Three Essays on Macroeconomics and Wealth Distribution

by

Xiaowen Lei

M.A., Simon Fraser University, 2012
B.A., Shanghai Jiao Tong University, 2011

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Economics
Faculty of Arts & Social Sciences

SIMON FRASER UNIVERSITY
Summer 2018

Approval

Name:	**Xiaowen Lei**
Degree:	**Doctor of Philosophy (Economics)**
Title:	**Three Essays on Macroeconomics and Wealth Distribution**
Examining Committee:	**Chair:** Alexander Karaivanov Professor
	Kenneth Kasa Senior Supervisor Professor
	David Andolfatto Supervisor Professor
	Paul Klein Supervisor Professor
	Chris Bidner Internal Examiner Associate Professor Department of Economics Simon Fraser University
	Martin Ellison External Examiner Professor Department of Economics University of Oxford
Date Defended:	**August 3,2018**

Abstract

My thesis focuses on macroeconomics and monetary policy, with a concentration on belief heterogeneity, household portfolio choice and wealth distribution. I also work on real option models in monetary policy.

The first chapter of my thesis, entitled **“Wait and See" Monetary Policy** , was co-authored with my classmate Michael Tseng, and is recently published in *Macroeconomic Dynamics*. The paper develops a model of the optimal timing of interest rate changes. With fixed adjustment costs and ongoing uncertainty, changing the interest rate involves the exercise of an option. Optimal policy therefore has a “wait-and-see" component, which can be quantified using option pricing techniques. We show that increased uncertainty makes the central bank more reluctant to change its target interest rate, and argue that this helps explain recent observed deviations from the Taylor Rule.

The second chapter is entitled **Risk, Uncertainty and the Dynamics of Inequality**, which is co-authored with my senior supervisor Professor Kenneth Kasa, and is recently published in *Journal of Monetary Economics*. That paper studies the dynamics of wealth inequality in a continuous-time Blanchard/Yaari model. Its key innovation is to assume that idiosyncratic investment returns are subject to (Knightian) uncertainty. In response, agents formulate ‘robust’ portfolio policies (Hansen and Sargent (2008)). These policies are non-homothetic; wealthy agents invest a higher fraction of their wealth in uncertain assets featuring higher mean returns. This produces an endogenous feedback mechanism that amplifies inequality. It also produces an accelerated rate of convergence, which resolves a puzzle recently identified by Gabaix, Lasry, Lions, and Moll (2016).

The third chapter, entitled **Information and Inequality**, studies wealth inequality in a continuous-time 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. If the coefficient of relative risk aversion exceeds unity, reduced estimation risk encourages investment in higher yielding risky assets. As a result, endogenous information acquisition amplifies wealth inequality. Relatively wealthy individuals buy more information, which leads them to invest more in higher yielding assets, which then makes them even wealthier.

Keywords: Inequality, Monetary Policy, Robust Control, Learning, Portfolio Choice, Continuous Time Model

Dedication

To My Parents

Acknowledgements

The last six years working towards my PhD degree has been intellectually challenging but also enormously fruitful. I could have not done it without the help of many people who gave me guidance, support, understanding and encouragement. I would like to take a moment to thank those who greatly contributed to my completion of the degree.

I am mostly indebted to my senior supervisor, Kenneth Kasa, who cares so much about my work, always respond to my inquiries promptly, taught me tremendously the knowledge in macroeconomics, guided me through all the ups and downs during the entire journey, provided me with financial support, and greatly inspired and reinforced my passion to be a good economist. I learnt from him the importance of doing original and rigorous research, focusing on priorities, and more importantly, being a generous person.

I also benefited from Professor Alexander Karaivanov, who gave lots of attention to my work on inequality and gave detailed comment. I am also grateful to Paul Klein for his advice and influence on me to become a rigorous researcher. David Andolfatto has always served as an inspiration for doing work that matters for the real world.

I benefited greattly from other faculty members too. Robert Jones taught me continuous time math and motivated me to do good research. His kindness will be remembered. Chris Muris and Irene Botosaru voluntarily spent effort on polishing my thesis. Luba Peterson gave me lots of feedback on my research in macroeconomics, and was always available for help and advice. John Knowles provided many useful tips in my work, and encouraged me to think outside of the box.

I am particularly humbled and privileged to have met many classmates and co-authors from whom I learnt a lot - among them are: Edouard Djeutem for his support and encouragement for my little academic achievement, Michael Tseng for his influence on me of mathy and scientific thinking, Yang Li for his constant inspiration of hard work, Vitor Hadad for his friendship and passion in economics. I thank Jing Wang for helping my transition to graduate study when I first landed on this continent, and I thank Omid Mousavi and Kevin Chen for helping me to work the Ph.D comprehensive exam. Many thanks also goes to my friends who gave me encouragement: Shelly Fei Yu, Dong Lu, Tong Thi Hanh, etc. I thank Christoph Luelfesmann and Simon Woodcock who gave me a chance at SFU. I am grateful to my dear friend Meiyu who is now resting in peace in the heaven. Her persistence and encouragement during my Ph.D will continue to inspire me to strive for academic excellence in every possible way.

I am deeply grateful for my parents for providing unconditional love and support. I thank J.S.Bach's music for providing me with spiritual guidance and concentration on my study. I gratefully acknowledge the financial support of the Richard Lipsey award and James Dean award, and thank all the staff and faculty of the department of economics at Simon Fraser University, especially Gwen Wild and Kathleen Vieira-Ribeiro. Finally, I want to take the chance to thank International Department at the Bank of Canada for providing me with the opportunity of doing a Ph.D dissertation internship there during 2018 summer. I benefited greatly from interaction with many good economists there. I want to particularly thank Geoffrey Dunbar, Stefano Gnocchi,Yoko Imura, Dmitry Matveev, Rhys R. Mendes, Gurnain Pasricha,Lin Shao, Walter Steingress, Shaofeng Xu etc. for taking the time and providng feedback for my research work.

Table of Contents

List of Tables

List of Figures

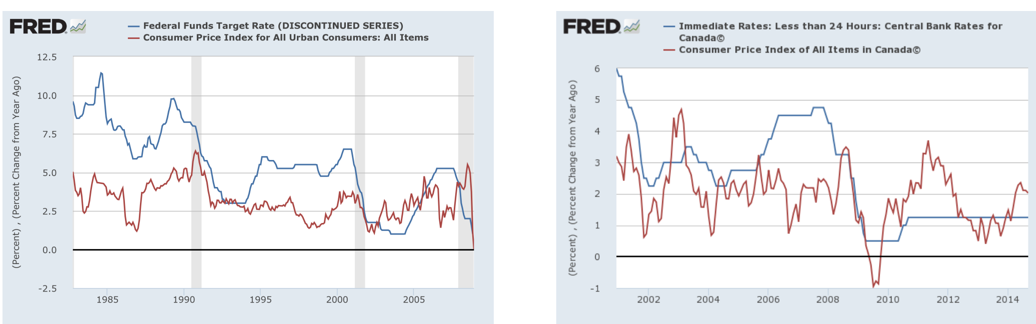
Chapter 1

"Wait and See" Monetary Policy

One possible FOMC strategy is to simply pocket the lower yields and continue to wait-and-see on the U.S. economic outlook. James Bullard, Federal Reserve Bank of St. Louis President, June 5, 2012 Speech.

1.1 Introduction

Interest rate adjustments occur at a lower frequency than macroeconomic data releases. For example, Figure 1.1(a) plots the monthly US federal funds target rate against the inflation rate. The frequency mismatch is apparent. Of course, in recent years interest rates have remained at zero for standard liquidity trap reasons, but rates exhibited inertia even before the financial crisis. Such inertia is not specific to the US Federal Reserve. The same plot for the Bank of Canada (BoC), Figure 1.1(b), shows that the BoC rates exhibits even more pronounced inertia.



(a) US Federal Fund Rate vs. Inflation

(b) Bank of Canada annual interest rate vs. Inflation

Figure 1.1: US Federal Fund Rate and Bank of Canada annual interest rate vs. inflation, monthly data.

There are several possible reasons why a central bank might be reluctant to change interest rates. The most obvious reason is that there is some sort of cost to changing the interest rate. However, standard convex adjustment cost models would not explain the observed long periods without change. Such models instead generate *continuous* gradual adjustment. Another possibility is that inertia arises for strategic reasons (see e.g. Woodford (1999)). Once again, however, strategic inertia produces continuous adjustment. A third possibility is insufficient data, as most national accounts data arrive only quarterly. Two facts run counter to this explanation. First, highly relevant financial market data is now available at virtually a continuous rate. Second, the plots in Figure 1.1 show that central banks respond even more haltingly than the frequency at which data is released.

Another information-based explanation comes from the demand-side rather than the supply-side. Perhaps the Fed simply cannot *process* all the information that is available. Therefore "rational inattention" produces inertia. This is also contradicted by the facts. First, standard linear-quadratic models of rational inattention produce attenuated adjustments, not discrete adjustments (see Sims (2010)). Second, the Fed employs *thousands* of economists and data analysts, whose primary job is to process data. Although rational inattention might be a plausible explanation for a lone central banker, it seems less plausible when applied to an entire institution of data processing specialists. Finally, probably the most common explanation is that wait-and-see is motivated by learning, i.e., the desire to reduce uncertainty. Despite its apparent plausibility, keep in mind standard learning models produce continuous adjustment, or at least adjustment at the same frequency at which the data is available. Even more problematic is the prediction that learning should generate trends in the data. As learning gradually reduces uncertainty, the central bank should become more willing to respond. Such trends are not apparent in the data.

Although all of these factors are no doubt important parts of monetary policy, they do *not* explain infrequent adjustment. Instead, in our view such inaction signals the presence of a *fixed cost* to taking action. What is this cost? Clearly, it is not a technical cost. The Fed could dictate a minute-by-minute target if it so desired. One possibility arises from the fact that monetary policy decisions are not made by a single individual, but by a *committee*. As anyone who has ever served on a committee can attest, committee decision-making has costs. We focus on the inertia induced by these costs. It is possible that committee decision-making introduces inertia by itself, independent of option value considerations, which arise from possible delays in reaching consensus.[1] However, Blinder (2009) cites evidence that it does not. Moreover, Blinder (2009) argues that such procedural delays are likely to be relatively low in the case of US and Canadian monetary policy, at least for the time period studied here, since they are both examples of what he calls "autocratically collegial" committees, in which a single chairman has a dominant influence.

The Federal Open Market Committee (FOMC) must meet at least four times each year in Washington, D.C. Since 1981, eight regularly scheduled meetings have been held each year at intervals of five to eight weeks. At each regularly scheduled meeting, the Committee votes on the policy to be carried out during the interval between meetings. Other central banks have similar meeting schedules. Although in theory the Fed can change rates between meetings, this rarely happens. According to FOMC statements, since 2002, 47 out of 52 target rate changes occurred at meeting times. Moreover, as the above plots reveal, it is not uncommon for rates to remain unchanged following a meeting of the FOMC. This suggests that there is more to infrequent adjustment than the costs of holding meetings.

[1] We thank an anonymous referee for pointing this out.

In this paper, we take the presence of fixed costs as given, and study their implications in a general equilibrium framework. We consider a continuous-time version of the standard New Keynesian model, in which the central bank attempts to balance inflation and output gap deviations. Without fixed costs, optimal policy would produce a version of the Taylor rule. Continuous evolution of macroeconomic data would produce continuous interest rate adjustments. Our primary goal is to show that the option value of waiting-to-see does not just explain infrequent adjustment. If Taylor rule predictions were simply wiggles around the implied wait-and-see policy, the value-added here would be second-order. Instead, we argue that an option value perspective introduces fundamentally new and important considerations into policy, which help to explain several widely noted and debated discrepancies between observed interest rates and Taylor rule predictions. In particular, we show that *uncertainty* becomes an important input into policy. In contrast, standard Taylor rule models are based on certainty-equivalence, and therefore predict that uncertainty does not influence policy.[2] Greater uncertainty increases the option value of waiting. We argue that post-2001 in the US is a period with higher uncertainty than pre-2001, thus policy makers react with more inertia in interest rate changes. In calibrating the model to US data, we find that the implied fixed cost corresponds to an approximately 1.89% of additional annual welfare loss.

Before proceeding, we note some important caveats to our analysis. First, although the novelty here is to incorporate option value considerations into macroeconomic policy, these option values are no doubt present in the private sector as well. In fact, studying these problems is where the analytical techniques were first developed (Dixit (1994), Stokey (2008)). For simplicity, we assume that only the policy maker faces fixed costs. There are reasons to believe, however, that waiting-to-see could interact in mutually destructive ways if both the government and the private sector has an option value to wait-and-see, especially if each side lacks information about the other (Caplin and Leahy (1994)). Second, the only uncertainty in our model is that regarding the future values of exogenous shocks. The central bank is fully informed about the underlying structural model and its parameter values. As noted above, we doubt whether learning about parameters would, by itself, explain observed inertia. However, if policy makers confront more diffusive forms of uncertainty, in combination with fixed costs, this might influence the implied option value of waiting-to-see (see e.g. Miao and Wang (2011) and Trojanowska and Kort (2010)).

The remainder of the paper is organized as follows. Section 2 relates our work on optimal monetary policy to the real options literature and to recent debates about Taylor Rule deviations. Section 3 derives a benchmark model without fixed costs of adjustment, which is essentially a version of the Taylor Rule. Section 4 solve for partial equilibrium where the private sector's expectation is exogenous. Section 5 proves that the central bank's optimal policy is part of a rational expectation equilibrium. Section 6 examines the properties of an optimal "wait-and-see" rule with respect to parameter changes. Section 7 takes the model's testable predictions to both US Fed and BoC data and studies its quantitative implications. Section 8 discusses future research in this framework and concludes. The Appendix contains a technical proof for sufficiency of appropriate Euler equations for optimality in terms of general stopping time problems.

[2]In fairness, some have argued that the Taylor rule is desirable not because it is optimal within the context of a given model, but rather because it is *robust* to the presence of model uncertainty (see, e.g. Levin, Wieland, and Williams (1999) and Levin and Williams (2003)).

1.2 Literature Review

Our study is motivated by recent debates about Taylor Rule deviations. Kahn (2010) examines various of versions of the Taylor Rule, and shows that although a Taylor Rule matches Federal funds target rate data well during 1980-1990s, there are large deviations from it in recent years. In particular, policy appears to have been too accommodative. Taylor (2014) argues that interest rate policy has become less predictable, and that the central bank has been "keeping rates too low for too long". In addition, Woodford (1999) shows that the optimal nominal interest rate rule should exhibit "inertia". In our model, the central bank exhibits a type of inertia different from that of Woodford. Here, higher inertia means the central bank widens the "no-action" band in which it allows the state to drift freely without applying interest rate adjustment. In contrast, Woodford shows that the ability to pre-commit to an interest rate path introduces an autoregressive component into the Taylor Rule. A larger autoregressive component corresponds to higher inertia. The idea is that small adjustments in the same direction improve the inter-temporal trade-off between inflation and the output-gap (Rudebusch (2006)). However, in practice, interest rates exhibit not only small adjustment but also large shifts, and, more recently, periods of no adjustment. Such time series does not have the typical sample path properties of an autoregressive process, suggesting possible misspecification. Rather than tying the current interest rate to the last adjustment via an autoregressive mechanism, in our model the central bank decides whether to adjust based on the current state of the economy. In this "wait-and-see" approach, apparent inertia arises from an option value. Periods of no adjustment and occasional large shifts, unaccounted for in an autoregressive specification, are naturally explained in our model by the central bank's reluctance to exercise an option.

In related work, Davig and Leeper (2008) and Svensson and Williams (2008) compute optimal Markov switching rules in which a policy maker sets different degrees of response to inflation depending on whether the state has crossed an *exogenous* threshold. Alba and Wang (2017) and Murray, Nikolsko-Rzhevskyy, and Papell (2015) fit empirical Markov switching models to US monetary policy. In these models, discrete adjustment reflects a switch between two sets of Taylor rules. In contrast, we study optimal policy with endogenously set optimal thresholds. Our model is also related to prior work on "sticky information". For example, Mankiw and Reis (2002) study the impact of discrete, optimally chosen, information updating on the Phillips curve. However, their focus is on the firm's side.

There has been considerable effort in bringing the real option effect into macroeconomic models (see e.g. Froot and Obstfeld (1991), Dixit (1993), Dixit (1994), and Stokey (2008)), which drives the key "wait and see" feature of our model. The techniques that we employ are also used in the menu cost literature. When firms face both first and second moment uncertainty about total-factor productivity, Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry (2012) shows that uncertainty increases during recessions. The non-convexities together with time variation in uncertainty imply that firms become more cautious in investing when uncertainty increases. Stokey (2013) develops a model which shows that uncertainty about future tax policy induces firms to temporarily stop investing. Both papers look at uncertainty from the firm side. A recent paper that examines uncertainty from the policy side is Alvarez and Dixit (2014), which analyzes the optimal timing of a break-up of the Euro zone using a real options framework, with the private sector modelled in a reduced form way. In contrast, our model casts a standard New Keynesian model into an impulse control framework and remain in a general equilibrium setting where both the central bank and private section have rational expectations.

1.3 Monetary Policy Without Fixed Costs

We briefly recall the optimal monetary policy under discretion in the absence of fixed costs. Assume time is discrete. Let π_t denote inflation and x_t denote the output gap. The Phillips curve and the dynamic IS curve in a standard New Keynesian model, are:

$$\pi_t = \kappa x_t + \beta E_t(\pi_{t+1}) + u_t \tag{1.1}$$

and

$$x_t = -\frac{1}{\gamma}(i_t - E_t(\pi_{t+1}) - r) + E_t(x_{t+1}) \tag{1.2}$$

where γ is the household's coefficient of risk-aversion, κ denotes the response of inflation to an output shock, i_t is the nominal interest rate, and r is the natural rate of interest, which is assumed to be constant here. The cost-push shock u_t is essential in our analysis. It represents the uncertainty the central banker is facing.

The central banker in our model conducts monetary policy with discretion, and therefore has the myopic goal of minimizing the deviations of current inflation and the output gap from target (normalized to zero).[3]Without adjustment costs, the central banker's problem is a static one:

$$\min_{x_t,\pi_t} \frac{1}{2}(\lambda x_t^2 + \pi_t^2) \tag{1.3}$$

subject to

$$\pi_t = \kappa x_t + v_t \tag{1.4}$$

.

where $v_t \equiv \beta E_t(\pi_{t+1}) + u_t$, and λ is the weight put on the output gap.[4] Assuming u_t is an AR(1) process with coefficient ρ_u, the optimal interest rate rule is given by

$$i_t^* = r + \Phi_i u_t \tag{1.5}$$

where $\Phi_i = \frac{\kappa\gamma(1-\rho_u)+\lambda\rho_u}{\kappa^2+\lambda(1-\beta\rho_u)}$. In the special case where u_t follows a random walk, we have $\Phi_i = \frac{\lambda}{\kappa^2+\lambda(1-\beta)}$. It is well-known that, without fixed costs, a linear-quadratic objective function implies certainty equivalence for the central bank's problem. *Uncertainty plays no role in optimal monetary policy, and only the current state matters.* As expected, the optimal interest rate is linear and *continuous* in the cost push shock. Under discretion, the output gap and inflation are also linear in the cost-push shock. Thus we can view this linear rule as a version of the Taylor Rule, which will be

[3]Such preference is a second-order approximation of a CRRA household's utility function, which is proved in Rotemberg and Woodford (1999) and Woodford (2003).

[4]With discretion, the central bank cannot influence the private sector's expectation in a systematic way. Thus, he treats expectations as an exogenous process when conducting policy.

compared with the central bank's optimal rule from our model in Section 4.

1.4 Model

We continue to assume the bank operates with discretion. The model is obtained by taking continuous time limits of the previous two standard discrete time New Keynesian equations (see e,g. Galí (2009)). Instead of solving a continuous time New Keynesian model from the outset (as is done, for example, in Fernández-Villaverde, Posch, and Rubio-Ramírez (2012)), we start by deriving discrete time counterpart, which summarizes the private sector's behavior under any policy, then take the continuous time limit. Our New Keynesian continuous time model admits analytic solutions, using techniques from stochastic control.

Rational Expectations requires the private sector to form expectations of the output gap and inflation that are consistent with central bank policy. Given such expectations, the central bank chooses optimal monetary policy taking the private sector's expectations as given. The resulting optimal policy must then conform to the private sector's expectations. In discrete time, it is reasonable to conjecture that the private sector's expectations follow a martingale:[5]

$$E_t \pi_{t+1} = \pi_t \tag{1.6}$$

$$E_t x_{t+1} = x_t. \tag{1.7}$$

In a discrete time setting, the private sector's information set at time t includes the current period interest rate. In our continuous time model, the interest rate is determined by the current state of the economy, which is evolving stochastically. The k-th adjustment i_k happens at a random, rather than a deterministic, time τ_k. Therefore the continuous time counterparts to the above conditions for stochastic processes π_t and x_t are that, conditional on the central bank's last adjustment, the private sector's expectation remains the same between interest rate adjustments:

$$E[\pi_\tau | \tau', i_k] = \pi_{\tau'} \text{ and } E[x_\tau | \tau', i_k] = x_{\tau'}.$$

where τ and τ' are any stopping times with $\tau_{k+1} > \tau \geq \tau' \geq \tau_k$.[6]

Substituting (6) and (7) into (1) and (2) respectively gives

[5]In general, we are not able to tell whether such expectations are unique. In fact, when banks' interest rate threshold is fixed, the private sector could have highly nonlinear expectations, as demonstrated by Davig and Leeper (2008). However, since we focus on an endogenous interest rate band rather than endogenous expectation, studying a version of martingale expectation and verifying that it is rational simplifies the analysis and helps us to focus on the model's inertia predictions.

[6]Recall that a *filtration* is an increasing sequence $\{\mathcal{G}_t\}$ of σ-algebras representing information flow. The σ-algebra $\mathcal{G}_t$, or *information set at time* t, contains events known up to time t. A *stopping time* τ is a random variable such that $\tau^{-1}(-\infty, t] \in \mathcal{G}_t$ for all t. A stopping time is therefore a random time that is known to the agent at time t. In our context, this means that the timing of interest rate adjustments that have occurred up to time t is known to the private sector at time t.

$$x_t = a_1\tilde{i} + b_1 u_t \tag{1.8}$$

$$\pi_t = a_2\tilde{i} + b_2 u_t \tag{1.9}$$

where $\tilde{i} = i_k - r$ for some k. The cost-push shock u_t is exogenous and is modelled by a Brownian motion.[7]

We derive a rational expectations equilibrium in which the central bank incorporates ex ante assumptions about the private sector's expectations into his cost minimization problem. Then variables x_t and π_t in the objective function in the central bank's problem can be summarized by a single state variable:

$$\begin{aligned}\frac{1}{2}(\lambda x_t^2 + \pi_t^2) &= \frac{1}{2}[\lambda(a_1\tilde{i} + b_1u_t)^2 + (a_2\tilde{i} + b_2u_t)^2] \\ &= \frac{1}{2}(\lambda a_1^2 + a_2^2)\tilde{i}^2 + (\lambda b_1^2 + b_2^2)u_t^2 + 2(a_1b_1\lambda + a_2b_2)\tilde{i}u_t \\ &= \frac{1}{2}\tilde{\lambda}(\tilde{i} - \theta u_t)^2 + \frac{\lambda^2\beta(2-\beta)}{(\lambda(1-\beta)^2 + \kappa^2)^2}u_t^2\end{aligned} \tag{1.10}$$

where $\theta = \frac{\lambda(1-\beta)}{\lambda(1-\beta)^2+\kappa^2}$, and $\tilde{\lambda} = \lambda a_1^2 + 1$.

Define $z_t \equiv \tilde{i}_t - \theta u_t$. Note that the second term in (10) is independent of monetary policy. It is the efficiency loss that cannot be mitigated when adjusting the interest rate. The first term implies a squared deviation of optimal interest rate. Assuming the above assumption holds, we could derive the conjectured coefficients $a_1 = \frac{\beta-1}{\kappa}, b_1 = -\frac{1}{\kappa}, a_2 = -1, b_2 = 0$.

Let K denote the fixed cost of adjustment. A discretionary policy maker's problem is to minimize the expected discounted sum of squared deviations of z_t and adjustment cost by choosing a sequence $\{\tau_k, i_k\}$, where τ_k and i_k are the time and size, respectively, of the k-th adjustment. The central bank's problem is then to find

$$V(z_0) = \inf_{\tau_k, i_k} E^{z_0}[\int_0^\infty e^{-\rho t}\left[\frac{1}{2}\tilde{\lambda}z_t^2\right]dt + \sum_{k=1}^{\infty} e^{-\tau_k}K] \tag{1.11}$$

where

- The expectation operator $E^{z_0}[\cdot]$ denotes expectation taken with respect to the law of the shock process z_t.
- The cost-push shock u_t is assumed to follow a Brownian motion $du_t = \sigma dB_t$ where σ is the volatility of the process. Increasing σ increases the uncertainty faced by the central bank.

[7] Evidence exists that inflation and output gap processes are persistent. In our model this translates to the choice of modelling the output gap and inflation as continuous-time random walks. Relaxing the unit root restriction, one could allow for mean reversion. In continuous-time, we can readily incorporate more general cost-push shocks by using an Ornstein-Uhlenbeck process, a sum of Brownian motion and a mean-reverting drift. Mean reversion in the cost-push shock would have two competing effects on the central bank's policy. The tendency of the process to revert to its long-run mean makes the central bank more reluctant to adjust. On the other hand, mean reversion makes the process stationary, which in turn decreases central bank's uncertainty, thereby increasing the frequency of adjustment.

- From (10), the summarized variable z_t then evolves according to $dz_t = -\theta\sigma dB_t = du_t$ with initial state z_0 almost surely.
- $\{\tau_k\}$ is a non-decreasing sequence of stopping times with respect to the natural filtration $\{\mathcal{F}_t\}$ generated by du_t. The timing of the k-th adjustment, which is a state-dependent random time τ_k, must be decided only using the central bank's information.
- Each i_k, the size of k-th adjustment, is a τ_k-measurable random variable. In other words, the central bank determines the size of an adjustment using information available at the (state-dependent random) time of adjustment.
- $z_{\tau_k} - z_{\tau_k^-} = i_{\tau_k} - i_{\tau_k^-}$. Since the cost-push shock is exogenous, changes in z_t correspond to changes in the interest rate.

Therefore the central bank's problem is to choose a random state-dependent sequence of adjustment times based on information available, along with the interest rate at those times. In the presence of a fixed cost K, continuous adjustment is clearly suboptimal.

The central bank's problem can be seen as a dynamic programming problem. Define the "best adjustment operator" $\mathcal{A}$, acting on bounded functions $W : \mathbb{R} \to \mathbb{R}$, by[8]

$$\mathcal{A}W(z) = \inf_{i \in \mathbb{R}} W(z - i) + K.$$

If $V(z)$ is the central bank's value function, then $\mathcal{A}V(z)$ is the resulting value from the best possible interest rate adjustment, if an adjustment is made at state z. Therefore we must have

$$V(z) = \inf_{\tau} E^z[\int_0^{\tau} e^{-\rho t} f(z_t) d_t + e^{-\rho\tau} \mathcal{A}V(z_{\tau^-})]$$

where, for ease of notation, we put $f(z) = \frac{1}{2}\tilde{\lambda} z^2$ and $\inf_\tau$ denotes the infimum over all finite stopping times. By optimality, $V(z) \leq \mathcal{A}V(z)$ since the central bank always has the choice of not exercising the option of adjusting. The central bank's problem can be therefore further rewritten as:

$$V(z) = \inf_{\tau} E^z[\int_0^{\tau} e^{-\rho t} f(z_t) d_t + e^{-\rho\tau} V(z_{\tau^-})].$$

Assuming now that $V(z) \in C^2(\mathbb{R})$, i.e. is twice-differentiable with continuous second derivative and the infimum $\inf_\tau$ is actually attained at a stopping rule τ^* given by an open region $U \subset \mathbb{R}$, then applying Dynkin's Formula to the right-hand side gives[9][10]

$$V(z) \quad = \quad V(z_0) + E^z[\int_0^{\tau^*} e^{-\rho t}(f(z_t) - \rho V(z_t) + \frac{1}{2} V''(z_t)\theta^2\sigma^2) dt]. \tag{1.12}$$

[8] More formally, $\mathcal{A}$ acts on bounded Borel-measurable functions on $\mathbb{R}$.

