ESSAYS ON ASYMMETRIC INFORMATION IN
MACROECONOMICS AND FINANCE

JASON GREGORY RAVIT

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
ECONOMICS
ADVISER: NOBUHIRO KIYOTAKI

MAY 2016
© Copyright by Jason Gregory Ravit, 2016.

All rights reserved.
Abstract

This thesis comprises three essays on the effects of asymmetric information on macroeconomic and financial market outcomes.

In Chapter 1, co-authored with Zongbo Huang and Michael Sockin, I embed imperfect substitutability across skill levels into a dynamic Mirrlees model and uncover a novel intertemporal wage compression channel in optimal labor taxation that can rationalize redistributive programs such as the Earned Income Tax Credit. This dynamic channel lowers the optimal tax rate at the bottom because it allows the planner to reduce the cost of providing insurance to unskilled workers while deterring skilled workers from misreporting. The optimal labor tax is progressive in the short-run and our channel is quantitatively significant compared to other channels highlighted in the literature.

In Chapter 2, I study the first-order approach in a class of continuous time dynamic mechanism design problems. This class of models has many applications in macroeconomics including insurance and optimal taxation. By working in continuous time, I take advantage of newly-developed techniques and I establish a novel sufficient condition for the optimal contract as implied by the first-order approach to be globally optimal: agents cannot overreport their shocks. This condition is simple, can be imposed ex ante, and is satisfied anyway in most settings.

In Chapter 3, I develop a simple model of the banking sector to explain three facts about the subprime mortgage boom: the unprecedented expansion of the non-agency mortgage-backed security market, securitized mortgages defaulted at a higher rate than did retained mortgages, and household income growth was negatively correlated with credit expansion and house prices. Commercial banks are regulated and raise funds via deposits while shadow banks are unregulated and fund themselves by securitizing mortgages. Commercial banks retain their loans, giving them skin in the game and an incentive to screen out subprime borrowers, while shadow banks do
not screen, thus lending to subprime borrowers. Unlike commercial banks, a shadow bank’s funding capacity depends on economic fundamentals and they supply more credit when it is easier to securitize mortgages. The economy with shadow banks is unstable and prone to large price and welfare jumps.
Acknowledgements

This dissertation, the culmination of an incredible five year journey, would not have been possible without the help of many people. My adviser, Nobu Kiyotaki, encouraged me to pursue those topics that interested me and he forced me to think hard about precisely what I was asking and why. Had I not experienced it with Nobu, I would not have believed it was possible for an adviser to be willing to discuss my work literally at any time for hours on end. Presenting in his reading group was an invaluable source of dynamic feedback on my research, and my favorite experience as a teaching assistant was working for Nobu.

Several other faculty members had a tremendous impact on my experience as a Ph.D. student. Mike Golosov delivered brutally honest but important feedback, usually so that I would clarify, simplify, and fully understand the mechanisms at work in my models. Ben Moll and Yuliy Sannikov showed me that everything really is nicer in continuous time and provided extremely helpful guidance with some of the more technical parts of my work. I also had the pleasure of being a teaching assistant for Yuliy’s terrific corporate finance course, which no doubt made preparing for my first job so much easier. Mark Aguiar and Richard Rogerson helped me to express my ideas as clearly as possible. I would also like to thank Saki Bigio, Markus Brunnermeier, Drew Fudenberg, Maryam Farboodi, Harrison Hong, Ilyana Kuziemko, Atif Mian, Stephen Morris, Galo Nuño, Harvey Rosen, Alp Simsek, Hyun Song Shin, and Motohiro Yogo for numerous helpful discussions. Special thanks to Larry Ball, my undergraduate adviser at Johns Hopkins, who was instrumental in my decision to study macroeconomics in the first place.

My time at Princeton would not have been the same had it not been for my classmates. My coauthors and friends, Zongbo Huang and Mike Sockin, were sources of great ideas and great conversations, and I could always count on them to challenge me. Kyle Chauvin, John Klopfer, and especially Yu Zhang started as my problem
set partners in our first year and became great friends with whom I could discuss whatever came to mind. My officemates on the third floor of BCF were always a great source of conversation and laughs. Qingqing Cao, Thomas Carter, Christian Moser, and Alejandro Van der Ghote made Nobu’s reading group the experience it was. I owe special thanks to Cristian Alonso for organizing EconCon with me.

