Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 184 (2019) 104937

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Information and Inequality ☆

Xiaowen Lei

Department of Economics, University of Oxford, Manor Road Building, Oxford, OX1 3UQ, United Kingdom

Received 27 June 2018; final version received 12 August 2019; accepted 15 August 2019
Available online 26 August 2019

Abstract

This paper studies wealth inequality in a Blanchard/Yaari model with idiosyncratic investment returns. Its key innovation is to assume that individuals can buy information. Information reduces uncertainty about the unknown mean investment return. Reduced estimation risk encourages investment in higher yielding risky assets. As a result, endogenous information acquisition amplifies wealth inequality. Wealthy individuals buy more information, which leads them to invest a higher share of their wealth in higher yielding assets, which then makes them even wealthier. The model's empirical implications are studied using Monte Carlo simulations and perturbation approximations. An empirically plausible decrease in information costs can explain about two-thirds of the observed increase in the top 1% wealth share.
Crown Copyright © 2019 Published by Elsevier Inc. All rights reserved.

JEL classification: D63; D82; D83

Keywords: Inequality; Information; Learning

☆ This paper was partially completed while the author was at Department of Economics, Simon Fraser University and the Bank of Canada, International Economic Analysis Department. I would like to thank Kenneth Kasa for guidance and support. I have also received helpful comments from Klaus Adam, Geoffrey Dunbar, Edouard Djeutem, Paul Beaudry, Martin Ellison, Jack Favilukis, Robert Jones, Alex Karaivanov, Katya Kartashova, Benjamin Moll, Jun Nie, Chris Muris, Galo Nuño, Michael Tseng, Laura Veldkamp, and Shenghao Zhu. I am especially grateful for constructive comments from two anonymous referees, which substantially improved the paper. I am responsible for remaining errors.

E-mail address: leixiaowen04@gmail.com.

https://doi.org/10.1016/j.jet.2019.08.007
0022-0531/Crown Copyright © 2019 Published by Elsevier Inc. All rights reserved.

1. Introduction

We live in a so-called 'information age'. We also live in an age of growing inequality. This paper shows these phenomena might be connected. It attempts to formalize and quantify an argument by Arrow (1987). Arrow noted that in financial markets the value of information is greater, the greater is the amount invested. At the same time, the cost is likely to be nearly independent of the amount invested. Consequently, wealthy individuals devote a higher fraction of their wealth to information. Arrow used a simple 2-period/1-agent example to argue that endogenous information acquisition amplifies inequality.

Although suggestive, Arrow (1987)'s example cannot address the quantitative significance of information acquisition in the dynamics of wealth inequality. This paper quantifies Arrow's example by incorporating learning and information acquisition into recent models of idiosyncratic investment risk. In contrast to models of idiosyncratic labor income risk (Aiyagari, 1994), investment risk models generate the sort of power laws that characterize observed wealth distributions.[1]

To focus on the role of information, this paper abstracts from all other sources of heterogeneity that create inequality. Agents have identical life expectancy, are born with identical initial wealth, and have access to private investment projects with identical mean returns. This mean return is unknown, however, and agents must learn about it by witnessing the history of their own investment returns. In addition, agents can pay a cost to acquire an additional signal. More precise signals are more expensive. In this environment, inequality is initially created by luck. Relatively wealthy agents experience relatively high investment returns. With learning, luck plays two important roles. First, it makes agents relatively 'optimistic', in the sense that high returns produce relatively high estimates of the mean return, which encourages risk-taking and wealth accumulation. Second, higher wealth allows agents to buy more information, which makes them more 'confident', where confidence is defined by the precision of their estimate.

Although it is well known that random growth over a random length of time is enough to generate fat-tailed Pareto wealth distributions (Reed, 2001), recent work by Gabaix et al. (2016) (henceforth GLLM) shows that these random growth models cannot explain observed inequality quantitatively. In particular, they show their transition dynamics are far too slow. GLLM argues that it is important to allow for either 'type dependence' or 'scale dependence', which generate deviations from Gibrat's Law. This paper shows that endogenous information acquisition produces a form of scale dependence.[2] The mean growth rate and volatility of an agent's wealth increase with wealth. This occurs for two reasons. First, relatively wealthy agents have higher savings rates. This is a widely documented feature of the data (Dynan et al., 2004). Second, relatively wealthy agents allocate a higher fraction of their portfolios to risky assets. Again, this is a widely documented feature of the data (Carroll, 2002). The key mechanism driving both these decisions is that wealthy agents buy proportionally more information. They do this because information is proportionally more valuable to them.

[1] Benhabib et al. (2017) note that idiosyncratic labor income risk models cannot generate wealth distributions with fatter right tails than the distribution of labor income. Investment risk models are examples of so-called 'random growth' models, which go back to the work of Champernowne (1953) and Simon (1955). Recent examples include Angeletos (2007), Benhabib et al. (2011), Benhabib et al. (2016), Cao and Luo (2017) and Toda (2014).

[2] Recent work by Cao and Luo (2017) shows that "type dependence" in entrepreneurial productivity, combined with a shift in tax policy and financial deregulation, produces persistent return and wealth differences.

One downside of studying a scale-dependent growth model is that we can no longer solve the model analytically. The model consists of a set of partial differential equations, which cannot be solved in closed-form.[3] In response, I employ a combination of Monte Carlo simulation techniques and classical perturbation approximations.

One might suspect that lowering the cost of information would exert an equalizing force on the distribution of wealth. Models based on asymmetric information and insider trading no doubt have this implication. However, here investment projects are agent-specific. One agent's information is of no value to anyone else.[4] As far as information choice is concerned, each agent is a Robinson Crusoe. From the perspective of an individual agent, information is simply a source of increasing returns, since it encourages risk-taking, which encourages growth. However, when these agent specific scale effects are combined with heterogeneous, non-diversifiable shocks, a powerful force for inequality is ignited. Agents who get lucky early in life use their good fortune to acquire information about future investment returns. In this way, wealth begets wealth.

To study the quantitative implications of the model, I assume the US economy started from a stationary distribution in 1981, with information being relatively costly. I calibrate the initial value of the information cost parameter to match the initial top 1% wealth share. Since information costs are not easily measured, to evaluate the ability of the model to explain changes in observed inequality, I compare two alternative strategies. To start, I simply pick the information cost parameters to replicate the top 1% wealth shares in both 1981 and 2014, and then evaluate their plausibility by calculating the implied shares of wealth and income spent on information. This strategy suggests that information costs declined by a factor of 24 for the median household. Although the initial shares of wealth and income spent on information seem plausible, the implied 2014 shares seem far too high. However, this should not be too surprising, since this strategy assumes that information costs can explain the *entire* increase in wealth inequality. In response, I then use fees from the hedge fund industry to directly measure the decline in information costs. These data suggest that, by 2014 costs had declined by a factor of 11 for the median household. Introducing this information cost reduction into the model, and assuming the economy had reached its steady state by 2014, I find that the top 1% wealth share increases from 24.4% to 32.7%. Since the actual share was 37.2%, hedge fund data suggest that the model can account for about two-thirds of the observed increase.

The analysis here is related to work by Peress (2004) and Kacperczyk et al. (2018). They too are motivated by Arrow (1987). However, there is an important difference between their work and mine. Endogenous information acquisition in these papers is about allocating attention to multiple assets for a given information capacity. Initially wealthy agents are *assumed* to have higher information capacity. These models focus on portfolio choice, and emphasize the general equilibrium effects on endogenous asset prices, with certain assets becoming less traded by unsophisticated investors due to strategic substitution. In my model, information capacity is endogenous, whereas asset returns are exogenous. By shutting down the general equilibrium channel and assuming a private investment technology, I am able to derive explicit expressions for the dynamic relationship between wealth and information. Note that in my model, agents do not initially differ in either wealth or information capacity. Lucky agents who are hit with a series

[3] For recent technical advances in macro models using this approach, see Achdou et al. (2017).

[4] Caveat: Since all projects are assumed to have the same (unknown) mean return, in principle agents could speed up their learning by observing the returns on other agents' projects. I assume these other returns are either unobserved, or that agents are unaware that all projects have the same mean return.

of positive shocks *become* more sophisticated because information is more valuable for wealthy agents, which makes them even wealthier.

The remainder of the paper is organized as follows. Section 2 motivates the discussion by providing background information on wealth inequality, hedge fund growth, and household portfolios. Section 3 outlines the model. Section 4 derives first-order perturbation approximations of an agent's policy functions. Sections 5 studies aggregation and the cross-sectional distribution of wealth and beliefs. Section 6 presents numerical solutions to the stationary wealth distribution. Section 7 compares top wealth shares in several economies with alternative information structures. Section 8 reports a sensitivity analysis by showing how top wealth share changes when the benchmark parameter values change. Section 9 briefly discusses related literature, and Section 10 offers a few concluding remarks on policy implications and possible extensions. Proofs and derivations are contained in an online technical appendix.

2. Motivation

This paper focuses on US wealth inequality. Saez and Zucman (2016) provide detailed data on top wealth shares. They use individual taxpayer data from the Internal Revenue Service, along with intermittent survey data from the Survey of Consumer Finances, and estate and foundations' tax records. They use capitalization methods to translate income to wealth. For assets that do not generate taxable income (e.g.: pensions, primary residences), they use other information, such as property taxes and pension distributions, to generate imputed levels of wealth in those asset categories.

The advantages and disadvantages of this data, as opposed to just using SCF or estate tax data, are discussed in Kopczuk (2015). He points out that for top wealth shares, the capitalization method produces higher estimates of top wealth shares than other approaches. However, the SCF data suffers from a well known nonresponse bias. For example, the response rate from the top percentile is only about 25%. This produces considerable underestimation for top wealth concentration. The estate tax data measures wealth concentration at the individual level, so it is not clear whether estimates are lower or higher than at the household level. Since most investment decisions are determined at the household level, this paper focuses on wealth concentration at the household level. Thus, the capitalization method is likely the more suitable. One caveat is that the capitalization method assumes that returns within asset categories are identical, which could generate bias if there exists correlation between wealth and returns (Fagereng et al., 2016).

Fig. 1 plots the top 1% wealth share from this dataset. This paper does not attempt to explain the reduction of top wealth concentration before the 1980s.[5] As one can see, a substantial increase in wealth inequality is observed beginning in the early 1980s. The top 1% wealth share was 'only' around 24.4% in 1981, but reached 37.2% by 2014.

There are many proposed explanations for this increase. Perhaps the most common one focuses on taxes (Aoki and Nirei, 2017). There were indeed shifts in tax policy in the early 1980s that favored the wealthy, and this paper does not dispute the potential role of taxes. However, one intriguing finding in Aoki and Nirei (2017) is that the *impact* of taxes on inequality seemed to increase as well around this time. This suggests that additional factors might be at play. This paper shows that the information technological change creates opportunities, by making information acquisition and information processing less costly. We can gauge the importance of these

[5] Piketty (2014) provides detailed explanations for this as a long process starting from WWI.

Fig. 1. Top 1% wealth share in the USA.

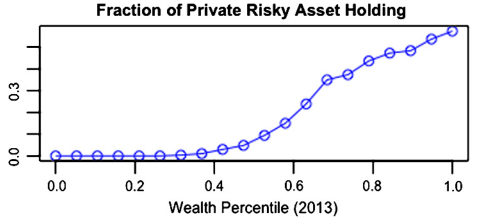


Fig. 2. Fraction of risky assets holding.

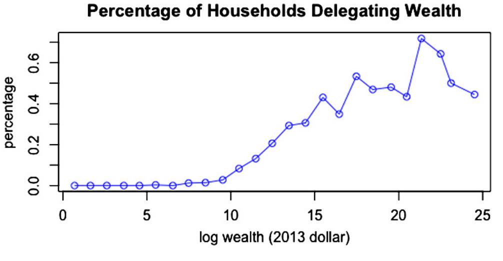


Fig. 3. Percentage of households delegating wealth.

changes indirectly by looking at changes in household portfolios and in the wealth management industry. Fig. 2 plots SCF data on the share of private risky assets in household portfolios in 2013. Households are grouped by relative log wealth, in increments of 5 percentage points. As one can see, relatively wealthy investors hold a higher *share* of risky assets in their portfolios. Simple dynamic consumption/portfolio models in the Merton tradition predict that portfolio shares are independent of wealth.

There are many potential reasons why risky portfolio shares might increase with wealth. This paper attributes it to the fact that wealthy investors have better information. They have better information because they can afford to buy it. Fig. 3 reports one piece of evidence to support this idea. It plots the percentage of households who pay either a hedge fund or a mutual fund for wealth management services as a function of log wealth. The data are from 2013.

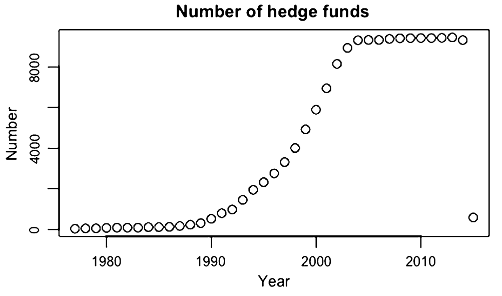


Fig. 4. Number of hedge funds.

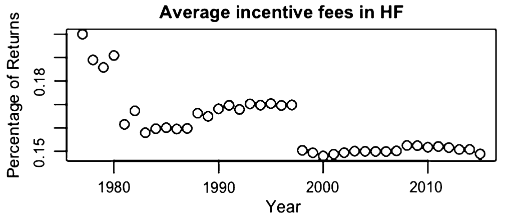


Fig. 5. Average incentive fees in hedge funds.

Evidently, if your wealth is less than $22,000 ($\approx e^{10}$), you are very unlikely to be delegating your wealth. Interestingly, the plot suggests that this represents something of a threshold. The likelihood of paying for wealth management rises rapidly after this point. This suggests the presence of a fixed cost. For simplicity, my model abstracts from fixed costs.[6]

Finally, from the perspective of my model, looking at simple stock holdings data is a bit misleading. Anyone can buy shares in Apple or Microsoft. My model presumes that risky investment is idiosyncratic. Hence, it is perhaps more consistent with recent work on private equity markets and entrepreneurship. (See Quadrini (2009) for a survey). However, even shares in public companies can become somewhat private if they are used in dynamic trading strategies based on private information and research. That's what the hedge fund industry is all about. Figs. 4 and 5 present data on the explosive growth of the hedge fund industry, using data from Lipper TASS hedge fund database. Detailed summary statistics of this dataset is provided in Appendix F.