[9] The value function V we construct below using smooth pasting is actually not twice-differentiable. The Appendix contains a proof why this derivation is nevertheless true.

[10] Dynkin's Formula is a classical result that generalizes the Fundamental Theorem of Calculus $V(z(\tau)) = V(z(0)) + \int_0^{\tau} V'(z(t))z'(t)dt$ from ordinary differential equations to Itô diffusions (Protter (2004), p356).

Therefore the state space $\mathbb{R}$ is divided into two regions

$$\begin{cases} \rho V = \frac{1}{2}\tilde{\lambda}z^2 + \frac{1}{2}V''\theta^2\sigma^2 & \text{on } U \\ \rho V < \frac{1}{2}\tilde{\lambda}z^2 + \frac{1}{2}V''\theta^2\sigma^2 & \text{on } \mathbb{R}\backslash U. \end{cases}$$

The inaction region U is therefore characterized by the central bank's first order condition[11]

$$\rho V = \frac{1}{2}\tilde{\lambda}z^2 + \frac{1}{2}V''\theta^2\sigma^2 \tag{1.13}$$

between adjustment and non-adjustment of the interest rate. It states that the expected loss of a central banker's value is equal to the sum of an immediate squared deviation of optimal interest rate, and the expected rate of capital loss of holding the option of not changing rates. Adjustment is applied immediately on $\mathbb{R}\backslash U$.

Conjecture now that U is an interval $U = (a, b)$ for some $a < b$. On U, the first order condition is an inhomogeneous ordinary differential equation with general solution

$$V = Az^2 + B_1 e^{\delta_1 z} + B_2 e^{\delta_2 z} + C$$

with unknown constants A, B, δ_1, δ_2, and C. The constants A, B, δ_1, δ_2 are found by matching coefficients:

$$A = \frac{\tilde{\lambda}}{2\rho}, C = \frac{\tilde{\lambda}\theta^2\sigma^2}{2\rho^2}, \delta_{1,2} = \pm\frac{\sqrt{2\rho}}{\theta\sigma}. \tag{1.14}$$

What remains is to solve for the free boundaries a and b, and to determine the adjustment amount i once the boundary is reached. We note that that $z - i = c$ must lie in (a, b). Furthermore, to satisfy the C^2 assumption on V in the above derivation, it is necessary that $V(a) = V(b) = V(c) - K$, $V'(a) = V'(b) = 0$, and $V''(a) = V''(b) = 0$. We prove in the Appendix that the first two set of conditions

$$V(a) = V(b) = V(c) - K \tag{1.15}$$

$$V'(a) = V'(b) = 0 \tag{1.16}$$

are sufficient for V to be the value function by generalizing Dynkin's Formula beyond C^2-functions. In other words, the optimizing central bank need only satisfy continuity, (15), and an Euler equation, (16), on the adjustment thresholds. The value function V is constant $V = V(c)$ on $\mathbb{R}\backslash(a, b)$. When facing a starting state $z_0 \in \mathbb{R}\backslash(a, b)$, the central bank immediately adjusts the state to c. Given the model's symmetry, it is reasonable to guess that (a, b) is of the form $(-S, S)$ and $c = 0$. Using this guess, the Euler equations become:

[11]The Appendix shows the necessary first order condition is in fact part of a sufficient condition, as assumption (iii) of the theorem.

$$AS^2 + B_1 e^{\delta_1 S} + B_2 e^{\delta_2 S} = B_1 + B_2 - K, \tag{1.17}$$
$$2AS + B_1 \delta_1 e^{\delta_1 S} + B_2 \delta_2 e^{\delta_2 S} = 0, \tag{1.18}$$

respectively. This is the same set of equations encountered in Dixit (1993). An approximate analytical solution for the threshold S is

$$S^* \approx \left(\frac{12\theta^2 \sigma^2 K}{\tilde{\lambda}} \right)^{\frac{1}{4}}. \tag{1.19}$$

This implies that a fixed cost K of 4^{th} order magnitude has first order importance in determining the adjustment threshold, while uncertainty of second order has first order effects on inertia.

1.5 Rational Expectations

It is clear that the inflation process π_t and output process x_t, under the central bank's optimal interest rate policy, is consistent with the private sector's expectations. Indeed, recall that the private sector's expectations regarding the two processes satisfy a local martingale-type condition:

$$E[\pi_\tau | \tau', i_k] = \pi_{\tau'} \text{ and } E[x_\tau | \tau', i_k] = x_{\tau'}.$$

where τ and τ' are any stopping times with $\tau_{k+1} > \tau \geq \tau' \geq \tau_k$. On the other hand, under the central bank's optimal policy,

$$x_\tau = a_1 i_k + b_1 u_\tau, \; x_{\tau'} = a_1 i_k + b_1 u_{\tau'}$$

and

$$\pi_\tau = a_2 i_k + b_2 u_\tau, \; \pi_\tau = a_2 i_k + b_2 u_{\tau'}.$$

By Doob's Optional Stopping Theorem (Lipster and Shiryayev (1989)), we have

$$E[x_\tau | \tau', i_k] = a_1 i_k + b_1 u_{\tau'} = x_{\tau'}$$

and

$$E[\pi_\tau | \tau', i_k] = a_2 i_k + b_2 u_{\tau'} = \pi_{\tau'}.$$

Hence, expectations are confirmed.

1.6 Expected Hitting Times and Adjustment

The properties of the optimal rule in our model shed some light on how different uncertainty regimes contributes to monetary policy inertia.[12] Given an arbitrary initial point z, the expected time to the first such resetting is found by[13]

$$T(z) = \frac{S^2 - z^2}{\sigma^2}. \tag{1.20}$$

Expressed in terms of the endogenous threshold $\pm S^*$ from Equation (18), expected hitting time is

$$T(z) = \frac{\sqrt{\frac{12\theta^2 K}{\tilde{\lambda}}}}{\sigma} - \frac{z^2}{\sigma^2}. \tag{1.21}$$

The derivative of $T(z)$ with respect to σ is

$$\frac{\partial T(z)}{\partial \sigma} = \frac{2(z^2 - \sigma\sqrt{3\theta^2 K/\tilde{\lambda}})}{\sigma^3}. \tag{1.22}$$

Thus, the effect of uncertainty on expected hitting times depends on the current state of the economy. There are two countervailing effects. On one hand, given a fixed threshold, a higher variance increases the probability that z hits the boundary, so the expected hitting time decreases. On the other hand, increased uncertainty also increases the optimal threshold, which makes the state less likely to reach the threshold. In our case, when $z^2 < \sigma\sqrt{3\theta^2 K/\tilde{\lambda}}$, the first effect dominates, and increased uncertainty decreases expected hitting times. This corresponds to the case where the initial interest rate is nearly optimal, but if the economy is hit with a severe negative shock, the central banker must react by immediately lowering interest rate.[14] If $z^2 > \sigma\sqrt{3\theta^2 K/\tilde{\lambda}}$, the second effect dominates, and increased uncertainty would increase expected hitting time. This could be the scenario where, given a mild level of inflation, the central banker would have increased the interest rate during normal times, but is less likely to do so during a recession. The two plots in Figure 1.1 show that both effects are present in the data. A large downward adjustment usually corresponds to an unexpected recession followed by a longer period of interest rate inertia. For example, consider the two recession periods of the US economy during early 90's and early 2000's. In both cases, the Fed lowered target rates swiftly at first, then waited for a long period (15 months and 9 months, respectively) before further adjustments.

[12]Incorporating stochastic volatility into our model, i.e. allowing the central bank to anticipate possible changes in uncertainty regime, would retain the general adjustment/non-adjustment regions of optimal policy. The non-adjustment region would depend on the specification of the volatility stochastic process. Although a similar comparison between different specifications and their resulting policies can be carried out, the simplifying assumption of an unanticipatory central bank highlights the effects of uncertainty in our setting and unburdens the discussion from technicalities.

[13]The hitting time distribution of Brownian motion can be derived using the strong Markov property. See, for example, Dixit (1993).

[14]Since the problem is symmetric, all the intuitions that follows also applies to a deflation shock.

Table 1.1: Parametrization (Monthly)

Description	Notation	Value
Risk aversion	γ	0.5
Discount rate	β	0.9967
Natural rate (US)	r_{USA}	0.2385%
Natural rate (Canada)	r_{Canada}	0.1570%
Sensitivity of inflation on output	κ	0.0076
Weight on output deviation	λ	1/2
Adjustment cost (Fed)	K_{Fed}	0.0126
Adjustment cost (BoC)	K_{BoC}	0.1300
US before 2001 volatility	σ_l	0.0126%
US after 2001 volatility	σ_h	0.0242%
Canada volatility	σ	0.0135%

1.7 Calibration

Our model has testable implications for observed interest rate targets. In particular, we take our wait-and-see rule to US Fed and Bank of Canada interest rate data, and examine whether the model can explain both the observed inertia and the timing of shifts in rates. For the US, monthly industrial production, inflation, and federal funds rate data for the period October 1982 to July 2008 are obtained from the Federal Reserve Bank of St. Louis website. Since monthly data on GDP is not available, industrial production serves as a proxy for the output variable.[15] To compute potential monthly output, a Hodrick-Prescott filter with $\lambda = 14400$ is used.

Using the parameters reported in Table 1.1, we calibrate the model to the observed federal funds target. The initial model interest rate is calibrated to match the actual initial interest rate in the sample period. The risk aversion parameter γ is set to be 0.5. The monthly discount factor β is 0.9967, to match an average annual discount rate of 4%. We calibrate the natural interest rate r from the sample average estimate in Holston, Laubach, and Williams (2016) of US from 1990 and 2007. A similar sample average estimate is computed for Canada from 2007 to 2015. This gives $r_{USA} = 2.9\%$ yearly, implying $r_{USA} = 0.2385\%$ monthly, and $r_{Canada} = 1.9\%$ yearly, implying $r_{Canada} = 0.1570\%$ monthly. The weight on output deviation is 0.5. We take monthly $\kappa = 0.0076$ from Galı and Gertler (1999).[16] We then break the US data into two periods, before and after 2001, since we view the

[15] The early 80's corresponded to a period where the Fed adopted a money supply target instead of targeting the interest rate. Therefore, it is appropriate to set that period as the starting time of our analysis.

[16] Galı and Gertler (1999) estimate the quarterly Philips curve

$$\pi_t = \kappa^q x_t + \beta^q \mathbb{E}_t(\pi_{t+1})$$

with $\kappa^q = 0.023$, where superscript q denotes the quarterly parameter. Using the martingale assumption that $\pi_t = \mathbb{E}_t(\pi_{t+1})$, we have

$$\pi_t = \frac{\kappa^q}{1-\beta^q} x_t = \frac{\kappa}{1-\beta} x_t.$$

This relationship between x_t and π_t holds regardless of data frequency. Thus, we can back out monthly κ by matching coefficients. With an annual discount rate of 4%, the implied monthly and quarterly discount rates are $\beta = 0.9967$, and $\beta^q = 0.99$, respectively. The implied monthly κ is 0.0076. For robustness check, post-2001 period as being more volatile (e.g., the tech stock boom and bust, the 9/11 attack, the financial crisis, etc). The parameters σ_l and σ_h are then computed from the standard deviation of the u_t series in these two periods. The u_t process is inferred from output gap and inflation data, along with the Phillips curve and model-implied private sector beliefs.[17][18] To be more concrete, we know that the New Keynesian Phillips curve gives us[19]

$$\pi_t = \kappa x_t + \beta \mathbb{E}_t(\pi_{t+1}) + u_t \tag{1.23}$$

We also know from the model that $\mathbb{E}_t(\pi_{t+1}) = \pi_t$. Therefore, we can infer the following cost shock process

$$u_t = (1-\beta)\pi_t - \kappa x_t \tag{1.24}$$

We then calibrate the adjustment cost K to match the weighted average interest rate duration (weighted by its duration). With US data, this produces $K_{Fed} = 0.0126$. With Canadian data, this produces $K_{BoC} = 0.13$. We will explain the quantitative interpretation of this shortly. Figures 1.2 and 1.3 demonstrate how the wait-and-see rule compares with the Taylor Rule. The main visual difference is that our rule exhibits policy inertia and large discrete jumps.

The adjustment cost parameters K_{Fed} or K_{BoC} can be translated to loss function equivalents. $K = 0.0126$ implies a fixed cost of 1.26% standard deviation from the steady state. The median interest rate duration in the US is 8 months, which implies the Fed changes rates about 1.5 times per year on average. Therefore, the annual cost is approximately an extra $1.5 \times 1.26\% = 1.89\%$ of welfare loss. Using the above parameters, we compare the goodness of the fit of the Taylor rule and the wait-and-see rule: The result is shown in Table 1.2. The result that wait-and-see rule has a better overall fitness is also robust to alternative parametrization of discount rate and elasticity of inflation to output parameter.

The previous results calibrate the adjustment cost K to match the weighted average interest rate duration, and study the overall fitness. For a robustness check, one can also calibrate K to minimize the mean squared error between the target rate and the model implied interest rate, and study how well they match the probability of interest rate change and the average size of interest

when we also take the alternative value of 5.355% as the annual discount rate, the implied discount rates are $\beta^q = 0.9870$, and $\beta = 0.9956$, with $\kappa = 0.0078$.

[17] Not only do we assume that volatility before and after 2001 is different, we also assume it is constant within each regime and not adapted to the central bank's filtration. Therefore, the central bank in our model cannot anticipate volatility changes *ex ante*. Incorporating stochastic volatility into the central bank's decision is in our framework is possible topic to be explored in detail in future research. More detailed remarks on this possibility are given in Section 8.

[18] Since the New Keynesian Phillips curve is widely used to study monetary policy, and our purpose is not to test the Phillips curve, we condition on the validity of the model and identify cost push shocks using the model. This back-solving strategy is also used in *The Conquest of American Inflation* (Sargent (1999)), where a series of residuals are inferred from the Phillips curve using observed data and the initial conditions for the government's beliefs.

[19] For the sake of tractability, the volatility here is conservatively estimated by that of the cost push shock, which is relatively smooth. One might expect that, in a richer model with more shocks, e.g. financial shocks, differences in volatility during recessions and normal regimes would be even larger, implying a stronger wait-and-see effect.

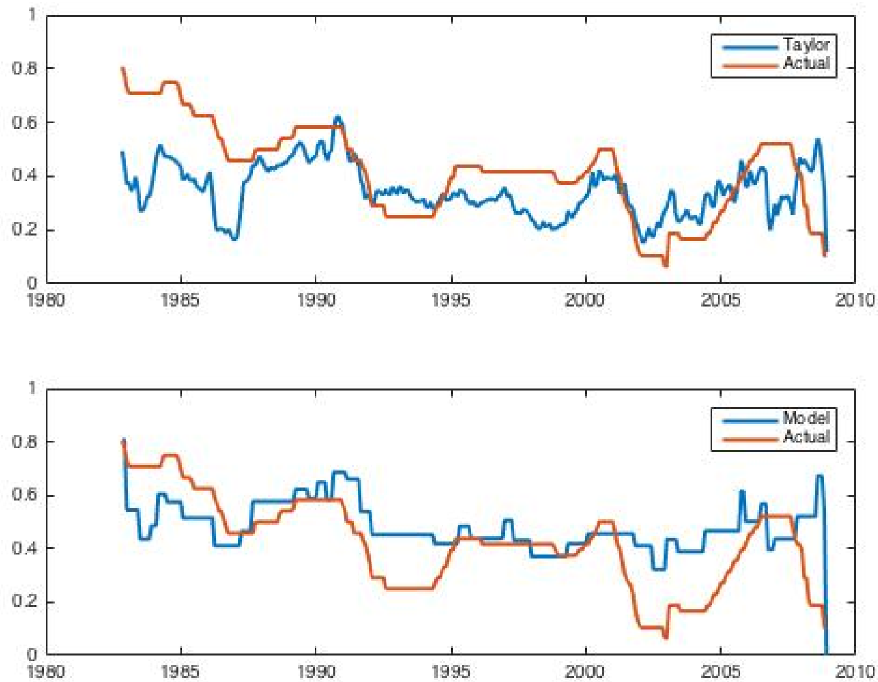


Figure 1.2: US: Model vs. Taylor Rule

Table 1.2: Mean Squared Error, Percent

Central banks	Model	Taylor
US Fed	2.04	2.17
BoC	1.11	1.20

rate changes. This produces $K_{Fed} = 0.017$, and $K_{BoC} = 0.1$ when keeping all other parameters with the same value as Table 1.1. Tables 1.3, 1.4, and 1.5 report the results.

We compare three statistics associated with the policy rule from our model and the Taylor rule.

a. Probability of non-adjustment Table 1.3 displays an obvious distinction between the Taylor rule and the wait-and-see rule. The Taylor rule produces a zero probability of keeping rates constant, while our model generates a reasonable probability of inaction. While it captures the possibility of not changing rates, which the Taylor Rule cannot explain, the wait-and-see rule generates an inaction probability that is larger than the data.

Table 1.3: Probability of Leaving Rates Unchanged

Central banks	Actual	Model	Taylor
US Fed	0.5	0.8949	0
BoC	0.7410	0.9398	0

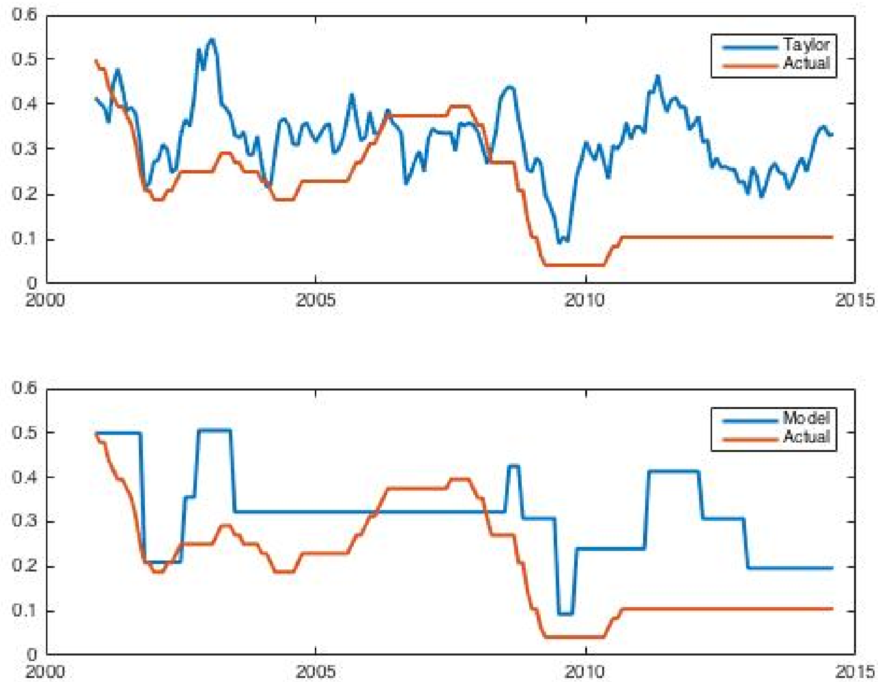


Figure 1.3: Canada: Model vs. Taylor Rule

Table 1.4: Conditional average absolute percentage change of rates

Central banks	Actual	Model	Taylor
US Fed	0.0531	0.1981	0.0676
BoC	0.1327	0.5843	0.1282

b. Conditional average absolute percentage change of rates Our model also has implications for the magnitude of rate changes. To see this, we take the absolute value of percentage change of rates conditional on changes occurring. The absolute value is taken both due to the symmetry of our model, and because the direction of change is not of immediate interest. We find that the wait-and-see rule generates larger changes than in the data.

c. Unconditional average absolute percentage change of rates An arguably better measure of overall fit is to consider the conjunction of the above two statistics. That is, taking into the account both the probability of adjustment and the magnitude of adjustment, we compute the unconditional average absolute percentage change of rates. The last column in Tables 1.4 and 1.5

Table 1.5: Unconditional average absolute percentage change of rates

Central banks	Actual	Model	Taylor
US Fed	0.0304	0.0208	0.0676
BoC	0.0344	0.0352	0.1282

are the same, since the probability of changing rates in the Taylor Rule is always one. However, the wait-and-see rule outperforms the Taylor rule in both US and Canada. One can see that although the Taylor rule matches the magnitude of changes fairly well, its inability to capture the timing of rate shifts forces it to produce a much larger deviation from the actual interest rate than the wait-and-see rule.

1.8 Conclusion

This is the first paper to study the optimal *timing* of interest rate changes. By combining uncertainty and adjustment costs, we are able to rationalize observed interest rate jumps and inertia using a simple impulse control model. The model is consistent with observed target interest rates of both the Fed and the BoC. By focusing on the question of *when* to change interest rates as opposed to *how much*, our framework sheds light on the essential role that uncertainty plays in policy. This motivates many possible directions for future research.

An immediate economic question is the welfare implications of a wait-and-see monetary policy. Without adjustment costs, a discretionary central bank is able to implement the first best outcome by minimizing output gap and inflation deviations instantaneously. With adjustment costs, it is clear that output and inflation deviations will persist, which generates potentially large efficiency losses.

Our technical framework admits several modifications and generalizations. Rather than one discrete adjustment at a time, the central bank may switch between regimes of continuous adjustment and no adjustment. The BoC time series in Figure 1.1 could very well reflect such a policy. To better match the level of interest rates shown by data, the exogenous policy-independent cost-push shock could be modelled by a more general Lévy process rather than Brownian motion.[20] One could also think about imposing a zero lower bound on interest rates. All such extensions fall under the umbrella of general optimal stopping/option exercise problems.

Of particular relevance is the question of how a zero lower bound would affect the central bank's policy in our model. A binding zero lower bound constrains the central bank's ability to react to shocks, since the only option would be an upward adjustment at the zero lower bound. The central bank might alter its behavior in response to this restriction. Near the zero lower bound, the inability, once the bound is reached, to set a negative nominal rate makes the central banker more sensitive to small negative shocks. The option of a large negative adjustment in response to the next large negative shock is no longer available. As a preventive measure, the central bank would apply downward adjustments more frequently near the bound. Away from the bound, the potential loss that could occur at the bound due to restricted ability to adjust could lead the central banker to make positive adjustments larger than she would in our model when positive shocks occur.

One could also argue that the central bank anticipates upturns and downturns of the economy in setting monetary policy, rather than just reacting to changes in uncertainty levels *ex post*. This could be addressed by a two-factor stochastic volatility model, where volatility of the cost-push shock follows a stochastic process, in the option exercise framework. The driving factor could be

[20]Lévy processes are continuous-time analogues of random walks, of which Brownian motion is a special case (Barndorff-Nielsen, Mikosch, and Resnick (2001)). One example is the independent sum of a Brownian motion and a compound Poission process. A general Lévy process can have infinitely many jumps during a finite time interval.

observable or latent. A latent stochastic volatility factor is often argued to be a major concern for monetary policy makers.

Also to be explored is the relationship between learning and a wait-and-see policy. For example, if the natural interest rate is unobservable, waiting not only brings the benefit of more information about the underlying shocks, but also refines estimates of parameter values. One could also envision a model where the private sector also faces fixed investment costs. Firms will delay investing due to policy uncertainty, which then feeds back to more uncertainty by the policy maker. This kind of feedback loop might be an impediment to recovery, with both the central bank and the firm waiting-to-see about each other. Lastly, future research could dig deeper into the way uncertainty is modelled here. For example, with fears of model misspecification, the central bank might want to react *more* aggressively with respect to the current state. So it might be important to examine how a preference for robustness would interact with the wait-and-see policy generated from our baseline model.

Chapter 2

Risk, Uncertainty and the Dynamics of Uncertainty

2.1 Introduction

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

Although investment risk models are successful in generating empirically plausible wealth distributions, they suffer from two drawbacks. First, existing applications focus on *stationary* distributions. However, what is notable about recent US wealth inequality is that it has increased. This suggests that some parameter characterizing the stationary distribution must have changed. It's not yet clear what changed. Second, Gabaix, Lasry, Lions, and Moll (2016) have recently shown that standard investment risk models based on Gibrat's Law cannot account for the *rate* at which inequality has increased. Top wealth shares have approximately doubled over the past 35-40 years. Standard model parameterizations suggest that this increase should have taken at least twice as long.

Our paper addresses both of these drawbacks. The key idea is to assume that agents confront (Knightian) uncertainty when investing. Following Hansen and Sargent (2008), agents have a benchmark model of investment returns. In standard random growth models, agents fully trust their benchmark model. That is, they confront risk, not uncertainty. In contrast, here agents distrust their model, in a way that cannot be captured by a conventional Bayesian prior. Rather than commit to single model/prior, agents entertain a *set* of alternative models, and then optimize against the worst-case model. Since the worst-case model depends on an agent's own actions, agents view

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

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

themselves as being immersed in a dynamic zero-sum game. Solutions of this game produce 'robust' portfolio policies. To prevent agents from being unduly pessimistic, in the sense that they attempt to hedge against empirically implausible alternatives, the hypothetical 'evil agent' who selects the worst-case model is required to pay a penalty that is proportional to the relative entropy between the benchmark model and the worst-case model.

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

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

Idiosyncratic uncertainty also helps resolve the transition rate puzzle of Gabaix, Lasry, Lions, and Moll (2016). They show that models featuring scale dependence, in which shocks to growth rates depend on the level of income or wealth, produce faster transition rates than traditional random growth models based on Gibrat's Law. Robust portfolio policies induced by uncertainty produce a form of scale dependence. Inequality dynamics are analytically characterized using the Laplace transform methods popularized by Moll and his co-authors. Although the model itself is nonlinear, this nonlinearity only arises when the inverse of the entropy penalty parameter is nonzero. For small degrees of uncertainty the parameter is close to zero. This allows us to employ classical perturbation methods to obtain approximate analytical solutions of the Laplace transform of the Kolmogorov-Fokker-Planck (KFP) equation, which then yield approximations of the transition rates.

To illustrate the quantitative significance of uncertainty induced inequality, we suppose the US economy was in a stationary distribution without uncertainty in 1980. Even without uncertainty wealth is concentrated at the top due to a combination of investment luck and longevity luck. Assuming agents live/work on average about 40 years, the wealth share of the top 1% is 24.3%,

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

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

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

roughly equal to the observed 1980 share. Uncertainty is then injected into the economy by setting the (inverse) entropy penalty parameter to a small nonzero value, while keeping all other parameters the same. This increases the top 1% wealth share to 36.9%, close to its current value of about 40%. If this increased inequality had been generated by a change in some other parameter, the transition rate at the mean level of wealth would be only 1.14%, implying a *half-life* of more than 60 years. Thus, assuming the economy is currently at least 90% of the way to a new stationary distribution, it should have taken 200 years to get here, rather than the observed 35-40 years! However, if increased inequality was instead generated by increased uncertainty, the transition rate at the mean more than triples, to 3.85%. This reduces the model implied transition time from 200 years to about 60 years; still longer than observed, but significantly closer.