I would like to thank Laura Hedden and Karen Neukirchen for their administrative support and management of the program, and the Graduate School, Department of Economics, and Woodrow Wilson School for generous financial support.

Finally, I want to thank my family: my parents, sister, and grandparents. I never would have studied economics in the first place had they not threatened to stop paying my tuition if I continued to take nothing but math courses! I do not know how I would have made it through graduate school without them and I will be forever grateful for their encouragement, support, and love.
To my parents and grandparents, who have always supported me.
# Contents

Abstract ................................................................. iii
Acknowledgements ....................................................... v
List of Figures .......................................................... xi

1 Dynamic Optimal Taxation with Endogenous Skill Premia 1

1.1 Introduction ......................................................... 1
1.2 Related Literature .................................................. 5
1.3 A Simple Model ..................................................... 8
1.4 Model ............................................................... 14
    1.4.1 Environment .................................................. 14
    1.4.2 Planner’s Problem .......................................... 20
1.5 Optimal Tax Policies .............................................. 30
1.6 Quantitative Analysis ............................................. 37
    1.6.1 Parametrization ............................................. 37
    1.6.2 Properties of Optimal Tax Policies ....................... 38
1.7 Discussion ......................................................... 46
    1.7.1 CES Production Function ................................ 46
    1.7.2 Skill Distribution .......................................... 47
1.8 Extensions .......................................................... 48
    1.8.1 Skill Bias in Production ................................ 48
    1.8.2 Endogenous Human Capital Formation .................... 50
3.4 Equilibrium ......................................................... 126
  3.4.1 Household’s Problem ........................................ 126
  3.4.2 Intermediaries’ Problems .................................... 128
  3.4.3 Equilibrium Contracts ....................................... 133
3.5 Properties of the Equilibrium .................................. 136
  3.5.1 Comparative Statics .......................................... 136
  3.5.2 Welfare Analysis ............................................ 145
3.6 Extension .......................................................... 148
3.7 Conclusion ........................................................ 159
3.8 Appendix: Omitted Proofs ........................................ 163
  A Section 3.4 Proofs ............................................... 163
  B Section 3.5 Proofs ............................................... 178
  C Section 3.6 Proofs ............................................... 184

Bibliography .......................................................... 185
List of Figures

1.1 Average income as a function of $\alpha$. .......................... 39
1.2 $\bar{\pi}^s_L(\theta)$ vs. $\bar{\pi}^M_L(\theta)$ for $\alpha = 2.5$ and $\alpha = 3.124$. ......................... 40
1.3 History-dependent labor wedge for low-skill worker for $\alpha = 2.5$ and $\alpha = 3.124$. ...................... 41
1.4 Impulse response function of the labor wedge. ......................... 44
1.5 Average intertemporal wedge as a function of $\alpha$. ................. 45
1.6 Labor wedge with skill bias consistent with $(\alpha, \varrho) = (2.5, 0)$. ............... 49
1.7 Labor wedge with skill bias consistent with $(\alpha, \varrho) = (3.124, 0)$. ............ 50

3.1 The rise and fall of the non-agency MBS market. .................. 107
3.2 The relationship between income growth and credit growth. ......... 109
3.3 A schematic representation of the economy. .......................... 116
3.4 The timeline of the game. ............................................. 125
3.5 The effect of higher credit worthiness on default probability. .......... 140
3.6 Welfare difference between commercial and shadow banking. ......... 148
3.7 Credit expansion/price feedback loop. ............................... 157
Chapter 1

Dynamic Optimal Taxation with Endogenous Skill Premia

1.1 Introduction

A primary objective of tax systems in advanced countries is to combat rising income inequality, a phenomenon that has become more pronounced in recent decades.\(^1\) Programs like the Earned Income Tax Credit (EITC) in the United States, for instance, aim to reduce inequality by targeting poverty, subsidizing low- and moderate-income workers through negative effective marginal tax rates.\(^2\) While such redistributive policies are ubiquitous in practice, standard models of optimal taxation have difficulty rationalizing these salient features of the tax code. Negative marginal tax rates in these settings are inefficient because they reduce incentives to work, going against the primary goal of the EITC, which is to incentivize low-income people to work more by providing a transfer only if they actually work. We develop a dynamic Mirrlees model of optimal taxation in which workers with different skills are imperfect substitutes in

---

\(^1\)See Piketty and Saez (2003) for an in-depth survey of the evidence.

