) denote the roots of $\Phi(s)$. After substituting s/ε for $\hat{s}$ and then multiplying numerator and denominator by ε^2 we get

$$\frac{\varepsilon^2}{(s - \varepsilon R_1)(s - \varepsilon R_2)} = \frac{-\varepsilon^2 \delta}{\varepsilon^3 + (s - \varepsilon\phi_1)(s - \varepsilon\phi_2) + \varepsilon\beta[s - \varepsilon r_1 + s - \varepsilon r_2]}$$

Taking limits we obtain $\lim_{\varepsilon\to 0}(R_1, R_2) = (\phi_1, \phi_2)$, which is the stated correspondence principle. □

Appendix F. Proof of Proposition 5.1

This follows directly from the proof of Proposition 4.4. Having solved for the stationary distribution, F_∞, the PDE in eq. (E.32) becomes

$$\tilde{F}_t = [\Lambda(s) + \varepsilon\Phi(s - \beta)]\tilde{F} - \varepsilon\beta\Phi(s - \beta)(G'/G)\tilde{F}$$

where $\tilde{F} \equiv HG$. Assuming $O(\beta) = O(\varepsilon)$, the last term can be dropped since it is second-order. □

Appendix G. Recursive Preferences II

To examine inequality dynamics with recursive preferences we just need to replace the expressions for (A_0, A_1) in the main text with the expressions for $(\hat{A}_0, \hat{A}_1)$ derived in Appendix C, and then replace the expressions for (a_0, a_1) stated in Proposition 4.1 with the following expressions for $(\hat{a}_0, \hat{a}_1)$:

$$\begin{aligned} \hat{a}_0 &= r - \varphi + \gamma\sigma^2\alpha_0^2 - \frac{1}{2}b_0^2 \\ \hat{a}_1 &= \frac{\frac{1}{2}\gamma^2\hat{A}_1 - (\sigma\gamma\alpha_0)^2(\hat{A}_0^2 + (\gamma - 1)\hat{A}_1)}{\hat{A}_0\gamma^2} - b_0 b_1 \end{aligned}$$

We can then follow the exact same approximation strategy as outlined in the main text. Figure 11 displays the resulting stationary distributions and convergence rates. We use the same parameter values as those in Table 1, with the exceptions noted in Appendix C. In particular, we increase ε from 0.045 to 0.45, we increase γ slightly from 1.31 to 1.5, and increase μ slightly, from 5.86% to 5.95%.

Given the higher value of ε and the reinforcing effect of the saving rate, it is perhaps not too surprising that we now see a greater increase in inequality. The top 1% wealth share increases from 13.1% to 39.1%. Although 39.1% is very close to current estimates, 13.1% is somewhat smaller than its 1980 value. Finally, and most importantly, we find that robust convergence rates are even higher than those reported in the main text. At the mean level of wealth, the nonrobust convergence rate is 1.6%, close to its original value of 1.14%. However, now the robust convergence rate becomes 6.4%, more than 200 basis points higher than before, and four times greater than its nonrobust value.

One concern with using a higher value of ε is that it induces an overly pessimistic drift distortion and implausibly small detection error probability. However, we find that $\varepsilon = 0.45$ still produces maximal drift distortions around 1%, and detection error probabilities above 40%.

[28]Also note that since $\hat{s} = s/\varepsilon$, as $\varepsilon \to 0$ we know $\hat{s} \to \infty$. From the 'initial value theorem' we know $\lim_{s\to\infty} F(s) = 0$ since $f(x)$ is bounded at the switch point.

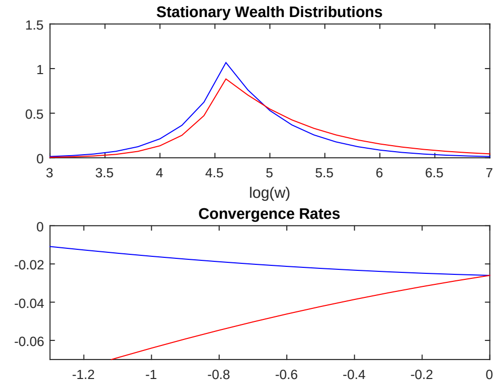

FIGURE 11. Recursive Preferences