Aoki and Nirei (2017) also study the dynamics of wealth inequality in a Blanchard-Yaari OLG model. A portfolio composition effect is also the key force behind increased inequality in their model. However, their paper features several important differences. First, they focus on income inequality rather than wealth inequality. Second, their model lacks a natural pertubation parameter, so they resort to numerical solutions of the KFP equation. They find that if the variance of idiosyncratic productivity shocks is calibrated to those of publicly traded firms, the model produces transition rates that are comparable to those in the data. However, if privately held firms are included, which is more consistent with the model, transition rates are too slow. Third, and most importantly, the underlying mechanism in their paper is different. They argue that reductions in top marginal income tax rates were the trigger that produced increased inequality. In support, they cite Piketty, Saez, and Stantcheva (2014), who report evidence on top income shares and tax rates from 18 OECD countries for the period 1960-2010. They show that countries experiencing the largest reductions in top marginal income tax rates also experienced the largest increases in top income inequality. We do not dispute the role that tax policy likely played in growing inequality. However, an interesting additional result in Piketty, Saez, and Stantcheva (2014) is that if you split the sample in 1980, the link between taxes and inequality increases markedly. This is exactly what you would expect to find if increased uncertainty coincided with tax reductions in 1980.[6]

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

[6] **?** compare the roles of taxes and technologically-induced changes in wage inequality in accounting for recent US wealth inequality. They find that wage inequality is more important. However, to generate fat tailed wealth distributions with idiosyncratic labor income risk, they follow **?** and introduce 'awesome states' in labor productivity. This procedure has been criticized by Benhabib, Bisin, and Luo (2017).

2.2 Motivation

This paper focuses on wealth inequality. To focus on the role of uncertainty, it abstracts from all other sources of heterogeneity that contribute to inequality, including labor income heterogeneity. Hence, this section only presents evidence on wealth inequality.

Arguably the best current estimates of US wealth inequality come from Saez and Zucman (2016a). They combine data from the Fed's Survey of Consumer Finances with IRS data on capital income. Wealth estimates are computed by capitalizing reported income data. Attempts are made to include the value of assets that do not generate capital income (e.g., pensions and life insurance), but Social Security is excluded. The value of owner occupied housing is computed from data on property taxes and mortgage interest payments. The major omission is human capital. One potential problem with this data is that the methodology assumes rates of return within asset categories are identical across households. This can produce biased estimates if returns are correlated with wealth within asset categories (Fagereng, Guiso, Malacrino, and Pistaferri, 2016b).[7]

In principle, we could look at the entire cross-sectional distribution, but since current interest (and our model) are focused on the right-tail, Figure 2.1 simply reports the top 1% wealth share.

Figure 2.1: Top 1% Wealth Share in the USA

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

Who are these top 1%? In 2012, the top 1% consisted of the wealthiest 1.6 million families, with a minimum wealth of about $4 million. Average wealth in this group is about $13.8 million. As noted by Saez and Zucman (2016a), wealth has become even more concentrated than indicated by

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

the conventional focus on the top 1%. The top 0.1% now own 22% of US wealth, more than triple its value in 1978. To belong to this exclusive club, one needs a *minimum* wealth of $20.6 million. Although we focus on the top 1%, it should become clear below that our model would likely do even better at accounting for the top 0.1%.

It is of course debatable whether an economy is ever in a stationary equilibrium, but our basic modeling strategy here is to suppose that the US economy was in such a state around 1980. We then ask whether an empirically plausible increase in (Knightian) uncertainty could have produced the sort of rapid increase in inequality observed in Figure 2.1. To address this question we need a model.

2.3 The Model

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

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

Agents can invest in three assets: (1) A risk-free technology.[8] The value of the risk-free asset follows the deterministic process

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

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

$$dS = \mu Sdt + \sigma SdB \tag{2.1}$$

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

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

Again, the key departure point of this paper is to suppose that agents have doubts about their idiosyncratic investment opportunities. In other words, they confront 'uncertainty', not risk. By assumption, these doubts cannot be captured by specifying a finite-dimensional Bayesian prior over alternative models. Instead, agents fear a broad spectrum of nonparametric alternatives. These alternatives might reflect omitted variables or complicated nonlinearities. As emphasized by Hansen and Sargent (2008), we want agents to be prudent, not paranoid, so that they only hedge against models that could have plausibly generated the historically observed data. To operationalize this, let q_t^0 be the probability measure defined by the Brownian motion process in the benchmark model (2.1), and let q_t be some alternative probability measure, defined by some competing model. The (discounted) relative entropy between q_t and q_t^0 is then defined as follows:[9]

$$\mathcal{R}(q) = \int_0^\infty e^{-\rho t} \left[\int \log \left(\frac{dq_t}{dq_t^0} \right) dq_t \right] dt \tag{2.2}$$

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

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

where $\tilde{E}$ denotes expectations with respect to the distorted measure q_t, and h_s represents a square-integrable process that is progressively measureable with respect to the filtration generated by q_t. Again from Girsanov's Theorem, we can view q_t as being induced by the following drift distorted Brownian motion[10]

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

which then defines the following conveniently parameterized set of alternative models

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

Agents are assumed to have time-additive CRRA preferences. Each agent wants to construct robust consumption and portfolio policies that perform adequately for all h's that lie within some convex set centered around the benchmark model in (2.1). To do this, agents enlist the services of a hypothetical 'evil agent', who is imagined to select models that *minimize* their utility. That is, agents view themselves as being engaged in the following dynamic zero-sum game:

$$V(w_0) = \max_{c,\alpha} \min_h E \int_0^\infty \left(\frac{c^{1-\gamma}}{1-\gamma} + \frac{1}{2\varepsilon} h^2 \right) e^{-(\rho+\delta)t} dt \tag{2.3}$$

subject to

$$dw = [(r + \alpha(\mu - r))w - c + \alpha\sigma w h]dt + \alpha w \sigma \cdot dB \tag{2.4}$$

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

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

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

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

$$(\rho+\delta)V(w) = \max_{c,\alpha} \min_{h} \left\{ \frac{c^{1-\gamma}}{1-\gamma} + \frac{1}{2\varepsilon}h^2 + [(r+\alpha(\mu-r))w - c + \alpha\sigma w h]V'(w) + \frac{1}{2}\alpha^2\sigma^2 w^2 V''(w) \right\} \quad (2.5)$$

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

$$h = -\varepsilon\sigma\alpha w V'(w) \quad (2.6)$$

$$\alpha = -\frac{(\mu-r)V'(w)}{[V''(w) - \varepsilon(V'(w))^2]w\sigma^2} \quad (2.7)$$

$$c = [V'(w)]^{-1/\gamma} \quad (2.8)$$

Even before solving the model, these policy functions reveal a lot about the equilibrium. First, notice that robustness takes the form of *pessimism*, i.e., by (2.6) the drift distortion h is negative since $\alpha > 0$ and $V'(w) > 0$. Second, notice that the distortion increases with volatility ($\sigma\alpha w$). When volatility increases, it becomes statistically more difficult to rule out alternative models. Third, notice that it is not *a priori* obvious how the magnitude of the distortion depends on wealth. If $V(w)$ is concave, so that wealthier agents have a lower marginal utility of wealth, then (ceteris paribus) weathier agents will be less pessimistic. Having money in the bank allows you to relax. However, the volatility term offsets this. For a given portfolio share, α, wealthier agents have more at stake. This makes them more worried, and triggers a greater preference for robustness and a more pessimistic drift distortion.[11] This volatility effect will be reinforced if the first effect leads wealthier agents to invest a higher fraction of their wealth in the risky technology.[12]

Turning to the portfolio policy in (2.7), we can see that robustness in some sense makes the investor more risk averse, in that it subtracts a positive term from $V''(w)$. However, as noted by Maenhout (2004), unless the utility function is quadratic ($\gamma = -1$) or logarithmic ($\gamma = 1$), the resulting portfolio policy is nonhomothetic, and α will depend on w. Observe that the implicit degree of risk aversion falls as w increases if $V(w)$ is concave, so that $V'(w)$ decreases with wealth. In the limit, as $V'(w) \to 0$, the preference for robustness completely dissipates, and the limiting

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

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

portfolio is the same as when $\varepsilon = 0$. Notice that this portfolio is just the classic Merton portfolio

$$\alpha_0 = \frac{\mu - r}{\gamma\sigma^2} \tag{2.9}$$

since one can readily verify that $V(w) \sim w^{1-\gamma}$ in this case. Finally, notice that the consumption function in (2.8) is the usual one. However, that does not mean uncertainty is irrelevant to saving decisions, since in general $V(w)$ will depend in complicated ways on the value of ε.

This analysis of the policy functions was premised on the properties of the unknown value function, $V(w)$. To verify and quantify these effects we must solve for $V(w)$. To do so we first substitute the policy functions in (2.6), (2.7), (2.8) into the HJB equation in (2.5), and then solve the resulting second-order ODE. Unfortunately, the resulting equation is nonlinear, and does not possess a known analytical solution. However, the fact that it *does* have a known analytical solution when $\varepsilon = 0$ opens the door to a classical perturbation approximation. That is, we look for a solution of the form[13]

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

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

Proposition 2.3.1. *A first-order pertubation approximation of the value function for the HJB equation in* (2.5) *is given by*

$$V(w) = A_0 \frac{w^{1-\gamma}}{1-\gamma} + \varepsilon A_1 \frac{w^{2(1-\gamma)}}{2(1-\gamma)} + O(\varepsilon^2) \tag{2.10}$$

with

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

where $\beta = 2(1-\gamma)$ *and* α_0 *is given by the nonrobust portfolio share in equation* (2.9).

Proof. See Online Appendix A. □

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

Corollary 2.3.2. *First-order perturbation approximations of the optimal portfolio policy,* $\alpha(w)$*, and saving rate,* $s(w) = 1 - c(w)/w$*, are given by*

$$\alpha(w) = \alpha_0 - \varepsilon\alpha_0\left(\frac{A_0^2 + (\gamma - 1)A_1}{\gamma A_0}\right) w^{1-\gamma} \tag{2.11}$$

$$s(w) = 1 - A_0^{-1/\gamma} + \varepsilon\frac{1}{\gamma}\left(A_0^{-1/\gamma - 1} A_1\right) w^{1-\gamma} \tag{2.12}$$

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

where α_0 is the nonrobust portfolio policy in (2.9)*, and* (A_0, A_1) *are constants defined in Proposition 2.3.1.*

Proof. See Online Appendix B. □

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

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

2.3.1 Comment on Units

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

2.3.2 Benchmark Parameterization

To illustrate the quantitative effects of model uncertainty on saving and portfolio choice the following benchmark parameter values are used

TABLE 1
BENCHMARK PARAMETER VALUES

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

These parameters are discussed in more detail in the following section. For now it is sufficient to focus on γ, since it determines the nature of the saving and portfolio policies. Empirical evidence yields a wide range of estimates, depending on data and model specifics. However, most estimates suggest $\gamma > 1$. We set $\gamma = 1.31$. Figure 2.2 plots the resulting portfolio policy:

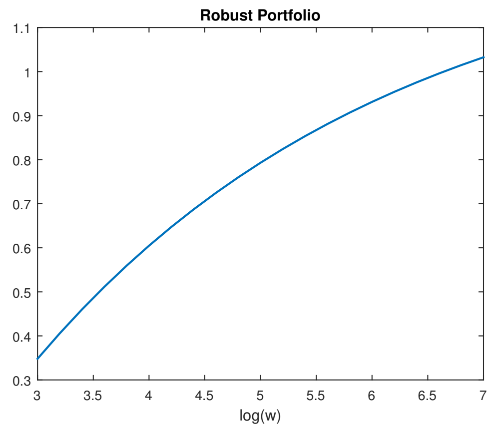


Figure 2.2: Robust Portfolio Share

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

Figure 2.2 reveals the key feedback mechanism in our model. Wealth begets wealth because it leads agents to invest in higher yielding technologies.[14] Notice, however, that inequality is not only a result of wealthy agents investing more than average in higher yielding assets. The effect is symmetric, in that bad luck and low wealth makes agents pessimistic, which leads them to hold most of their wealth in the safe asset. The very poorest agents invest only about 30% in the risky technology. Our model therefore provides a novel explanation for why poor households hold most, if not all, of their wealth in safe low-yielding assets. It is also broadly consistent with the empirical evidence in Carroll (2002), who examines portfolio data from the Survey of Consumer Finances for the period 1962-95. He finds that the top 1% hold about 80% of their wealth in risky assets, while the bottom 99% hold (on average) about 40%. (See his Table 4).

The other force driving inequality is saving. Empirical evidence suggests wealthy households have higher saving rates (Saez and Zucman, 2016). Wealth increasing saving rates also produce deviations

[14]From eq. (2.11) and the expressions for (A_0, A_1) in Proposition 3.1, we can see that $A_0 > 0$ whenever $\gamma > 1$. Hence, a sufficient condition for $\alpha'(w) > 0$ is that $A_1 > 0$. A necessary and sufficient condition for $A_1 > 0$ is that $\gamma^{-1}[r - (\rho + \delta)] + \frac{1}{2}\frac{\rho+\delta}{\gamma-1} + \frac{1}{2}\sigma^2\alpha_0^2[3(1-\gamma)+1] > 0$. A sufficient condition for this is that $\gamma < 1.33$. Alternatively, if $\gamma > 1.33$, then the condition will be satisfied if $r \geq \rho + \delta$ and σ^2 is sufficiently small.

from Gibrat's Law, which amplifies inequality. Figure 2.3 plots individual saving rates as a function of log wealth:

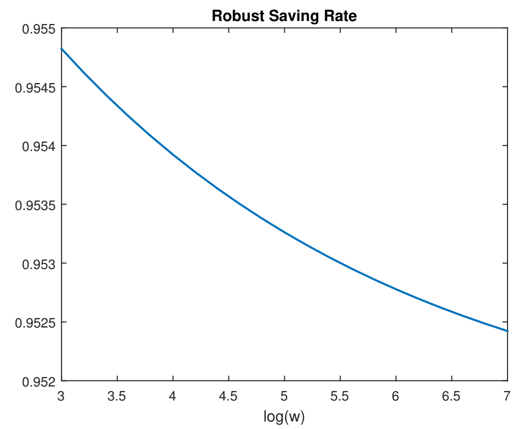


Figure 2.3: Robust Saving Rate

Evidently, the model's predictions along this dimension are counterfactual. Saez and Zucman (2016a) report 'synthetic' saving rate estimates that increase with wealth. Dynan, Skinner, and Zeldes (2004) also find increasing saving rates in a variety of US data sets. Note that a saving rate that decreases with wealth actually *reduces* inequality, since poorer households save a larger fraction of their wealth. What's happening here is that, for a given portfolio allocation, wealthy agents are less pessimistic and expect higher mean returns. Because the intertemporal elasticity of substitution ($1/\gamma$) is less than one, higher expected returns reduce the savings rate, since the income effect dominates the substitution effect. This suggests that we could avoid this problem by employing recursive preferences, which delivers a separation between risk aversion and intertemporal substitution. Online Appendix C outlines such an extension using a Duffie-Epstein (1992) aggregator with an intertemporal elasticity of substitution of one. It shows that for very similar parameter values to those in Table 1, saving rates actually increase with wealth.[15]

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

The mechanism driving these portfolio and saving dynamics is a pessimistic drift distortion. Earlier it was noted that it is not at all clear whether wealthy agents are more or less pessimistic in equilibrium. On the one hand, wealth provides financial security, so wealthy agents can afford to be less robust. This effect operates via the decline in $V'(w)$ in the evil agent's policy function. On

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

the other hand, greater financial security enables you to take on more risk. As a result, wealthier agents have more to lose from model misspecification. This effect operates via the scaling term, αw, in the evil agent's policy function. The following result shows that for small values of ε the first effect dominates if $\gamma > 1$.

Corollary 2.3.3. *To a first-order approximation, the equilibrium distortion function, $h(w)$, is increasing if and only if $\gamma > 1$.*

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

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

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

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

□

To interpret the $h(w)$ function it is convenient to map it into an implied drift distortion. Notice that if $h(w)$ is substituted into the HJB equation in (2.5) the agent's problem becomes equivalent to the standard one, but with an endogenous, wealth-dependent, drift equal to $\mu + \sigma h(w)$. Figure 2.4 plots this implied drift distortion using the benchmark parameter values given in Table 1:

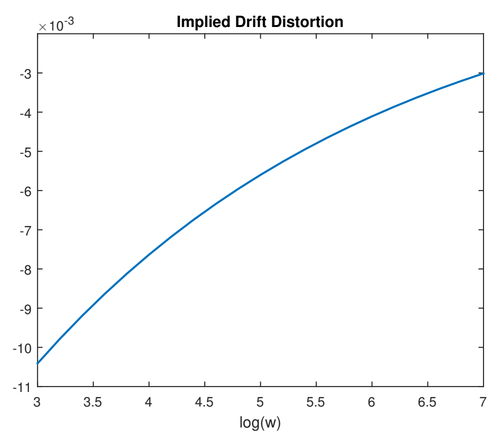


Figure 2.4: Drift Distortion

Perhaps not surprisingly, the drift distortion mirrors the portfolio policy depicted in Figure 2.2. Agents begin life with a small degree of pessimism, equivalent to about a 60 basis point negative drift distortion. However, if they are unlucky, and they move into the left tail of the wealth distribution, their pessimism grows to more than a 1.0 percentage point negative drift distortion. As seen in Figure

2.2, this is enough to nearly keep them out of the market. Hence, robustness generates something like a 'poverty trap', in that pessimism and low wealth become self-fulfilling.[16]

2.3.3 Comment on Preferences vs. Environment

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

2.4 Stationary Distributions

This analysis thus far has focused on the problem and decision rules of a single agent. Our primary interest, however, is on wealth inequality. To address this we must aggregate the individual wealth dynamics and compute the cross-sectional distribution of wealth. The first step is to substitute the optimal policy functions into the individual's budget constraint in equation (2.4). When doing this it is important to observe that, as in Hansen and Sargent (2008), fears of model misspecification are solely in the mind of the agent. The agent's benchmark model is the true data-generating process, the agent just doesn't know it. Uncertainty still matters, however, because it influences behavior.

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

Proposition 2.4.1. *To a first-order approximation in ε, individual log wealth follows the diffusion process*

$$dx = (a_0 + \varepsilon a_1 e^{(1-\gamma)x})dt + (b_0 + \varepsilon b_1 e^{(1-\gamma)x})dB \tag{2.13}$$

with

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

where (A_0, A_1) are constants defined in Proposition 2.3.1 and α_0 is the nonrobust portfolio share defined in equation (2.9)*.*

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

Proof. See Online Appendix D. □

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

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

$$\frac{\partial f}{\partial t} = -\frac{\partial[(a_0 + \varepsilon a_1 e^{(1-\gamma)x})f]}{\partial x} + \frac{1}{2}\frac{\partial^2[(b_0 + \varepsilon b_1 e^{(1-\gamma)x})^2 f]}{\partial x^2} - \delta f + \delta\zeta_0 \tag{2.14}$$

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

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

Applying the Laplace transform to equation (2.14) and dropping $O(\varepsilon^2)$ terms yields the following approximation of the cross-sectional wealth dynamics:[17]

$$\frac{\partial F}{\partial t} = \Lambda(s)F(t, s) + \varepsilon\Phi(s - \beta)F(t, s - \beta) + \delta e^{-sx_0} \tag{2.15}$$

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

$$\Lambda(s) = \frac{1}{2}b_0^2 s^2 - a_0 s - \delta \tag{2.16}$$

$$\Phi(s) = b_0 b_1 s^2 + (2b_0 b_1 \beta - a_1)s + \beta(b_0 b_1 \beta - a_1) \tag{2.17}$$

The $\Lambda(s)$ and $\Phi(s)$ functions pick-up the derivatives of the KFP equation, since $\mathcal{L}\{f'(x)\} = sF(s)$ and $\mathcal{L}\{f''(x)\} = s^2 F(s)$. The fact that derivatives are converted to a simple multiplication is what makes Laplace transforms so useful. This method is less commonly applied to problems with variable coefficients, as is the case here, since the Laplace transform of the product of two functions is not the product of their Laplace transforms.[18] However, the particular form of variable coefficients here, which involves multiplication by an exponential function, is one case that works out nicely. In particular, the so-called 'shift theorem' implies $\mathcal{L}\{e^{\beta x} f(x)\} = F(s - \beta)$. This result follows rather obviously from the defintion of a Laplace transform after a simple change of variables.

The $F(t, s - \beta)$ term on the right-hand side of (2.15) makes the KFP equation nonstandard. Later we shall obtain an approximate solution by approximating the discrete shift with a derivative. For now, notice that in the nonrobust case, when $\varepsilon = 0$, the problem becomes entirely standard, and we get:

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

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

Proposition 2.4.2. *The nonrobust stationary distribution of log wealth is double exponential, with a mode at* $x_0 = \log(w_0)$*,*

$$f(x) = c_1 h(x_0 - x)e^{\phi_1 x} + c_2 h(x - x_0)e^{\phi_2 x} \tag{2.18}$$

where $h(\cdot)$ *is the Heaviside (unit-step) function, and* (c_1, c_2) *are constants of integration chosen to ensure continuity at the switch point,* x_0*, and satisfaction of the adding-up constraint* $\int_{-\infty}^{\infty} f(x) = 1$*. The exponents* (ϕ_1, ϕ_2) *are the positive and negative roots, respectively, of the quadratic* $\Lambda(s) = 0$*, where* $\Lambda(s)$ *is given by equation* (2.16)*.*

Proof. Since this result is standard, we merely sketch a proof. First, the Laplace transform of the stationary distribution can be obtained by setting $\partial F/\partial t = 0$ in equation (2.15), and then solving the resulting algebraic equation for $F(s)$. When $\varepsilon = 0$, this just gives $F(s) = -\delta e^{-sx_0}/\Lambda(s)$. Next, we invert $F(s)$ to get $f(x)$. This can either be accomplished by expanding $\Lambda(s)^{-1}$ into partial fractions, and using the result that $\mathcal{L}^{-1}\{(s - \phi)^{-1}\} = e^{\phi x}$, or by using contour integration and the residue calculus, noting that the singularity in the left-half plane corresponds to $x > 0$, while the singularity in the right-half plane corresponds to $x < 0$. □

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

Corollary 2.4.3. *In the nonrobust (*$\varepsilon = 0$*) case, the stationary distribution of wealth is double Pareto*

$$f(w) = c_1 h(w_0 - w)w^{\phi_1 - 1} + c_2 h(w - w_0)w^{\phi_2 - 1} \tag{2.19}$$

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

The fact that the conventional (nonrobust) distribution of wealth is Pareto is no accident. This is a well documented feature of the data, and the main attraction of constant parameter idiosyncratic investment risk models is that they quite naturally generate a steady-state Pareto distribution.[19] The key parameter here is ϕ_2,

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

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

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

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

$$\Lambda(s)F(s) + \varepsilon\Phi(s-\beta)F(s-\beta) + \delta = 0 \tag{2.20}$$

Although a perturbation approximation with respect to ε seems natural, the distribution changes quite quickly, even for $\varepsilon \approx 0$, making a first-order approximation unreliable. This will become apparent later. So instead we approximate (2.20) by converting it to an ODE in s, using β as a step-size. This presumes β is 'small', so that $\gamma \approx 1$. Solving this ODE yields the following result,

Proposition 2.4.4. *To a first-order approximation in β, the robust stationary distribution of log wealth is double exponential, with a mode at $x_0 = \log(w_0)$,*

$$f_r(x) = c_{1r}h(x_0 - x)e^{\psi_1 x} + c_{2r}h(x - x_0)e^{\psi_2 x} \tag{2.21}$$

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

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

where $\Phi'(s-\beta)$ is the derivative of the $\Phi(s)$ function in (2.17) evaluated at $(s-\beta)$.

Proof. See Online Appendix E. □

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

Although it is not obvious from the statement of Proposition 2.4.4, one can also show that the robust stationary distribution satisfies the following correspondence principle,

Corollary 2.4.5. *As $\varepsilon \to 0$, the robust stationary distribution in eq.* (2.21) *converges pointwise to the nonrobust stationary distribution in eq.* (2.18)*.*

Proof. See Online Appendix F. □

2.4.1 Top Wealth Shares

It should be emphasized that Propositions 2.4.2 and 2.4.4 characterize the *entire* cross-sectional distribution of wealth. They could be used to compute Gini coefficients, Theil/entropy indices, or any other inequality measure of interest. However, in line with recent debate, we focus here on top wealth shares, in particular, on the top 1% wealth share. Given the stationary distribution $f_\infty(w)$, this is defined as follows

$$\text{top 1\% share} = \frac{\int_{\hat{w}}^{\infty} w f_{\infty}(w)dw}{\int_{0}^{\infty} w f_{\infty}(w)dw}$$

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

Lemma 2.4.6. *Given a double Pareto distribution as in eq.* (2.19)*, the top 1% wealth share is given by*

$$\textit{top 1\% share} = \frac{c_2 \frac{\hat{w}^{1+\phi_2}}{1+\phi_2}}{c_2 \frac{w_0^{1+\phi_2}}{1+\phi_2} - c_1 \frac{w_0^{1+\phi_1}}{1+\phi_1}} \tag{2.22}$$

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

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

2.4.2 Benchmark Parameterization

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

As discussed earlier, initial wealth, w_0, is arbitrary, as long as we are careful to adjust ε to maintain a 'reasonable' drift distortion. (Later we discuss how we define reasonable, but Figure 2.4 shows that it is only around 1%, even for the poorest agents). We simply normalize initial wealth to $w_0 = 100$. Moskowitz and Vissing-Jorgensen (2002) provide evidence from the 1989 SCF on the distribution of private equity returns, and report a value of 6.9% for the median return, so we set $\mu = 5.86\%$, to be conservative. The more important parameters are δ, σ^2, and γ. It is natural to calibrate δ to the mean work life, keeping in mind that this has been changing over time and that the model abstracts from important life-cycle effects. We simply set $\delta = 0.026$, which implies a mean work life of about 38 years.[20] Calibrating σ^2 is more difficult. The model is based on idiosyncratic investment risk, so using values from the stock market would be misleading. Benhabib and Bisin (2017) cite evidence on the returns to owner-occupied housing and private equity which suggest a standard deviation in the range of 10-20%, so to be conservative we set $\sigma = 9.2\%$. Finally, the basic mechanism in our paper hinges on people becoming less pessimistic as they become wealthier.

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

For this we need $\gamma > 1$. Fortunately, this is consistent with a wide range of evidence. However, there is less agreement on the magnitude of γ. Asset market data imply large values. However, as noted by Barillas, Hansen, and Sargent (2009), these large values are based on models that abstract from model uncertainty. They show that a modest amount of uncertainty can substitute for extreme degrees of risk aversion, and their argument applies here as well. In addition, our approximate solution strategy presumes that $1 - \gamma$ is small, so we set $\gamma = 1.31$.