Fig. 4 shows that the hedge fund industry grew slowly during the 1980s, and then took off during the 1990s. This explosive growth coincided with a rapid fall in their fees during the 1980s, as seen in Fig. 5.[7]

Although the evidence in these figures is suggestive, the real question is whether plausible changes in the market for information can *quantitatively* account for the dramatic and rapid rise in inequality depicted in Fig. 1. To address this question we need a model.

[6] Fixed information costs are likely to be more important in the left-tail of the wealth distribution. Here I focus on the right-tail.

[7] The average hedge fund entry cost has gone down. However, this does not imply that it is cheap. Hedge funds have always been a club for the wealthy, even now. For example, for a typical hedge fund in 2013, one still needs at least $1.26 mill. to enter, which implies one must be within the top 12.5% to take advantage of the hedge fund industry.

3. The model

The model here combines two existing literatures. The first extends the workhorse Blanchard/Yaari continuous-time OLG model by incorporating idiosyncratic investment returns. Benhabib et al. (2016) show that this extension produces a double Pareto wealth distribution, with tail exponents that are easily interpretable functions of the model's underlying parameters. However, their model is a traditional random growth model without scale dependence, i.e., it obeys Gibrat's Law. Rather than study scale dependence, they focus on the role of bequests and fiscal policy in amplifying inequality. The second consists of a single paper, an unpublished Ph.D thesis by Turmuhambetova (2005). She incorporates endogenous information acquisition into a traditional Merton portfolio problem. However, her model consists of a single agent, so obviously there isn't much discussion of inequality.

3.1. The setup

The economy consists of a measure 1 continuum of agents with exponentially distributed lifetimes. Death occurs at Poisson rate δ. When an agent dies, he is instantly replaced by a new agent with initial wealth w_0. One can interpret this initial wealth as the present value of an agent's (riskless) lifetime labour income. Agents have no bequest motive. They can invest in three assets: a risk free technology, a competitively supplied annuity, and a risky technology. The value of the risk free asset follows the deterministic process

$$dQ(t) = \tilde{r}Q(t)dt \tag{3.1}$$

with constant rate of return $\tilde{r}$. Since agents face idiosyncratic death risk, there is a gain from setting up an annuities market, which allows agents to die in debt. By no arbitrage and free entry, the rate of return on these annuities is $r = \tilde{r} + \delta$. Therefore, in equilibrium, no rational agent has an incentive to use the risk free technology. The value of the private risky technology obeys a geometric Brownian motion

$$dS(t) = \mu S(t)dt + (1-\lambda)\sigma S(t)dB(t) \tag{3.2}$$

That is, the private risky technology has a constant mean growth rate μ and a return volatility of σ. To be more realistic, I also assume that λ fraction of investment risk can be diversified away in the economy, so that the actual return volatility becomes $(1-\lambda)\sigma = \tilde{\sigma}$. To focus on scale dependence rather than type dependence (GLLM), I assume the mean and volatility are identical across agents. What is important is that the shocks, dB, are uncorrelated across agents.

A novel feature of my model is the assumption that agents do not know the mean return, μ, of their investment project.[8] They must learn about it over time. As noted in the Introduction, I assume they do this by observing their own returns. Agents are unaware their ancestors used a technology with the same mean, so history does not matter to them. Likewise, they are unaware that other currently alive agents have the same mean return, so there is no perceived gain from observing other agents. Interestingly, even when agents observe a common stochastic pro-

[8] As noted by Merton (1980), uncertainty about σ decreases as sampling frequency increases. It disappears in the continuous time limit.

cess, evidence suggests they weight their own experience more heavily (Malmendier and Nagel, 2016).[9]

Uncertainty is represented by a filtered probability space $(\Omega, \mathbb{P}, \{\mathcal{F}_t\}_{t\geq 0}, \mathcal{F})$, induced by an (unobserved) one-dimensional standard Brownian motion $B(t)$, which satisfies the usual conditions. Each agent has an equivalent probability measure $\hat{\mathbb{P}}$ that generates his own observable filtration $\{\hat{\mathcal{F}}_t\} \subset \{\mathcal{F}_t\}$. This filtration defines the following conditional mean, $\hat{\mu}(t)$, and variance, $\gamma(t)$, of an agent's estimate of μ

$$\hat{\mu}(t) = E[\mu \mid \hat{\mathcal{F}}_t] \tag{3.3}$$

$$\gamma(t) = E[(\hat{\mu}(t) - \mu)^2 \mid \hat{\mathcal{F}}_t] \tag{3.4}$$

At birth, the investor has a prior mean $\hat{\mu}(0)$ and a prior estimation variance $\gamma(0)$.

By Girsanov's theorem, $\hat{B}(t)$ is a Brownian motion under the investor's own filtration.[10] Following standard filtering theory [Liptser and Shiryaev (2000)], the innovation process $\hat{B}(t)$ induced by the investor's own filtration is related to the unobserved $B(t)$ by

$$d\hat{B}(t) = \frac{1}{\tilde{\sigma}}[\frac{dS(t)}{S(t)} - \hat{\mu}(t)dt] = dB(t) + \frac{\mu - \hat{\mu}(t)}{\tilde{\sigma}}dt \tag{3.5}$$

3.2. *Filtering and information*

In addition to observed returns, at each instant of time t the investor observes a noisy signal y that correlates with μ,

$$dy(t) = \mu dt + \sigma_y(t)dB_y(t) \tag{3.6}$$

where $\{B_y(t)\}$ are standard Brownian motions, independent of $\{B(t)\}$. This generates a stream of additional information about the unobserved μ, which can be used by the investor to update his beliefs in Bayesian fashion.

The investor's Kalman filtering equations can be written in innovations form

$$d\hat{\mu}(t) = \frac{\gamma(t)}{\tilde{\sigma}}d\hat{B}_s(t) + \frac{\gamma(t)}{\sigma_y(t)}d\hat{B}_y(t) \tag{3.7}$$

$$d\gamma(t) = -\gamma(t)^2\left(\frac{1}{\tilde{\sigma}^2} + \frac{1}{\sigma_y(t)^2}\right)dt \tag{3.8}$$

where $d\hat{B}_y$ is related to $dB_y(t)$ according to

$$d\hat{B}_y(t) = \frac{1}{\sigma_y(t)}[dy(t) - \hat{\mu}(t)dt] = dB_y(t) + \frac{\mu - \hat{\mu}(t)}{\sigma_y(t)}dt \tag{3.9}$$

Now, the key innovation of my model is to allow agents to reduce the variance of the noisy signal, $\sigma_y^2(t)$, by paying an information cost. Note, this is a monetary cost, expressed in units of wealth, not an information processing cost, expressed in bits per unit of time. However, I assume

[9] Ehling et al. (2016) and Nakov and Nuño (2015) incorporate experiential learning into an OLG framework. They focus on asset pricing.

[10] Girsanov's theorem delivers a stochastic process analog of a Jacobian. It describes how the dynamics of a stochastic process change when the original measure is changed to an equivalent (mutually absolutely continuous) probability measure. For details, see Liptser and Shiryaev (2000) or Øksendal (2003).

this monetary cost is a function of the informativeness of the signal, as measured in conventional information-theoretic terms. In particular, I suppose the investor chooses an instantaneous channel capacity, $\kappa(t)$.[11] It provides a measurement of uncertainty reduction in a random variable. Formally, it is defined by the Kullback-Leibler divergence

$$\kappa = \int \log \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{P} \tag{3.10}$$

where $d\mathbb{P}$ and $d\mathbb{Q}$ represent the posterior and prior conditional distributions of μ. In this particular case with a Gaussian information structure, $\sigma_y(t)$ and $\kappa(t)$ are related by

$$\frac{1}{2}\left(\log \gamma(t) - \log \sigma_y^2(t))\right) \leq I(t) \tag{3.11}$$

where $I(t) \equiv \frac{1}{2}\log(2\kappa(t))$. This relates signal precision to the rate of information. Given the current estimation variance, a higher channel capacity allows a more precise signal (i.e., a lower signal variance). This accelerates learning.

The investor chooses instantaneous capacity, $\kappa(t)$, by paying a monetary cost. I assume the cost function, $q(\kappa)$, is quadratic:

$$q(\kappa(t)) = \frac{1}{2\theta}(\kappa(t))^2 \tag{3.12}$$

where θ is a cost parameter. When θ is small, information is costly. This function captures the increasing marginal complexity of processing financial information accurately. The increasing marginal cost structure also ensures the investor can never perfectly learn μ. One interpretation of this cost function is that there is a competitive industry producing and selling financial information at its marginal cost.[12]

3.3. Optimization problem

Each agent is assumed to have time-additive constant relative risk aversion (CRRA) preferences. She needs to make continuous decisions on consumption, $c(t)$, the fraction of wealth invested in the risky technology, $\pi(t)$, and how much information to acquire, $\kappa(t)$. Formally, she faces the following dynamic optimization problem:

$$V(w, \hat{\mu}, \gamma) = \max_{c,\kappa,\pi} \mathbb{E}\left[\int_t^\infty \exp^{(-(\rho+\delta)s)} \frac{c(s)^\alpha}{\alpha} ds | \hat{\mathcal{F}}_t\right] \tag{3.13}$$

s.t.:

$$dw(t) = [w(t)(r + \pi(t)(\hat{\mu}(t) - r)) - c(t) - q(\kappa(t))]dt + \pi(t)\tilde{\sigma} w(t) d\hat{B}(t) \tag{3.14}$$

$$d\hat{\mu}(t) = \frac{\gamma(t)}{\tilde{\sigma}} d\hat{B}(t) + \frac{\gamma(t)}{\sigma_y(t)} d\hat{B}_y(t) \tag{3.15}$$

[11] In rational inattention models (e.g.: Sims (2003)), it is conventional to use base 2 logs, so that the entropy of a discrete distribution with equal weight on two points, $-0.5\log(0.5) - 0.5\log(0.5) = 1$, and this unit of information is called one "bit".

[12] A general convex information cost structure is considered in Turmuhambetova (2005). As here, Andrei and Hasler (2014) use a quadratic cost function.

$$d\gamma(t) = -\gamma(t)^2 \left(\frac{1}{\tilde{\sigma}^2} + \frac{1}{\sigma_y(t)^2} \right) dt \tag{3.16}$$

$$\frac{1}{2} \left(\log \gamma(t) - \log \sigma_y^2(t) \right) \leq I(t) \tag{3.17}$$

where $I(t) \equiv \frac{1}{2} \log(2\kappa(t))$. Here, $1 - \alpha$ is the coefficient of relative risk aversion, and ρ denotes the rate of time preference. The capacity constraint on information is binding for an optimizing investor. Therefore, the choice of $\sigma_y^2(t)$ is equivalent to the choice of $\kappa(t)$. Thus we can substitute $\kappa(t)$ for $\sigma_y^2(t)$ in what follows. Using Ito's lemma, the HJB equation for this optimization problem can be written as

$$\beta V(w, \hat{\mu}, \gamma) = \max_{c,\kappa,\pi} \frac{c^\alpha}{\alpha} + V_w \left(rw + \pi \left(\hat{\mu} - r \right) w - c - q(\kappa) \right) + \frac{1}{2} V_{ww} \pi^2 \tilde{\sigma}^2 w^2 + \left(\frac{1}{2} V_{\hat{\mu}\hat{\mu}} - V_\gamma\right)\left(\frac{\gamma^2}{\tilde{\sigma}^2} + 2\kappa\gamma\right) + V_{w\hat{\mu}} \pi w \gamma \tag{3.18}$$

where $\beta = \rho + \delta$. The first-order conditions deliver the following policy functions, expressed in terms of the unknown value function $V(w, \hat{\mu}, \gamma)$[13]:

$$c^* = V_w^{\frac{1}{\alpha - 1}} \tag{3.19}$$

$$\pi^* = \frac{-V_w(\hat{\mu} - r) - V_{w\hat{\mu}}\gamma}{V_{ww} w \tilde{\sigma}^2} \tag{3.20}$$

$$\kappa^* = \theta\gamma \left(\frac{V_{\hat{\mu}\hat{\mu}} - 2V_\gamma}{V_w} \right) \tag{3.21}$$

Notice the optimal consumption function is the same as in the full information case, although consumption is indirectly influenced through changes in the value function. The risky portfolio share, π^*, has two terms. The first term reflects myopic asset demand as in a standard Merton portfolio problem, where the investor trades off between excess expected returns and its volatility. The second term reflects a hedging demand, which comes from learning about the parameter μ. Its sign depends on the level of risk aversion, α, and its intuition will become clearer after we derive the perturbation approximations in the next section. The κ^* policy function shows that less information will be purchased when its cost is high (θ is low), and when uncertainty about returns is low (γ is small). Notice that the effect of uncertainty is greater when agents are more averse to it (i.e., when V_γ is more negative). More interestingly, it shows that, *ceteris paribus*, wealthy agents will buy more information, since their marginal utility of wealth, V_w, is relatively low. The role of the numerator will become clear in the next section.