\(^2\)The United Kingdom’s version is called the Working Tax Credit, and Austria, Belgium, Canada, Denmark, Finland, France, the Netherlands, New Zealand, and Sweden all have similar programs.
the production process and uncover a novel “intertemporal wage compression” channel that provides a new rationale for the EITC, even in the absence of idle workers: negative marginal rates for low-income workers relax incentive constraints in earlier periods, making it cheaper for the planner to provide insurance against shocks earlier in life while simultaneously deterring high-income workers from shirking.

Similar to Mirrlees (1971) and its dynamic counterparts in the “new dynamic public finance” (NDPF) literature, we cast the optimal (nonlinear) tax function as the outcome of a mechanism design problem. In our model, agents receive privately-observed skill shocks over time and sell their labor to a representative firm that aggregates their effective labor (labor supplied multiplied by skill) into a final numéraire good. Importantly, we build on Stiglitz (1982) and Ales, Kurnaz, and Sleet (2015) and assume that agents with different skills are imperfect substitutes in this production process. As a consequence, changes in relative labor supply change relative wages across workers, giving rise to skill premia. Our model is motivated by the findings in Katz and Murphy (1992), Acemoğlu and Autor (2011), and Autor (2014) that not only have skill premia risen over the last few decades, but this trend is a large contributor to growing inequality, especially in the bottom and middle of the distribution.3

In static settings, imperfect substitutability gives rise to the “intratemporal wage compression” channel, first introduced in Stiglitz (1982), that reduces skill premia by encouraging (discouraging) labor that reduces (increases) the wages of high-skill (low-skill) workers. In particular, these wages make it costlier for high-skill workers to mimic low-skill workers. The planner can achieve this by subsidizing, i.e., a negative marginal rate, high-skill labor and taxing low-skill labor. Therefore, imperfect substitutability, in itself, cannot rationalize redistributive programs like the EITC.

---

3Inequality between the bottom 99% and the top 1% has been driven by “superstar effects” and increases in CEO compensation driven by capital gains; see Gabaix, Lasry, Lions, and Moll (2015) for a reduced form dynamic model and Ales and Sleet (2016) and Scheuer and Werning (2016) for static Mirrlees models with superstar effects.
Our dynamic model introduces a new, “intertemporal wage compression” channel.\textsuperscript{4} Here, the planner relaxes high-skill agents’ incentive constraints today by reducing their utility in future states that are unlikely to occur. In particular, if skill is persistent, a skilled agent who misreports makes the planner believe he will be unskilled in the future. To deter this agent from misreporting, the planner makes these future states very painful for the lying agent and instead redistributes heavily towards states that low-skill agents are likely to face, i.e., provides more insurance to low-skill agents. The planner does this by subsidizing low-skill labor (the EITC) in the future, which raises the supply of low-skill labor and lowers the low-skill wage (and raises skill premia). The threat of low wages and long hours in the future reduces variation in future utility and ensures that the dynamic loss from mimicking low-skill agents outweighs the static gain. Therefore, in our model, the purpose of the EITC is to lower the cost of providing valuable (dynamic) insurance to unskilled workers by relaxing skilled agents’ incentive constraints in earlier periods. Negative rates are possible precisely because wages adjust downward to offset the higher labor supply. We are not making any statements about the EITC’s current focus on the extensive margin; rather, by focusing on only the intensive margin, we highlight new, complementary dynamic insurance and incentive motives. Further, since incentivizing high-skill agents to report truthfully is more important than insuring them, the old “intratemporal” channel dominates our “intertemporal” channel in the right tail, as in Golosov, Troshkin, and Tsyvinski (2016b). Since more skilled agents receive a higher wage, they can more easily mimic less skilled workers through excessive saving. This causes the intertemporal wedge, or implicit savings tax, to be higher for more skilled agents and lower for less skilled agents.