Figure 2.5 displays the two distributions, using the results in Propositions 2.4.2 and 2.4.4,

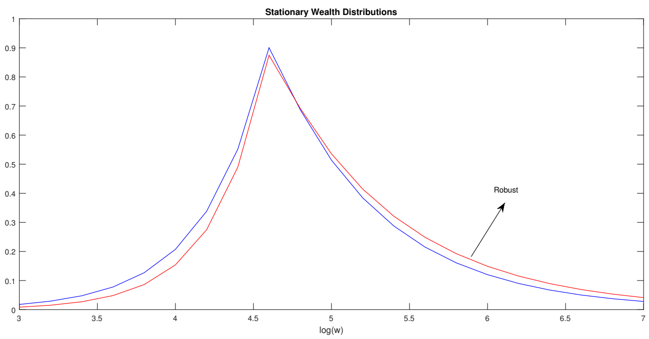


Figure 2.5: Stationary Distributions

The main thing to notice is that the robust distribution has a fatter right tail, reflecting greater wealth inequality. The nonrobust distribution has a right-tail exponent of $\phi_2 = -1.45$, whereas the robust distribution has a right-tail exponent of $\psi_2 = -1.28$. Using Lemma 2.4.6 we find that the top 1% share in the nonrobust economy is 24.3%, roughly equal to the actual value in 1980 depicted in Figure 2.1. In contrast, the top 1% share in the robust economy is 36.9%, which represents a significant increase, but is still 5 percentage points lower than the data. We could of course increase robust inequality to more closely match recent data simply by increasing ε. We argue later that higher values of ε would still be plausible in terms of drift distortions and detection error probabilities. However, higher values of ε start to produce implausible portfolio policies. We should also note that the robust distribution comes somewhat closer to matching the growth in inequality in the extreme right-tail which, according to Saez and Zucman (2016a), is where most of the action has been. They find that the share of the top 0.1% increased from 7% in 1978 to 22% in 2012. Our model generates an increase from 11.9% to 22.3%.

It is easier to visualize the extreme right tail using a log-log scale, as depicted in Figure 2.6. Since the distribution of log wealth is exponential, these are simply linear.

These plots extend to the top 0.1% in the nonrobust case (i.e., $e^{-7} \approx .001$), and to the top 0.2% in the robust case (i.e., $e^{-6} \approx .002$). Since median wealth is about 111 in the nonrobust case, and about 123 in the robust case, these plots suggest that the threshold top 0.2% individual is about 66 times wealthier than the median (i.e., $e^9 = 8,103 \approx 66 \times 123$). For comparison, Saez and Zucman's (2016) data show that the threshold 0.1% individual is more than 400 times wealthier than the median, whose wealth is about \$50,000 (i.e., $20.6\text{mill}/50,000 = 412$). Although a 0.2% threshold

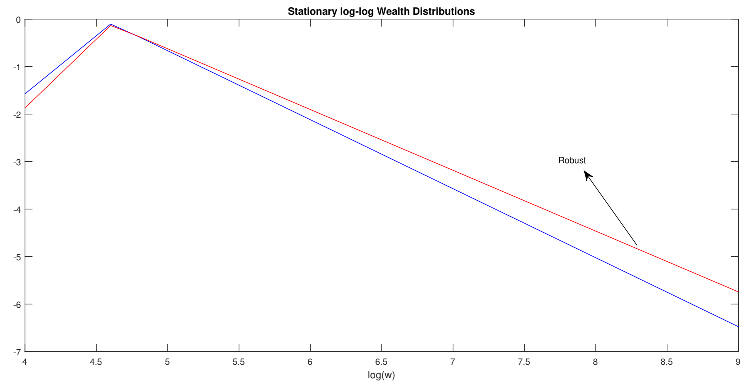


Figure 2.6: Stationary Distributions: Log/Log Scale

would produce a smaller number, it seems safe to conclude that the model does not fully capture the extreme wealth levels that occur in the extreme right tail of the US wealth distribution. This is not too surprising given that the model abstracts from labor income inequality and bequests.

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

2.5 Transition Dynamics

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

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

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

To characterize the transition rate the differential equation in (2.15) must be solved. Although this describes the evolution of the Laplace transform, remember the Laplace transform embodies all the information in the distribution itself. In fact, it does so in an especially revealing way. In particular, the convergence rate of $F(s)$ at a particular value of $(-s)$ provides the convergence rate of the $(-s)th$ moment of $f(x)$.[21] Without uncertainty ($\varepsilon = 0$), the solution of (2.15) is straightforward, and is the same as in GLLM

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

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

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

implying a *half-life* of 60 years! Convergence in the tails is even slower.[22]

Unfortunately, solution of equation (2.15) when $\varepsilon \neq 0$ is more complicated. The exponential functions in the drift and diffusion coefficients of (2.13) produce a discrete shift in the Laplace transform, so that both $F(s,t)$ and $F(s-\beta,t)$ appear on the right-hand side of the differential equation. While a first-order perturbation approximation of the stationary distribution is defensible, a first-order perturbation approximation of $F(s,t)$ along the entire convergence path is more problematic, given the discrete shift, which essentially creates a 'singular perturbation' problem.

Our strategy for solving (2.15) is to approximate the discrete shift with a differential equation. This yields,

$$\frac{\partial F}{\partial t} = \Lambda(s)F(t,s) + \varepsilon\Phi(s-\beta)\left[F(t,s) - \beta\frac{\partial F}{\partial s}\right] + \delta e^{-sx_0} \tag{2.23}$$

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

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

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

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

and obtain separate ODEs for $H(t)$ and $F_\infty(s)$. Solving the ODE for $F_\infty(s)$ produces the stationary distribution, characterized in Propostion 2.4.4. Solving the ODE for $H(t)$ then gives us the convergence rate. The $G(s)$ function is determined by boundary conditions, and it turns out to be $G(s) = F^0_\infty(s) - F^1_\infty(s)$, where $F^0_\infty(s)$ is the Laplace transform of the initial stationary distribution (given in Propostion 2.4.2), and $F^1_\infty(s)$ is the Laplace transform of the new stationary distribution (given in Proposition 2.4.4). We relegate the details to the Appendix, and merely state the following result

Proposition 2.5.1. *Let $\tilde{F}(t,s) \equiv H(t)G(s)$ be the solution of the homogeneous part of the PDE in* (2.23)*, which describes the dynamics of the cross-sectional distribution of wealth. Then to a first-order approximation in β, $\tilde{F}(t,s)$ is given by*

$$\tilde{F}(t,s) = \tilde{F}(0,s)e^{[\Lambda(s)+\varepsilon\Phi(s-\beta)]t} \tag{2.24}$$

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

Proof. See Online Appendix G. □

The convergence rate is determined by the function $\Lambda(s) + \varepsilon\Phi(s-\beta)$. Naturally, when $\varepsilon = 0$, the convergence rate is the same as in the nonrobust case. Note that the convergence rate depends on s, implying that different moments converge at different rates. Figure 2.7 plots this function. It extends from $s = 0$ to $s = -1.3$. Since the robust Pareto exponent is 1.28, the largest moment that exists for both distributions is 1.28.

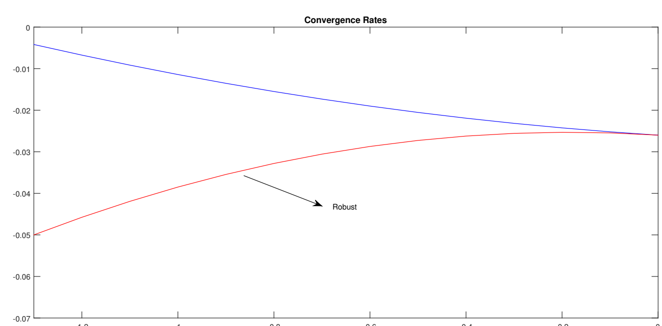


Figure 2.7: Convergence Rates

The top line plots $\Lambda(s)$, the nonrobust convergence rate. As in GLLM, convergence is slower in the tails. At the mean ($s = -1$), the convergence rate is 1.14%. Slower tail convergence reflects the fact that parametric disturbances move along the distribution like a 'traveling wave', first hitting young individuals with low wealth, then spreading to older, more wealthy individuals. With exponentially distributed lifetimes, most individuals are young, and possess little wealth.

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

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

Finally, one should remember that robustness induced amplification and acceleration occurs despite the fact that with additive preferences a counterfactually declining saving rate acts as a stabilizing force. As noted earlier in Section 2.3, and discussed in more detail in Online Appendix C, we can avoid this by using recursive preferences. Online Appendix H presents plots comparing stationary distributions and convergence rates when preferences are assumed to be recursive, as in Duffie and Epstein (1992). It shows that recursive preferences reinforce the paper's main findings concerning steady state inequality and convergence rates.

2.5.1 Numerical Solution

Analytical solutions are attractive because they reveal the underlying mechanisms at work. Still, it is important to keep in mind that all the previous results are just approximations. These approximations can be avoided by solving the model numerically. In principle, even the agent's optimization problem could be solved numerically, but here we focus on the accuracy of our key result, related to convergence rates. Proposition 5.1 is based on approximating a shifted Laplace transform with a differential equation, and it is not entirely clear how good this approximation is. So here we numerically solve an $O(\varepsilon)$ approximation of the KFP equation in (2.14) using a standard discretization algorithm, based on central-difference approximations of the derivatives.[23] Even this $O(\varepsilon)$ approximation is nonstandard, due to the dependence of the drift and diffusion coefficients on x.

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

[23] Our matlab code is available upon request.

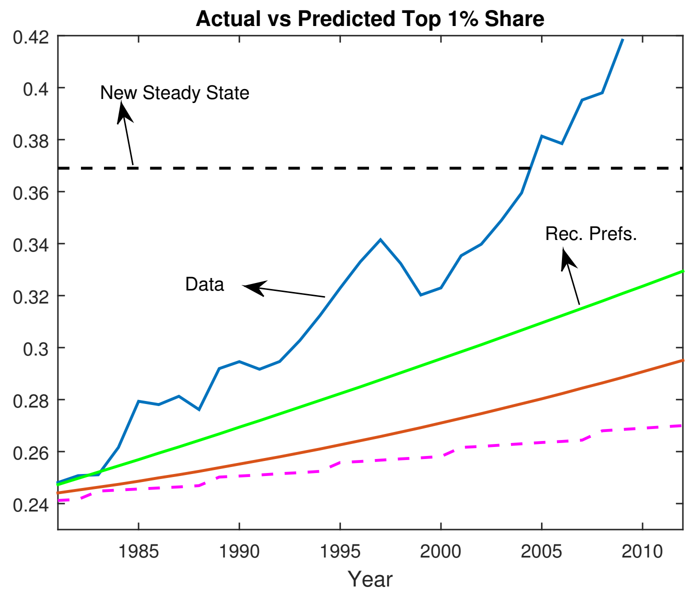


Figure 2.8: Actual vs Predicted Top Wealth Shares: Numerical Solution

The path of the benchmark time-additive model lies in the middle, showing an increase in in the top 1% share from about 24% to 30%. Although significant, this clearly understates the observed increase. Part of the apparent discrepancy simply reflects the fact that the new stationary distribution understates long-run inequality by about 5 percentage points. As noted earlier, although we could increase steady state inequality by increasing ε, this would produce counterfactual portfolio policies. However, the path discrepancy reflects more than a discrepancy in steady states. The convergence rate appears to be too slow as well, both in terms of the data and the model's prediction. Proposition 5.1 suggests the top 1% share should be converging at a rate of at least 3.85%, since that is the predicted mean convergence rate, and our model predicts the tails should converge faster than the mean. In contrast, Figure **??** suggests the top 1% share is converging at about a 1.9% (average) rate. Hence, we are off by at least a factor of 2. This suggests that $\gamma - 1 = .31$ might be too large for the ODE to accurately approximate the shifted Laplace transform.

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

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

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

2.6 Detection Error Probabilities

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

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

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

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

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

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

where T is the sample length. This bound is difficult to evaluate since x is stochastic. As a result, we compute the following (state-dependent) approximation, $\text{avg} \quad DEP(x) \leq \frac{1}{2}\exp\left\{-\frac{1}{8}Th^2(x)\right\}$.

By fixing x, this overstates the bound at values of x where $h(x)$ is small, while understating it at values of x where $h(x)$ is large. Figure 2.9 plots this approximation using the benchmark parameter values and assuming $T = 200$.

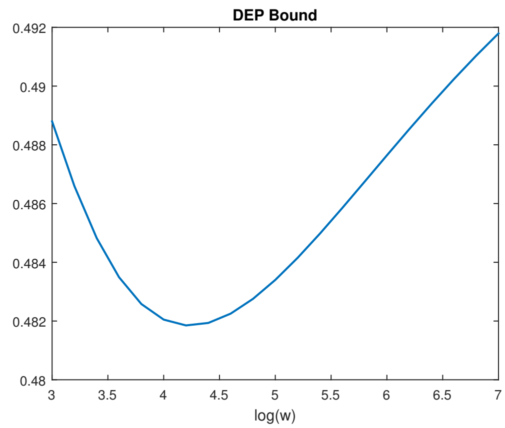


Figure 2.9: Detection Error Probability Bound

Notice the DEP bound is nonmonotonic. This reflects the nonmonotonicity of $h(x)$, which itself reflects the endogeneous nature of pessimism in the model. The agents who are least pessimistic in equilibrium are the very poor, because they choose to avoid the risky technology, and the very rich, because they have a low marginal utility of wealth. Agents in the middle have the most to worry about. When interpreting this plot one should keep in mind that a fixed T might not be appropriate given the model's OLG structure. In particular, one could argue that T should be age-specific. This would be the case if agents weight their own experience more heavily than the entire past history of data, as evidence suggests they do.[24] If this were the case, the DEP would be smaller than Figure 2.9 portrays for large values of x, while being larger for small values of x.

The main point to notice, however, is that the DEPs are quite large, for all values of x. This is true despite the fact that setting $T = 200$ overstates the data that is actually available to agents. (Remember, our time unit here is a year). Although AHS suggest the bound is not always tight, the numbers in Figure 2.9 are so large that we suspect more refined estimates would still preserve the basic conclusion that the degree of pessimism here is plausible. In fact, another way to see this is to simply suppose investment outcomes are i.i.d coin flips. With a 10% standard deviation and 200 years of data, a 95% confidence interval for the mean would be $\pm 2 \times \frac{.10}{\sqrt{200}} \approx 1.4\%$. The implied drift distortions reported in Figure 2.4 are well within this interval.

2.7 Conclusion

The recent increase in US wealth inequality raises three important questions: (1) What caused it?, (2) Why did it start around 1980?, and (3) Why has it been so rapid? There are already many answers

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

to the first two questions. Although we think our proposed explanation - increased uncertainty - provides an interesting and novel answer to these two questions, we view it more as a complement than a substitute to existing explanations. However, as recently emphasized by Gabaix, Lasry, Lions, and Moll (2016), answering the third question is more challenging. Standard random growth models cannot explain why top wealth shares more than doubled in little more than a single generation. To account for the rate of increase in inequality, they argue it is necessary to extend random growth models by incorporating either 'type dependence' or 'scale dependence', which produce heterogeneity in growth rates. Using the tools of robust control theory, we show that increased uncertainty produces a natural form of scale dependence, which generates significantly more rapid convergence rates.

More work needs to be done to make increased uncertainty a fully convincing explanation of recent US wealth inequality. First, the notion that the world became more uncertain around 1980 seems plausible, but should be more carefully documented. There is abundant evidence that idiosyncratic labor income risk increased around this time (Ljungqvist and Sargent, 1998), but risk is not the same as uncertainty, and wealth is not the same as labor income. Robust control provides a link between objective risk and subjective uncertainty by viewing distortions as 'hiding behind' objective risk. Hence, it would be interesting to consider an extension to stochastic volatility, which would open the door to stochastic uncertainty as well. In fact, there is no reason to view risk and uncertainty as competing explanations. Using panel data on the Forbes 400, Gomez (2017) finds that inequality during 1980-95 was primarily driven by increased risk, while return heterogeneity (possibly driven by increased uncertainty) was primarily responsible for increased top wealth shares during the period 1995-2015. Combining both increased risk and increased uncertainty might help to close the gap in Figure 2.8. Second, the key source of scale dependence in our model is that wealthy individuals earn higher mean returns. Again, there is abundant evidence in support of this (although the evidence is less clear using after-tax returns). However, it is not clear whether this return differential reflects portfolio allocation decisions, as in our model, or whether wealthy individuals receive higher returns within asset categories as well. Using tax records from Norway, Fagereng, Guiso, Malacrino, and Pistaferri (2016a) argue that wealthy investors receive higher returns within asset categories. Hence, it would be interesting to consider a model where both forces are at work. Finally, we have argued that wealth dependent pessimism is empirically plausible because the drift distortions it produces could not be statistically rejected based on historical data. However, we suspect our detection error probability bound is not very tight, and one could easily obtain tighter bounds using simulations.

Chapter 3

Information and Inequality

3.1 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's (1987) 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 *can* 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 expectancies, are born with identical initial wealth, and have access to private investment projects that feature identical mean returns. In this environment, inequality is initially created by luck. Relatively wealthy agents have good luck with their investments, and are blessed with long lifetimes. Interestingly, we shall see that learning magnifies the importance of luck, since good luck makes agents relatively optimistic (at least temporarily), and this encourages risk-taking.

Although it is now well known that random growth over a random length of time are enough to generate fat-tailed Pareto wealth distributions (Reed (2001)), recent work by Gabaix, Lasry, Lions, and Moll (2016) (henceforth GLLM) shows that these random growth models fall short quantitatively in accounting for observed inequality. In particular, they show their transition dynamics are far too

[1] Benhabib, Bisin, and Luo (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, Bisin, and Zhu (2011), Benhabib, Bisin, and Zhu (2016), Cao and Luo (2017) and Toda (2014).

slow. GLLM argue that it is important to allow for either 'type dependence' or 'scale dependence', which generate deviations from Gibrat's Law.

This paper shows that learning and information acquisition combine to produce a form of scale dependence. 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, Skinner, and Zeldes (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. In fact, we shall see that if learning were based on exogenous information, wealthy agents would have *lower* savings rates.

One downside of studying a scale-dependent growth model is that we can no longer solve the model analytically. The model's dynamics are characterized by a coupled pair of partial differential equations (PDEs). A Hamilton-Jacobi-Bellman (HJB) equation describes optimal decisions by a single agent, while a Kolmogorov-Fokker-Planck (KFP) equation aggregates these decisions to describe the evolution of the cross-sectional distribution of wealth and beliefs.[2] Neither PDE can be solved in closed-form. In response, I employ a combination of monte carlo simulation techniques and classical perturbation approximations.

Perturbation approximations of an individual agent's decision rules are based on the fact that we *can* analytically solve the HJB equation for a log utility investor who faces an infinite cost of information. Hence, the analysis here presumes agents have preferences that are 'close' to logarithmic and that information is 'expensive'. After substituting these approximate decision rules into the agent's wealth accumulation equation, monte carlo simulation methods are used to numerically compute the stationary distribution of wealth.

To shed some analytical light on this distribution, I then turn to the KFP equation. This is a 3-dimensional PDE, describing the distribution of wealth and beliefs (as summarized by the conditional mean and variance of expected returns). I study this system using classical time-scale separation methods. These methods convert the problem to one of studying the interaction between two lower dimensional 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. I first solve a 1-dimensional KFP equation for wealth, holding beliefs fixed. I 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 is used to characterize the right-tail of the marginal distribution of wealth. I show that endogenous information acquisition reduces the absolute value of the right-tail Pareto exponent, thereby increasing top wealth shares.[3]

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

[2]For recent technical advances in macro models using this approach, see Achdou, Han, Lasry, Lions, and Moll (2017).

[3]This time-scale separation approach is closely related to the 'mean ODE' method pioneered by Marcet and Sargent (1989) in the macroeconomic learning literature.

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, nondiversifiable 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.

Although the 'information revolution' did not occur overnight, for the purposes of this paper I assume that it did. Specifically, I assume that until 1980 information was prohibitively expensive. Agents could passively learn from observed data, but they could not speed up learning by buying information. I also assume the US economy had settled into a stationary distribution. I calibrate the model's parameters to replicate this distribution. Then all of a sudden the cost of information declines. Of course, information is not easily measured, nor is its cost easy to quantify. Given this, I proceed indirectly. I use my model along with observed changes in household portfolios to infer an information cost reduction. Since the units of the information cost are not independent of the units of wealth, one has to be a little careful when interpreting the magnitude of the information cost parameter. From a behavioral standpoint, what matters is how much wealth is spent on information. I show that this share is quite small, even for the wealthiest agents, amounting to less than 1% of wealth. Hedge fund data are used to support this calibration. Using the previously described simulation and approximation strategies, I find that the steady state top 1% wealth share increases from 23.9% to 32.9%. Although this is less than the increase reported by Saez and Zucman (2016b), who find the top 1% wealth share is now around 42%, it is still quite significant, given that the model abstracts from labor income heterogeneity, bequests, and other factors generating inequality.

The analysis here is related to recent work by Kacperczyk, Nosal, and Stevens (2014). They too are motivated by Arrow (1987). However, there is an important difference between their work and mine. Their paper is based on the 'rational inattention' literature (Sims (2003)). They study an economy in which some agents are *endowed* with greater 'sophistication', as measured by channel capacity. Sophisticated investors can process more information. They argue that observed wealth is a good proxy for sophistication. They then show that financial markets enable sophisticated investors to increase their relative wealth. In my model, causation runs the other way. Wealthy agents *become* sophisticated (perhaps by hiring agents with large channel capacities). Beyond this difference in cause and effect, the approach here is arguably more consistent with the critique of Sims (2006), who points out that financial markets might not be the best application of rational inattention. In financial markets, information is certainly *not* free, and the relevant constraints are not on the ability to process free information that is lying around. Instead, the relevant constraint is being able to pay others to gather and process information for you. The entire hedge fund industry owes its existence to this difference.

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

[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.

policy functions. I show that policy functions can be decomposed into an exogenous learning component and an endogenous learning component. Section 5 studies aggregation and the cross-sectional distribution of wealth and beliefs. I first compute monte carlo simulations of the cross-sectional wealth distribution. I then use the KFP equation to derive analytical approximations of the stationary marginal distributions of wealth and beliefs. Section 6 examines data from the hedge fund industry in more detail. It shows that more expensive funds earn higher returns. Section 7 provides a brief discussion of related literature, while Section 8 offers some concluding remarks and suggestions for future research. Proofs and derivations are contained in the Appendix.

3.2 Motivation

This paper focuses on wealth inequality. Arguably the best current estimates of US wealth inequality come from Saez and Zucman (2016b). They combine data from the Fed's Survey of Consumer Finances with IRS data on capital income. Wealth estimates are computed by capitalizing reported income data. Some attempts are made to include assets that do not generate capital income (e.g., pensions and and life insurance). However, Social Security is excluded. Housing values are imputed from data on property taxes and mortgage interest payments. One potential problem with these data is that the methodology presumes rates of return within asset categories are identical across households. This can produce biased estimates if returns are correlated with wealth within asset categories (Fagereng, Guiso, Malacrino, and Pistaferri (2016b)).[5]

Although the model could be used to study the entire cross-sectional distribution of wealth, following convention, Figure 3.1 simply reports the top 1% wealth share.[6] In 2012, the top 1% consisted of the wealthiest 1.6 million households, with a minimum wealth of about $4 million.

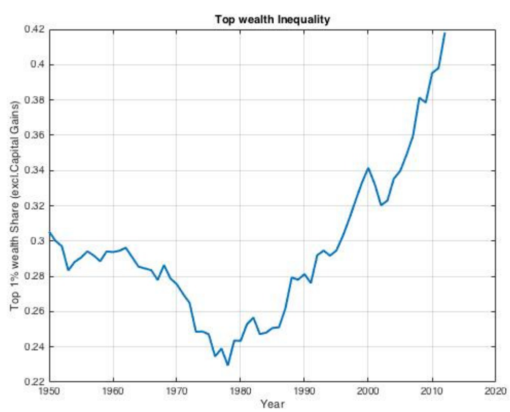


Figure 3.1: Top 1% Wealth Share in the USA

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

[6] Saez and Zucman (2016b) note that wealth has become even more concentrated than indicated by the conventional focus on the top 1%. The top 0.1% now own 22% of US wealth, more than triple its 1978 value.

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

There are many proposed explanations for this increase. Perhaps the most common 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. For example, Kasa and Lei (2017) argue that increased (Knightian) uncertainty, perhaps due to technological change, introduction of new investment opportunities, like financial securitization, financial derivatives, emerging market assets, etc. and globalization, led investors to adopt 'robust' portfolio policies, which then contributed to inequality, since robust portfolio policies imply that wealthy agents invest relatively more in risky assets.

This paper focuses on a different implication of technological change. Instead of creating doubts about returns, here technological change creates opportunities, by making information acquisition and information processing less costly. We can gauge the importance of these changes indirectly by looking at changes in household portfolios and in the wealth management industry. Figure 3.2 plots SCF data on the share of risky assets (as proxied by stocks) in household portfolios. Households are grouped by relative log wealth, in increments of 5 percentage points. The blue line pertains to 1983, and the red line pertains to 2013. Two features stand out. First, relatively wealthy investors

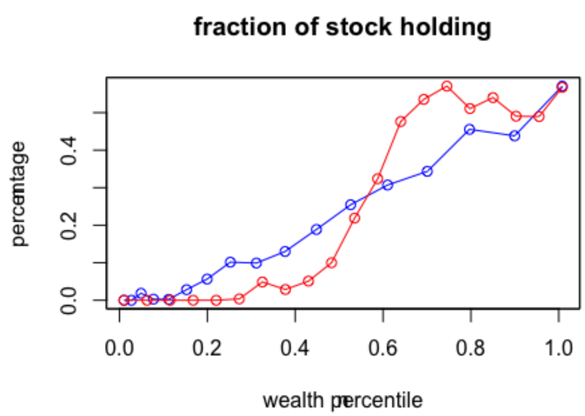


Figure 3.2: Fraction of Risky Assets Holding

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. Second, portfolio disparities across households have increased over time. This simply reflects the growth in inequality over time. Notice that households below the median are now investing a *smaller* share of their portfolios in risky assets than they were in 1983!

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. Figure 3.3 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.

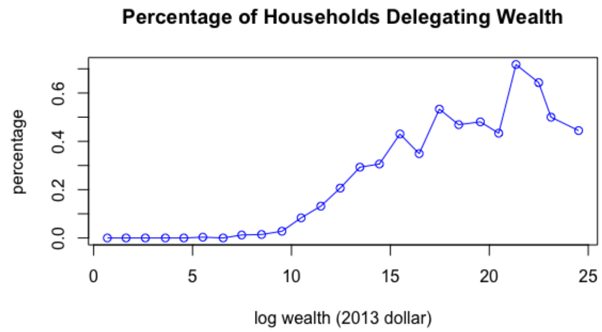


Figure 3.3: Percentage of households delegating wealth

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. This is one feature of reality my model will abstract from.

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. Figures 3.4 and 3.5 present data on the explosive growth of the hedge fund industry.

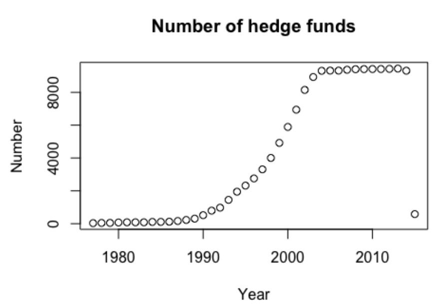


Figure 3.4: Number of Hedge Funds

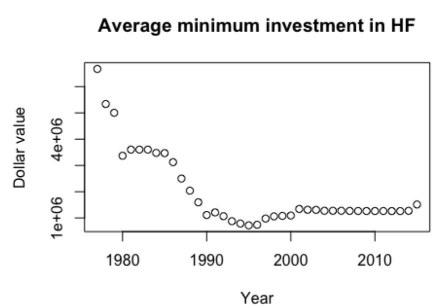


Figure 3.5: Average Min. Investment in Hedge Funds

Figure 3.4 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 Figure 3.5.[7] As discussed above, in a model with scale effects, lower fees do not necessarily translate into greater equality. In Section 6 I shall explore the performance of these funds in greater detail. For now, the key feature of these funds is that one can obtain better performance by paying higher fees.