4. Policy function approximations

Due to the interdependence of the value function $V(w, \hat{\mu}, \gamma)$ and the policy function equations (3.19)–(3.20), we need to solve a 3-dimensional, highly nonlinear PDE. A closed-form solution is wishful thinking. However, it turns out that when information is prohibitively expensive ($\theta = 0$), and the investor has log preferences ($\alpha = 0$), the PDE for the value function can be

[13] Given the recursive structure, I now drop the time t notation for convenience.

solved analytically.[14] This opens the door to a classical perturbation approximation. The derivation of value function is derived in the Appendices A and B. Here, I focus on the implied policy functions:

Proposition 1. *To an $O(\theta, \alpha)$ approximation, the policy functions for savings rate, risky portfolio share, and information choice are given by*

$$\tilde{s}^* \approx \underbrace{(1-\beta)}_{\text{benchmark saving rate}} + \underbrace{\frac{\alpha}{2\tilde{\sigma}^2}\left[(\hat{\mu}-r)^2 + \frac{\gamma^2}{\beta\tilde{\sigma}^2}\right]}_{\text{exogenous learning}}$$
$$+ \underbrace{\theta\beta w \exp\left(\frac{2\beta\tilde{\sigma}^2}{\gamma}\log(\hat{\mu}-r) - \frac{(\hat{\mu}-r)^2}{2\gamma}\right)}_{\text{endogenous learning}} \tag{4.22}$$

$$\pi^* \approx \frac{1}{\tilde{\sigma}^2}[\underbrace{(1+\alpha)(\hat{\mu}-r)}_{\text{myopic portfolio}} + \underbrace{\alpha\frac{(\hat{\mu}-r)\gamma}{\beta\tilde{\sigma}^2}}_{\text{exogenous learning}}$$
$$+ \underbrace{\theta w \exp\left(\frac{2\beta\tilde{\sigma}^2}{\gamma}\log(\hat{\mu}-r) - \frac{(\hat{\mu}-r)^2}{2\gamma}\right)[(\hat{\mu}-r) + \frac{2\beta\tilde{\sigma}^2 - (\hat{\mu}-r)^2}{(\hat{\mu}-r)}]]}_{\text{endogenous learning}} \tag{4.23}$$

$$\kappa^* \approx \frac{\theta w}{\beta}\left[1 - \frac{\beta\tilde{\sigma}^2}{\gamma}e^{\beta\tilde{\sigma}^2/\gamma}\Gamma(0, \beta\tilde{\sigma}^2/\gamma)\right] \approx \frac{\theta w \gamma}{\beta^2\tilde{\sigma}^2} \tag{4.24}$$

where $\Gamma(a, x) = \int_x^\infty t^{a-1}e^{-t}dt$ is the incomplete Gamma function.

Proof. See Appendix A and B. □

4.1. Comments on information choice

Like any other good, the optimal amount of information is chosen so that its marginal cost equals its marginal benefit. Recall that the information cost function is $q(\kappa(t)) = \frac{1}{2\theta}(\kappa(t))^2$, which implies that the marginal cost of information at the optimum is equal to $\frac{\kappa^*}{\theta}$. Therefore, another way to interpret equation (4.24) is that, to a first order approximation, the agent is willing to spend her wealth on acquiring information up to a point where her marginal cost of acquiring it at the moment, $\frac{\kappa^*}{\theta}$, equals the perceived expected discounted lifetime marginal benefit from acquiring it, (i.e.: $\frac{w\gamma}{\beta^2\tilde{\sigma}^2}$). It is not surprising that the marginal benefit rises when either wealth or uncertainty increase. On one hand, increased wealth enhances the marginal benefit of acquiring more information for given uncertainty level. On the other hand, most rich people will already have a relatively low estimation uncertainty, which implies that acquiring more information probably has less marginal benefit. To see the net effect of these two opposing forces, one can calculate the drift of the optimal choice of information capacity. We know that

[14] In this case, learning is based on exogenous information, and the learning problem decouples from the saving and portfolio problems. This is due to the fact that a log investor does not need to hedge against perceived changes in the investment opportunity set. Income and substitution effects offset each other (Gennotte, 1986).

$$
\begin{aligned}
E_t(d\kappa) &= \frac{\theta}{\beta^2\tilde{\sigma}^2}E_t(d(w\gamma)) \\
&= \frac{\theta}{\beta^2\tilde{\sigma}^2}E_t(\gamma dw + wd\gamma + \frac{1}{2}dwd\gamma) \\
&= \frac{\theta}{\beta^2\tilde{\sigma}^2}(\gamma\mu(w)w + w(-\gamma^2(\frac{1}{\tilde{\sigma}^2} + \frac{2\theta w}{\beta^2\tilde{\sigma}^2})))dt \\
&\approx \frac{\theta}{\beta^2\tilde{\sigma}^2}\gamma\mu(w)wdt \qquad (4.25)
\end{aligned}
$$

where the second equality applies Ito's lemma, and the last approximation takes a first order approximation of γ around $\gamma = 0$. Since the mean wealth growth rate, $\mu(w) > 0$, we know that up to an $O(\gamma^2)$ approximation, the wealth effect dominates. Note that this analysis takes into account the fact that rich people learn faster, but this effect is of order $O(\gamma^2)$, while the wealth effect is of order $O(\gamma)$. This holds as long as γ is small. Of course, one might question whether this approximation is accurate. I shall return to this question in the numerical section.

4.2. Portfolio choice

From equation (4.23), one can see that to a first-order approximation, one can decompose optimal portfolio choice into three pieces. The first part is the myopic demand of a Merton agent. The second piece is a portfolio hedging demand coming from learning (Gennotte, 1986). Notice this is independent of wealth, so learning by itself is not a source of scale effects. It is worthwhile to examine the sign of this hedging demand. With a small initial estimation error, $(\hat{\mu} - r)$ is almost always positive. Therefore, the sign of the hedging demand is determined by one's risk aversion coefficient α. Since agents invest in their own private technology and receive private signals, this source of idiosyncratic shocks in learning already creates heterogeneity in portfolios. However, it is not a source of scale dependence. The last component of portfolio demand is what creates scale dependence. The presence of information choices increases portfolio demand directly, whose effect is stronger the wealthier an agent is.

4.3. Savings rate

Just like portfolio choice, from equation (4.22), one can also decompose the savings rate into three pieces. The first piece $(1 - \beta)$ is the saving rate of a benchmark log utility agent without learning. The agent adheres to the Permanent Income Hypothesis in this case, and consumes a fixed percentage of wealth equal to their effective discount rate $(\rho + \delta)$. The interpretation of the second piece is the additional saving rate of a CRRA agent, whose sign may or may not be positive, depending on the level of risk aversion. The last component reflects information choice. Observe that it is wealth dependent. *Ceteris paribus*, wealthy agents have higher saving rates. This is consistent with the data.[15] In this model, it is due to the fact that wealthy investors prefer to save their wealth for acquiring information and invest in risky assets. However, I shall return to this point and show in the numerical section that the scale dependence in the savings channel has a quantitatively smaller effect than the portfolio choice channel.

[15] See, for example, Dynan et al. (2004).

Corollary 4.1. *Ceteris paribus, as information cost decreases, the demand for information increases. This effect is stronger when agents get richer. Finally, increased information leads to an increased risky portfolio share. That is,* $\frac{\partial \kappa^*}{\partial \theta} > 0$, $\frac{\partial \kappa^*}{\partial \theta \partial w} > 0$, $\frac{\partial \pi^*}{\partial \kappa} > 0$.

Proof. This comes directly from taking partial derivatives of the optimal information policy function with respect to information cost, i.e.: $\frac{\partial \kappa^*}{\partial \theta} = \frac{w\gamma}{\beta^2 \tilde{\sigma}^2} > 0$; $\frac{\partial \kappa^*}{\partial \theta \partial w} = \frac{\gamma}{\beta^2 \tilde{\sigma}^2} > 0$. Finally, since $\frac{\partial \pi^*}{\partial w} > 0$ and $\frac{\partial \kappa^*}{\partial w} > 0$, we then have $\frac{\partial \pi^*}{\partial \kappa} > 0$. □

Not surprisingly, when information becomes less expensive ($\theta \uparrow$), agents demand more information. However, notice the effect interacts with wealth. This implies that even though everyone buys more information when information is cheaper, the marginal effect is greater for wealthy agents. As will be discussed in the next section, this give them a greater incentive to accumulate wealth through a portfolio channel. It is important to keep in mind that these results are just partial derivatives, holding other endogenous variables constant. Full responses are depicted in the Numerical Results section.

5. The distribution of wealth

Thus far we have studied the problem and decision rules of a single agent. Since our main goal is to study inequality, we must now aggregate these decision rules and characterize the cross-sectional distribution of wealth. Readers not interested in analytic results can jump directly to the Numerical Results section, where the quantitative results are presented.

5.1. Individual wealth and belief dynamics

I begin by describing the wealth and belief dynamics of an individual agent. Wealth dynamics are obtained by substituting the policy functions into the wealth accumulation equation in (3.14).

Lemma 5.1. *To a first-order approximation, an individual's wealth dynamics under filtration* $\{\mathcal{F}_t\}$ *are governed by the diffusion*

$$dw = (a + \theta w b) w dt + (c + \theta w \hat{d}) w dB \tag{5.26}$$

where

$$a = \frac{(\mu - r)}{\tilde{\sigma}^2}\left[(1+\alpha)(\hat{\mu} - r) + \alpha \frac{(\hat{\mu} - r)\gamma}{\beta \tilde{\sigma}^2}\right] + \frac{\alpha}{\tilde{\sigma}^2}\left[(\hat{\mu} - r)^2 + \frac{\gamma^2}{\beta \tilde{\sigma}^2}\right] \tag{5.27}$$

$$b = \exp(f^0(\hat{\mu}, \gamma))(r + \frac{(\hat{\mu} - r)}{\tilde{\sigma}^2}((\hat{\mu} - r) + f^0_{\hat{\mu}}\gamma)) \tag{5.28}$$

$$c = \frac{1}{\tilde{\sigma}}\left[(1+\alpha)(\hat{\mu} - r) + \alpha \frac{(\hat{\mu} - r)\gamma}{\beta \tilde{\sigma}^2}\right] \tag{5.29}$$

$$\hat{d} = \frac{\exp(f^0(\hat{\mu}, \gamma)}{\tilde{\sigma}}(\hat{\mu} - r + f^0_{\hat{\mu}}\gamma) \tag{5.30}$$

where $f^0(\hat{\mu}, \gamma) = \frac{2\beta\sigma^2}{\gamma} \log(\hat{\mu} - r) - \frac{(\hat{\mu} - r)^2}{2\gamma}$

Proof. See Appendix C. □

There are two effects when information cost goes down. They work in the same direction on wealth accumulation: First, for a given level of information demand, cheaper information relaxes the investor's budget constraint, allowing more savings and investment. Second, cheaper information triggers a substitution effect toward more information, allowing investors to be more confident about asset returns, thus taking on more risk. From Lemma 5.1, one can see that if we fix the two endogenous variables, an increase in θ leads to an increase in the growth rate and volatility of wealth. Of course, since wealth and uncertainty themselves are endogenous, one can't tell from the above comparative statics the total effect of an increase of θ on growth and volatility. To study the total derivative, I rely on numerical simulations.

For some questions it is convenient to characterize the wealth distribution analytically. For example, doing so clarifies the interaction between belief heterogeneity and wealth inequality. For more technical readers, here we shall see that learning and endogenous information produces a right-tail Pareto exponent that is smaller in absolute value than without learning, and we can interpret this exponent in terms of the model's underlying parameters. Analytic characterization of the cross-sectional distribution of wealth and beliefs can be accomplished by studying the properties of the Kolmogorov-Fokker-Planck (KFP) equation. Although the KFP equation can be used to study the transition dynamics of inequality, here I focus on stationary distributions.[16]

We begin by collecting together the stochastic differential equations describing an individual's wealth and beliefs.

Proposition 2. *Let* $x = \log(w)$*, and let* $f(x, \hat{\mu}, \gamma)$ *denote the stationary cross-sectional distribution of (log) wealth and beliefs. It obeys the following KFP partial differential equation (subscripts denote partial derivatives)*

$$
\begin{aligned}
0 = {} & f\left[\frac{3\gamma}{\tilde{\sigma}^2} - \delta + \theta e^x \big(c\hat{d} - b + \frac{2\gamma}{\beta^2\tilde{\sigma}^2} + \frac{4\gamma}{\beta^2\tilde{\sigma}^2} + \frac{1}{2}\hat{d}_{\hat{\mu}}\frac{\gamma}{\tilde{\sigma}} + \frac{1}{2}\hat{d}_{\hat{\mu}}\frac{\gamma}{\tilde{\sigma}}\big)\right] \\
& + f_x\left[\frac{1}{2}c^2 - a + \theta c\hat{d} + \theta e^x\big(c\hat{d} - b + \hat{d}_{\hat{\mu}}\frac{\gamma}{\tilde{\sigma}}\big)\right] \\
& + f_{\hat{\mu}}\left[\frac{\gamma(\hat{\mu}-\mu)}{\tilde{\sigma}^2} + \theta e^x\left(\frac{2\gamma(\hat{\mu}-\mu)}{\beta^2\tilde{\sigma}^2} + \frac{\gamma\hat{d}}{\tilde{\sigma}} + \frac{\hat{d}}{2}\frac{\gamma}{\tilde{\sigma}}\right)\right] + f_\gamma\gamma^2\left(\frac{1}{\tilde{\sigma}^2} + \frac{2\theta e^x}{\beta^2\tilde{\sigma}^2}\right) \\
& + \left(\frac{1}{2}c^2 + \theta e^x c\hat{d}\right) f_{xx} + \frac{1}{2}f_{\hat{\mu}\hat{\mu}}\left(\frac{\gamma^2}{\tilde{\sigma}^2} + \frac{2\theta e^x\gamma^2}{\beta^2\tilde{\sigma}^2}\right) + \frac{1}{2}(c + \theta e^x\hat{d})\frac{\gamma}{\tilde{\sigma}} f_{x\hat{\mu}} \\
& + \delta\zeta(X - X_0)
\end{aligned}
\tag{5.31}
$$

where $X = [x, \hat{\mu}, \gamma]^T$ *represents the state vector, and* $X_0 = [x_0, \mu_0, \gamma_0]^T$ *represents initial (log) wealth and beliefs at birth. The function* $\zeta(.)$ *is a Dirac delta function.*

Proof. See Appendices C and D. □

The KFP equation characterizing the stationary distribution is a 3-dimensional PDE, describing the distribution of wealth and beliefs, as summarized by the conditional mean and variance of expected returns. One can study this system using classical time-scale separation methods. These methods convert the problem to one of studying the interaction between two lower dimensional

[16] Gabaix et al. (2016) focus on transition dynamics. Kasa and Lei (2018) use the time-dependent KFP equation to argue that ambiguity accelerates transition dynamics.

Table 1
Benchmark parameter values.

μ	ρ	δ	α	γ_0	σ	λ	θ_1	θ_2	w_0
0.11	0.015	0.0246	−0.4	0.002	0.5	0.47	0.000016	0.00077	1

subsystems. In particular, when uncertainty about mean returns is 'small', learning is 'slow'. As a result, wealth and beliefs evolve on different time-scales. Wealth evolves quickly, while beliefs evolve slowly. One can first solve a 1-dimensional KFP equation for wealth, holding beliefs fixed, then use the implied stationary distribution of wealth to average over wealth dependence in the equations describing the cross-sectional distribution of beliefs. Finally, Laplace's method can be used to characterize the right-tail of the marginal distribution of wealth.