\textsuperscript{4}Besides these two channels, optimal labor taxes are shaped by two additional forces. First, the planner must insure agents against shocks in the current period while incentivizing them to work, similar to Diamond (1998). Second, the planner can reduce the cost of insurance against earlier shocks by relaxing earlier incentive constraints. This is the “intertemporal” force from Golosov, Troshkin, and Tsyvinski (2016b). Decreasing the substitutability between skill types lowers these two terms for low-skill workers and raises them for high-skill workers.
After characterizing the optimal tax system, we perform several numerical exercises to demonstrate that for empirically reasonable parameter values, our intertemporal wage compression channel is both qualitatively and quantitatively significant. While the wage compression term is monotonically decreasing from above to below zero in static models, its dynamic counterpart is hump-shaped and is negative at both ends of the skill distribution. This pattern reflects the dynamic insurance and incentives motives highlighted above that are absent from static models. We illustrate that for a range of reasonable parameter values, the marginal labor tax rate in the left tail is negative, consistent with redistributive programs like the EITC. We also show that the labor tax is progressive in the short-run, in contrast to models with linear production functions in which skill premia are absent. In the cross-section, the optimal labor tax’s shape is determined by both the traditional forces from standard Mirrlees models and the wage compression channel and exhibits the hump-shaped pattern described above. Welfare gains relative to simple tax policies are small.

Finally, we explore two extensions of our baseline analysis. We first introduce skill bias in the production process, similar to Katz and Murphy (1992), and show that skill bias weakens the EITC. The reason is that the effects of exogenous skill bias can dominate those of endogenous wages and weaken the intertemporal wage compression channel. Second, we allow for endogenous human capital formation as in Stantcheva (2015) in that the planner can directly subsidize its accumulation. In contrast to static models that imply subsidies for high-skill agents and taxes for low-skill agents, as long as the most skilled agents do not benefit disproportionately, we show that wage compression forces in a dynamic setting raise subsidies for the least and most skilled agents relative to what they would be if the planner ignored skill premia. For the former, this reduces the cost providing valuable dynamic insurance while for the latter this lowers skill premia.
In addition to characterizing the intertemporal wage compression channel, we also make a methodological contribution to the literature. We are among the first to extend the techniques developed for heterogeneous agent models in continuous time by, for instance, Lasry and Lions (2007) and Nuño and Moll (2015), to a dynamic public finance setting that features imperfect substitutability in the production process. By casting our mechanism design problem in continuous time, we reduce the planner’s problem to a coupled system of partial differential equations, the Hamilton-Jacobi-Bellman equation to solve for policy functions and the Kolmogorov Forward Equation to solve for the (stationary) distribution of agents in the economy.\footnote{In addition to making the planner’s problem more tractable, the tools we employ make the planner’s problem more computationally manageable. Strong nonlinearities near the corners of the parameter space, however, impose limits on our ability to solve the model numerically.} Nuño and Moll (2015) develop a method to solve a social planner’s problem with heterogeneous agents in settings in which the market clearing conditions are “linear.” We adopt and extend their techniques to our setting with nonlinear aggregation.

The remainder of the paper is organized as follows. Section 1.2 discusses the related literature. Section 1.3 outlines a simplified version of the model to highlight its key features while 1.4 introduces the full model and derives the planner’s mechanism design problem. Section 1.5 highlights the key properties of the optimal tax system while Section 1.6 provides a quantitative analysis. Section 1.7 discusses some of the model’s main assumptions and Section 1.8 explores two extensions of the model (one with skill bias and another with endogenous human capital formation). Finally, Section 1.9 concludes.

### 1.2 Related Literature

The benchmark static Mirrlees models include Mirrlees (1971), Diamond (1998), and Saez (2001). The NDPF literature adds dynamics to these models by allowing skill to be persistent over time. Kocherlakota (2005), Albanesi and Sleet (2006), and Golosov,
Tsyvinski, and Werning (2006) are early examples; Farhi and Werning (2013) focus on time series properties of the tax system while Golosov, Troshkin, and Tsyvinski (2016b) investigate the cross-sectional properties, including the top tax rate under various skill distributions. All of these models focus on a linear production technology, and consequently abstract from endogenous skill premia. As such, they lack the forces we highlight that give rise to the optimality of programs like the EITC.

Stiglitz (1982) was the first paper to embed imperfect substitutability into a static Mirrlees model, using the simplest possible setting: static with two types of agents. Jacobs (2012) extends this to include endogenous human capital accumulation and Ales, Kurnaz, and Sleet (2015) include many types and a skill-to-task assignment framework. All three find that the wage compression channel is a force for a positive marginal rate at the bottom and negative rate at the top, which is the opposite of what is implied by programs like the EITC.\(^6\) Rothschild and Scheuer (2013, 2014) show that task choice partially undoes the wage compression force in a Mirrlees model in which workers have different skills in different tasks and choose how much to work on each task. Our framework focuses on the interaction between insurance in a dynamic setting and this wage compression channel.