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 Figure 2.1. To address this question we need a model.

3.3 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, Bisin, and Zhu (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 effects, i.e., it obeys Gibrat's Law. Rather than study scale effects, 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.

3.3.1 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. Agents 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

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

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.

[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.

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 section 3.1, 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 process, 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 $\widehat{\mu(0)}$ 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 (2001)], the innovation process $\hat{B}(t)$ induced by the investor's own filtration is related to the unobserved $B(t)$ by

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

3.3.2 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

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

[9] Ehling, Graniero, and Heyerdahl-Larsen (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 (2001) or Øksendal (2003).

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

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

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 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, $\kappa(t)$. This constrains the maximum rate of information flow that can be noiselessly transmitted, measured in bits.[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

$$\frac{\gamma(t)}{\sigma_y(t)^2} \leq \kappa(t) \tag{3.11}$$

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 speeds up 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 the parameter μ. One interpretation of this cost function is that there is a competitive industry producing and selling financial information at its marginal cost.[12]

[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.

3.3.3 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:

$$\begin{aligned} dw(t) &= [w(t)(r+\pi(t)(\hat{\mu}(t)-r)) - c(t) - q(\kappa(t))]dt + \pi(t)\sigma w(t) d\hat{B}(t) && (3.14)\\ d\hat{\mu}(t) &= \frac{\gamma(t)}{\sigma} d\hat{B}(t) + \frac{\gamma(t)}{\sigma_y(t)} d\hat{B}_y(t) && (3.15)\\ d\gamma(t) &= -\gamma(t)^2\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_y(t)^2}\right)dt && (3.16)\\ \frac{\gamma(t)}{2\sigma_y(t)^2} &\leq \kappa(t) && (3.17) \end{aligned}$$

where $1-\alpha$ 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

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

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]:

$$\begin{aligned} c^* &= V_w^{\frac{1}{\alpha-1}} && (3.19)\\ \pi^* &= \frac{-V_w(\hat{\mu}-r) - V_{w\hat{\mu}}\gamma}{V_{ww} w \sigma^2} && (3.20)\\ \kappa^* &= \theta\gamma\left(\frac{V_{\hat{\mu}\hat{\mu}} - 2V_\gamma}{V_w}\right) && (3.21) \end{aligned}$$

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

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

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.

3.4 Policy Function Approximations

To compute the policy functions, we need to know the value function, $V(w, \hat{\mu}, \gamma)$. To compute V, we first substitute the policy functions in eqs. back into the HJB eq. (**??**). This produces a 3-dimensional, highly nonlinear PDE. A closed-form solution is wishful thinking. However, in a particular limiting case we *can* obtain a closed-form solution. This opens the door to a classical perturbation approximation. It turns out that when information is prohibitively expensive ($\theta = 0$), and the investor has log preferences ($\alpha = 0$), the PDE for V can be solved analytically. 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)).

Since the value function is not of interest by itself, details of its derivation are relegated to Appendix C.2. Here we focus on the implied policy functions:

Proposition 3.4.1. *To an $O(\theta, \alpha)$ approximation, the policy functions for consumption, risky portfolio share, and information choice are given by*

$$c^* \approx \beta w[1 - \frac{\alpha}{2\beta\sigma^2}\left[(\hat{\mu}-r)^2 + \frac{\gamma^2}{\beta\sigma^2}\right] - \frac{1}{2}\theta w \lambda \gamma^2] \tag{3.22}$$

$$\pi^* \approx \frac{1}{\sigma^2}[(1+\alpha)(\hat{\mu}-r) + \alpha\frac{(\hat{\mu}-r)\gamma}{\beta\sigma^2} + \frac{1}{2}\theta w \lambda(\hat{\mu}-r)\gamma^2] \tag{3.23}$$

$$\kappa^* \approx \frac{\theta w}{\beta}\left[1 - \frac{\beta\sigma^2}{\gamma}e^{\beta\sigma^2/\gamma}\Gamma(0, \beta\sigma^2/\gamma)\right] \approx \frac{\theta w \gamma}{\beta^2\sigma^2} \tag{3.24}$$

where $\Gamma(a,x) = \int_x^\infty t^{a-1}e^{-t}dt$ is the incomplete Gamma function and $\lambda = \left[\beta^2\sigma^2[2\beta\sigma^2 - (\mu-r)^2]\right]^{-1}$.

Proof. See Appendix C.2. □

3.4.1 Comments on Information Choice

Like any other good, the optimal amount of information is chosen so that its marginal cost equals its marginal benefit. With the above approximate solution for κ^*, the marginal cost and marginal benefit of information can be seen quite clearly. 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 eq. (3.24) 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, $\frac{w\gamma}{\beta^2\sigma^2}$. It is not surprising that the marginal benefit rises when either wealth or uncertainty increase. An essential observation is that β and σ are the same for everyone. Thus, heterogeneity in information choice comes from the multiplication of w and γ. A potential conflict arises between w and γ. On one hand, increased wealth enhances the marginal benefit of acquiring more information. On the other hand, most rich people will already have a relatively low γ (except for those who were poor and are suddenly hit with a large positive return shock), which implies that acquiring more information pays off less in further reducing estimation variance. Since both w and γ are endogenous outcomes of the model, and are interdependent, one needs to resort to numerical solutions to know which effect dominates. Policy functions are plotted using the parameter values in Table 3.1. The choice of these parameter values will be discussed in detail in the following section.

Table 3.1: Benchmark Parameter Values

μ	β	δ	γ_0	σ	θ	w_0
0.15	0.06	0.04	0.001	0.26	0.0016	1

For now, let's focus on what determines optimal information choice by inspecting Figure 3.6. It plots κ^* as a function of wealth w and estimation variance γ, which are the only two state variables that differ across agents concerning κ. It is obvious from this figure that wealthy agents buy more information. Although κ also increases with γ, it does so to a much less extent quantitatively. Finally, it is useful to summarize how other parameter values influence information choice

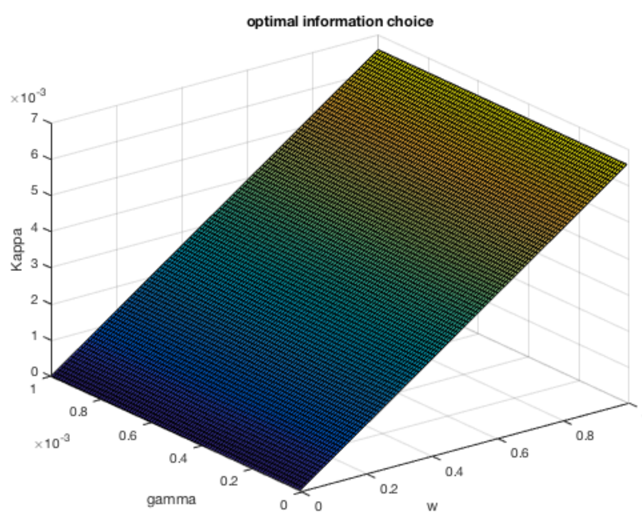


Figure 3.6: Optimal Information Choice

Corollary 3.4.2. *Information choice κ increases when w increases, θ increases, σ decreases, δ decreases, ρ decreases and γ increases. In the limit, when uncertainty γ vanishes, agents don't buy information.* ($\lim_{\gamma \to 0} \kappa = 0$).

Proof. See Appendix C.3. □

Not surprisingly, when the cost of information goes down ($\theta \uparrow$), 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. Second, there is an interesting interplay between the volatility parameter σ and information choice. On the one hand, increased σ implies a noisier investment environment, which reduces the marginal benefit of paying money to understand mean growth rate μ. On the other hand, an increased σ also implies that the asset return is a less reliable signal. For better estimation, one would demand more external information, which implies that κ should increase. Interestingly, the latter effect turns out to be at most of second order importance. To a first-order approximation, the first effect dominates. Third, when people live longer ($\delta \downarrow$), the lifetime benefit of acquiring information increases. Therefore, agents acquire more information. This effect is enhanced with rising patience ($\rho \downarrow$), since agents with more patience are more willing to sacrifice current consumption for the long-term benefit of acquiring information. Finally, the benefit from information increases with greater uncertainty. The intuition for this result will be clearer in the next section discussing belief dynamics. Briefly, the same amount of information reduces the *level* of uncertainty more when uncertainty is high, thus stimulating more demand for information.

3.4.2 Saving Rate

Let $\tilde{s}^*$ denote an agent's optimal saving rate. From equation (3.22), we know that

$$\tilde{s}^* \approx \underbrace{(1-\beta)}_{\textit{benchmark saving rate}} + \underbrace{\frac{\alpha}{2\sigma^2}\left[(\hat{\mu}-r)^2 + \frac{\gamma^2}{\beta\sigma^2}\right]}_{\textit{exogenous learning}} + \underbrace{\frac{1}{2}\theta\beta w\lambda\gamma^2}_{\textit{endogenous learning}} \tag{3.25}$$

Therefore, to a first-order approximation, one can decompose the saving 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 percentage of wealth equivalent 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. With CRRA utility, when $\alpha < 0$, the intertemporal elasticity of substitution is less than unity, so that agents are less willing to substitute consumption across time. The reverse is true when $\alpha > 0$.

The last component of saving reflects information choice. An important observation is that this component is wealth dependent. *Ceteris paribus*, wealthy agents have higher saving rates. This is consistent with the data.[14] 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. Figure 3.7 illustrates this effect quantitatively. 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 information choice model says something interesting on saving rate qualitatively, its important quantitative implications are not about saving.

[14]See, for example, Carroll (2001).

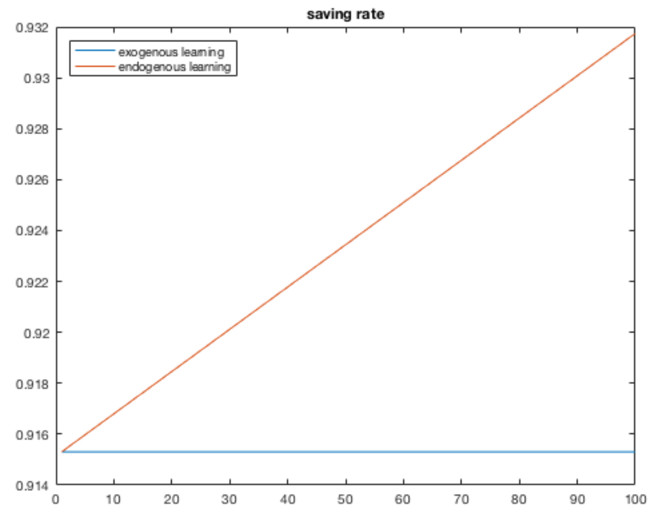


Figure 3.7: Saving Rate

3.4.3 Portfolio Choice

Just like saving rate, one can also decompose optimal portfolio choice into three pieces.

$$\pi^* \approx \frac{1}{\sigma^2}\ [\underbrace{(1+\alpha)(\hat{\mu}-r)}_{myopic\ portfolio} + \underbrace{\alpha\frac{(\hat{\mu}-r)\gamma}{\beta\sigma^2}}_{exogenous\ learning} + \underbrace{\frac{1}{2}\theta w\lambda(\hat{\mu}-r)\gamma^2}_{endogenous\ learning}\ ] \tag{3.26}$$

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. Going back to equation (3.20), we will discover the source of hedging demand, which comes from the term $\frac{-V_{w\hat{\mu}}\gamma}{V_{ww}w\sigma^2}$. We know that $V_{ww} < 0$. Therefore, the sign of hedging demand is determined by $V_{w\hat{\mu}}$, which in turn depends on the sign of α. On the one hand, a higher mean estimate makes the agent want to take the opportunity of investing more in risky assets. On the other hand, estimates are mean reverting, which implies a potential downward adjustment in future estimates. Uncertainty of the mean estimates makes the agent to invest less than the myopic portfolio. For a log utility agent, these two effects cancel out exactly. When $\alpha > 0$, the first effect dominates, and when $\alpha < 0$, the second effect dominates. 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. The presence of information choices increases portfolio demand directly, whose effect is stronger the wealthier an agent is.

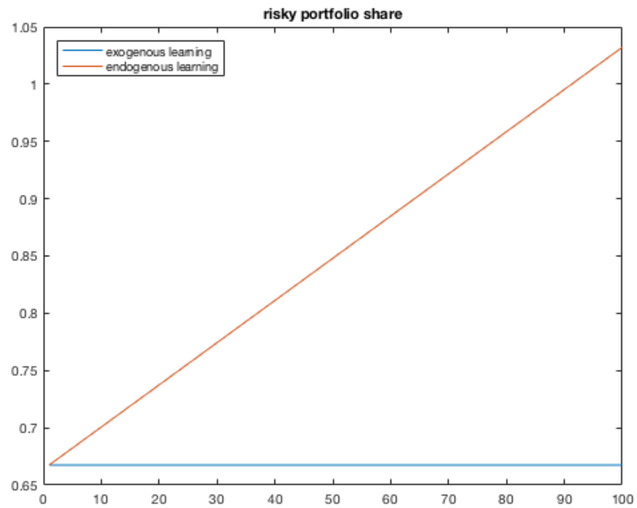


Figure 3.8: Risky Portfolio Share

Figure 3.8 illustrates the portfolio share quantitatively. 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. In some sense, they might appear to be less risk averse than poorer investors. However, such apparent “risk-loving" behavior is driven by endogenous information choice.

3.4.4 Optimism vs. Confidence

From above, we know that endogenous information introduces more risk taking among wealthy investors. However, what does it do for beliefs? With a Gaussian information structure, beliefs are fully summarized by the conditional mean and variance. We can define “optimism" by $\hat{\mu}$, and confidence by γ^{-1}. Figure 3.9 compares the effect of optimism in both the exogenous and endogenous learning cases by setting estimation variance and wealth at its initial value. Greater optimism leads to more risk taking in both cases, and the effect is stronger in the endogenous information case, making all agents appear less risk averse.

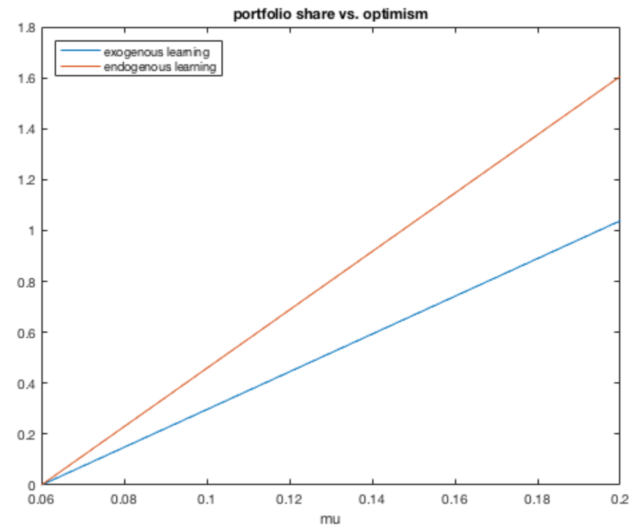


Figure 3.9: Portfolio Share vs. Optimism

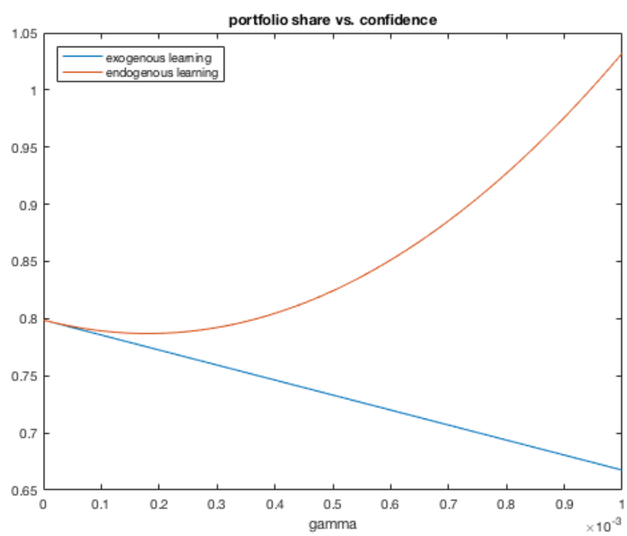


Figure 3.10: Portfolio Share vs. Confidence

It is interesting to examine how the level of confidence affects portfolio choice, and how it depends on the nature of information. Without information choice, a higher γ implies less confidence in one's estimate. A relatively risk-averse agent ($\alpha < 0$) would want to reduce her risky portfolio. Although this effect is still present in the endogenous information case, it is dominated by the other effect: *Ceteris Paribus*, lower confidence raises the marginal benefit of acquiring more information (equation (3.24)), which in turn increases confidence, and stimulates more risk taking. In general, the overall effect is ambiguous, depending on agents' risk preference and the learning environment. Nevertheless, with reasonable parameters, Figure 3.10 illustrates that the second effect dominates when γ is high, and *vice versa*. This is because the marginal benefit of acquiring information is higher when the level of uncertainty is higher.

3.5 The Distribution of Wealth and Beliefs

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.

3.5.1 Individual Wealth and Belief Dynamics

We 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 3.5.1. *To a first-order approximation, an individual's wealth dynamics under filtration* $\{\mathcal{F}_t\}$ *are governed by the diffusion*

$$dw = (a + \theta wb)wdt + (c + \theta w\hat{d})wdB \tag{3.27}$$

where

$$a = \frac{(\mu - r)}{\sigma^2}\left[(1+\alpha)(\hat{\mu} - r) + \alpha\frac{(\hat{\mu} - r)\gamma}{\beta\sigma^2}\right] + \frac{\alpha}{2\sigma^2}\left[(\hat{\mu} - r)^2 + \frac{\gamma^2}{\beta\sigma^2}\right] \tag{3.28}$$

$$b = \frac{\lambda\gamma^2(\mu - r)(\hat{\mu} - r)}{2\sigma^2} + \frac{\beta}{2}\lambda\gamma^2 - \frac{\gamma^2}{2\beta^4\sigma^4} \tag{3.29}$$

$$c = \frac{1}{\sigma}\left[(1+\alpha)(\hat{\mu} - r) + \alpha\frac{(\hat{\mu} - r)\gamma}{\beta\sigma^2}\right] \tag{3.30}$$

$$\hat{d} = \frac{1}{2\sigma}\lambda(\hat{\mu} - r)\gamma^2 \tag{3.31}$$

Proof. See Appendix C.4. □

Notice that the drift term collapses to $a(\hat{\mu}, \gamma)$ as $\theta \to 0$. This is the case when information is infinitely costly. Thus, agents learn exogenously. Notice also from equation (3.28) that the effect of learning on wealth depends on α. For the knife-edge case of log utility ($\alpha = 0$), parameter uncertainty does not affect the growth rate. To a first order approximation, we know that risk aversion does not affect information choice. However, risk aversion does alter an agent's portfolio composition. As long as $\alpha < 0 (\alpha > 0)$, learning decreases (increases) the growth rate of wealth because hedging causes the agent to invest less (more) in risky assets. The drift term also has a wealth-dependent piece as long as $\theta > 0$. The benchmark parameters ensure that $b(\hat{\mu}, \gamma) > 0$. Therefore, the growth rate is higher when either θ or w is higher. Since higher θ implies cheaper information, agents acquire more information, and invest more in risky assets. However, wealthy agents invest even more. This generates "scale dependent" growth.

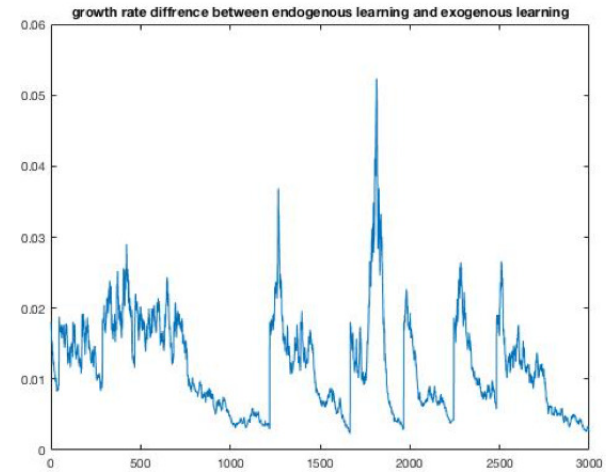


Figure 3.11: Sample Path: Growth Rate Difference

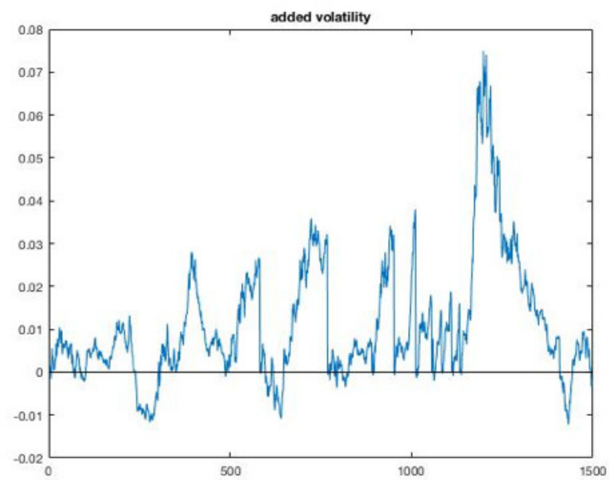


Figure 3.12: Sample Path: Volatility Difference

This scale dependence is illustrated in Figure 3.11, which plots a sample path of the growth rate difference between the exogenous learning and endogenous learning cases. Although it fluctuates between $0 - 6\%$ due to shocks, the difference is always positive. A byproduct of the information model is that the volatility of wealth is also wealth dependent. $c(\hat{\mu}, \gamma)$ can be higher or lower than in the no learning case, again depending on the sign of α. However, a positive $d(\hat{\mu}, \gamma)$ ensures that volatility increases with wealth. Figure 3.12 plots the added volatility from endogenous information. Although occasionally the number goes negative due to noise, it is most of the time positive, due to a higher risky portfolio share. This is consistent with the data because wealthy people are known to have more risky portfolios, and thus are more vulnerable to financial shocks.[15] This could be an important channel on its own, which says something interesting about the second moment of wealth inequality. However, I will focus on the mean rather than volatility of wealth inequality for now.

Lemma 3.5.2. *The following equations characterize approximate belief dynamics under the data-generating filtration $\{\mathcal{F}_t\}$. The conditional variance follows*

$$d\gamma = -\gamma^2 \left(\frac{1}{\sigma^2} + \frac{2\theta w}{\beta^2 \sigma^2} \right) dt \tag{3.32}$$

and the conditional mean follows

$$d\hat{\mu} = \frac{\gamma(\mu - \hat{\mu})}{\beta^2 \sigma^2}[\beta^2 + 2\theta w]dt + \frac{\gamma}{\sigma} dB + \frac{\sqrt{2\theta w}\gamma}{\beta\sigma} dB_y \tag{3.33}$$

Proof. This comes from substituting the optimal information choice eq.(3.24) into the Kalman filter eqs. (3.15) and (3.16) after using κ to substitute out for σ_y^2. □

From eq. (3.32), we see that the estimation variance of wealthy investors decreases faster. Another way of expressing eq. 3.32 is $\frac{d\gamma}{\gamma} = -\left(\frac{\gamma}{\sigma^2} + 2\kappa^*\right) dt$. That is, higher information capacity increases the *rate* at which uncertainty is reduced. Therefore, for a fixed κ, its marginal benefit in reducing γ is higher when the level of γ is higher. The conditional mean $\hat{\mu}$ is mean-reverting. Higher wealth tilts the agent's attention from the free signal dB towards the costly private signal dB_y due to its increased relative informativeness. Therefore, the drift of this mean-reversion process also increases with wealth in response to the increased sensitivity to the new external signal.

3.5.2 Numerical Solution

A stationary distribution is computed numerically using Monte Carlo simulation. 5000 agents are endowed with initial wealth $w = w_0$. Wealth and belief dynamics are simulated using eqs. (3.14), (3.15) and (3.16), with time step size discretized such that $dt \approx 0.3$. Agents experience birth and death with probability δdt at each time step. After simulating the economy for 300 years, a stationary distribution is achieved. Constraints on parameter values constrain the parameter space. These constraints are discussed in detail in Appendix C.9. Moskowitz and Vissing-Jorgensen (2002) provide evidence using samples from the Survey of Consumer Finances up until 1998 that the mean annual return to private equity is about 13.2%. Kartashova (2014) extends the SCF sample till 2010, and using her Table 5 results, I compute its average annual mean return to be equal to 16.53%. I choose the closest possible combinations of μ and σ to these empirical values, which implies that $\mu = 0.15$.[16] Benhabib and Bisin (2016) cite evidence on the returns to owner-occupied housing and private equity which suggest a standard deviation in the range of $10 - 20\%$. I set the closest value to this range, that also satisfies parameter space constraints , which implies that $\sigma = 0.26$.

[15] Think of the sharp decrease in wealth inequality following the financial crisis, where very rich people got very unlucky because their wealth is mostly in unsafe assets.

[16] Following Ehling, Graniero, and Heyerdahl-Larsen (2014), I set prior mean $\hat{\mu}_0 = \mu$, so that all agents start with the true mean estimate.

To satisfy the constraints, β is chosen to be 0.06, and $\delta = 0.04$. This implies a mean (work) life of 25 years. I pose the following thought experiment: Suppose that information is infinitely costly to acquire before 1980 ($\theta = 0$). Then suddenly investors have access to information by paying a finite but large information cost ($\theta > 0$). How does this change the stationary distribution of wealth? To answer this, I first jointly select the two remaining free parameters, risk aversion α and the initial estimation variance, γ_0, such that the exogenous learning economy ($\theta = 0$) matches the observed top 1% share in 1980, which is about 23.9%. I then chose the one combination of γ_0, α such that they are as close as possible to log utility, consistent with previous analytical approximation. This implies that $\alpha = -0.4$ and $\gamma_0 = 0.001$. This is in keeping with the analytic assumption of a "small" α. Although γ_0 might seem quite small, we shall see that even this modest degree of uncertainty can generate quantitatively significant effects. Lastly, the choice of θ is clearly important. However, since all the action comes from the multiplication of θ and w, not θ per se, we must calibrate θw jointly. This is pinned down by matching the average observed portfolio difference between 1983 and 2013, which is around 30% in SCF data. From the portfolio equation, this implies that the average $\theta w = 0.0016$ at the mass point. Finally, I choose $w_0 = 1$ and $\theta = 0.0016$ for convenience.

Figure 3.13 compares the stationary distributions of wealth in these two cases. The blue line pertains to the distribution in 1980. The red line pertains to the endogenous learning economy. We can see that the endogenous learning economy generates a fatter right tail, implying higher concentration of wealth at the top. The implied top 1% wealth share increases to 32.9%, nine percentage points higher than in the exogenous learning economy.