Proposition 3. *Let ϕ_L and ϕ_0 denote the tail Pareto exponent of stationary distribution of wealth in endogenous and exogenous information economy respectively. Let $\phi_{\hat{\mu}}$ and $\phi_{\hat{\mu}\hat{\mu}}$ be the first and second derivative of the tail exponent w.r.t. $\hat{\mu}$ evaluated at the $\hat{\mu} = \mu$. For large x, the (right) tail Pareto exponent with endogenous information, ϕ_L, is approximately*

$$\phi_L \approx \phi_0 + (1 + \frac{\phi_{\hat{\mu}}^2}{2|\phi_{\hat{\mu}\hat{\mu}}|}) \qquad \Rightarrow \qquad |\phi_L| < |\phi_0|$$

Hence, endogenous information decreases the tail exponent, and endogenous information increases top wealth shares.

Proof. See Appendix E. □

Using this time-scale separation approximation, one can then see that endogenous information acquisition reduces the absolute value of the right-tail Pareto exponent, thereby increasing top wealth shares.

6. Numerical solution

In this section, I first present the policy functions using calibrated parameters. I then compute the stationary distribution using Monte Carlo simulation, and study the quantitative implications of endogenous information acquisition in amplifying wealth inequality.

Table 1 summarizes the benchmark parameters I use, with the choice of each parameter discussed below.

1). *Death rate δ.* The mean working life is $\delta^{-1} = 40.7$ years. It is chosen to match average duration in the US labor force. According to published data from Bureau of Labor Statistics, average labor force duration in the United States in 1996–2016 is 40.7 years,[17] implying an annual death rate of 2.46%.

2). *Volatility, σ, and diversification, λ.* The model emphasizes idiosyncratic investment risk. Of course, investment risk is both idiosyncratic and aggregate. To capture this, I follow the approach in Angeletos and Panousi (2011) (later referred to as AGP). AGP consider a risky asset investment process s.t.:

[17] See https://www.bls.gov/emp/tables/median-age-labor-force.htm.

$$dS(t) = \mu S(t)dt + (1-\lambda)\sigma S(t)dB(t) \tag{6.32}$$

where λ captures the fraction of investment risk that can be diversified in the public equity market. One can think of a household owning a private business, where λ fraction of it goes public, while the rest $(1-\lambda)$ remains private. This interpretation suggests that one can simply look at the ratios of corporate and proprietors profits to total profits as an estimate of the fraction of diversifiable risk. As in AGP, the gross idiosyncratic volatility of risky investment implies that $\sigma = 0.5$, where they combine several estimates of private sector volatility in US from the literature. Using the ratio of proprietors' profits over total profits as an estimate of diversifiable risk, they find using National Income and Product Account (NIPA) data that during 1981–2006, the fraction is 47%. This suggests that $\lambda = 0.47$. Therefore, the "pure" risk $(1-\lambda)\sigma$ would be around 0.26. This is used for my benchmark calibration. AGP then note that using the estimated idiosyncratic variance of consumption growth in the US, the number is much smaller, with "after diversification" risk $(1-\lambda)\sigma = 0.2$ instead. Therefore, I will experiment with a range of values in the subsequent sensitivity analysis.

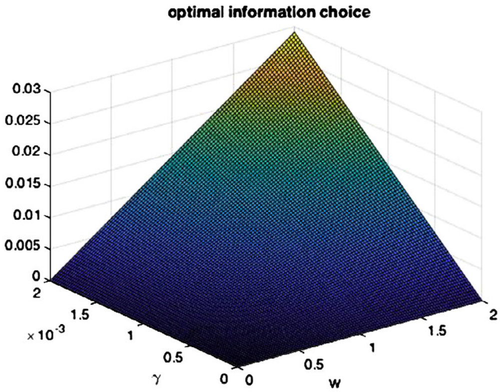
3). *Time preference, ρ*. Empirical estimates of (annual) time preference are around 1% to 2%. I take the average estimate here so that $\rho = 1.5\%$.

4). *Risk aversion, α*. I set the benchmark value to $\alpha = -0.4$, implying a coefficient of relative risk aversion of 1.4. Although this might appear to be implausibly low from an asset pricing perspective, we know from the work of Barillas, Hansen, and Sargent (2009), for example, that it can be misleading to interpret risk aversion coefficients in environments featuring different information structures. Moreover, there are constraints in choosing this parameter. For example, if α is too low, poor people in the economy would either start shorting assets, or start saving too much. But since this is an economy populated by Robinson Crusoes, one needs to keep the savings rate and portfolio shares between zero and one. Parameter search using these criteria shows that -0.4 is the lowest that α can go. However, we will still see the effect of changing α in changing wealth distribution in the subsequent sensitivity analysis.

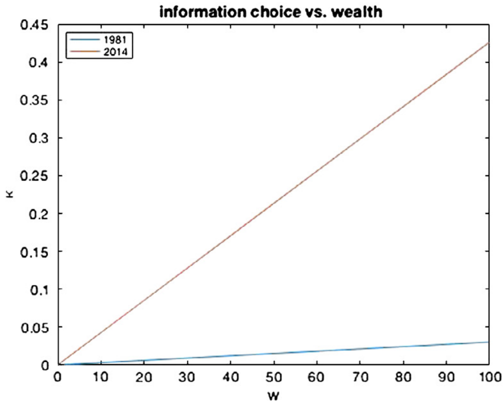
5). *Mean growth rate, μ*. Since now I take into account that a fraction of investment risk is diversified in the financial market, the mean growth rate μ has the interpretation of being a weighted average of private and public equity mean growth rates. The average annual real return in the US stock market is 6.5%, from Robert Shiller's stock market index data from 1980 to 2018. With private equity mean return being 15% (as the average of Moskowitz et al. (2002) and Kartashova (2014)), and using $\lambda = 0.47$ as the weight on public and private equity, I find that new mean estimate is 0.11.

6). *Prior estimation variance, γ_0*. This parameter is estimated from the learning and portfolio choice literature *a la* Brennan (1998) and Xia (2001). The idea is that the prior variance should correspond roughly to the variance of the sample mean. Brennan uses Ibbotson and Sinquefield data (1926–1977) and finds that the prior variance is $(0.0452)^2$, while Xia (01) uses CRSP data (1950–1997) and finds that the prior variance should be equal to $(0.0433)^2$. This gives us γ_0 to be roughly 0.002. However, one might argue that the variance of annual public equity returns might not correspond to its private equity counterpart, with the latter being potentially more volatile and harder to estimate. However, as we will see in the next section, changes in γ_0 do not change the results much quantitatively.

7). *Information cost parameter, θ*. I explore two ways of calibrating information cost changes. The first calibrates θ to observed changes in inequality, and then asks whether the implied expenditure share is 'reasonable' on a priori grounds. I assume that information cost in 1981 was originally high (θ_1). Then information cost suddenly decreases to θ_2 to produce higher inequality. I assume that the economy has since then transitioned to reach a new steady state in 2014, so



((A)) .Optimal information choice



((B)) Information choice vs. wealth

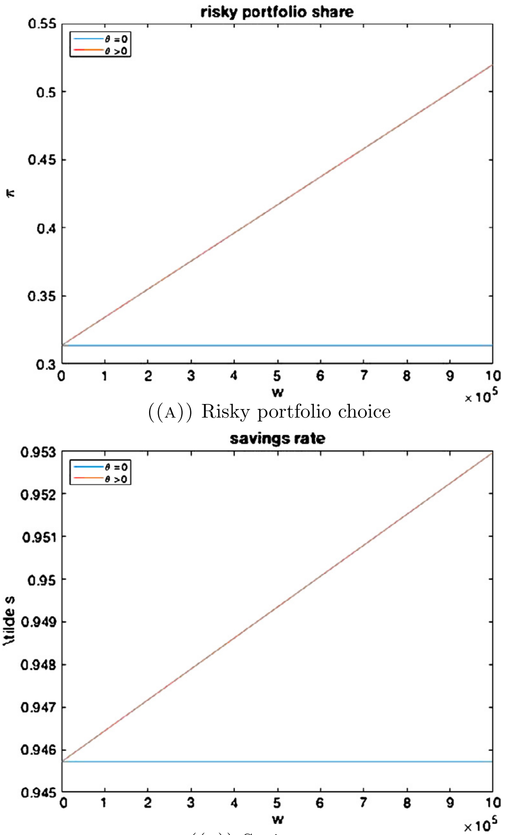
Fig. 6. Optimal information choice.

that we see the observed inequality in 2014. This strategy implies that $\theta_1 = 0.000016$ in 1981, while $\theta_2 = 0.00077$ in 2014. This then implies that for the median household, information costs declined by a factor of 24.3. It also implies that the share of wealth and income spent on information by the median and mean household was about 0.12% and 1.73%, respectively, in 1981. These seem like plausible values. However, in response to the large decline in information costs, the model predicts that these shares increased to 2.92% and 39.7%. These information expenditures are clearly too large. In fact, if we look at the mean rather than the median household, the shares are even bigger.[18]

I plot the information choice functions using these parameters in Fig. 6.

Fig. 6 (A) plots information choice as a function of wealth w and uncertainty γ in 2014. As discussed in Section 4, information demand increases with both uncertainty and wealth, with the

[18] In computing information expenditures, one needs to take into account the fact that average estimation uncertainty is also different with different information costs. To do this, I use $\bar{\gamma}^2_{2014}(\theta_2)$, which is the median person's estimation variance given the 2014 information cost parameter in computing the average demand and spending on information, similarly $\gamma^2_{1981}(\theta_1)$ for 1981. The same method is used when computing at the mean.



((A)) Risky portfolio choice

((B)) Savings rate

Fig. 7. Risky portfolio share and savings rate.

wealth effect being the dominant force. Fig. 6 (B) plots the response of information to wealth in both years, fixing uncertainty at its initial level. As one can see, in both years, capacity choice increases with wealth, but the response is much stronger in 2014.

Fig. 7 (A) illustrates the risky portfolio share as a function of wealth. We can see that not only do wealthy investors invest more in risky assets, but they also invest a higher *fraction* of wealth in risky assets. Although they might appear to be less risk averse than poorer investors, this behavior is driven by endogenous information choice. Fig. 7 (B) illustrates the savings rate as a function of wealth. From the graph, one can see that although the saving rate increases with wealth, which contributes to inequality, the increase is quite mild. Although the endogenous information has interesting qualitative effects on saving, its important quantitative implications are on portfolio choice.

Next, I simulate the economy using the above benchmark parameters to study the quantitative implications of endogenous information acquisition on inequality. 3000 agents are endowed with initial wealth $w = w_0$. Wealth and belief dynamics are simulated, with time step size discretized such that $dt = 0.5$. Agents experience birth and death with probability δdt at each time step. After simulating the economy for 500 years, a stationary distribution is achieved. I first use empirically plausible parameters from the benchmark calibration. Later, in section 8, I report

Table 2
Effect of $\Delta\theta$.

Mean growth rate change	Volatility change	Estimation variance decrease rate change
30.00%	6.54%	105%

how the results change in response to small perturbations of each parameter, holding fixed all other parameters at their benchmark values.

We should not be too surprised that the model implies implausibly large information expenditures when calibrated to match the *entire* increase in inequality. After all, the model completely ignores taxes, bequests, and labor income! Thus, it is useful to calibrate the model a second way, using data on actual information expenditures. Since wealthy households are more likely to delegate their wealth to wealth management firms, I utilize the Lipper TASS hedge fund dataset to estimate information cost reductions.[19] To do this, I interpret fees paid to fund managers as the cost of information. The data set spans from 1977 to 2005, and I compare changes in fee structure by splitting the sample into two periods, i.e.: 1977–1980 and 1981–2005. Most hedge funds charge two types of fees: management fees and incentive fees. Although the management fee is charged as a percentage of wealth annually, the incentive fee is not. Instead, the incentive fee is estimated as the percentage of profits taken by the fund manager. Thus, adjustment is needed to convert incentive fees measured as percentage of wealth. I adjust the estimates by multiplying the percentage of incentive fee with the average profit to fund size measured by net asset value ratio. Using the median as the average measure, I find that information costs declined by a factor of 10.88. At the mean, they declined by a factor of 15.00. Not surprisingly, this is smaller than the above value of 24.3, which was implied by the observed increase in inequality. The question therefore becomes – How much inequality does this measured cost reduction produce? To answer this question, I first set θ_1 at its 1981 level so that it matches the 1981 top 1% share. I then search θ_2 such that the implied information cost is reduced by a factor of 10.88. This generates a value of $\theta_2 = 0.000226$ in 2014. Using this estimate of θ_2, I re-run the Monte Carlo simulations. I find that the resulting top 1% wealth share becomes 32.67%, which implies an increase of 8.27 percentage points. Since top 1% share actually increased by 12.8 percentage points from 1981 to 2014, this suggests that endogenous information acquisition can explain about 64.6% of the rise in top 1% wealth share.

To identify the contributing factors behind the rise in inequality, I compute the cross sectional mean percentage changes of the drift and volatility of wealth, along with the rate at which estimation variance decreases in response to a change from $\theta_1(0.000016)$ to $\theta_2(0.00077)$.

As one can see, since the units here are percentage changes relative to their θ_1 values, the interpretation is that the rate at which average wealth grows increases by 30%, while the same measure for volatility increases by 6.54% (Table 2). That is, a reduction in the cost of information increases both the average growth rate and volatility of wealth, with the effect stronger on the growth rate. Finally, one can also see that when one moves from the θ_1 to θ_2 economy, estimation variance decreases roughly twice as fast.

[19] The Lipper TASS hedge fund dataset contains high quality panel data over several decades. It reports detailed information on hedge fund fee structure, return performance and various characteristics of over 7,500 actively reporting hedge funds and funds of hedge funds, plus over 11,000 graveyard funds. A detailed summary of the data set can be found in Appendix F (Table 5).

Table 3
Top 1% share.

θ	Economy A	Economy B
$\theta = 0.00077$	28.37%	37.20%
$\theta = 0.000226$	25.08%	32.67%

7. Inequality in two economies

The above section examined how much inequality would have increased from 1981 to 2014 if the only change in the US economy had been a reduction in the cost of information, assuming we live in a world of endogenous information acquisition. Note that a reduction in the cost of information has *two* effects: (1) an average effect whereby everyone buys more information than before, and (2) a distributional effect, which creates scale dependence. To isolate these two different effects, it is useful to decompose the mechanism in more detail. To do this, I consider the following two hypothetical economies:

Economy A: Exogenous information economy.
Economy B: Endogenous information economy.