To our knowledge, the only dynamic model of nonlinear taxation with imperfect substitutability is Heathcote, Storesletten, and Violante (2016). Though they use the same production technology as us and are motivated by the skill premium literature, they solve for the optimal tax system within a particular class of functions that they argue closely approximates the actual U.S. system. This also means that they take the EITC as given. As far as we are aware, ours is the first dynamic Mirrlees model with endogenous skill premia. While our optimal tax functions are qualitatively similar

\(^6\)The reason for the negative rate is that they assume skill is bounded, which Diamond and Saez (2011) argue is not practically useful for thinking about top tax rates. Even with unbounded support and a fat-tailed distribution so that the top rate is positive, however, it is still lower than what it would be with a linear production function
away from the right tail, ours also allows for history-dependence whereas theirs is a restricted function of only current income.

There are several studies, such as Diamond (1980), Saez (2002), Choné and Laroque (2005), Laroque (2005), Beaudry, Blackorby, and Szalay (2009), and Choné and Laroque (2010) that show that redistributive programs like the EITC are desirable. Saez (2002) shows that the optimality of the EITC hinges on two key ingredients: labor supply responses must be concentrated on the extensive margin (whether or not to work instead of how much) and the social planner must put disproportionately-large welfare weights on low-skill workers. Even with a powerful extensive margin, without sufficiently strong redistributive preferences there is no EITC in his model. In contrast, our model focuses exclusively on the intensive margin and the planner is utilitarian, which allows us to isolate the effects of the insurance value. Heckman, Lochner, and Cossa (2003) argue that the EITC can be justified on the grounds that it helps boost human capital accumulation; Stantcheva (2015) demonstrates that the planner indeed should subsidize human capital investment for unskilled workers in a dynamic Mirrlees model, but the subsidy acts as a force for intertemporal savings and the labor tax for these workers is still positive; our model does not require human capital to justify the EITC and hence, like the actual EITC, is not earmarked for specific expenditures.

We solve our heterogeneous agent model by using techniques from the theory of mean field games developed by Lasry and Lions (2007), Achdou, Han, Lasry, Lions, and Moll (2015), and Nuño and Moll (2015). Gabaix, Lasry, Lions, and Moll (2015) use these techniques to study the dynamics of inequality over time while Kaplan, Moll, and Violante (2016) use them to study monetary policy. To our knowledge we are the first to use them to study optimal contracts and tax policy.

Choné and Laroque (2010), unlike the other authors listed, do not require an extensive margin or strong redistributive preferences; they do, however, require multidimensional heterogeneity.
An advantage of working with these techniques is that characterizing the economy’s stationary equilibrium is straightforward, and we introduce stochastic retirement shocks in a perpetual youth framework to ensure that a stationary distribution exists. This also simplifies the tax system because it removes age/time as a state variable.

1.3 A Simple Model

Before developing the full model, we begin with a simplified version to illustrate the main forces at work and how negative rates for low-income agents can emerge. We defer to the next section for more details on the specifics of the model. There are two dates, $t = 0, 1$, and a unit measure of agents at each date. At $t = 0$, a fraction $\pi$ of the agents have high skill, $\theta_H$, while the remaining fraction $1 - \pi$ have low skill, $\theta_L < \theta_H$. Skill is stochastic across dates and evolves according to the Markov transition matrix

$$\Pi = \begin{bmatrix} 1 - \pi_L & \pi_L \\ 1 - \pi_H & \pi_H \end{bmatrix},$$

where $\pi_i = P(\theta_2 = \theta_H | \theta_1 = \theta_i)$. It follows that $\Pi_{L,1} = \pi (1 - \pi_H) + (1 - \pi) (1 - \pi_L)$ is the low-skill population size at $t = 1$ and $\Pi_{H,1} = \pi \pi_H + (1 - \pi) \pi_L$ is the high-skill population size at $t = 1$. We can think of the $\pi_i$ as capturing the “persistence” of the skill process: when $\pi_L = \pi_H = \pi$ skill is i.i.d. and when $0 = \pi_L < \pi_H = 1$ skill has perfect persistence.