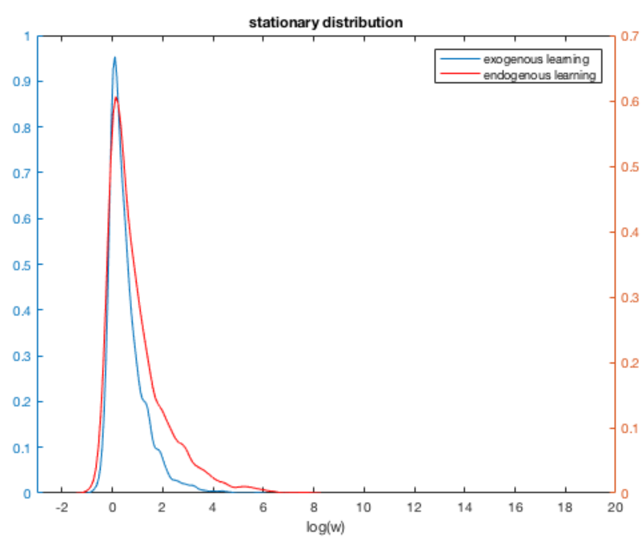


Figure 3.13: Stationary Wealth Distributions

3.5.3 The KFP Equation

Although we have just seen that it is straightforward to simulate the stationary cross-sectional distribution of wealth, for some questions it is convenient to characterize this distribution analytically. For example, doing so will clarify the interaction between belief heterogeneity and wealth inequality. We shall also see that learning induced inequality 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. Analytical characterization of the cross-sectional distribution of wealth and beliefs can be done 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.[17]

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

Proposition 3.5.3. *Let $x = \log(w)$. To an $O(\theta)$ approximation, each agent's states evolves according to the diffusion*

$$\begin{bmatrix} dx \\ d\hat{\mu} \\ d\gamma \end{bmatrix} = \begin{bmatrix} a - \frac{1}{2}c^2 + \theta(b - cd)e^x \\ \frac{\gamma(\mu-\hat{\mu})}{\beta^2\sigma^2}[\beta^2 + 2\theta e^x] \\ -\gamma^2(\frac{1}{\sigma^2} + \frac{2\theta e^x}{\beta^2\sigma^2}) \end{bmatrix} dt + \begin{bmatrix} c + \theta d e^x & 0 & 0 \\ \frac{\gamma}{\sigma} & \frac{\sqrt{2\theta e^x}\gamma}{\beta\sigma} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} dB \\ dB_y \\ 0 \end{bmatrix} \tag{3.34}$$

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)

$$\begin{aligned} 0 \quad = \quad & f\left[\frac{3\gamma}{\sigma^2} - \delta + \theta e^x(\frac{6\gamma}{\beta^2\sigma^2} + \frac{\lambda\gamma^2}{\sigma^2} + 2cd - b)\right] + f_x\left[\frac{1}{2}c^2 - a + \frac{\gamma}{\sigma}c_{\hat{\mu}} + \theta e^x(3cd - b + \frac{\lambda\gamma^2}{\sigma^2})\right] \\ & + f_{\hat{\mu}}\left[\frac{\gamma(\hat{\mu}-\mu)}{\sigma^2} + \theta e^x\left(\frac{2\gamma(\hat{\mu}-\mu)}{\beta^2\sigma^2} + \frac{\gamma d}{\sigma}\right)\right] + f_\gamma\gamma^2\left(\frac{1}{\sigma^2} + \frac{2\theta e^x}{\beta^2\sigma^2}\right) + \left(\frac{1}{2}c^2 + \theta e^x cd\right) f_{xx} \\ & + \frac{1}{2}f_{\hat{\mu}\hat{\mu}}\left(\frac{\gamma^2}{\sigma^2} + \frac{2\theta e^x\gamma^2}{\beta^2\sigma^2}\right) + (c + \theta e^x d)\frac{\gamma}{\sigma}f_{x\hat{\mu}} + \delta\zeta(X - X_0) \end{aligned} \tag{3.35}$$

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 Appendix C.5. □

A mere glance at eq. (3.35) will send shivers down the spine of anyone who has tried to solve PDEs. Although it 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 (3.34), 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.[18]

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

[17] Gabaix, Lasry, Lions, and Moll (2016) focus on transition dynamics. Kasa and Lei (2017) use the time-dependent KFP equation to argue that ambiguity accelerates transition dynamics.

[18] **?** provide a good textbook description of averaging and time-scale separation methods.

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 (5.35). 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 3.5.4. *To an $O(\gamma, \theta)$ approximation, the stationary distribution of $x(> 0)$ is*

$$f_1(x|\hat{\mu}, \gamma) = A_0(\hat{\mu}, \gamma) e^{\phi(\hat{\mu}, \gamma)x} + \theta A_1(\hat{\mu}, \gamma) e^{[1+\phi(\hat{\mu}, \gamma)]x} \tag{3.36}$$

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. See Appendix C.6. ☐

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

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

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

$$\mathcal{L}[h] = \left(\frac{3\gamma}{\sigma^2} - \delta\right) h + \frac{\gamma}{\sigma^2}(\hat{\mu} - \mu) h_{\hat{\mu}} + \frac{\gamma^2}{\sigma} h_\gamma + \frac{1}{2}\frac{\gamma^2}{\sigma^2} h_{\hat{\mu}\hat{\mu}} \tag{3.37}$$

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

$$h(\hat{\mu}, \gamma) \equiv N^{IG}(\hat{\mu}, \gamma) = \frac{1}{\sqrt{2\pi}\sigma\gamma^3} \exp\left[-\frac{(\hat{\mu} - \mu)^2}{2\gamma} - \frac{\delta\sigma^2}{\gamma}\right] \tag{3.38}$$

Proof. Take derivatives and substitute into eq. (3.37). □

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. The details are relegated to the Appendix C.7.

Proposition 3.5.7. *To an $O(\gamma, \theta)$ approximation, the stationary cross-sectional distribution of beliefs is*

$$f_2(\hat{\mu}, \gamma) = B_0 N^{IG}(\hat{\mu}, \gamma) + \theta B_1 G(\hat{\mu}, \gamma) \tag{3.39}$$

where (B_0, B_1) are normalizing constants, and $G(\hat{\mu}, \gamma)$ is a function defined in the Appendix C.7.

Proof. See Appendix C.7. □

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 3.5.4 with the marginal distribution for beliefs derived in Proposition 3.5.7, we get the following characterization of the marginal distribution of (log) wealth

Proposition 3.5.8. *To an $O(\gamma, \theta)$ approximation, the stationary cross-sectional marginal distribution of (log) wealth is (for $x > 0$)*

$$\begin{aligned} \Lambda(x) &= \int_0^{\gamma_0} \int_{-\infty}^{\infty} f_1(x|\hat{\mu}, \gamma) f_2(\hat{\mu}, \gamma) d\hat{\mu} d\gamma \quad (3.40) \\ &= \int_0^{\gamma_0} \int_{-\infty}^{\infty} \left\{ A_0 B_0 e^{\phi x} \cdot N^{IG}(\hat{\mu}, \gamma) + \theta \left[A_0 B_1 e^{\phi x} \cdot G(\hat{\mu}, \gamma) + A_1 B_0 e^{(\phi+1)x} \cdot N^{IG}(\hat{\mu}, \gamma) \right] \right\} d\hat{\mu} d\gamma \end{aligned}$$

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 $(\hat{\mu}, \gamma) = (\mu, 0)$,

$$(\phi + 1)x \approx \phi_0 x + [1 + \phi_{\hat{\mu}}(\hat{\mu} - \mu) + .5\phi_{\hat{\mu}\hat{\mu}}(\hat{\mu} - \mu)^2 + \phi_\gamma \gamma]x \tag{3.41}$$

Note that ϕ_0 is the coefficient in a standard scale independent random growth model without learning. Hence, to assess the impact of endogenous information and learning 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,

Corollary 3.5.9. *For large x, the (right) tail Pareto exponent with learning, 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 and learning increase top wealth shares.

Proof. See Appendix C.8. □

Substituting the parameter values from Table 1 into the expression $\phi_0 = \frac{\hat{a} - \sqrt{\hat{a}^2 \pm 2c^2\delta}}{c^2}$, where $\hat{a} = a - \frac{c^2}{2}$, we find $\phi_0 = -1.49$. This is the Pareto exponent without learning or endogenous information. It implies a top 1% wealth share of 23.9%. Then, with learning and endogenous information, Corollary 5.5 implies $\phi_L = -0.43$. One cannot easily translate this into an implied wealth share, since the distribution is no longer Pareto, and this exponent pertains to an exponential that is scaled by the small parameter θ. Still, such a small (absolute) value points to a high level of inequality.

3.6 Additional Evidence

The key mechanism in this paper is that wealthier households have more information, and this information leads them to invest more in riskier, higher yielding assets. In this section I use data from the hedge fund industry and the Survey of Consumer Finances (SCF) to lend support to this mechanism.

3.6.1 Results

To empirically examine how wealth affects risk taking and the decision to pay for wealth management, I use SCF data to run two regressions. Both regressions attempt to control for life-cycle effects and demographic characteristics that are absent from my model. The first regresses the percentage of wealth invested in risky assets on log wealth,
IV regression:

$$\begin{aligned} Percentagewealthinriskyassets_i =& \alpha + \beta_1 logwealth_i + \beta_2 age_i + \beta_3 age_i^2 \\ &+ \sum_s \beta_{4s} I(race_s) + \sum_s \beta_{5s} I(education_s) + \epsilon_i \end{aligned} \tag{3.42}$$

A potential endogeneity problem arises here since the dependent variable is deflated by wealth, which also appears as a regressor. In response, I use log wage income and log housing values as instruments for log wealth. Both are clearly correlated with wealth, but should be less correlated with portfolio shares.

The second regression examines the link between wealth an information. I run a logit regression of the probability of paying for wealth management on log wealth, again controlling for life-cycle effects and demographics. Although in terms of my model it would be preferable to use the amount spent on information as the dependent variable, rather than the 0/1 choice of paying for wealth management, this data is not available. *Logit regression:*

$$\begin{aligned} Probability of delegating wealth_i =& \alpha + \beta_1 logwealth_i + \beta_2 age_i + \beta_3 age_i^2 \\ &+ \sum_s \beta_{4s} I(race_s) + \sum_s \beta_{5s} I(education_s) + \epsilon_i \end{aligned} \tag{3.43}$$

Table 2 displays the results from both regressions. All asset values are expressed in 2013 dollars, and robust standard errors are used.

Table 3.2: Are the rich more likely to delegate wealth? Do they invest more in stocks?

	(1983 p.stock)	(2013 p.stock)	(prob.delegate wealth(2013))
	(1)	(2)	(3)
log wealth	6.084 (0.495)	8.280 (0.150)	1.337 (0.078)

Notice that in both 1983 and 2013 an increase in wealth produces a highly significant increase in the share invested in risky assets. As predicted, the effect is larger in 2013. The dependent variable has been scaled by 100, so a coefficient of 8.28 suggests that a 10% increase in wealth produces a 8.28 percentage point increase in the share invested in risky assets. This is a stronger effect than implied by the benchmark parameterization of my model (see, e.g., Figure **??**). The coefficient in the logit regression suggests that in 2013 a 10% increase in wealth produces a highly significant 13% increase in odds of paying for wealth management.

What does this increased risk taking and wealth delegation imply about rates of return on wealth? To investigate this, I regress after tax and after fee private equity returns on log wealth in both 1983 and 2013.[19] Table 3.3 reports results of the following cross-sectional IV regression of returns on log wealth, again controlling for individual household characteristics.

[19]The same endogeneity issue arises here, since only previous survey year income/profit and current survey year market value of private equity are reported in the survey, but not previous survey year market value. Therefore, if we define the dependent variable return on equity as the ratio of previous year income/profit to current year market value, the denominator appears on both sides, which creates a potential bias of the estimate. Again, log wage and log housing values are used as instruments.

IV regression

$$\begin{aligned} Returnonequity_i = & \alpha + \beta_1 logwealth_i + \beta_2 risktolerance_i \\ & + \beta_3 age_i + \sum_s \beta_{4s} I(year_s) + \sum_s \beta_{5s} I(education_s) + \epsilon_i \end{aligned} \tag{3.44}$$

Table 3.3: Do the rich have higher returns on wealth?

	ROR 1983(percent)	ROR2013(percent)
	(1)	(2)
log wealth	3.350e-02 (8.038e-3)	4.049e-02 (2.589e-03)

Although the rich have always enjoyed higher returns, we can see the effect is stronger in 2013. In 1983, a 10% increase in wealth was associated with a 33.5 basis point increase in the rate of return. In 2013, this same increase corresponded to a 40.49 basis point increase. Clearly, these estimates overstate return differentials in the extreme tails of the distribution, where wealth can different by several orders of magnitude.

Next, I turn to the Lipper Hedge fund dataset to examine the relationship between hedge fund performance and their fee structures. We know wealthy investors are more likely to delegate their wealth to these hedge funds, but how did they perform? Figure 3.14 compares the average annualized after fee rate of return on investment in different hedge funds with a passive investment strategy of holding the $S\&P$ 500 index. The red vertical line represents the annualized mean return of the $S\&P$ 500, and the histogram shows the distribution of average annualized returns for a sample of hedge funds. The sample runs from 1977 January to 2015 January, which is a reasonable indicator of their long-term performance. A common perception is that hedge fund performance has been declining, with average performance even worse than holding a passive stock index. In the data, I do indeed observe this pattern for the *average* hedge fund. However, if we look at the whole distribution of these funds, more than 30% of them have superior after fee returns than holding a passive $S\&P$ 500 index.

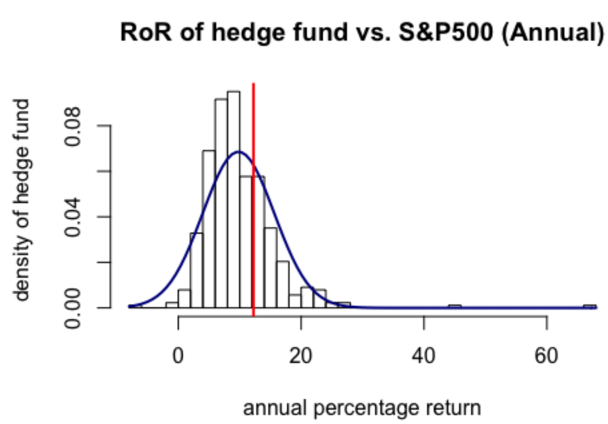


Figure 3.14: Distribution of Hedge Fund ROR vs. $S\&P$500

Notes: This histogram plots the density of annualized rate of returns. The $S\&P500$ historical return data is taken from NYU Stern school (Aswath Damodaran's website).

It is therefore of interest to identify those hedge funds that have superior returns. Hedge funds differ substantially in terms their investment type, leverage ratio, lock up period, and high water marks. Our focus is on the fee structure and return performance. A typical hedge fund usually charges two types of fees: a management fee as a percentage of total wealth delegated to the fund, and an incentive fee charged as a percentage of the return. I run the following random effects panel regression to study the effect of these two fees on hedge fund monthly return performance.

TASS random-effect panel regression

$$\begin{aligned} RateofReturn_{i,t} =& \alpha + \beta_1 incentivefee_{i,t} + \beta_2 managementfee_{i.t} + \beta_3 netassetvalue_{i,t} \\ &+ \beta_4 highwatermark_{i,t} + \beta_5 lockupperiod_{i,t} + \beta_6 averageleverage_{i,t} \\ &+ \sum_s \beta_{7s} I(primarycategory_s) + \sum_s \beta_{8s} I(year_s) + \epsilon_{i,t} \end{aligned} \tag{3.45}$$

Here a random effects model makes more sense than a fixed effect model because fees are typically set at the inception date, and do not usually change over time. This makes it more reasonable in understanding how heterogeneity of fees *between* different funds affects their return performance, rather than how changes in fees *within* a fund affect its time-series return performance . For now, let's focus on the main results presented in Table 3.4.

Table 3.4: Do hedge fund fees affect hedge fund performance (all funds, annual)?

	RoR(all years)	RoR(before 2008)	RoR(after 2008)
incentive fee	0.092 (0.016)	0.099(0.023)	0.087 (0.019)

The key result here is that better performance is associated with higher fees.[20] This is consistent with my model, which features a quadratic information cost. Improved return estimates can be purchased, but the marginal cost is increasing.

3.7 Related Literature

As discussed earlier, this paper is a direct descendent of work by Arrow (1987), Benhabib, Bisin, and Zhu (2016), and Turmuhambetova (2005). However, it can also be viewed as part of larger literature attempting to incorporate endogenous information into macroeconomics and finance. This literature is nicely 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

[20] Liang (1999) finds the same positive effect of incentive fees on return performance using Hedge Fund Research data.

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, Nosal, and Stevens (2014), work by Van Nieuwerburgh and Veldkamp (2010) and Batchuluun, Luo, and Young (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, as noted by Sims (2006), rational inattention is perhaps not a great fit to financial markets, where information is scarce and costly. 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.[21]

Finally, this paper is related to recent work on 'financial literacy' and asset market participation.[22] 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.

3.8 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

[21] 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.

[22] Examples include Lusardi, Michaud, and Mitchell (2017), Vissing-Jorgensen (2003), Van Rooij, Lusardi, and Alessie (2011), and Luo, Nie, and Wang (2017).

long ago, information naturally lends itself to increasing returns, and when increasing returns are combined with idiosyncratic shocks, wealth inequality emerges.

Some people argue 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 in my model, 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. **?** 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 in my model as well, if households were allowed to pool risks. However, since my model features finite lifetimes and no secular growth in per capita initial 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, Lasry, Lions, and Moll (2016) suggests that the real challenge is to understand transition dynamics. Why has inequality grown so *rapidly*? Transition dynamics could easily be studied here numerically, but the time-scale separation approach I used to characterize the stationary distributions does not extend easily to transition dynamics. 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.

[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.

Bibliography

Abramowitz, M., and I. A. Stegun (1964): *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, vol. 55. Courier Corporation.

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

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

Alba, J. D., and P. Wang (2017): "Taylor rule and discretionary regimes in the United States: Evidence from a k-state Markov regime-switching model," *Macroeconomic Dynamics*, 21, 817–833.

Alvarez, F., and A. Dixit (2014): "A real options perspective on the future of the Euro," *Journal of Monetary Economics*, 61, 78–109.

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

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

Andrei, D., and M. Hasler (2014): "Optimal asset and attention allocation," *Available at SSRN 2516984*.

Angeletos, G.-M. (2007): "Uninsured Idiosyncratic Investment Risk and Aggregate Saving," *Review of Economic Dynamics*, 10, 1–30.

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

Arrow, K. J. (1987): "The demand for information and the distribution of income," *Probability in the Engineering and Informational Sciences*, 1(01), 3–13.

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

Barndorff-Nielsen, O. E., T. Mikosch, and S. I. Resnick (2001): *Lévy Processes: Theory and Applications*. Springer.

Batchuluun, A., Y. Luo, and E. Young (2017): "Portfolio Choice with Information-Processing Limits," working paper, Hong Kong University.

Benhabib, J., and A. Bisin (2016): "Skewed wealth distributions: Theory and empirics," Discussion paper, National Bureau of Economic Research.

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

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

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

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

Blinder, A. S. (2009): "Making monetary policy by committee," *International Finance*, 12(2), 171–194.

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. J. Terry (2012): "Really uncertain business cycles," Discussion paper, National Bureau of Economic Research.

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

Cao, D., and W. Luo (2017): "Persistent heterogeneous returns and top end wealth inequality," *Review of Economic Dynamics*, 26, 301–326.

Caplin, A., and J. Leahy (1994): "Business as usual, market crashes, and wisdom after the fact," *American Economic Review*, 84(3), 548–565.

Carroll, C. D. (2001): "A theory of the consumption function, with and without liquidity constraints," *The Journal of Economic Perspectives*, 15(3), 23–45.

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

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

Davig, T., and E. M. Leeper (2008): "Endogenous monetary policy regime change," in *NBER International Seminar on Macroeconomics 2006*, pp. 345–391. University of Chicago Press.

Dixit, A. K. (1993): *The Art of Smooth Pasting*. Taylor & Francis.

——— (1994): *Investment Under Uncertainty*. Princeton University Press.

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

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

Dynkin, E. B. (1965): *Markov Processes*. Springer.

Ehling, P., A. Graniero, and C. Heyerdahl-Larsen (2014): "Asset prices and portfolio choice with learning from experience," *Available at SSRN 2378330*.

——— (2016): "Asset prices and portfolio choice with learning from experience," .

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

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

——— (2016b): "Heterogeneity and persistence in returns to wealth," Discussion paper, National Bureau of Economic Research.

Favilukis, J. (2013): "Inequality, stock market participation, and the equity premium," *Journal of Financial Economics*, 107(3), 740–759.

Fernández-Villaverde, J., O. Posch, and J. F. Rubio-Ramírez (2012): "Solving the New Keynesian Model in Continuous Time," *University of Hamburg, Mimeo*.

Froot, K. A., and M. Obstfeld (1991): "Stochastic process switching: some simple solutions," Discussion paper, National Bureau of Economic Research.

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

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

Galí, J. (2009): *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton University Press.

Gali, J., and M. Gertler (1999): "Inflation dynamics: A structural econometric analysis," *Journal of Monetary Economics*, 44(2), 195–222.

Gel'fand, I. M., and N. Y. Vilenkin (2014): *Generalized Functions: Applications of Harmonic Analysis*. Academic Press.

Gennotte, G. (1986): "Optimal portfolio choice under incomplete information," *The Journal of Finance*, 41(3), 733–746.

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

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

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

Grossman, S. J., and J. E. Stiglitz (1980): "On the impossibility of informationally efficient markets," *The American economic review*, pp. 393–408.

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

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

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

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

Holston, K., T. Laubach, and J. Williams (2016): "Measuring the natural rate of interest: International trends and determinants," in *NBER International Seminar on Macroeconomics 2016*. Elsevier.

Kacperczyk, M., J. B. Nosal, and L. Stevens (2014): "Investor Sophistication and Capital Income Inequality," Discussion paper, National Bureau of Economic Research.

Kahn, G. A. (2010): "Taylor rule deviations and financial imbalances," *Federal Reserve Bank of Kansas City Economic Review, Second Quarter*, 2010, 63–99.

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

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

Kartashova, K. (2014): "Private equity premium puzzle revisited," *The American Economic Review*, 104(10), 3297–3334.

Kasa, K., and X. Lei (2017): "Risk, Uncertainty, and the Dynamics of Inequality," *Simon Fraser University Working Paper*, 17, 06.

KLIBANOFF, P., M. MARINACCI, AND S. MUKERJI (2005): “A Smooth Model of Decision Making Under Ambiguity,” *Econometrica*, 73, 1849–92.

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

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

LEVIN, A. T., V. WIELAND, AND J. WILLIAMS (1999): “Robustness of simple monetary policy rules under model uncertainty,” in *Monetary Policy Rules*, pp. 263–318. University of Chicago Press.

LEVIN, A. T., AND J. C. WILLIAMS (2003): “Robust monetary policy with competing reference models,” *Journal of Monetary Economics*, 50(5), 945–975.

LIANG, B. (1999): “On the performance of hedge funds,” *Financial Analysts Journal*, 55(4), 72–85.

LIPSTER, R. S., AND A. N. SHIRYAYEV (1989): *Theory of Martingales*. Kluwer.

LIPTSER, R. S., AND A. N. SHIRYAEV (2001): *Statistics of Random Processes II: II. Applications*, vol. 2. Springer Science & Business Media.

LITTMAN, W., G. STAMPACCHIA, AND H. F. WEINBERGER (1963): “Regular points for elliptic equations with discontinuous coefficients,” *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 17(1-2), 43–77.

LJUNGQVIST, L., AND T. J. SARGENT (1998): “The European Unemployment Dilemma,” *Journal of Political Economy*, 106, 514–550.

LUO, Y. (2016): “Robustly Strategic Consumption-Portfolio Rules with Information Frictions,” forthcoming in *Management Science*.

LUO, Y., J. NIE, AND H. WANG (2017): “Ignorance, Uncertainty, and Strategic Consumption-Portfolio Decisions,” working paper, Hong Kong University.

LUSARDI, A., P.-C. MICHAUD, AND O. S. MITCHELL (2017): “Optimal financial knowledge and wealth inequality,” *Journal of Political Economy*, 125(2), 431–477.

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

MALMENDIER, U., AND S. NAGEL (2016): “Learning from Inflation Experiences,” *Quarterly Journal of Economics*, 131, 53–87.

MANKIW, N. G., AND R. REIS (2002): “Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve,” *The Quarterly Journal of Economics*, 117(4), 1295–1328.

MARCET, A., AND T. J. SARGENT (1989): “Convergence of least-squares learning Mechanisms in Self-Referential Linear Stochastic Models,” *Journal of Economic Theory*, 48, 337–68.

MERTON, R. C. (1980): “On estimating the expected return on the market: An exploratory investigation,” *Journal of Financial Economics*, 8(4), 323–361.

MIAO, J., AND N. WANG (2011): “Risk, uncertainty, and option exercise,” *Journal of Economic Dynamics and Control*, 35(4), 442–461.

MOSKOWITZ, T., AND A. VISSING-JORGENSEN (2002): “The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?,” *American Economic Review*, 92, 745–78.

MURRAY, C. J., A. NIKOLSKO-RZHEVSKYY, AND D. H. PAPELL (2015): “Markov switching and the Taylor principle,” *Macroeconomic Dynamics*, 19, 913–930.

NAKOV, A., AND G. NUÑO (2015): “Learning from experience in the stock market,” *Journal of Economic Dynamics and Control*, 52, 224–239.

Øksendal, B. (1985): "Stochastic processes, infinitesimal generators and function theory," in *Operators and Function Theory*, pp. 139–162. Springer.

——— (2003): *Stochastic differential equations*. Springer.

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

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

Protter, P. E. (2004): *Stochastic Integration and Differential Equations*. Springer.

Quadrini, V. (2009): "Entrepreneurship in macroeconomics," *Annals of Finance*, 5(3), 295–311.

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

Rotemberg, J. J., and M. Woodford (1999): "Interest rate rules in an estimated sticky price model," in *Monetary Policy Rules*, pp. 57–126. University of Chicago Press.

Rudebusch, G. D. (2006): "Monetary policy inertia: fact or fiction?," *International Journal of Central Banking*, 2(4), 85–135.

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

——— (2016b): "Wealth inequality in the United States since 1913: Evidence from capitalized income tax data," *The Quarterly Journal of Economics*, 131(2), 519–578.

Sargent, T. J. (1999): *The Conquest of American Inflation*. Princeton University Press.

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

Sims, C. A. (2003): "Implications of rational inattention," *Journal of monetary Economics*, 50(3), 665–690.

——— (2006): "Rational inattention: Beyond the linear-quadratic case," *American Economic Review*, 96(2), 158–163.

——— (2010): "Rational inattention and monetary economics," *Handbook of Monetary Economics*, 3, 155–181.

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

Stokey, N. L. (2008): *The Economics of Inaction: Stochastic Control Models with Fixed Costs*. Princeton University Press.

——— (2013): "Wait-and-See: Investment Options under Policy Uncertainty," Discussion paper, National Bureau of Economic Research.

Svensson, L. E., and N. Williams (2008): "Optimal monetary policy under uncertainty: a Markov jump-linear-quadratic approach," *Federal Reserve Bank of St. Louis Review*, 90(July/August 2008).

Taylor, J. B. (2014): "The role of policy in the Great Recession and the weak recovery," *American Economic Review*, 104(5), 61–66.

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

Trojani, F., and P. Vanini (2002): "A Review Of Perturvative Approaches For Robust Optimal Portfolio Problems," *Computational methods in decision-making, economics and finance*, 74, 109.

Trojanowska, M., and P. M. Kort (2010): "The worst case for real options," *Journal of Optimization Theory and Applications*, 146(3), 709–734.

Turmuhambetova, G. (2005): "Decision making in an economy with endogenous information," University of Chicago Ph.D thesis.

Van Nieuwerburgh, S., and L. Veldkamp (2010): "Information acquisition and under-diversification," *The Review of Economic Studies*, 77(2), 779–805.

Van Rooij, M., A. Lusardi, and R. Alessie (2011): "Financial literacy and stock market participation," *Journal of Financial Economics*, 101(2), 449–472.

Veldkamp, L. L. (2011): "Information choice in macroeconomics and finance," pp. 17 – 27. Princeton University Press.