In Economy A, everyone is constrained to have the same channel capacity, although this common capacity is allowed to respond to changes in information costs. In Economy B, people can choose to buy more capacity, and this choice will in general depend on wealth. Naturally, inequality is higher in Economy B than in Economy A. This is the central point of the paper. Still, it is useful to know how much of this increase is due to the pure effect of more (common) information, which encourages *everyone* to take more risks, and how much additional inequality is created by the feedback between wealth and information acquisition. To accomplish this decomposition we need to hold constant the average information capacity across the two economies. Recall that a typical agent's estimation uncertainty evolves according to

$$d\gamma(t) = -\left(\frac{\gamma(t)^2}{\tilde{\sigma}^2} + 2\kappa(t)^*\gamma(t)\right)dt \tag{7.33}$$

where $\kappa(t)^* = \frac{\theta w\gamma}{\beta^2\sigma^2}$ indicates the level of that common capacity. Substitute in, this implies that

$$d\gamma(t) = -\gamma(t)^2\left(\frac{1}{\tilde{\sigma}^2} + \frac{2\theta w(t)}{\beta^2\tilde{\sigma}^2}\right)dt \tag{7.34}$$

In Economy A, I feed in everyone's estimation uncertainty in the above way when they happen to be the median person, i.e.: $w(t) = w(median)$. I report the implied top 1% share in Table 3. I again report results for two different θ values. When $\theta = 0.00077$, Economy B perfectly matches the top 1% share in 2014 by construction. We can see that Economy A produces a top 1% share of 28.37%. Hence, the feedback between wealth and information accounts for about two-thirds of the total 12.4 percentage point increase in inequality. For comparison, I then use $\theta = 0.000226$, which is the empirically implied value from the hedge fund data. In this case, feedback accounts for even a larger share of the total increase, roughly 90%.

8. Sensitivity analysis

In this section, I study how changes in the parameters affect changes in inequality. Formally, I define Δ *top 1%* as the added top 1% wealth share when the economy moves from 1981 to 2014

Table 4
Sensitivity analysis.

μ	0.08	0.09	0.10	0.11	0.12	0.13	0.14
Δ top 1% share	1.18%	2.68%	6.4%	12.79%	19.51%	10.41%	3.26%
ρ	0.013	0.014	0.015	0.016	0.017	0.018	0.019
Δ top 1% share	15.14%	13.92%	12.79%	11.66%	10.72%	9.91%	9.22%
δ	0.024	0.0246	0.025	0.026	0.027	0.028	0.029
Δ top 1% share	13.49%	12.79%	12.35%	11.23%	10.37%	9.67%	8.93%
α	−0.40	−0.39	−0.38	−0.37	−0.36	−0.35	−0.34
Δ top 1% share	12.79%	12.92%	12.7%	12.4%	11.76%	10.95%	9.64%
γ_0	0.0008	0.001	0.0012	0.0014	0.0016	0.0018	0.002
Δ top 1% share	8.78%	9.69%	10.46%	11.15%	11.73%	12.27%	12.79%
$\tilde{\sigma}$	0.20	0.21	0.22	0.23	0.24	0.25	0.26
Δ top 1% share	13.25%	16.82%	18.53%	18.83%	17.54%	14.44%	12.79%

with the decline of information cost. The most important observation is that, for all parameter values, top 1% wealth share increases from 1981 to 2014. This is the essential message of the paper. Next, let's discuss results of each parameter changes. First, the added inequality increases when the mean return μ increases, up to a certain level. This is because higher mean return implies more wealth invested in the risky assets, which produces more demand for information, and amplifies the scale dependence effect coming from endogenous information acquisition. However, when μ increases even further, this effect is dampened. This is because at higher level of mean return, agents have already taken on lots of portfolio risk, and the economy has already experienced high inequality. For example, when $\mu = 0.14$, top 1% wealth share is already 92.14. Therefore, the added effect of endogenous information acquisition is smaller.

Second, when the rate of time preference, ρ, increases, the added inequality decreases. This is understandable, because a higher ρ implies less patience, thus reducing agents' incentive to save, invest and acquire information.

Next, an increase in the birth and death rate δ is similar to an increase in ρ, since they both increase the effective discount rate, thus dampening the amplification effect on inequality.

Further, let's turn to risk aversion. As argued in the calibration section, there are restrictions on how high risk aversion can go. However, one can still study sensitivity within certain range. As the Table 4 shows, higher risk aversion results in more added inequality. Endogenous information acquisition amplifies inequality more when agents have higher risk aversion. In other words, information acquisition leads more risk averse agents to be more confident, which leads them to take on more risk, thus generating more inequality.

I proceed by discussing the effects of γ_0. As one can see, an increase of initial uncertainty implies larger inequality increases. This comes directly from endogenous information acquisition. Higher initial uncertainty leads to higher demand of information, which amplifies inequality.

Last but not least, one needs to understand the volatility parameter. Since $(1 - \lambda)$ and σ always show up together, the only volatility that matters is $\tilde{\sigma}$. Thus it suffices to experiment around changes in $\tilde{\sigma}$. As one can see, the added inequality responds non-monotonically to an increase in $\tilde{\sigma}$. This could be surprising to some readers. This can be understood with the following two opposing forces: At the beginning, increases in volatility reduces agents exposure to risky portfolios. Therefore, endogenous information has a bigger role in building investors' confidence in risk taking. However, when volatility increases even further, the investment environment is so volatile that learning about the mean return is a much less attractive business, therefore the amplification effect of endogenous information acquisition is dampened.

9. Related literature

As discussed earlier, this paper is a direct descendent of work by Arrow (1987), Benhabib et al. (2016), and Turmuhambetova (2005). However, it can also be viewed as part of a larger literature that incorporates endogenous information into macroeconomics and finance. This literature is summarized by Veldkamp (2011). However, there are a couple of important differences between this recent literature and this paper. First, most of this literature focuses on either asset pricing or portfolio choice. Asset pricing is not an issue here, since projects are private and returns are (implicitly) generated by linear technologies. On the other hand, portfolio choice *is* important here, but rather than being an end in itself, I use heterogeneous portfolio choice as an input to the study of inequality. A second important difference is that this literature focuses on either rational inattention or noisy rational expectations equilibria. In addition to the previously discussed work of Kacperczyk et al. (2018), work by Van Nieuwerburgh and Veldkamp (2010) and Batchuluun et al. (2017) shows that rational inattention has interesting implications for both asset pricing and portfolio choice, generally in the direction of discouraging risky investment and diversification, and increasing risk premia. However, the traditional rational attention approach with *fixed* information capacity is perhaps not a great fit to financial markets. Rather than impose exogenous capacity constraints on agents, this paper views causality as going in the other direction. Here wealth determines an agents ability to *pay* for information processing. Besides being more plausible, viewing information as costly makes it easier to quantify. For example, Luo (2016) finds that capacity constraints must be implausibly tight before rational inattention begins to exert significant effects on portfolio choice.

Probably the most common approach to incorporating endogenous information into macroeconomics and finance is to follow Grossman and Stiglitz (1980) and study a noisy rational expectations environment. In fact, the extension by Verrecchia (1982) is quite similar to this paper, since he allows agents to buy more precise return signals. However, these models are typically static, and so are ill-suited to a quantitative analysis of inequality trends. They are static for a good reason. In these models, agents invest in a *common* asset, and information is private. The focus of the analysis in then on a difficult signal extraction problem from endogenously determined prices. Dynamic versions of these models are notoriously difficult to solve (Wang, 1993). This paper sidesteps this problem entirely, by assuming investment projects are linear and agent specific. Signal extraction takes place, but it only influences agent specific decisions, not market-clearing prices.[20]

Further, this paper is related to recent work on ‘financial literacy’ and asset market participation.[21] However, this literature focuses more on the left-tail of the distribution, as opposed to top wealth shares. My paper shows why endogenous information leads the rich to get richer. In contrast, the financial literacy literature helps explain why the poor stay poor.

[20] Ziegler (2012) reviews a closely related, but distinct, literature on ‘heterogeneous beliefs’ based on the assumption that agents have different priors. This literature might have interesting implications for inequality, but it begs the question of why agents have different priors.

[21] Examples include Lusardi et al. (2017), Vissing-Jorgensen (2003), Van Rooij et al. (2011), and Luo et al. (2017).

Finally, this paper is also related to Kasa and Lei (2018), who assume investors form robust portfolios in response to an increasingly ambiguous financial environment since the 1980s. They show robust portfolios also produce scale-dependent wealth dynamics. In reality, it is likely the case that both Knightian uncertainty has increased, while at the same time, information cost have declined. Household portfolio choices likely reflect a combination of both forces. In fact, Luo (2016) argues that both frictions are needed to explain observed portfolio behavior quantitatively. However, this paper differs from Kasa and Lei (2018) in important ways. The key mechanism here is through the second moment of beliefs (by building confidence), while Kasa and Lei (2018) works through the first moment (optimism/pessimism). By doing this, the information approach is consistent with *other* evidence we observe in the financial markets, and not just observed increase in inequality.[22] First, lower information costs stimulates the demand for information, which then leads to increased asset price 'informativeness'. Bai et al. (2016) show that price informativeness has indeed increased in recent decades, with the increase being stronger among stocks with greater institutional ownership. Second, with endogenous asset returns, the information mechanism in this paper would predict a decreased equity premium, since investment becomes less risky. Lettau et al. (2008) and Jagannathan et al. (2001) provide evidence in support of this. Finally, the information approach generates more inertia in portfolio behavior, which is consistent with evidence pointed out by Agnew et al. (2003) and Bilias et al. (2010). This is because with learning, new information take time to be reflected in policies. These are all features of the data that Kasa and Lei (2018) cannot generate.

10. Conclusion

Arguably the two dominant trends in the global economy in recent decades have been the explosive growth in information technology, which has reduced the cost of information, and the widening gap between rich and poor. This paper shows why these trends might be related. Perhaps surprisingly, the model here suggests that when individuals can buy information, reduced information costs can *increase* inequality. Increased access to information makes investment less risky for everyone, which encourages everyone to take greater risks, which encourages growth. However, it encourages wealthier individuals relatively more, and this exacerbates inequality. As Arrow (1987) surmised long ago, information naturally lends itself to increasing returns, and when increasing returns are combined with idiosyncratic shocks, wealth inequality emerges.

Some have argued that inequality isn't necessarily a bad thing. What about in my model? Here inequality emerges from a Pareto improving decrease in the cost of information. Since investment projects are idiosyncratic, there is no sense in which the rich are benefiting at the expense of the poor. If you want to make a case against inequality here, you must lay blame on the idiosyncratic nature of investment. The right way to address inequality in my model is not by taxing information, but by encouraging risk-sharing and the pooling of investment projects. Greenwood and Jovanovic (1990) develop a model in which there are fixed costs of risk sharing. Their model generates a Kuznets Curve in the distribution of wealth. At low levels of wealth, idiosyncratic shocks dominate, and inequality increases as the economy develops. Once it becomes economical to pay the fixed cost, however, risk-sharing emerges and inequality decreases. A similar dynamic would likely emerge here as well, if households were allowed to pool risks. However, since my model features finite lifetimes and no secular growth in per capita initial

[22] I thank an anonymous referee for clarifying the differences between this paper and Kasa and Lei (2018).

wealth, a fixed cost to risk sharing could actually exacerbate inequality, since only the wealthy would find it advantageous to pool risks.[23]

There are (at least) two important avenues for future work. First, this paper has focused on stationary distributions. However, recent work by Gabaix et al. (2016) suggests that the real challenge is to understand transition dynamics. Why has inequality grown so *rapidly*? Second, one might argue that the 1% get too much attention. Perhaps more important are the bottom 50%, who for a variety of reasons that are outside my model, do not participate in financial markets at all. It would be interesting to combine endogenous participation with both endogenous information and endogenously determined prices. It seems possible that encouraging more market participation, as is typically done in the 'financial literacy' literature, could actually backfire when information is endogenous, since it encourages entry by relatively uninformed investors, who in equilibrium end up losing money on average to the informed investors.

Appendix A. Value function with log utility

When $\alpha = \theta = 0$, the HJB equation becomes

$$\beta V = \max_{c,\pi} \left[\log(c) + [rw + \pi(\hat{\mu} - r)w - c]V_w + \frac{1}{2} V_{ww} \pi^2 \tilde{\sigma}^2 w^2 + \frac{1}{2} \frac{\gamma^2}{\tilde{\sigma}^2} V_{\hat{\mu}\hat{\mu}} - \frac{\gamma^2}{\tilde{\sigma}^2} V_\gamma \right] \tag{A.35}$$

With log utility, we know $c = \beta w$ and $\pi = (\hat{\mu} - r)/\tilde{\sigma}^2$ as in the Merton consumption portfolio choice problem. Comparing the above HJB equation with the HJB with general risk aversion coefficient α, the term $V_{w\hat{\mu}}$ is omitted, because the changes in estimate $\hat{\mu}$ doesn't affect a log agent's portfolio decision, thus has zero effect on his marginal utility of wealth. If we assume $\beta = r$,[24] we can guess the following wealth-separable functional form for the value function

$$V^{00} = A \log(w) + g^0(\hat{\mu}, \gamma) + K \tag{A.36}$$

Substituting this guess into the original HJB equation (A.35), we find that $A = 1/\beta$, $K = \log(\beta)/\beta$, and that $g^0(\hat{\mu}, \gamma)$ solves the following 2D PDE

$$g^0 = \frac{1}{2\beta^2} \frac{(\hat{\mu} - r)^2}{\tilde{\sigma}^2} + \frac{\gamma^2}{\beta \tilde{\sigma}^2} \left(\frac{1}{2} g^0_{\hat{\mu}\hat{\mu}} - g^0_\gamma \right) \tag{A.37}$$

Let's further guess that

$$g^0(\hat{\mu}, \gamma) = \frac{1}{2\beta^2} \frac{(\hat{\mu} - r)^2}{\tilde{\sigma}^2} + \tilde{g}^0(\hat{\mu}, \gamma) \tag{A.38}$$

This implies that $\tilde{g}$ satisfies the PDE

$$\tilde{g}^0_\gamma = \frac{1}{2} \tilde{g}^0_{\mu\mu} + \frac{1}{2\beta^2 \tilde{\sigma}^2} - \frac{\beta \tilde{\sigma}^2}{\gamma^2} \tilde{g}^0 \tag{A.39}$$

The boundary condition requires that $\tilde{g}^0(\hat{\mu}, 0) = 0$, because the value function converges to its no-learning counterpart as long as learning finishes (i.e.: $\gamma = 0$). We can eliminate the last term using the change of variables

[23] Favilukis (2013) shows that a fixed cost not only helps explain trends in inequality and market participation, it also helps explain asset price dynamics.