An agent’s utility in state $s \in \{L, H, LL, LH, HL, HH\}$ when his skill is $\theta_i$ is

$$\tilde{u}(c_s, \ell_s) = u(c_s) - \phi \left( \frac{y_s}{\theta_i w_i} \right) = u(c_s) - \phi (\ell_s),$$

This contrasts with life cycle models, where agents know exactly when they will retire and this date is the same for everyone. Our model is thus a middle ground between the classic life cycle models and the endogenous retirement model of Shourideh and Troshkin (2015) in which production is linear and skill evolves deterministically over the life cycle.
where \( y_s = \theta_i w_i \ell_s \) is income, \( w_i \) is the wage, and \( \ell_s \) is labor. An agent’s expected lifetime utility, then, is

\[
U(\{c_s, \ell_s\}) = E_0 [\tilde{u}(c_{s_0}, \ell_{s_0}) + \beta \tilde{u}(c_{s_1}, \ell_{s_1})].
\]

Agents sell their labor to a representative firm at each date, which aggregates all labor into final output,

\[
Y_t = F(\Pi_{L,t}\theta_L\ell_{L,t}, \Pi_{H,t}\theta_H\ell_{H,t}),
\]

where \( F \) is a constant returns to scale production function with \( F_{jj} < 0 < F_{ij} \). Crucially, \( F \) need not be linear, as is the case in standard Mirrlees models. The firm’s profit-maximization problem imposes the wage constraints

\[
w_i \geq F_i \quad (1.1)
\]

for \( i \in \{L, H\} \) at each date. In particular, the firm’s problem is static so the wage depends on only current skill, not histories.

An agent’s skill at each date is his own private information so the planner designs a mechanism to induce agents to truthfully reveal their skills. An allocation specifies consumption and income at each date as a function of reported skill. Consider the planner’s problem at \( t = 0 \): he chooses an allocation to minimize costs subject to the wage-setting constraints (1.1), promise-keeping\(^9\)

\[
u(c_L) - \phi(\ell_L) + \beta \omega_L \geq U_0, \quad (1.2)
\]

\(^9\)There is also a promise-keeping constraint for high-skill agents but we do not need it in this section.
and incentive-compatibility

\[ u(c_H) - \phi(\ell_H) + \beta \omega_H \geq u(c_L) - \phi\left( \frac{\theta_L w_L}{\theta_H w_H} \ell_L \right) + \beta \omega_L|H, \]  \tag{1.3} \]

where \( \omega_i \) is promised utility at \( t = 1 \) for an agent with skill \( i \) at \( t = 0 \) and \( \omega_L|H \) is promised utility at \( t = 1 \) for an agent with high skill at \( t = 0 \) but who reported low skill.\(^{10}\) Since the promise-keeping constraint for the low-skill agent binds, we can combine (1.2) with (1.3) to get

\[ u(c_H) - \phi(\ell_H) + \beta \omega_H \geq U_0 + \phi(\ell_L) - \phi\left( \frac{\theta_L w_L}{\theta_H w_H} \ell_L \right) + \beta (\omega_L|H - \omega_L). \]

In static models such as Stiglitz (1982), only the “static incentives” terms are present and the planner sets \( \tau_L > 0 \) because this reduces the supply of low-skill labor, \( \ell_L \). As long as \( \theta_L w_L < \theta_H w_H \), this slackens (1.3) because high-skill agents must work more to mimic low-skill workers. This is the “intratemporal wage compression” channel. Now, however, the planner can control dynamic incentives and in particular, can relax the incentive constraint at \( t = 0 \) by affecting \( \omega_L \) and \( \omega_L|H \) with policy at \( t = 1 \).

To see how this works, consider the planner’s problem at \( t = 1 \) for an agent who reported low skill at \( t = 0 \).\(^{11}\) The planner must still respect promise-keeping and the incentive constraint, but there is a new threat-keeping constraint to ensure that a deviating agent does not receive utility above \( \omega_L|H \). Formally, the planner solves

\[
\min_{\{c_L, c_H, \ell_L, \ell_H, w_L, w_H\} \in (L, H)} \{ \pi_L [c_{LH} - \ell_{LH} \theta_H w_H] + (1 - \pi_L) [c_{LL} - \ell_{LL} \theta_L w_L] \}
\]

\(^{10}\)There is an additional incentive constraint that rules out low-skill agents misreporting as high-skill agents but this constraint is slack.