Verrecchia, R. E. (1982): "Information acquisition in a noisy rational expectations economy," *Econometrica: Journal of the Econometric Society*, pp. 1415–1430.

Vissing-Jorgensen, A. (2003): "Perspectives on behavioral finance: Does" irrationality" disappear with wealth? Evidence from expectations and actions," *NBER macroeconomics annual*, 18, 139–194.

Wang, J. (1993): "A model of intertemporal asset prices under asymmetric information," *The Review of Economic Studies*, 60(2), 249–282.

Woodford, M. (1999): "Optimal monetary policy inertia," *The Manchester School*, 67(s1), 1–35.

——— (2003): *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press.

Ziegler, A. C. (2012): *Incomplete information and heterogeneous beliefs in continuous-time finance*. Springer Science & Business Media.

Appendix A

"Wait and See" Monetary Policy

This section contains a proof that the derivation contained in Section 4 holds when the value function V only satisfies the Euler equations, or so called smooth pasting condition, rather than being C^2, i.e. twice differentiable with continuous second derivative. The difficulty is that the classical Itô's Lemma, and its corollary Dynkin's Formula, are applicable only for C^2-functions. A value function V constructed via smooth pasting is clearly not C^2 everywhere, in particular at points where the pasting is done. On its state space $\mathbb{R}$, the central bank's value function V is only C^1; differentiating twice as was done in (12) in Section 4 might not make sense. Here we show that our results remain true nevertheless. In addition to the current model, this formal argument is provided also with a view towards extending our analysis to more general settings discussed in the Conclusion, such as one where the central bank is allowed to switch between periods of continuous adjustment and no adjustment, or is restricted by a zero lower bound, or faces a cost-push shock that is a more general Lévy process with jumps, or a stochastic volatility model. In practice, our theorem is a general verification theorem, i.e. one that provides a sufficient condition—the Euler equation—for optimality.

We recall the following notation: For $\mathbb{R}^n$-valued Itô diffusion z_t with infinitesimal generator L_z and a bounded domain $W \subset \mathbb{R}^n$, consider the expectations operator $f(z) \mapsto E^z[\int_0^{\tau_V} L_z f(z_t)dt]$[1]. If $E^z[\tau_V] < \infty$ for all $z \in W$, then a finite measure, called the *Green measure*, can be defined on W by[2]

$$\int_V \phi(z')d\mu_z(z') = E^z[\int_0^{\tau_V} \phi(z_t)dt], \ \ \forall \phi \in C(\overline{W}).$$

In other words, the Green measure μ_z is the length of time that z_t, given the initial state z, is expected to spend in W. In our model, the W of interest is the conjectured inaction region $(a, b) \subset \mathbb{R}$ of the central bank.

[1]The *infinitesmal generator* L_z of an Itô diffusion z_t is the operator defined by, for $f : \mathbb{R}^n \to \mathbb{R}$ sufficiently smooth, $L_z f(z_0) = \lim_{\Delta t \to} \frac{E^{z_0}[f(z_{\Delta t})] - f(z_0)}{\Delta t}$. For example, L_z of the cost-push shock $dz_t = \theta\sigma dB_t$ is the second-order differential operator $\frac{1}{2}\theta^2\sigma^2\frac{d^2}{dz^2}$. For a central bank with value function V, $L_z V(z_0) = \frac{1}{2}\theta^2\sigma^2 V''(z_0)$ is therefore the expected marginal change of option value V under the law of dz_t with initial state z_0. This term, plus the flow cost, appears in the central bank's first order condition Equation (13).

[2]See, for example, Øksendal (1985).

If the differential operator L_z is uniformly elliptic on D, then for all $z \in D$, μ_z is absolutely continuous with respect to Lebesgue measure, in which case we denote its density by $d\mu_z(x') = G(z, z')dz'$. In fact, one has[3]

$$G(z, z') \in L^q_{loc}(\mathbb{R}^n), \text{ if } q < 1 + \frac{n}{2},$$

where $L^q_{loc}(\mathbb{R}^n)$ denotes the Lebesgue space of functions with finite L^q-norm when restricted to compact subsets.

As stated in Section 4, the central bank's problem can be reduced to an optimal stopping problem. Therefore we prove sufficiency of the Euler equation in terms of the general optimal stopping problem. Let $g : \mathbb{R}^n \to \mathbb{R}$ be Borel-measurable. The problem is to find

$$V(z) = \inf_{\tau} E^z[g(z_\tau)],$$

where $\inf_\tau$ denotes the infimum over all stopping times. In our context, g is the central bank's time discounted value function plus total flow cost due to deviations driven by the cost-push shock.

Let $V : \mathbb{R}^n \to \mathbb{R}$ that satisfy the following conditions:

(i) There exists a region $D \subset \mathbb{R}^n$ with C^1-boundary δD and $V \in C^2_b(\mathbb{R}^n \backslash \delta D)$.

(ii) $h|_D \geq g|_D$. (Continuation)

(iii) $L_z h = 0$ on D where L_z is the infinitesimal generator of z_t. (First order condition)

(iv) $h|_{\mathbb{R}^n - D} = g|_{\mathbb{R}^n - D}$ and $L_X h \leq 0$ on $\mathbb{R}^n - \overline{D}$. (Non-continuation)

(v) $h \in C^1(\mathbb{R}^n)$. (*Euler equation/C^1-smooth pasting condition*)

Then V solves the optimal stopping problem, with the infimum being attained by the first exit time of D.

Proof. Using Assumptions (i) (C^1-boundary and C^2 almost everywhere), and (v) (C^1-pasting), integration by parts shows that the second mixed partials $\delta_{z_i z_j} h$, defined almost everywhere (outside δD), is the distributional second derivative of V.[4][5]

By the boundedness assumption on V (part of (i)) and existence of distribution derivatives, V lies in the L^p-Sobolev space $W^{2,p}$ for any p. Since $C^2(\mathbb{R}^n)$ is dense in $W^{2,p}$, there exists a sequence $\{f_k\} \subset C^2(\mathbb{R}^n)$ such that

[3]See Theorem 5.4 in Littman, Stampacchia, and Weinberger (1963).

[4]The notion of *distributional derivative* generalizes the classical derivative to *tempered distributions*, or *generalized functions* (see Gel'fand and Vilenkin (2014)).

[5]Our proof strategy is as follows: First observe that V is C^2 except on a set of measure zero, where smooth pasting is applied. Its classical second derivative, defined almost everywhere, is in fact its distributional second derivative. Use the Sobolev Embedding Theorem to argue that the distributional second derivative can be well approximated by C^2-functions. Apply Dynkin's Formula to the approximating C^2-sequence and show that this is well-behaved with respect to taking appropriate limit.

$$\|f_k - h\|_{W^{2,p}} \to 0.$$

On the other hand, by the Sobolev Embedding Theorem, if $p > \frac{n}{2}$,[6]

$$\|f_k - h\|_\infty \to 0.$$

Now Dynkin's Formula holds for each f_k:

$$E^z[f_k(z_\tau)] = f_k(z) + \int_V L_z f_k(z')G(z,z')dz'.$$

So

$$\lim_{k\to\infty} E^z[f_k(z_\tau)] = \lim_{k\to\infty} f_k(z) + \lim_{k\to\infty} \int_V L_z f_k(z')G(z,z')dz'.$$

Choose $p > \frac{n}{2} + 1$, then its conjugate exponent q satisfies $q < 1 + \frac{2}{n}$. The boundedness assumption means that the Sobolev norm can be approximated by the uniform norm:

$$\lim_{k\to\infty} f_k(z) = V(z)$$

As a by-product, the approximating sequence $\{f_k\}$ can be chosen as regular, in the sense of large p, to accommodate possible singularities of the Green density: by choice of $\{f_k\}$,

$$\|L_z f_k(z') - L_z V(z')\|_{L^p} \to 0.$$

By Hölder's inequality,

$$|\int_V L_z f_k(z') - L_z V(z')G(z,z')dz'| \le \|L_z f_k(z') - L_z V(z')\|_{L^p} \cdot \|G(z,z')\|_{L^q} \to 0.$$

(Strictly speaking, we have assumed D is compact, but this is without loss of generality). So, for all z,

$$\lim_{k\to\infty} E^z[f_k(z_\tau)] = V(z) + \int_V L_z V(z')G(z,z')dz'.$$

By the Dominated Convergence Theorem,

$$E^z[V(z_\tau)] = \lim_{k\to\infty} E^z[f_k(z_\tau)].$$

So we have that a Dynkin's Formula holds for V:

$$E^z[V(z_\tau)] = V(z) + \int_W L_z V(z')G(z,z')dz'.$$

[6]See, for example, Chapter 4 of **?**.

It follows that h is superharmonic locally on δD, therefore everywhere (see Dynkin (1965), p22). This proves the theorem.

$\square$

Appendix B

Risk, Uncertainty and the Dyanamics of Uncertainty

B.1 Proof of Proposition 2.3.1

After substituting the h policy function in eq. (2.6) into the HJB equation in eq. (2.5) we get

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

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

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

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

$$\begin{aligned} c(w) &= V_w^{-\frac{1}{\gamma}} \\ &\approx [V_w^0 + \varepsilon V_w^1]^{-\frac{1}{\gamma}} \\ &\approx \left(V_w^0\right)^{-\frac{1}{\gamma}} - \varepsilon\frac{1}{\gamma}\left(V_w^0\right)^{-\frac{1}{\gamma}-1} V_w^1 \\ &\equiv c_0 + \varepsilon c_1 \end{aligned} \tag{B.2}$$

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

$$\begin{aligned} \alpha(w) &= -\frac{(\mu-r)}{\sigma^2}\left[\frac{V_w}{w(V_{ww} - \varepsilon V_w^2)}\right] \\ &\approx -\frac{(\mu-r)}{\sigma^2}\left[\frac{wV_w^0}{w^2 V_{ww}^0} + \varepsilon\frac{wV_w^1 V_{ww}^0 - wV_w^0(V_{ww}^1 - (V_w^0)^2)}{w^2 (V_{ww}^0)^2}\right] \\ &\equiv \alpha_0 + \varepsilon\alpha_1 \end{aligned} \tag{B.3}$$

Substituting these approximations into HJB equation in (B.1) gives us

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

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

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

Note this system is recursive. We can first solve the ε^0 equation for V^0, and then substitute this into the ε^1 equation. Solving for V^0 just gives the Merton solution. In particular, we conjecture $V^0(w)=\frac{A}{1-\gamma}w^{1-\gamma}$. Note, this implies $c_0=A^{-1/\gamma}w$ and $\alpha_0=\frac{(\mu-r)}{\gamma\sigma^2}$. After canceling the common $w^{1-\gamma}$ term we can solve for A. This produces the expression for A_0 stated in Proposition 2.3.1.

To solve the ε^1 equation we conjecture $V^1(w)=\frac{B}{\epsilon}w^{\epsilon}$, and try to solve for B and ϵ. From (A.26) and (A.27), this guess implies $c_1=-\gamma^{-1}A^{-1/\gamma-1}Bw^{2-\gamma}$ and $\alpha_1=-\left(\frac{\mu-r}{\sigma^2}\right)\left(\frac{A^2+(\gamma-1)B}{A\gamma^2}\right)w^{1-\gamma}$. Note, these now depend on w. Substituting these into the ε^1 equation, we find that if $\epsilon=2(1-\gamma)$, we can cancel out the terms in w and solve for B. Doing so produces the expression for A_1 stated in Proposition 2.3.1. □

B.2 Proof of Corollary 2.3.2

This follows immediately from the proof of Proposition 2.3.1. Here we fill in some of the details omitted in the above proof. Substituting the expressions for V^0 and V^1 into the expression for $\alpha(w)$ in eq. (B.3) gives

$$\begin{aligned}
\alpha(w) &\approx \alpha_0-\varepsilon\left(\frac{\mu-r}{\sigma^2}\right)\left[\frac{-\gamma A_0A_1w^{-3\gamma}-A_0A_1(1-2\gamma)w^{-3\gamma}+A_0^3w^{-3\gamma}}{\gamma^2A_0^2w^{-2\gamma-1}}\right] \\
&= \alpha_0-\varepsilon\alpha_0\left[\frac{-\gamma A_1-(1-2\gamma)A_1+A_0^2}{\gamma A_0}\right]w^{1-\gamma}
\end{aligned}$$

One can readily verify this is the same expression as stated in Corollary 2.3.2. Next, substituting the expressions for V^0 and V^1 into the expressions for $c(w)$ in eq. (B.2) gives

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

This then implies the expression for $s(w)=1-[c(w)/w]$ stated in Corollary 2.3.2. □

B.3 Recursive Preferences I

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

This observational equivalence has sparked a more recent literature that attempts to distinguish risk aversion from both ambiguity aversion and intertemporal substitution. Hansen and Sargent (2008, chpts. 18 and 19) note that the key to separating risk aversion from ambiguity aversion is to introduce hidden state variables, which the agent attempts to learn about. Early robust control methods were criticized because they abstracted from learning. Ambiguity is then defined by distortions of the agent's estimates of the hidden states.[1]

Hidden states can be used to represent a wide range of unobservables. For example, time invariant hidden states can index alternative models. Here we assume the hidden state is an unobserved, potentially time-varying, mean investment return. In principle, we could allow the agent to be uncertain about both the dynamics conditional on a particular mean growth rate, as well as the mean itself. However, for our purposes it is sufficient to assume the agent is only uncertain about the mean.[2]

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

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

where $\mathbb{E}$ is the expectation operator, π is a probability measure over outcomes conditional on a model, and μ is a probability measure over models. Ambiguity aversion is characterized by the properties of the ϕ function, while risk aversion is characterized by the properties of the u function. If ϕ is concave, the agent is ambiguity averse. KMM refer to $\mathbb{E}_\pi$ as 'first-order beliefs', while $\mathbb{E}_\mu$ is referred to as 'second-order beliefs'. Note that when ϕ is nonlinear, the implicit compound lottery defined by selecting a model with unknown parameters cannot be reduced to a single lottery over a 'hypermodel', as in Bayesian decision theory, so the distinction between models and parameters becomes important. Also note that from the perspective of smooth ambiguity aversion, evil agents and entropy penalized drift distortions are just a device used to produce a particular distortion in second-order beliefs about continuation values, i.e., where $\phi(V) \approx -\exp(-\varepsilon V)$.

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

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

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

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

With recusive preferences, the agent's problem becomes

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

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

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

and where for simplicity we've assumed the the elasticity of intertemporal substitution is unity. The effective rate of time preference is $\varphi = \rho + \delta$, and the coefficient of relative risk aversion is $\gamma \neq 1$.[3] The budget constraint is the same as before

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

The HJB equation is

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

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

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

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

If these are substituted into the HJB equation we get:

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

Our goal is to compute the following first-order approximation

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

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

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

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

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

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

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

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

where

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

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

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

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

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

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

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

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

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

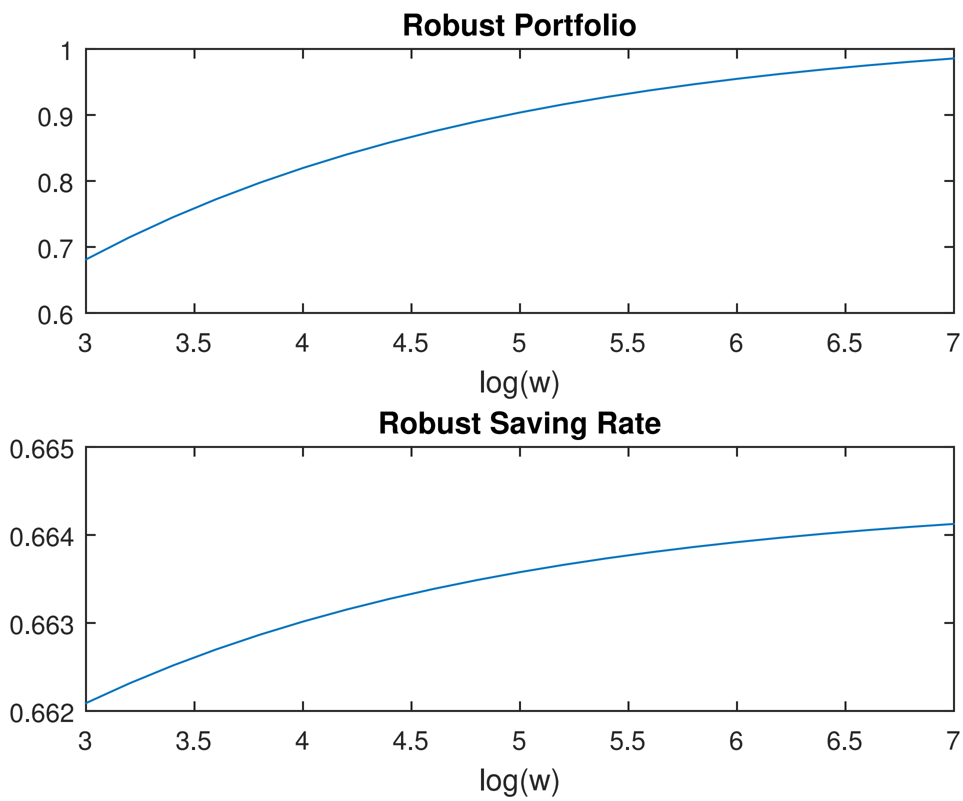


Figure B.1: Policy functions

The portfolio shares are similar to those in Figure 2.2, though wealth dependence is somewhat weaker. The key difference here is the saving rate. Now it is increasing in wealth. However, as in Figure 2.3, wealth dependence is very weak.

B.4 Proof of Proposition 2.4.1

Substituting the policy functions into the budget constraint gives

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

where $\alpha(w)$ and $c(w)$ are the approximate policy functions given in Corollary 2.3.2. Collecting terms, individual wealth dynamics can be described

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

where the drift coefficient is given by

$$\begin{aligned} a(w;\varepsilon) &= r + \frac{(\mu - r)^2}{\gamma\sigma^2} - A_0^{-1/\gamma} + \varepsilon\left\{\frac{-\gamma^2\sigma^2\alpha_0^2[A_0^2 + (\gamma-1)A_1)] + \gamma A_1 A_0^{-1/\gamma}}{\gamma^2 A_0}\right\} w^{1-\gamma} \\ &\equiv \bar{a}_0 + \varepsilon\bar{a}_1 w^{1-\gamma} \end{aligned} \tag{B.4}$$

and the diffusion coefficient is given by

$$\begin{aligned} b(w;\varepsilon) &= \sigma\alpha_0 - \varepsilon\left\{\alpha_0\sigma\gamma\frac{A_0^2 + (\gamma-1)A_1}{\gamma^2 A_0}\right\} w^{1-\gamma} \\ &\equiv b_0 + \varepsilon b_1 w^{1-\gamma} \end{aligned} \tag{B.5}$$

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

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

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

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

Finally, defining $a_0 \equiv \bar{a}_0 - \frac{1}{2}b_0^2$ and $a_1 = \bar{a}_1 - b_0 b_1$ gives the result stated in Proposition 2.4.1. □

B.5 Proof of Proposition 2.4.4

For convenience, we start by reproducing the KFP equation in (2.14)

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

Evaluating the derivatives gives

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

where $a(x)$ and $b(x)$ are the drift and diffusion coefficients defined in equations (B.4) and (B.5) after the change of variables $w = e^x$. In general, these sorts of partial differential equations are not fun to solve. However, this PDE is *linear*, which opens the door to transform methods. The first step is to evaluate the derivatives of the $a(x)$ and $b(x)$ functions, and then drop the $O(\varepsilon^2)$ terms. Then we take the Laplace transform of both sides, with x as the transform variable. When doing this we use the following facts

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

where $\mathcal{L}\{f(x)\} \equiv F(t,s) \equiv \int_{-\infty}^{\infty} f(t,x)e^{-sx}dx$ defines the (two-sided) Laplace transform. The first result follows from interchanging differentiation and integration, while the second and third results follow from integration by parts (using the boundary conditions $f(-\infty) = f(\infty) = 0$). The fourth result is called the 'shift theorem', and follows from the change of variable $s \to s - \beta$. The last result is more subtle. The fact that the Laplace transform of a delta function is just equal to 1 uses results from the theory of generalized functions.[4]

Following these steps produces equation (2.15) in the text, which we repeat here for convenience

$$\frac{\partial F}{\partial t} = \Lambda(s)F(t,s) + \varepsilon\Phi(s-\beta)F(t,s-\beta) + \delta \tag{B.6}$$

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

$$\begin{aligned}\Lambda(s) &= \frac{1}{2}b_0^2 s^2 - a_0 s - \delta \\ \Phi(s) &= b_0 b_1 s^2 + (2b_0 b_1\beta - a_1)s + \beta(b_0 b_1 \beta - a_1)\end{aligned}$$

To solve equation (B.6) we use the approximation $F(t,s-\beta) \approx F(t,s) - \beta\frac{\partial F}{\partial s}$. This gives us

$$F_t = \Lambda(s)F(t,s) + \varepsilon\Phi(s-\beta)\left[F(t,s) - \beta F_s\right] + \delta \tag{B.7}$$

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

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

Loosely speaking, $H(t)$ captures transition dynamics, $G(s)$ captures initial and boundary conditions, and $F_\infty(s)$ captures the new stationary distribution. Here we focus on this last component. Plugging this guess into (B.7) gives

$$H'G = \Lambda[HG] + \varepsilon\Phi[HG - \beta HG'] + \{[\Lambda + \varepsilon\Phi]F_\infty - \varepsilon\Phi\beta F_\infty' + \delta\} \tag{B.8}$$

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

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

variable we can write the ODE in parentheses as follows

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

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

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

The general solution of this linear ODE is the sum of a particular solution to the nonhomogeneous equation and the solution of the homogeneous equation. However, we can ignore the homogeneous solution, since we know that a stationary distribution does not exist when $\delta = 0$.[5] Stated in terms of F_∞, the particular solution is

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

After expanding the denominator polynomial into partial fractions we obtain the result stated in Proposition 2.4.4. To prove the correspondence principle stated in Corollary 2.4.5 we can just reverse the change of variables back to s. Let (R_1, R_2) denote the two roots of $F(\hat{s})$, (ϕ_1, ϕ_2) denote the roots of $\Lambda(s)$, and (r_1, r_2) denote the roots of $\Phi(s)$. After substituting s/ε for $\hat{s}$ and then multiplying numerator and denominator by ε^2 we get

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

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

B.6 Proof of Proposition 2.5.1

This follows directly from the proof of Proposition 2.4.4. Having solved for the stationary distribution, F_∞, the PDE in eq. (B.8) becomes

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

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

B.7 Recursive Preferences II

To examine inequality dynamics with recursive preferences we just need to replace the expressions for (A_0, A_1) in the main text with the expressions for $(\hat{A}_0, \hat{A}_1)$ derived in Appendix B.3, and then replace

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

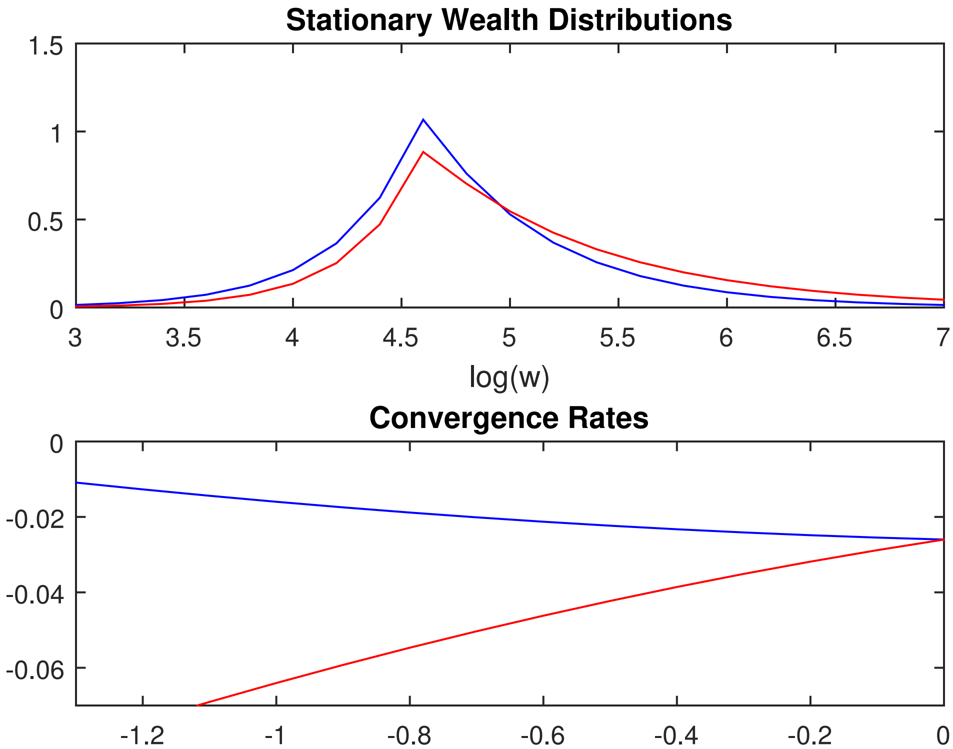


Figure B.2: Distribution and convergence rate

the expressions for (a_0, a_1) stated in Proposition 4.1 with the following expressions for $(\hat{a}_0, \hat{a}_1)$:

$$\hat{a}_0 = r - \varphi + \gamma\sigma^2\alpha_0^2 - \frac{1}{2}b_0^2$$

$$\hat{a}_1 = \frac{\frac{1}{2}\gamma^2\hat{A}_1 - (\sigma\gamma\alpha_0)^2(\hat{A}_0^2 + (\gamma-1)\hat{A}_1)}{\hat{A}_0\gamma^2} - b_0b_1$$

We can then follow the exact same approximation strategy as outlined in the main text. Figure B.2 displays the resulting stationary distributions and convergence rates. We use the same parameter values as those in Table 1, with the exceptions noted in Appendix B.3. In particular, we increase ε from 0.045 to 0.45, we increase γ slightly from 1.31 to 1.5, and increase μ slightly, from 5.86% to 5.95%.

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

1.6%, close to its original value of 1.14%. However, now the robust convergence rate becomes 6.4%, more than 200 basis points higher than before, and four times greater than its nonrobust value.