[24] This assumption is used throughout the rest of the paper.

$$\tilde{g}^0 = e^{\beta\tilde{\sigma}^2/\gamma} w(\hat{\mu}, \gamma) \tag{A.40}$$

which gives us the following heat equation for w

$$w_\gamma = \frac{1}{2} w_{\hat{\mu}\hat{\mu}} + \frac{1}{2\beta^2\tilde{\sigma}^2} e^{-\beta\tilde{\sigma}^2/\gamma} \tag{A.41}$$

The above equation resembles a heat equation with a source term, and has the following solution

$$\begin{aligned} w(\hat{\mu}, \gamma) &= \frac{1}{2\beta^2\tilde{\sigma}^2} \int_0^\gamma \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(\gamma - s)}} \exp\left(\frac{-(\hat{\mu} - y)^2}{2(\gamma - s)}\right) e^{-\beta\tilde{\sigma}^2/s} dy ds \\ &= \frac{1}{2\beta^2\tilde{\sigma}^2} \int_0^\gamma e^{-\beta\tilde{\sigma}^2/s} \left[\frac{1}{\sqrt{2\pi(\gamma - s)}} \exp\left(\frac{-(\hat{\mu} - y)^2}{2(\gamma - s)}\right) dy\right] ds \\ &= \frac{1}{2\beta^2\tilde{\sigma}^2} \int_0^\gamma e^{-\beta\tilde{\sigma}^2/s} ds \end{aligned} \tag{A.42}$$

Unwinding the change of variables gives us

$$g^0(\hat{\mu}, \gamma) = \frac{1}{2\beta^2} \frac{(\hat{\mu} - r)^2}{\tilde{\sigma}^2} + e^{\beta\tilde{\sigma}^2/\gamma} \left(\frac{1}{2\beta^2\tilde{\sigma}^2} \int_0^\gamma e^{-\beta\tilde{\sigma}^2/s} ds\right) \tag{A.43}$$

Finally, the value function follows

$$V^{00}(w, \hat{\mu}, \gamma) = \frac{1}{\beta} \log(w) + \frac{\log(\beta)}{\beta} + \frac{1}{2\beta^2\tilde{\sigma}^2} \left[(\hat{\mu} - r)^2 + e^{\beta\tilde{\sigma}^2/\gamma} \int_0^\gamma e^{-\beta\tilde{\sigma}^2/s} ds\right] \tag{A.44}$$

The last term can be expressed in terms of the exponential integral function, i.e.:

$$e^{\beta\tilde{\sigma}^2/\gamma} \int_0^\gamma e^{-\beta\tilde{\sigma}^2/s} ds = \beta\tilde{\sigma}^2 e^{\beta\tilde{\sigma}^2/\gamma} E_i(-\beta\tilde{\sigma}^2/\gamma) + \gamma \tag{A.45}$$

where the exponential integral function is defined as

$$E_i(-\beta\tilde{\sigma}^2/\gamma) = -\int_{\beta\tilde{\sigma}^2/\gamma}^{\infty} \frac{e^{-s}}{s} ds \tag{A.46}$$

By exploiting the relationship between exponential integral equation and incomplete gamma function, $E_i(-\beta\tilde{\sigma}^2/\gamma) = -\Gamma(0, \beta\tilde{\sigma}^2/\gamma)$, we could be ready for further policy function approximation.

Appendix B. Perturbation approximation

Proof. Let's posit the following value function, consisting of a first order expansion in θ around $\theta = 0$.

$$V(w,\hat{\mu},\gamma)=\frac{1}{\beta}\left[\frac{[we^{g(\hat{\mu},\gamma)}]^{\alpha}-1}{\alpha}+\theta(\frac{[we^{f(\hat{\mu},\gamma)}]^{\alpha+1}-1}{\alpha+1}+C)\right] \tag{B.47}$$

Further, let's do a first order expansion of the solution of $g(\hat{\mu},\gamma)$ and $f(\hat{\mu},\gamma)$ around $\alpha=0$,[25] i.e.:

$$g=g^0+\alpha g^1+O(\alpha^2) \tag{B.48}$$
$$f=f^0+\alpha f^1+O(\alpha^2) \tag{B.49}$$

Next, substitute this functional form into the policy functions. To an O(α, θ) approximation, one gets the approximate policy functions

$$c^*\approx\beta w[1-\alpha(g^0-\log\beta)-\theta we^{f^0}] \tag{B.50}$$
$$\pi^*\approx\frac{1}{\tilde{\sigma}^2}[(\hat{\mu}-r)+\alpha(g^0_{\mu}\gamma+(\hat{\mu}-r))+\theta we^{f^0}(\hat{\mu}-r+f^0_{\mu}\gamma)] \tag{B.51}$$
$$\kappa^*\approx\theta\gamma w(g^0_{\mu\mu}-2g^0_{\gamma}) \tag{B.52}$$

As it turns out, we do not need to fully solve functions $g(\hat{\mu},\gamma)$ and $f(\hat{\mu},\gamma)$ to get the policy functions. All we need to solve are g^0 and f^0. Note that

$$\lim_{\alpha\to 0,\theta\to 0}V(w,\hat{\mu},\gamma)=\frac{1}{\beta}(\log w+g^0(\hat{\mu},\gamma)) \tag{B.53}$$

which is the same $g^0(\hat{\mu},\gamma)$ in eq. (A.43).

One can further approximate the solution:

$$\begin{aligned}g^0&=\log\beta+\frac{1}{2\beta\tilde{\sigma}^2}[(\hat{\mu}-r)^2+\gamma-\beta\tilde{\sigma}^2e^{\beta\tilde{\sigma}^2/\gamma}\Gamma(0,\beta\tilde{\sigma}^2/\gamma)]\\&\approx\log\beta+\frac{1}{2\beta\tilde{\sigma}^2}[(\hat{\mu}-r)^2+\frac{\gamma^2}{\beta\tilde{\sigma}^2}]\end{aligned} \tag{B.54}$$

The approximation takes a second order Taylor expansion of g^0 around $\gamma=0$.

Next, let's solve for the $f^0(\hat{\mu},\gamma)$ term using perturbation. To do this, we first plug in policy functions (B.50), (B.51), (B.52) into the HJB equation, and get

$$\begin{aligned}\beta V=&\frac{V_w^{\frac{\alpha}{\alpha-1}}-1}{\alpha}+V_w[rw-(\hat{\mu}-r)w\frac{V_w(\hat{\mu}-r)+V_{w\mu}\gamma}{V_{ww}w\tilde{\sigma}^2}-V_w^{\frac{1}{\alpha-1}}-\frac{\theta}{2}(\frac{V_{\mu\mu}-2V_{\gamma}}{V_w}\gamma)^2]\\&+\frac{1}{2}V_{ww}\tilde{\sigma}^2\left(\frac{V_w(\hat{\mu}-r)+V_{w\mu}\gamma}{V_{ww}\tilde{\sigma}^2}\right)^2+(\frac{1}{2}V_{\mu\mu}-V_{\gamma})\gamma^2(\frac{1}{\tilde{\sigma}^2}+2\theta\frac{V_{\mu\mu}-2V_{\gamma}}{V_w})\\&-V_{w\mu}\gamma(\frac{V_w(\hat{\mu}-r)+V_{w\mu}\gamma}{V_{ww}\tilde{\sigma}^2})\end{aligned} \tag{B.55}$$

Note that eq. (B.47) can be rewritten into

$$V=V^0+\theta V^1 \tag{B.56}$$

where V^0 denotes the solution when $\theta=0$ and V^1 is the value function's derivative w.r.t. θ when $\theta=0$. I take derivative of eq. (B.55) w.r.t. θ, and evaluate it at $\theta=0$. After some algebra, we get

[25] Similar perturbation techniques are used in robust portfolio problems for a CRRA agent in Trojani and Vanini (2002).

$$\begin{aligned}\beta V^1 &= wV^1_w\left(\frac{(\hat\mu-r)^2}{(1-\alpha)\tilde\sigma^2}+\frac{\alpha g_\mu\gamma(\hat\mu-r)}{\tilde\sigma^2(1-\alpha)}-(\frac{1}{\beta})^{\frac{1}{\alpha-1}}e^{\frac{\alpha}{\alpha-1}g}\right)\\ &\quad + w^2V^1_{ww}\left(\frac{(\hat\mu-r)^2}{2\tilde\sigma^2(1-\alpha)^2}+\frac{\gamma^2\alpha^2g_\mu^2}{2\tilde\sigma^2(1-\alpha)^2}+\frac{\gamma(\hat\mu-r)\alpha g_\mu}{\tilde\sigma^2(1-\alpha)^2}\right)\\ &\quad + wV^1_{w\mu}\left(\frac{\gamma^2\alpha g_\mu}{\tilde\sigma^2(1-\alpha)}+\frac{\gamma(\hat\mu-r)}{\tilde\sigma^2(1-\alpha)}\right)+V^1_{\mu\mu}\frac{\gamma^2}{2\tilde\sigma^2}\\ &\quad - V^1_\gamma\frac{\gamma^2}{\tilde\sigma^2}+\frac{1}{2\beta}e^{\alpha g}\left((g_{\mu\mu}+\alpha g_\mu^2-2g_\gamma)\gamma\right)^2 w^{\alpha+1}\end{aligned} \tag{B.57}$$

Note that the non-homogeneous term is proportional to $w^{\alpha+1}$. Therefore, the previous conjecture of the functional form on V^1 is confirmed. Evaluating the above at $\alpha=0$ and simplifying, we get

$$\begin{aligned}\beta[we^{f^0}+C-1] &= we^{f^0}\left(\frac{(\hat\mu-r)^2}{\tilde\sigma^2}-\beta\right)+wf^0_\mu e^{f^0}\frac{\gamma(\hat\mu-r)}{\tilde\sigma^2}\\ &\quad + \frac{\gamma^2}{\tilde\sigma^2}\left(\frac{1}{2}e^{f^0}(f^0_{\mu\mu}+(f^0_\mu)^2)-f^0_\gamma e^{f^0}\right)w+\frac{\gamma^2}{2}w\left(g^0_{\mu\mu}-2g^0_\gamma\right)^2\end{aligned} \tag{B.58}$$

By matching the constant term and the we^{f^0} terms, we get

$$C=1 \tag{B.59}$$

and a PDE that $f^0(\hat\mu,\gamma)$ needs to satisfy

$$2\beta\tilde\sigma^2=(\hat\mu-r)^2+(\hat\mu-r)\gamma f^0_\mu+\frac{\gamma^2}{2}\left[f^0_{\mu\mu}+(f^0_\mu)^2-2f^0_\gamma+\tilde\sigma^2(g^0_{\mu\mu}-2g^0_\gamma)^2e^{-f^0}\right] \tag{B.60}$$

To an $O(\gamma^2)$ approximation, we can get

$$f^0(\hat\mu,\gamma)=\frac{2\beta\tilde\sigma^2}{\gamma}\log(\hat\mu-r)-\frac{(\hat\mu-r)^2}{2\gamma}\quad\square \tag{B.61}$$

Appendix C. Proof of Lemma 5.1

Proof. Combining policy functions and individual wealth dynamics (3.14), one can write down each agents' perceived law of motion for wealth as

$$dw=\hat\mu(w)wdt+\sigma(w)w\hat{d}B \tag{C.62}$$

Note that this is the agent's perceived law of motion of his wealth. To study wealth distribution, it is useful to relate agents' perceived law of motion to actual law of motion using, again, the Girsanov theorem. Recall that $\hat{d}B=dB-\frac{(\hat\mu-\mu)}{\tilde\sigma}dt$. Therefore,

$$\mu(w)=a(\hat\mu,\gamma)+\theta wb(\hat\mu,\gamma) \tag{C.63}$$

$$\sigma(w)=c(\hat\mu,\gamma)+\theta wd(\hat\mu,\gamma) \tag{C.64}$$

s.t.:

$$a = \frac{(\mu - r)}{\tilde{\sigma}^2}\left[(1+\alpha)(\hat{\mu} - r) + \alpha\frac{(\hat{\mu}-r)\gamma}{\beta\tilde{\sigma}^2}\right] + \frac{\alpha}{2\tilde{\sigma}^2}\left[(\hat{\mu}-r)^2 + \frac{\gamma^2}{\beta\tilde{\sigma}^2}\right] \tag{C.65}$$

$$b = \exp(f^0(\hat{\mu},\gamma))(r + \frac{(\hat{\mu}-r)}{\tilde{\sigma}^2}((\hat{\mu}-r) + f^0_{\hat{\mu}}\gamma)) \tag{C.66}$$

$$c = \frac{1}{\tilde{\sigma}}\left[(1+\alpha)(\hat{\mu}-r) + \alpha\frac{(\hat{\mu}-r)\gamma}{\beta\tilde{\sigma}^2}\right] \tag{C.67}$$

$$\hat{d} = \frac{\exp(f^0(\hat{\mu},\gamma)}{\tilde{\sigma}}(\hat{\mu} - r + f^0_{\hat{\mu}}\gamma) \tag{C.68}$$

where $f^0(\hat{\mu},\gamma) = \frac{2\beta\tilde{\sigma}^2}{\gamma}\log(\hat{\mu}-r) - \frac{(\hat{\mu}-r)^2}{2\gamma}$. □

Appendix D. Proof of Proposition 2

Proof. Let $x = \log w$. Applying Ito's lemma, one can then write the log wealth dynamics as

$$dx = [a + \theta b e^x - \frac{1}{2}(c + \theta\hat{d}e^x)^2]dt + (c + \theta\hat{d}e^x)dB \tag{D.69}$$

Dropping the second order term of θ, we can simplify the law of motion to

$$dx = [(a - \frac{1}{2}c^2) + \theta e^x(b - c\hat{d})]dt + [c + \theta\hat{d}e^x]dB \tag{D.70}$$