\(^{11}\)The problem for an agent who reported high skill at \( t = 0 \) is essentially static.
subject to the wage constraints (1.1) at $t = 1$, promise-keeping

$$\pi_L \tilde{u} (c_{LH}, \ell_{LH}) + (1 - \pi_L) \tilde{u} (c_{LL}, \ell_{LL}) \geq \omega_L, \quad (1.4)$$

incentive-compatibility

$$\tilde{u} (c_{LH}, \ell_{LH}) \geq \tilde{u} \left( c_{LL}, \frac{\theta_L w_L}{\theta_H w_H} \ell_{LL} \right), \quad (1.5)$$

and threat-keeping

$$\pi_H \tilde{u} (c_{LH}, \ell_{LH}) + (1 - \pi_H) \tilde{u} (c_{LL}, \ell_{LL}) \leq \omega_{L|H}. \quad (1.6)$$

Let $\lambda_L$ denote the Lagrange multiplier on (1.4) and $\lambda_{L|H}$ the multiplier on (1.6). Then the Lagrangian for this problem, ignoring (1.1) and (1.5), is

$$L^* = \min_{\{c_{L,H}, \ell_{L,H}, \omega_{L,H}\} \in \{(L,H)\}} \left\{ \pi_L (c_{LH} - \ell_{LH} \theta_H w_H) + (1 - \pi_L) (c_{LL} - \ell_{LL} \theta_L w_L) + \lambda_L \left( 1 - \frac{\lambda_{L|H}}{\lambda_L} \pi_H \right) \tilde{u} (c_{LH}, \ell_{LH}) + (1 - \pi_L) \left( 1 - \frac{\lambda_{L|H}}{\lambda_L} \frac{1 - \pi_H}{1 - \pi_L} \right) \tilde{u} (c_{LL}, \ell_{LL}) \right\}$$

Just as in models with linear production functions, $1 - \frac{\lambda_{L|H}}{\lambda_L} \frac{1 - \pi_H}{1 - \pi_L}$ is a “pseudo-welfare weight” on agents who report low skill in both periods (and $1 - \frac{\lambda_{L|H}}{\lambda_L} \frac{\pi_H}{\pi_L}$ is the weight on agents who report low then high skill). As long as $\pi_L < \pi_H$ then the former exceeds the latter and the planner would like to redistribute from the unlikely state, $(L, H)$, to the likely one, $(L, L)$ and provide more insurance to the poorest agents. It is easy to see from (1.4) and (1.6) that this transfer makes it easier to satisfy not only promise-keeping but also threat-keeping. The reason is that $(L, H)$ is an unlikely state so less utility in that state is outweighed by the benefit of more utility in the more likely state, $(L, L)$. 

11
By relaxing (1.4) and (1.6) the planner can increase $\omega_L$ and decrease $\omega_{L|H}$ at $t = 0$. But this lowers the right side of (1.3) through the “dynamic incentives” term and thus relaxes the $t = 0$ incentive constraint, making it cheaper to provide insurance at $t = 0$. This is the key difference between static and dynamic models. While in both cases, the planner would like to relax the $t = 0$ incentive constraint, in the static model his only tool is the tax function at $t = 0$ while in the dynamic model he can also use the tax function at $t = 1$.

As explained earlier, in static models the planner relaxes the $t = 0$ incentive constraint by raising the rate on low-skill agents (and lowering rates on high-skill agents). Now, however, the planner can accomplish the same goal by tightening the incentive constraint at $t = 1$, (1.5). In other words, he front-loads incentives by giving more information rent at $t = 0$ and less at $t = 1$. Relaxing the incentive constraint is more valuable at $t = 0$ than at $t = 1$ because this is when misreporting skill gives the agent the best information about his shocks in the future relative to the planner. But since (1.5) is completely static, the planner tightens it in exactly the same way he would in a static model: reduce the rate on low-skill agents. This increases the supply of low-skill labor, lowering their wages and making it easier for high-skill agents to mimic them. Indeed, whereas the Lagrange multiplier on (1.1) is positive for low-skill workers in a static model, it may be negative in a dynamic model, meaning the planner would like to pay them a lower wage in the second period. This is the “intertemporal wage compression” channel: the planner lowers the tax rate on low-skill workers in the future because doing so allows him to more cheaply provide insurance against shocks today.