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

Appendix C

Information and Inequality

C.1 Value function with log utility

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

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

With log utility, we know $c = \beta w$ and $\pi = (\hat{\mu} - r)/\sigma^2$ 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 affect a log agent's portfolio decision, thus has zero effect on his marginal utility of wealth. If we assume $\beta = r$ [1], we can guess the following wealth-separable functional form for the value function

$$V^{00} = A\log(w) + g^0(\hat{\mu}, \gamma) + K \tag{C.2}$$

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

$$g^0 = \frac{1}{2\beta^2}\frac{(\hat{\mu} - r)^2}{\sigma^2} + \frac{\gamma^2}{\beta\sigma^2}\left(\frac{1}{2}g^0_{\hat{\mu}\hat{\mu}} - g^0_\gamma\right) \tag{C.3}$$

Let's further guess that

$$g^0(\hat{\mu}, \gamma) = \frac{1}{2\beta^2}\frac{(\hat{\mu} - r)^2}{\sigma^2} + \tilde{g}^0(\hat{\mu}, \gamma) \tag{C.4}$$

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

$$\tilde{g}^0_\gamma = \frac{1}{2}\tilde{g}^0_{\mu\mu} + \frac{1}{2\beta^2\sigma^2} - \frac{\beta\sigma^2}{\gamma^2}\tilde{g}^0 \tag{C.5}$$

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

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

$$\tilde{g}^0 = e^{\beta\sigma^2/\gamma} w(\hat{\mu}, \gamma) \tag{C.6}$$

which gives us the following heat equation for w

$$w_\gamma = \frac{1}{2} w_{\hat{\mu}\hat{\mu}} + \frac{1}{2\beta^2\sigma^2} e^{-\beta\sigma^2/\gamma} \tag{C.7}$$

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

$$\begin{aligned} w(\hat{\mu}, \gamma) &= \frac{1}{2\beta^2\sigma^2} \int_0^\gamma \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi(\gamma - s)}} \exp\left(\frac{-(\hat{\mu} - y)^2}{2(\gamma - s)}\right) e^{-\beta\sigma^2/s} dy ds \\ &= \frac{1}{2\beta^2\sigma^2} \int_0^\gamma e^{-\beta\sigma^2/s} \left[\frac{1}{\sqrt{2\pi(\gamma - s)}} \exp\left(\frac{-(\hat{\mu} - y)^2}{2(\gamma - s)}\right) dy\right] ds \\ &= \frac{1}{2\beta^2\sigma^2} \int_0^\gamma e^{-\beta\sigma^2/s} ds \end{aligned} \tag{C.8}$$

Unwinding the change of variables gives us

$$g^0(\hat{\mu}, \gamma) = \frac{1}{2\beta^2} \frac{(\hat{\mu} - r)^2}{\sigma^2} + e^{\beta\sigma^2/\gamma} \left(\frac{1}{2\beta^2\sigma^2} \int_0^\gamma e^{-\beta\sigma^2/s} ds\right) \tag{C.9}$$

Finally, the value function follows

$$V^{00}(w, \hat{\mu}, \gamma) = \frac{1}{\beta} \log(w) + \frac{\log(\beta)}{\beta} + \frac{1}{2\beta^2\sigma^2} \left[(\hat{\mu} - r)^2 + e^{\beta\sigma^2/\gamma} \int_0^\gamma e^{-\beta\sigma^2/s} ds\right] \tag{C.10}$$

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

$$e^{\beta\sigma^2/\gamma} \int_0^\gamma e^{-\beta\sigma^2/s} ds = \beta\sigma^2 e^{\beta\sigma^2/\gamma} E_i(-\beta\sigma^2/\gamma) + \gamma \tag{C.11}$$

where the exponential integral function is defined as

$$E_i(-\beta\sigma^2/\gamma) = -\int_{\beta\sigma^2/\gamma}^\infty \frac{e^{-s}}{s} ds \tag{C.12}$$

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

C.2 Perturbation Approximation

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

$$V(w, \hat{\mu}, \gamma) = \frac{1}{\beta} \left[\frac{[we^{g(\hat{\mu},\gamma)}]^\alpha - 1}{\alpha} + \theta(\frac{[we^{f(\hat{\mu},\gamma)}]^{\alpha+1} - 1}{\alpha + 1} + C)\right] \tag{C.13}$$

Further, let's do a first order expansion of the solution of $g(\hat{\mu},\gamma)$ and $f(\hat{\mu},\gamma)$ around $\alpha = 0$ [2], i.e:

$$g = g^0 + \alpha g^1 + O(\alpha^2) \tag{C.14}$$

$$f = f^0 + \alpha f^1 + O(\alpha^2) \tag{C.15}$$

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

$$c^* \approx \beta w[1 - \alpha(g^0 - \log\beta) - \theta w e^{f^0}] \tag{C.16}$$

$$\pi^* \approx \frac{1}{\sigma^2}[(\hat{\mu} - r) + \alpha(g^0_\mu \gamma + (\hat{\mu} - r)) + \theta w e^{f^0}(\hat{\mu} - r + f^0_\mu \gamma)] \tag{C.17}$$

$$\kappa^* \approx \theta\gamma w(g^0_{\mu\mu} - 2g^0_\gamma) \tag{C.18}$$

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

$$\lim_{\alpha\to 0,\theta\to 0} V(w,\hat{\mu},\gamma) = \frac{1}{\beta}(\log w + g^0(\hat{\mu},\gamma)) \tag{C.19}$$

which is the same $g^0(\hat{\mu},\gamma)$ in eq. (C.9).

One can further approximate the solution:

$$g^0 = \log\beta + \frac{1}{2\beta\sigma^2}[(\hat{\mu} - r)^2 + \gamma - \beta\sigma^2 e^{\beta\sigma^2/\gamma}\Gamma(0,\beta\sigma^2/\gamma)] \approx \log\beta + \frac{1}{2\beta\sigma^2}[(\hat{\mu} - r)^2 + \frac{\gamma^2}{\beta\sigma^2}] \tag{C.20}$$

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 (C.16), (C.17), (C.18) into the HJB equation, and get

$$\begin{aligned}\beta V = {} & \frac{V_w^{\frac{\alpha}{\alpha-1}} - 1}{\alpha} + V_w[rw - (\hat{\mu} - r)w\frac{V_w(\hat{\mu} - r) + V_{w\mu}\gamma}{V_{ww} w\sigma^2} - V_w^{\frac{1}{\alpha-1}} - \frac{\theta}{2}(\frac{V_{\mu\mu} - 2V_\gamma}{V_w}\gamma)^2] \\ & + \frac{1}{2}V_{ww}\sigma^2\left(\frac{V_w(\hat{\mu} - r) + V_{w\mu}\gamma}{V_{ww}\sigma^2}\right)^2 + (\frac{1}{2}V_{\mu\mu} - V_\gamma)\gamma^2(\frac{1}{\sigma^2} + 2\theta\frac{V_{\mu\mu} - 2V_\gamma}{V_w}) \\ & - V_{w\mu}\gamma(\frac{V_w(\hat{\mu} - r) + V_{w\mu}\gamma}{V_{ww}\sigma^2})\end{aligned} \tag{C.21}$$

Note that eq. (C.13) can be rewritten into

$$V = V^0 + \theta V^1 \tag{C.22}$$

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. (C.21) w.r.t θ, and evaluate it at $\theta = 0$. After some algebra, we get

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

$$
\begin{aligned}
\beta V^1 = & w V_w^1 \left(\frac{(\hat{\mu}-r)^2}{(1-\alpha)\sigma^2} + \frac{\alpha g_\mu \gamma(\hat{\mu}-r)}{\sigma^2(1-\alpha)} - (\frac{1}{\beta})^{\frac{1}{\alpha-1}} e^{\frac{\alpha}{\alpha-1}g} \right) \\
& + w^2 V_{ww}^1 \left(\frac{(\hat{\mu}-r)^2}{2\sigma^2(1-\alpha)^2} + \frac{\gamma^2\alpha^2 g_\mu^2}{2\sigma^2(1-\alpha)^2} + \frac{\gamma(\hat{\mu}-r)\alpha g_\mu}{\sigma^2(1-\alpha)^2} \right) \\
& + w V_{w\mu}^1 \left(\frac{\gamma^2 \alpha g_\mu}{\sigma^2(1-\alpha)} + \frac{\gamma(\hat{\mu}-r)}{\sigma^2(1-\alpha)} \right) + V_{\mu\mu}^1 \frac{\gamma^2}{2\sigma^2} \\
& - V_\gamma^1 \frac{\gamma^2}{\sigma^2} + \frac{1}{2\beta} e^{\alpha g} \left((g_{\mu\mu} + \alpha g_\mu^2 - 2g_\gamma)\gamma \right)^2 w^{\alpha+1}
\end{aligned}
\tag{C.23}
$$

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

$$
\begin{aligned}
\beta[we^{f^0} + C - 1] = & we^{f^0} \left(\frac{(\hat{\mu}-r)^2}{\sigma^2} - \beta \right) + w f_\mu^0 e^{f^0} \frac{\gamma(\hat{\mu}-r)}{\sigma^2} \\
& + \frac{\gamma^2}{\sigma^2} \left(\frac{1}{2} e^{f^0} (f_{\mu\mu}^0 + (f_\mu^0)^2) - f_\gamma^0 e^{f^0} \right) w + \frac{\gamma^2}{2} w \left(g_{\mu\mu}^0 - 2g_\gamma^0 \right)^2
\end{aligned}
\tag{C.24}
$$

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

$$
C = 1 \tag{C.25}
$$

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

$$
2\beta\sigma^2 = (\hat{\mu}-r)^2 + (\hat{\mu}-r)\gamma f_\mu^0 + \frac{\gamma^2}{2} \left[f_{\mu\mu}^0 + (f_\mu^0)^2 - 2f_\gamma^0 + \sigma^2 (g_{\mu\mu}^0 - 2g_\gamma^0)^2 e^{-f^0} \right] \tag{C.26}
$$

Let $f^0(\hat{\mu}, \gamma) = \log(p(\hat{\mu}, \gamma))$, s.t. $p(\hat{\mu}, \gamma)$ needs to satisfy

$$
2\beta\sigma^2 p = (\hat{\mu}-r)^2 p + (\hat{\mu}-r)\gamma p_\mu + \frac{\gamma^2}{2} [p_{\mu\mu} - 2p_\gamma + \sigma^2 (g_{\mu\mu}^0 - 2g_\gamma^0)^2] \tag{C.27}
$$

There is no direct analytic solution to the above. However, one can approximate it when estimation risk is small. Guess a second order approximate solution s.t:

$$
p(\hat{\mu}, \gamma) \approx D_0 + D_1(\hat{\mu}-\mu) + D_2\gamma + \frac{1}{2} D_3 (\hat{\mu}-\mu)^2 + \frac{1}{2} D_4 \gamma^2 + D_5 (\hat{\mu}-\mu)\gamma \tag{C.28}
$$

Plug back into the PDE (C.27) and match coefficients, we will get $D_0 = D_1 = D_2 = D_3 = D_5 = 0$, and that $D_4 = \left[\beta^2\sigma^2 [2\beta\sigma^2 - (\mu - r)^2] \right]^{-1}$. Let $D_4 = \lambda$, the complete solution of f^0 is given by

$$
f^0(\hat{\mu}, \gamma) \approx \log(\frac{1}{2}\lambda\gamma^2) \tag{C.29}
$$

□

C.3 Proof of Corollary 3.4.2

Proof. It is directly observed from eq. (3.24) how κ varies with θ, w, σ and γ. Recall that we have $\beta = \rho + \delta$. Therefore, κ decreases with both ρ and δ. Finally, let $x = \beta\sigma^2/\gamma$. With the small γ

assumption, we have when $\gamma \to 0$, $x \to \infty$. Since $\frac{\Gamma(a,x)}{x^{a-1}e^{-x}} \to 1$ as $x \to \infty$. (See Abramowitz and Stegun (1964)). From equation 3.24, we have $\lim_{\gamma \to 0} \kappa = 0$. □

C.4 Proof of Lemma 3.5.1

Proof. Combining policy functions (3.22), (3.23) and (3.24) and individual wealth dynamics (3.14), one can write down each agents' perceived law of motion for wealth as

$$dw = \hat{\mu}(w)wdt + \sigma(w)w\hat{d}B \tag{C.30}$$

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 $\hat{d}B = dB - \frac{(\hat{\mu}-\mu)}{\sigma}dt$. Therefore,

$$\begin{aligned} \mu(w) &= \hat{\mu}(w) - \sigma(w)\frac{(\hat{\mu}-\mu)}{\sigma} \\ &= \frac{(\hat{\mu}-r)}{\sigma^2}\left[(1+\alpha)(\hat{\mu}-r) + \alpha\frac{(\hat{\mu}-r)\gamma}{\beta\sigma^2} + \frac{\theta w}{2}\lambda(\hat{\mu}-r)\gamma^2\right] \\ &+ \frac{\alpha}{2\sigma^2}\left[(\hat{\mu}-r)^2 + \frac{\gamma^2}{\beta\sigma^2}\right] + \frac{\theta\beta w}{2}\lambda\gamma^2 - \frac{\theta w\gamma^2}{2\beta^4\sigma^4} \end{aligned} \tag{C.31}$$

and that

$$\sigma(w) = \frac{1}{\sigma}[(1+\alpha)(\hat{\mu}-r) + \alpha\frac{(\hat{\mu}-r)\gamma}{\beta\sigma^2} + \frac{1}{2}\theta w\lambda(\hat{\mu}-r)\gamma^2] \tag{C.32}$$

To simplify, one would see

$$\mu(w) = a(\hat{\mu},\gamma) + \theta w b(\hat{\mu},\gamma) \tag{C.33}$$

$$\sigma(w) = c(\hat{\mu},\gamma) + \theta w d(\hat{\mu},\gamma) \tag{C.34}$$

s.t:

$$a(\hat{\mu},\gamma) = \frac{(\mu-r)}{\sigma^2}\left[(1+\alpha)(\hat{\mu}-r) + \alpha\frac{(\hat{\mu}-r)\gamma}{\beta\sigma^2}\right] + \frac{\alpha}{2\sigma^2}\left[(\hat{\mu}-r)^2 + \frac{\gamma^2}{\beta\sigma^2}\right] \tag{C.35}$$

$$b(\hat{\mu},\gamma) = \frac{\lambda\gamma^2(\mu-r)(\hat{\mu}-r)}{2\sigma^2} + \frac{\beta}{2}\lambda\gamma^2 - \frac{\gamma^2}{2\beta^4\sigma^4} \tag{C.36}$$

$$c(\hat{\mu},\gamma) = \frac{1}{\sigma}\left[(1+\alpha)(\hat{\mu}-r) + \alpha\frac{(\hat{\mu}-r)\gamma}{\beta\sigma^2}\right] \tag{C.37}$$

$$d(\hat{\mu},\gamma) = \frac{1}{2\sigma}\lambda(\hat{\mu}-r)\gamma^2 \tag{C.38}$$

□

C.5 Proof of Proposition 3.5.3

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

$$dx = [a + \theta b e^x - \frac{1}{2}(c + \theta \hat{d} e^x)^2]dt + (c + \theta \hat{d} e^x)dB \tag{C.39}$$

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

$$dx = [(a - \frac{1}{2}c^2) + \theta e^x (b - c\hat{d})]dt + [c + \theta \hat{d} e^x]dB \tag{C.40}$$

Therefore, the system can be written as

$$\begin{bmatrix} dx \\ d\hat{\mu} \\ d\gamma \end{bmatrix} = \begin{bmatrix} a - \frac{1}{2}c^2 + \theta(b - c\hat{d})e^x \\ \frac{\gamma(\mu - \hat{\mu})}{\beta^2\sigma^2}[\beta^2 + 2\theta e^x] \\ -\gamma^2(\frac{1}{\sigma^2} + \frac{2\theta e^x}{\beta^2\sigma^2}) \end{bmatrix} dt + \begin{bmatrix} c + \theta\hat{d}e^x & 0 & 0 \\ \frac{\gamma}{\sigma} & \frac{\sqrt{2\theta e^x}\gamma}{\beta\sigma} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} dB \\ dB_y \\ 0 \end{bmatrix} \tag{C.41}$$

The above state dynamics can be written in the matrix form

$$dX = G(X)dt + \Sigma(X)d\bar{B} \tag{C.42}$$

The stationary KFP equation is thus written

$$0 = -\sum_{i=1}^{3} \frac{\partial}{\partial X_i}[G_i(X)f] + \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3} \frac{\partial^2}{\partial X_i \partial X_j}\left[\left(\sigma(X)\sigma^T(X)\right)_{i,j} f\right] - \delta f + \delta\zeta(X - X_0) \tag{C.43}$$

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. □

C.6 Proof of Proposition 3.5.4

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

$$f_1(x|\hat{\mu}, \gamma) = f^0(x|\hat{\mu}, \gamma) + \theta f^1(x|\hat{\mu}, \gamma) \tag{C.44}$$

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

$$\delta f^0 = -\hat{a} f_x^0 + \frac{1}{2}c^2 f_{xx}^0 + \delta\zeta(x_0) \tag{C.45}$$

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 $x_0 = 0$.

Next, combining eq. (3.35) and the solution of $f^0(x|\hat{\mu},\gamma)$, we know that the PDE for f^1 is

$$\begin{aligned}\delta f^1 &= -\hat{a} f^1_x + \frac{1}{2}c^2 f^1_{xx} + e^x\left[(c\hat{d}-\hat{b})f^0 + (2c\hat{d}-\hat{b})f^0_x + c\hat{d}f^0_{xx}\right] \\ &+ e^{x(\phi+1)}\left[(c\hat{d}-\hat{b}) + \phi(2c\hat{d}-\hat{b}) + \phi(\phi-1)c\hat{d}\right] + \delta\zeta(0)\end{aligned} \tag{C.46}$$

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

$$\delta A_1 = -\hat{a}(\phi+1)A_1 + \frac{1}{2}c^2(\phi+1)\phi A_1 + A_0\left[(c\hat{d}-\hat{b}) + \phi(2c\hat{d}-\hat{b}) + \phi(\phi-1)c\hat{d}\right] \tag{C.47}$$

This gives

$$A_1 = \frac{A_0\left[(c\hat{d}-\hat{b}) + \phi(2c\hat{d}-\hat{b}) + \phi(\phi-1)c\hat{d}\right]}{\delta + \hat{a}(\phi+1) - \frac{1}{2}c^2(\phi+1)\phi} \tag{C.48}$$

□

C.7 Proof of Proposition 3.5.7

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

$$\begin{aligned}\psi(\hat{\mu},\gamma) \equiv Ee^x &= \int_0^\infty e^x\left[A_0 e^{\phi x} + \theta A_1 e^{(\phi+1)x}\right]dx \\ &= -\frac{1}{\phi+1}A_0 - \theta A_1\frac{1}{\phi+2}\end{aligned} \tag{C.49}$$

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

$$\psi(\hat{\mu},\gamma) \approx \frac{-A_0}{\phi+1} \tag{C.50}$$

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:

$$q(\hat{\mu},\gamma) = q^0 + \theta q^1 \tag{C.51}$$

where q^1 solves

$$\frac{2}{\beta^2}\psi\mathcal{L}[q^1] + \frac{2}{\beta^2}\left[\frac{(\hat{\mu}-\mu)}{\sigma^2}\psi_{\hat{\mu}} + \frac{\gamma^2}{\sigma^2}\psi_\gamma\right]q^0 = 0 \tag{C.52}$$

This is equivalent to

$$\mathcal{L}[g^1] + \left[\frac{(\hat{\mu}-\mu)}{\sigma^2}\frac{\psi_{\hat{\mu}}}{\psi} + \frac{\gamma^2}{\sigma^2}\frac{\psi_\gamma}{\psi}\right]N(\hat{\mu},\gamma) = 0 \tag{C.53}$$

which could be written into the following form:

$$\mathcal{L}[q^1] + Q(\hat{\mu},\gamma) = 0 \tag{C.54}$$

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

$$u_t + Au = f(t), u(0) = u_0 \tag{C.55}$$

We know the solution is

$$u(t) = e^{-tA}u_0 + \int_0^t e^{(s-t)A} f(s) ds \tag{C.56}$$

Now suppose we have the diffusion problem

$$u_t(x,t) = \mathcal{L}[u] + \lambda(x,t) \tag{C.57}$$

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

$$\begin{aligned} u(x,t) &= \mathcal{L}(t)u(x,0) + \int_0^t \mathcal{L}(t-s)\lambda(s)ds \\ &= \int_{-\infty}^{\infty} N(x-y,t)u_0(y)dy + \int_0^t \int_{-\infty}^{\infty} N(x-y,t-s)\lambda(y,s)dyds \end{aligned} \tag{C.58}$$

where in the standard diffusion problem

$$N(x,t) = \frac{1}{2\sqrt{\pi k t}} e^{\frac{-x^2}{4kt}} \tag{C.59}$$

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 3.5.7 given in abstract terms is

$$G(\hat{\mu},\gamma) = \int_0^{\infty} \int_0^{\infty} N(\hat{\mu}-y,\gamma-s)Q(y,s)ds \tag{C.60}$$

□

C.8 Proof of Corollary 3.5.9

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

$$\int_a^b h(s)e^{Mf(s)}ds \approx \sqrt{\frac{2\pi}{Mf''(s_0)}}h(s_0)e^{Mf(s_0)} \tag{C.61}$$

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

The dominant piece of the marginal density

$$\int_0^{\gamma_0} \int_{-\infty}^{\infty} A_1(\hat{\mu},\gamma)B_0(\hat{\mu},\gamma)e^{(\phi+1)x}d\hat{\mu}d\gamma \tag{C.62}$$

can thus be approximated using this method. Using the approximation

$$(\phi+1)x \approx \phi_0 x + [1 + \phi_{\hat{\mu}}(\hat{\mu}-\mu) + .5\phi_{\hat{\mu}\hat{\mu}}(\hat{\mu}-\mu)^2 + \phi_\gamma \gamma]x \tag{C.63}$$

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

$$(\phi+1)x \approx \phi_0 x + (1 + \frac{\phi_{\hat{\mu}}^2}{2|\phi_{\hat{\mu}\hat{\mu}}|})x \tag{C.64}$$

Therefore,

$$\phi_L \approx \phi_0 + (1 + \frac{\phi_{\hat{\mu}}^2}{2|\phi_{\hat{\mu}\hat{\mu}}|}) \tag{C.65}$$

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

C.9 Constraints on Parameter Space

The combination of parameters $(\hat{\mu}, \beta, \gamma_0, \theta, \sigma, \alpha, w_0)$ need to satisfy the following constraints:
1. Benchmark mean growth rate $a(\hat{\mu}, \gamma)$ and volatility parameter $c(\hat{\mu}, \gamma)$ are positive.
2. Portfolio share (3.23) is between 0 and 1.
3. Saving rate (3.25) is between 0 and 1.
4. Information cost share between 0 and 1.
5. Parameter $\lambda > 0$ to ensure existence of perturbation solution of the value function.

C.10 Supplementary Tables

Table C.1: Cross-sectional households wealth characteristics

Statistic	N	Mean	St. Dev.	Min	Max
2013 SCF					
Stock	11,782	71,490.890	913,754.600	0	27,000,000
Bond	11,782	21,629.760	405,470.800	0	13,010,000
Cash	11,782	45,046.150	399,630.200	0	10,096,000
Market Value of Private Equity	11,782	1,275,563.000	22,109,295.000	0	795,730,000
Total Wealth	11,782	1,413,730.000	22,521,950.000	1	802,855,000
If Delegate Wealth	11,782	0.057	0.232	0	1
Risk Tolerance	11,782	1.805	0.809	1	4
Capital Income	11,782	6,004.741	116,560.500	0	4,160,000
1983 SCF					
Stock	4,009	297,654.700	2,972,495.000	0.000	125,434,000.000
Bond	4,009	85,243.550	846,403.000	0.000	35,400,000.000
Cash	4,009	73,046.270	346,524.100	0.000	9,847,100.000
If Delegate Wealth	4,009	1.853	0.886	1	4

Units: Value are adjusted to 2013 dollar value using CPI data from Federal Reserve St. Louis. *Notes:* This table shows the summary statistics of various household wealth characteristics in *Survey of Consumer Finance* imputed full public data, weighted using the SCF weights. *Stock* variable is defined as the summation of direct holding and delegated holding of stocks. Direct holding of stocks include holding of households' own company stock, other companies' stocks and direct holding of foreign stocks. Delegated stock holding consists of stock mutual funds, saving stock, managed stocks and stock holding in all retirement accounts, adjusted by the reported fraction of investment in stocks. Likewise, *Bond* is defined as the summation of households' direct holding of bonds, bond mutual funds, government bond funds, other bond funds and all retirement accounts bonds value. *Cash* is defined as the summation of reported checking account, saving and money market accounts and certificate of deposit values. Following Moskowitz and Vissing-Jorgensen (2002), I focus on the largest private business when computing private equity values. *Market Value of Private Equity* is computed by subtracting net worth of the private business in 2013 with its 2012 income, further adjusted by the difference between net debt owed by the private business to households. *Pre-Tax Profit of Private Equity* is defined as the reported 2012 pre-tax profits. *Wealth* is the aggregation of all stock, bond, cash and private equity value. *Percentage of Risky Assets* is derived by dividing the sum of households stock holding and private equity value by total wealth. *If Delegate Wealth* equals 1 is households report having invested in any mutual funds or hedge funds. Finally, the survey asks the question "Which of the statements on this page comes closest to the amount of financial risk that you (and your husband/wife/partner) are willing to take when you save or make investments? Households can choose one of the following: 1. *Take substantial financial risks expecting to earn substantial returns, 2. *Take above average financial risks expecting to earn above average returns, 3. *Take average financial risks expecting to earn average returns, 4. *Not willing to take any financial risks. The variable *Risk Tolerance* is thus defined by recoding "1 = 4; 2 = 3; 3 = 2; 4 = 1".

Table C.2: Summary statistics of Lipper TASS hedge fund panel data

Statistic	N	Mean	St. Dev.	Min	Max
incentive fee	158,158	15.621	7.492	0.000	35.000
rate of return	158,158	0.853	5.944	−89.922	934.480
net asset value	158,158	16,193.760	227,304.700	0.861	6,003,175.000
lock up period	158,158	4.049	8.675	0	90
minimum investment	158,003	1,239,778.000	6,436,964.000	0	100,000,000
high water mark	158,158	0.756	0.430	0	1
average leverage	146,618	60.959	179.984	0.000	2,000.000

Notes: This table shows the summary statistics of variables used for analysis from Lipper TASS hedge fund data. Time-series fund monthly performance data is merged with fee history and fund product details data set into one panel. The final data set spans from 1977 January to 2015 January. Variable *incentive fee* and *management fee* are measured in percentage terms. The original *rate of return* data is computed in the end of each month. *net asset value* denotes the overall size of the hedge fund. *lock up period* serves as a measure of liquidity of the investment. *minimum investment* denotes minimum dollar value one needs to put in front when entering the fund. *high water mark* is equal to 1 if the fund only charges incentive fee based on net historical positive returns, 0 if otherwise. Finally, *average leverage* controls the average risk the fund takes.

Table C.3: Does the rich have higher returns on wealth?

	ROR 1983	ROR2013
	(1)	(2)
log wealth	3.35e-02***	3.912e-02***
	(8.038e-03)	(0.029)
risk tolerance	$-1.116e-02$***	-1.404***
	(4.489e-03)	(0.128)
age	-2.868e-04	0.054***
	(3.482e-03)	(0.005)
age^2	3.637e-06***	
	(3.721e-05)	
Constant	$-2.178e-01$**	-2.886***
	(7.809e-02)	(0.730)
Observations	573	24,221
R^2	0.1521	0.098
Adjusted R^2	0.1339	0.097
Residual Std. Error	0.1494 (df = 560)	12.457 (df = 24197)

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table C.4: Does the rich delegate more wealth and invest more in stocks

	(1983 p.stock)	(2013 p.stock)	(prob.delegate wealth(2013))
	(1)	(2)	(3)
log wealth	6.084***	7.0367***	1.3369e+00***
	(0.557551)	(1.2712e-01)	(7.8434e-02)
logwealth^2			-3.6865e-02***
			(2.8847e-03)
age	0.496739***	4.6650e-01	3.7808e-02*
	(0.327165)	(8.1163e-02)	(1.5831e-02)
age^2	$-3.7753e-03$***	3.252***	-3.5246e-04*
	(8.8104e-04)	(0.390)	(1.4829e-04)
constant	-2.5182e+01 **	-50.559***	-34.221***
	(2.3041e+00)	(1.161e+01)	(8.6822e+00)
Observations	573	4009	4009
R^2	0.4306	0.5532	0.264
Adjusted R^2	0.5526	0.224	0.263
Residual Std. Error	25.32 (df = 17100)	21.858 (df = 2695)	29.529 (df = 16192)
Notes:	***Significant at the 1 percent level.		
	**Significant at the 5 percent level.		
	*Significant at the 10 percent level.		