Therefore, the system can be written as

$$\begin{bmatrix} dx \\ d\hat{\mu} \\ d\gamma \end{bmatrix} = \begin{bmatrix} a - \frac{1}{2}c^2 + \theta(b - c\hat{d})e^x \\ \frac{\gamma(\mu-\hat{\mu})}{\beta^2\tilde{\sigma}^2}[\beta^2 + 2\theta e^x] \\ -\gamma^2(\frac{1}{\tilde{\sigma}^2} + \frac{2\theta e^x}{\beta^2\tilde{\sigma}^2}) \end{bmatrix} dt + \begin{bmatrix} c + \theta\hat{d}e^x & 0 & 0 \\ \frac{\gamma}{\tilde{\sigma}} & \frac{\sqrt{2\theta e^x}\gamma}{\beta\tilde{\sigma}} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} dB \\ dB_y \\ 0 \end{bmatrix} \tag{D.71}$$

The above state dynamics can be written in the matrix form

$$dX = G(X)dt + \Sigma(X)d\bar{B} \tag{D.72}$$

The stationary KFP equation is thus written

$$0 = -\sum_{i=1}^{3}\frac{\partial}{\partial X_i}[G_i(X)f] + \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\frac{\partial^2}{\partial X_i\partial X_j}\left[\left(\sigma(X)\sigma^T(X)\right)_{i,j} f\right] - \delta f + \delta\zeta(X - X_0) \tag{D.73}$$

where $X = [x, \hat{\mu}, \gamma]^T$ represents the vector of the state variables, $X_0 = [x_0, \mu_0, \gamma_0]^T$ represents initial endowment and beliefs at the mass point, and that $\zeta(.)$ represents the Dirac delta function. □

Appendix E. Analytical approximation

The time-scale separation approach is closely related to the 'mean ODE' method pioneered by Marcet and Sargent (1989) in the macroeconomic learning literature. Although eq. (5.31) is *linear*, the coefficients are complicated functions of the state, which rules out a transform function approach. The key to making analytical headway here is to notice that the bottom two equations in (system), describing the evolution of beliefs, are scaled by the conditional variance,

γ. Moreover, we know γ is monotonically decreasing, so that if γ_0 is small, γ remains small. This means that we can use time-scale separation methods to effectively decouple the evolution of wealth from the evolution of beliefs. Since x is 'fast' and $(\hat{\mu}, \gamma)$ are 'slow', we first characterize the x dynamics while keeping beliefs fixed. Next, we use the implied stationary distribution of x (which depends on $(\hat{\mu}, \gamma)$) to average over the wealth dependence in the equations describing beliefs. Finally, we can substitute the resulting stationary distribution of beliefs back into the conditional distribution for x to get the stationary marginal distribution of wealth.[26]

In other words, remember the solution of the (steady state) KFP equation gives us the stationary *joint* distribution of wealth and beliefs. Our averaging solution strategy is based on decomposing this joint distribution into the product of a conditional and a marginal,

$$f(x, \hat{\mu}, \gamma) = f_1(x|\hat{\mu}, \gamma) f_2(\hat{\mu}, \gamma)$$

The conditional density f_1 is described by a nonlinear ODE, which can be solved using an $O(\theta)$ perturbation approximation. In general, the marginal distribution of beliefs would be difficult to handle, since beliefs interact with wealth. However, when beliefs are relatively 'slow', we can use averaging to obtain a simpler 2-dimensional PDE for $(\hat{\mu}, \gamma)$ which can again be solved relatively easily using an $O(\theta)$ perturbation approximation.

Yet another way of describing this solution strategy can be seen by looking directly at the KFP equation in (5.31). Notice that when $\gamma = 0$ all terms involving partials of $\hat{\mu}$ and γ disappear. Hence, to an $O(\gamma)$ approximation, we can view $\hat{\mu}$ and γ as constants. Hence, we can describe our time-scale separation strategy as an $O(\gamma)$ perturbation approximation,

Proposition 4. *To an $O(\gamma, \theta)$ approximation, the stationary distribution of $x (> 0)$ is*

$$f_1(x|\hat{\mu}, \gamma) = A_0(\hat{\mu}, \gamma) e^{\phi(\hat{\mu}, \gamma) x} + \theta A_1(\hat{\mu}, \gamma) e^{[1+\phi(\hat{\mu}, \gamma)]x} \tag{E.74}$$

where the normalizing constants (A_0, A_1) are chosen to maintain continuity at $x = 0$ and ensure adding up, $\int f_1 = 1$. The exponent $\phi(\hat{\mu}, \gamma)$ is the negative root of the quadratic $\frac{1}{2} c^2 \phi^2 - \hat{a}\phi - \delta = 0$, where $\hat{a} \equiv a - \frac{1}{2} c^2$.

Proof. We can write $f_1(x|\hat{\mu}, \gamma)$ into

$$f_1(x|\hat{\mu}, \gamma) = f^0(x|\hat{\mu}, \gamma) + \theta f^1(x|\hat{\mu}, \gamma) \tag{E.75}$$

where $f^0(x|\hat{\mu}, \gamma)$ solves the following quadratic function

$$\delta f^0 = -\hat{a} f_x^0 + \frac{1}{2} c^2 f_{xx}^0 + \delta \zeta(x_0) \tag{E.76}$$

where $\zeta(x_0)$ is the Dirac delta function at the mass point x_0 where $x_0 = 0$. Therefore, the solution becomes

$$f^0(x|\hat{\mu}, \gamma) = \begin{cases} A_0 e^{\phi_1 x}; & x > 0 \\ \tilde{A}_0 e^{\phi_2 x}; & x < 0 \end{cases}$$

where coefficients A_0 and $\tilde{A}_0$ are determined by integrating the distribution to one, and the continuity condition at $x_0 = 0$.

---

[26] Pavliotis and Stuart (2008) provide a good textbook description of averaging and time-scale separation methods.

Next, combining eq. (5.31) and the solution of $f^0(x|\hat{\mu},\gamma)$, we know that the PDE for f^1 is

$$\begin{aligned}\delta f^1 = -\hat{a} f^1_x + \frac{1}{2}c^2 f^1_{xx} + e^x\left[(c\hat{d}-\hat{b})f^0 + (2c\hat{d}-\hat{b})f^0_x + c\hat{d} f^0_{xx}\right] \\ + e^{x(\phi+1)}\left[(c\hat{d}-\hat{b}) + \phi(2c\hat{d}-\hat{b}) + \phi(\phi-1)c\hat{d}\right] + \delta\zeta(0)\end{aligned} \tag{E.77}$$

where $\hat{a} \equiv a - \frac{1}{2}c^2$ and $\hat{b} \equiv b - c\hat{d}$. The homogeneous part of the solution is the same as for f^0. The particular solution is $A_1 e^{(\phi+1)x}$ where A_1 must solve

$$\delta A_1 = -\hat{a}(\phi+1)A_1 + \frac{1}{2}c^2(\phi+1)\phi A_1 + A_0\left[(c\hat{d}-\hat{b}) + \phi(2c\hat{d}-\hat{b}) + \phi(\phi-1)c\hat{d}\right] \tag{E.78}$$

This gives

$$A_1 = \frac{A_0\left[(c\hat{d}-\hat{b}) + \phi(2c\hat{d}-\hat{b}) + \phi(\phi-1)c\hat{d}\right]}{\delta + \hat{a}(\phi+1) - \frac{1}{2}c^2(\phi+1)\phi} \quad \square \tag{E.79}$$

Several points are worth noting here. First, since we are only interested in top wealth shares, this result only characterizes the distribution for $x > 0$. However, a completely analogous and symmetric result applies for $x < 0$. Second, remember that $(\hat{\mu},\gamma)$ are being treated as fixed parameters. The notation here reminds us that the Pareto exponents and normalizing constants depend on these slowly varying parameters. To fully characterize the distribution over long time-scales, we need the distribution of beliefs. Third, notice that when learning is exogenous ($\theta = 0$) the distribution is exactly Pareto, with an exponent that solves the same sort of quadratic that applies in the geometric Brownian motion case. Still, this exponent depends on $(\hat{\mu},\gamma)$ since beliefs influence the portfolio allocation, which influences the drift and volatility of wealth. Finally, and most importantly, notice that when learning is endogenous ($\theta > 0$), the distribution of wealth is only approximately Pareto, with an extra component that dies out more slowly. Asymptotically, for large x, this piece will dominate top wealth shares, and we obtain

Corollary E.1. *To an $O(\gamma,\theta)$ approximation, endogenous information increases top wealth shares.*

Proof. This follows simply from the fact that $|\phi+1| < |\phi|$. $\square$

We can now turn to the distribution of beliefs. When doing this, it is convenient to define the following linear operator

$$\mathcal{L}[h] = \left(\frac{3\gamma}{\tilde{\sigma}^2} - \delta\right)h + \frac{\gamma}{\tilde{\sigma}^2}(\hat{\mu}-\mu)h_{\hat{\mu}} + \frac{\gamma^2}{\tilde{\sigma}}h_\gamma + \frac{1}{2}\frac{\gamma^2}{\tilde{\sigma}^2}h_{\hat{\mu}\hat{\mu}} \tag{E.80}$$

To an engineer, this is a 'diffusion operator', with source and convection terms. This is useful for us, since learning and diffusion are inverses of each other. In particular, when $\theta = 0$, the evolution of beliefs satisfies the PDE, $h_t = \mathcal{L}[h]$, plus appropriate delta functions capturing the prior. Stationary beliefs are then the solution of $\mathcal{L}[h] = 0$.

Lemma E.2. *When $\theta = 0$ the stationary cross-sectional distribution of beliefs is*

$$h(\hat{\mu}, \gamma) \equiv N^{IG}(\hat{\mu}, \gamma) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}\gamma^3} \exp\left[-\frac{(\hat{\mu}-\mu)^2}{2\gamma} - \frac{\delta\tilde{\sigma}^2}{\gamma}\right] \tag{E.81}$$

Proof. Take derivatives and substitute into eq. (E.80). □

The stationary joint density of beliefs is the product of a normal distribution (i.e., a 'heat kernel'), and an inverse gamma density. The notation N^{IG} is used as a mnemonic for this. It is the *product* of two separate densities because when $\theta = 0$, the conditional variance for an individual agent is deterministic, and evolves independently from the conditional mean $\hat{\mu}$. The stationary cross-sectional conditional variance reflects the balancing of individual learning with exponential lifetimes. In terms of γ, this generates an inverse-gamma density.

Of course, when $\theta \neq 0$, wealth influences beliefs, so in principle we need to account for this. However, since wealth evolves on a faster time-scale, we can simply average out this dependence.

Proposition 5. *To an $O(\gamma, \theta)$ approximation, the stationary cross-sectional distribution of beliefs is*

$$f_2(\hat{\mu}, \gamma) = B_0 N^{IG}(\hat{\mu}, \gamma) + \theta B_1 G(\hat{\mu}, \gamma) \tag{E.82}$$

where (B_0, B_1) are normalizing constants, and $G(\hat{\mu}, \gamma)$ is a function defined in eqn. (E.94) in the following proof.

Proof. The "fast" dynamics of x can be averaged out by

$$\begin{aligned}\psi(\hat{\mu}, \gamma) \equiv Ee^x &= \int_0^\infty e^x \left[A_0 e^{\phi x} + \theta A_1 e^{(\phi+1)x}\right] dx \\ &= -\frac{1}{\phi+1} A_0 - \theta A_1 \frac{1}{\phi+2}\end{aligned} \tag{E.83}$$

Since $O(\theta)$ term is of second order, we can approximate the above as

$$\psi(\hat{\mu}, \gamma) \approx \frac{-A_0}{\phi+1} \tag{E.84}$$

Let $q(\hat{\mu}, \gamma) = f_2(\hat{\mu}, \gamma)$. Again, let's look for a first order perturbation solution for $(\hat{\mu}, \gamma)$, i.e.:

$$q(\hat{\mu}, \gamma) = q^0 + \theta q^1 \tag{E.85}$$

where q^1 solves

$$\frac{2}{\beta^2}\psi \mathcal{L}[q^1] + \frac{2}{\beta^2}\left[\frac{(\hat{\mu}-\mu)}{\sigma^2}\psi_{\hat{\mu}} + \frac{\gamma^2}{\sigma^2}\psi_\gamma\right] q^0 = 0 \tag{E.86}$$

This is equivalent to

$$\mathcal{L}[g^1] + \left[\frac{(\hat{\mu}-\mu)}{\sigma^2}\frac{\psi_{\hat{\mu}}}{\psi} + \frac{\gamma^2}{\sigma^2}\frac{\psi_\gamma}{\psi}\right] N(\hat{\mu}, \gamma) = 0 \tag{E.87}$$

which could be written into the following form:

$$\mathcal{L}[q^1] + Q(\hat{\mu}, \gamma) = 0 \tag{E.88}$$

where $Q(.)$ can be interpreted as a "source" term in a standard diffusion problem. Although this is an operator equation, we can exploit an analogy from the theory of ODEs. Suppose we have the non-homogeneous ODE

$$u_t + Au = f(t), \qquad u(0) = u_0 \tag{E.89}$$

We know the solution is

$$u(t) = e^{-tA}u_0 + \int_0^t e^{(s-t)A} f(s)ds \tag{E.90}$$

Now suppose we have the diffusion problem

$$u_t(x,t) = \mathcal{L}[u] + \lambda(x,t) \tag{E.91}$$

where $\mathcal{L}[.]$ is some diffusion operator, and $\lambda(.)$ is a source. It turns out, the solution is analogous to the ODE case

$$\begin{aligned} u(x,t) &= \mathcal{L}(t)u(x,0) + \int_0^t \mathcal{L}(t-s)\lambda(s)ds \\ &= \int_{-\infty}^{\infty} N(x-y,t)u_0(y)dy + \int_0^t \int_{-\infty}^{\infty} N(x-y,t-s)\lambda(y,s)dyds \end{aligned} \tag{E.92}$$

where in the standard diffusion problem

$$N(x,t) = \frac{1}{2\sqrt{\pi k t}} e^{\frac{-x^2}{4kt}} \tag{E.93}$$

is called the heat kernel, and u_0 is the delta function at $\hat{\mu}_0, \gamma_0$. Therefore, the $G(.)$ function defined in Proposition 5 given in abstract terms is

$$G(\hat{\mu}, \gamma) = \int_0^{\infty}\int_0^{\infty} N(\hat{\mu} - y, \gamma - s)Q(y,s)ds \quad \square \tag{E.94}$$

The process of averaging over x is reflected in the properties of the $G(\hat{\mu}, \gamma)$ function. If we let $\psi(\hat{\mu}, \gamma) = Ee^x$, where expectations are computed with respect to the stationary conditional distribution derived in Proposition 3, then G depends on $\psi_{\hat{\mu}}$ and ψ_{γ}. These determine how changes in beliefs affect mean wealth.

Finally, by combining the conditional distribution of x derived in Proposition 4 with the marginal distribution for beliefs derived in Proposition 5, we get the following characterization of the marginal distribution of (log) wealth

Proposition 6. *To an $O(\gamma, \theta)$ approximation, the stationary cross-sectional marginal distribution of (log) wealth is (for $x > 0$)*

$$\Lambda(x) = \int_0^{\gamma_0} \int_{-\infty}^{\infty} f_1(x|\hat{\mu}, \gamma) f_2(\hat{\mu}, \gamma) d\hat{\mu} d\gamma \tag{E.95}$$

$$= \int_0^{\gamma_0} \int_{-\infty}^{\infty} \Big\{ A_0 B_0 e^{\phi x} \cdot N^{IG}(\hat{\mu}, \gamma)$$

$$+ \theta \Big[A_0 B_1 e^{\phi x} \cdot G(\hat{\mu}, \gamma) + A_1 B_0 e^{(\phi+1)x} \cdot N^{IG}(\hat{\mu}, \gamma) \Big] \Big\} d\hat{\mu} d\gamma$$

where for notational convenience the dependence of ϕ and (A_i, B_i) on $(\hat{\mu}, \gamma)$ has been suppressed.

Proof. By direct substitution. □

As long as we confine our attention to top wealth shares, we can focus on the x coefficients in the integrand. They are nonlinear functions of $(\hat{\mu}, \gamma)$. Existence of a well defined distribution requires $\phi < -1$. By inspection, the dominating term will be the last one. We can approximate this around $(\hat{\mu}, \gamma) = (\mu^*, \gamma^*)$, with the RHS being its common capacity learning counterpart.

$$(\phi + 1)x \approx \phi_0 x + [1 + \phi_{\hat{\mu}}(\hat{\mu} - \mu^*) + .5\phi_{\hat{\mu}\hat{\mu}}(\hat{\mu} - \mu^*)^2 + \phi_\gamma(\gamma - \gamma^*)]x \tag{E.96}$$

Note that ϕ_0 is the coefficient in an economy with common information capacity. Hence, to assess the impact of endogenous information on top wealth shares, we can focus on the term in brackets. It turns out that $\phi_{\hat{\mu}\hat{\mu}} < 0$, so that the bracket term is a concave function of $\hat{\mu}$, with a unique maximum. As $x \to \infty$, this maximum point will dominate the value of the integral. Hence, for large x we can just focus on it. This can be formalized using ‘Laplace’s Method’. Applying Laplace’s Method yields,

Proposition 7. *For large x, the (right) tail Pareto exponent with endogenous information, ϕ_L, is approximately*

$$\phi_L \approx \phi_0 + (1 + \frac{\phi_{\hat{\mu}}^2}{2|\phi_{\hat{\mu}\hat{\mu}}|}) \quad \Rightarrow \quad |\phi_L| < |\phi_0|$$

Hence, endogenous information increase top wealth shares.

Proof. The integral is approximated using Laplace’ method, which is used to approximate the following functional form

$$\int_a^b h(s) e^{Mf(s)} ds \approx \sqrt{\frac{2\pi}{Mf''(s_0)}} h(s_0) e^{Mf(s_0)} \tag{E.97}$$

as $M \to \infty$, where $f(s) < 0$ is maximized at s_0.

The dominant piece of the marginal density

$$\int_0^{\gamma_0} \int_{-\infty}^{\infty} A_1(\hat{\mu}, \gamma) b(\hat{\mu}, \gamma) e^{(\phi+1)x} d\hat{\mu} d\gamma \tag{E.98}$$

can thus be approximated using this method. Using the approximation

$$(\phi + 1)x \approx \phi_0 x + [1 + \phi_{\hat{\mu}}(\hat{\mu} - \mu^*) + .5\phi_{\hat{\mu}\hat{\mu}}(\hat{\mu} - \mu^*)^2 + \phi_\gamma(\gamma - \gamma^*)]x \tag{E.99}$$

Since $\phi_\gamma < 0$, we know that $(\hat{\mu}, \gamma) = (-\phi_{\hat{\mu}}/\phi_{\hat{\mu}\hat{\mu}} + \mu^*, \gamma^*)$ maximizes the exponent. The exponent evaluated at the maximum thus becomes

$$(\phi + 1)x \approx \phi_0 x + (1 + \frac{\phi_{\hat{\mu}}^2}{2|\phi_{\hat{\mu}\hat{\mu}}|})x \tag{E.100}$$

Therefore,

$$\phi_L \approx \phi_0 + (1 + \frac{\phi_{\hat{\mu}}^2}{2|\phi_{\hat{\mu}\hat{\mu}}|}) \tag{E.101}$$

which implies that $|\phi_L| < |\phi_0|$. □

Appendix F. Summary of Lipper TASS hedge fund dataset

Table 5
Lipper TASS.

Statistic	N	Mean	St. Dev.	Min	Max
Year Month	741,993	2,005.405	6.212	1,977.100	2,015.100
incentive fee	741,993	15.506	7.573	0.000	35.000
management fee	741,993	0.537	1.114	0	8
monthly rate of return	741,993	0.865	6.206	−89.922	934.480
net asset value	741,993	9,163.227	160,497.000	0.861	6,003,175.000
lock up period	741,993	3.799	8.320	0	90
minimum investment size	741,683	1,500,857.000	8,086,231.000	0	100,000,000
high water mark	741,993	0.675	0.468	0	1
average leverage	712,410	63.061	203.639	0.000	2,000.000

References

Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., Moll, B., 2017. Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach. Discussion paper. National Bureau of Economic Research.
Agnew, J., Balduzzi, P., Sunden, A., 2003. Portfolio choice and trading in a large 401 (k) plan. Am. Econ. Rev. 93 (1), 193–215.
Aiyagari, S.R., 1994. Uninsured idiosyncratic risk and aggregate saving. Q. J. Econ. 109, 659–684.
Andrei, D., Hasler, M., 2014. Optimal asset and attention allocation. Available at SSRN 2516984.
Angeletos, G.-M., 2007. Uninsured idiosyncratic investment risk and aggregate saving. Rev. Econ. Dyn. 10, 1–30.
Angeletos, G.-M., Panousi, V., 2011. Financial Integration, Entrepreneurial Risk and Global Imbalances. Discussion paper. National Bureau of Economic Research.
Aoki, S., Nirei, M., 2017. Zipf's law, Pareto's law, and the evolution of top incomes in the U.S. Am. Econ. J. Macroecon. 9, 36–71.
Arrow, K.J., 1987. The demand for information and the distribution of income. Probab. Eng. Inf. Sci. 1 (1), 3–13.
Bai, J., Philippon, T., Savov, A., 2016. Have financial markets become more informative? J. Financ. Econ. 122 (3), 625–654.
Batchuluun, A., Luo, Y., Young, E., 2017. Portfolio Choice with Information-Processing Limits. working paper. Hong Kong University.
Benhabib, J., Bisin, A., Luo, M., 2017. Earnings inequality and other determinants of wealth inequality. Am. Econ. Rev. Pap. Proc. 107, 593–597.
Benhabib, J., Bisin, A., Zhu, S., 2011. The distribution of wealth and fiscal policy in economies with finitely-lived agents. Econometrica 79, 123–157.

Benhabib, J., Bisin, A., Zhu, S., 2016. The distribution of wealth in the Blanchard-Yaari model. Macroecon. Dyn. 20, 466–481.
Bilias, Y., Georgarakos, D., Haliassos, M., 2010. Portfolio inertia and stock market fluctuations. J. Money Credit Bank. 42 (4), 715–742.
Brennan, M.J., 1998. The role of learning in dynamic portfolio decisions. Rev. Finance 1 (3), 295–306.
Cao, D., Luo, W., 2017. Persistent heterogeneous returns and top end wealth inequality. Rev. Econ. Dyn. 26, 301–326.
Carroll, C.D., 2002. Portfolios of the rich. In: Guiso, L., Haliassos, M., Jappelli, T. (Eds.), Household Portfolios. MIT Press, pp. 389–430.
Champernowne, D., 1953. A model of income distribution. Econ. J. 63, 318–351.
Dynan, K.E., Skinner, J., Zeldes, S.P., 2004. Do the rich save more? J. Polit. Econ. 112, 397–444.
Ehling, P., Graniero, A., Heyerdahl-Larsen, C., 2016. Asset Prices and Portfolio Choice with Learning from Experience.
Fagereng, A., Guiso, L., Malacrino, D., Pistaferri, L., 2016. Heterogeneity and Persistence in Returns to Wealth. Discussion paper. National Bureau of Economic Research.
Favilukis, J., 2013. Inequality, stock market participation, and the equity premium. J. Financ. Econ. 107 (3), 740–759.
Gabaix, X., Lasry, J.M., Lions, P.L., Moll, B., 2016. The dynamics of inequality. Econometrica 84.
Gennotte, G., 1986. Optimal portfolio choice under incomplete information. J. Finance 41 (3), 733–746.
Greenwood, J., Jovanovic, B., 1990. Financial development, growth, and the distribution of income. J. Polit. Econ. 98 (5, Part 1), 1076–1107.
Grossman, S.J., Stiglitz, J.E., 1980. On the impossibility of informationally efficient markets. Am. Econ. Rev., 393–408.
Jagannathan, R., McGrattan, E.R., Scherbina, A., 2001. The Declining US Equity Premium. Discussion paper. National Bureau of Economic Research.
Kacperczyk, M., Nosal, J., Stevens, L., 2018. Investor sophistication and capital income inequality. J. Monet. Econ. https://doi.org/10.1016/j.jmoneco.2018.11.002.
Kartashova, K., 2014. Private equity premium puzzle revisited. Am. Econ. Rev. 104 (10), 3297–3334.
Kasa, K., Lei, X., 2018. Risk, uncertainty, and the dynamics of inequality. J. Monet. Econ. 94, 60–78.
Kopczuk, W., 2015. What do we know about the evolution of top wealth shares in the United States? J. Econ. Perspect. 29, 47–66.
Lettau, M., Ludvigson, S.C., Wachter, J.A., 2008. The declining equity premium: what role does macroeconomic risk play? Rev. Financ. Stud. 21 (4), 1653–1687.
Liptser, R.S., Shiryaev, A.N., 2000. Statistics of Random Processes II: Applications, vol. 2. Springer.
Luo, Y., 2016. Robustly strategic consumption–portfolio rules with informational frictions. Manag. Sci 63 (12), 4158–4174.
Luo, Y., Nie, J., Wang, H., 2017. Ignorance, Uncertainty, and Strategic Consumption-Portfolio Decisions. Working paper. Hong Kong University.
Lusardi, A., Michaud, P.-C., Mitchell, O.S., 2017. Optimal financial knowledge and wealth inequality. J. Polit. Econ. 125 (2), 431–477.
Malmendier, U., Nagel, S., 2016. Learning from inflation experiences. Q. J. Econ. 131, 53–87.
Marcet, A., Sargent, T.J., 1989. Convergence of least-squares learning mechanisms in self-referential linear stochastic models. J. Econ. Theory 48, 337–368.
Merton, R.C., 1980. On estimating the expected return on the market: an exploratory investigation. J. Financ. Econ. 8 (4), 323–361.
Moskowitz, T.J., Vissing-Jorgensen, A., et al., 2002. The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle? Discussion paper. National Bureau of Economic Research, Inc.
Nakov, A., Nuño, G., 2015. Learning from experience in the stock market. J. Econ. Dyn. Control 52, 224–239.
Øksendal, B., 2003. Stochastic Differential Equations. Springer.
Pavliotis, G., Stuart, A., 2008. Multiscale Methods: Averaging and Homogenization. Springer Science & Business Media.
Peress, J., 2004. Wealth, information acquisition, and portfolio choice. Rev. Financ. Stud. 17 (3), 879–914.
Piketty, T., 2014. Capital in the Twenty-First Century. Harvard University Press.
Quadrini, V., 2009. Entrepreneurship in macroeconomics. Ann. Finance 5 (3), 295–311.
Reed, W.J., 2001. The Pareto, Zipf and other power laws. Econ. Lett. 74, 15–19.
Saez, E., Zucman, G., 2016. Wealth inequality in the United States since 1913: evidence from capitalized income tax data. Q. J. Econ. 131 (2), 519–578.
Simon, H.A., 1955. On a class of skew distribution functions. Biometrika 44, 425–440.
Sims, C.A., 2003. Implications of rational inattention. J. Monet. Econ. 50 (3), 665–690.
Toda, A.A., 2014. Incomplete markets, dynamics and cross-sectional distributions. J. Econ. Theory 154, 310–348.
Trojani, F., Vanini, P., 2002. A review of perturbative approaches for robust optimal portfolio problems. In: Computational Methods in Decision-Making, Economics and Finance, vol. 74, p. 109.

Turmuhambetova, G., 2005. Decision Making in an Economy with Endogenous Information.
Van Nieuwerburgh, S., Veldkamp, L., 2010. Information acquisition and under-diversification. Rev. Econ. Stud. 77 (2), 779–805.
Van Rooij, M., Lusardi, A., Alessie, R., 2011. Financial literacy and stock market participation. J. Financ. Econ. 101 (2), 449–472.
Veldkamp, L.L., 2011. Information Choice in Macroeconomics and Finance. Princeton University Press.
Verrecchia, R.E., 1982. Information acquisition in a noisy rational expectations economy. Econometrica: J. Econ. Soc., 1415–1430.
Vissing-Jorgensen, A., 2003. Perspectives on behavioral finance: does "irrationality" disappear with wealth? Evidence from expectations and actions. NBER Macroecon. Annu. 18, 139–194.
Wang, J., 1993. A model of intertemporal asset prices under asymmetric information. Rev. Econ. Stud. 60 (2), 249–282.
Xia, Y., 2001. Learning about predictability: the effects of parameter uncertainty on dynamic asset allocation. J. Finance 56 (1), 205–246.
Ziegler, A.C., 2012. Incomplete Information and Heterogeneous Beliefs in Continuous-Time Finance. Springer Science & Business Media.