Why does the planner want to increase wage inequality at $t = 1$ by subsidizing low-skill labor? The reason is that the threat of low wages in the future deters high-skill agents from misreporting today. If an agent with high skill reports low income today, the planner actually does not want to compress wages at $t = 1$, “wage expansion” is probably a better name for this channel.
then because skill is very persistent, the planner believes this agent will be low-skill in the future. For an agent who really has low skill, this is not a problem because the planner will provide him with more insurance anyway. But for an agent who really has high skill, this is costly because he will have to work more at a lower wage to reach the same level of income he would have received had he reported truthfully at $t = 0$. However, if this agent now decides to report truthfully at $t = 1$, he will still receive low utility because the planner redistributes away from that state and he thus has no incentive to misreport at $t = 0$ either.

In the full model, we assume that the production function $F$ is a constant elasticity of substitution aggregator with elasticity $\alpha$; when $\alpha \to \infty$ the model resembles a standard dynamic Mirrlees model while when $\alpha$ is close to 1 skills are more complementary. When skills are more complementary, $w_L$ can be fairly large if the supply of low-skill labor is low so the planner has more room to lower $w_L$ by encouraging low-skill labor without having to worry about wage inequality being too high. In addition, the wages are sensitive to changes in the labor supply so the planner can manipulate wages more easily by subsidizing labor; when $\alpha$ is very large, the planner does not subsidize labor because doing so has minimal impact on wages. In other words, negative rates are less likely when $\alpha$ is higher because subsidizing labor simply allows agents to earn more while working less since there are smaller opposing wage movements. In the full model we confirm numerically that the EITC is more generous when $\alpha$ is closer to 1.

We should stress that the optimality of the EITC in our setting is not a theorem and that for some parameter values, the intertemporal wage compression channel may not be strong enough to generate lower rates. However, when we solve the full model with empirically reasonable parameter values, the EITC usually does emerge.
1.4 Model

1.4.1 Environment

Time is continuous and is indexed by $t \in [0, \infty)$. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ denote a complete, filtered probability space, where the filtration $\{\mathcal{F}_t\}$ satisfies the usual conditions. The economy has a unit measure of agents who work until they retire. When an agent retires, a new agent is born to maintain a constant population size. Let $i \in [0, 1]$ index individual agents.

Agents are born with heterogeneous earning ability/skill $\theta^i_0$. Ability at each date is private information and evolves according to a diffusion process

$$d\theta^i_t = \mu_\theta(\theta^i_t) \, dt + \sigma(\theta^i_t) \, dZ^i_t,$$  \hspace{1cm} (1.7)

where $Z^i_t$ is a standard Wiener process adapted to the filtration for all $(i, t)$ in the economy\footnote{Going forward, we will omit the $i$-superscript when doing so does not create confusion.} and $\mu(\theta), \sigma(\theta) \in C^2(\Omega)$ are measurable functions that satisfy the standard Lipschitz and growth conditions:

$$\|\mu(t, x) - \mu(t, y)\| \leq A |x - y|, \quad \|\sigma(t, x) - \sigma(t, y)\| \leq A |x - y|,$$

$$\|\mu(t, x)\| \leq B (1 + |x|), \quad \|\sigma(t, x)\| \leq B (1 + |x|),$$

for some constants $A, B > 0$. Going forward, let $\mathcal{F}_t$ denote the natural filtration with respect to $\{\theta_t\}$. The specification in (1.7) is general enough to accommodate many standard functional forms. Unless otherwise stated, however, we take $\sigma(\theta) = \sigma_0 \theta$.

While most of our results go through for more general processes, this specification keeps the algebra simple and nests some special cases that we discuss below. Going forward, all functions are such that, when applied to $\{\theta_t\}$, they are $\mathcal{F}$-adapted.
We restrict $\theta_t \in \Theta \subseteq \mathbb{R}^+$, and define reflecting boundaries $\underline{\theta} = \inf \Theta > 0$, $\overline{\theta} = \sup \Theta < \infty$. The lower bound assumption is crucial for generating long-run dynamics since otherwise, all agents will have zero ability with probability one over the course of their lifetimes since zero is an absorbing boundary. Let $\theta^t = \{\theta_s\}_{s \in [0,t]}$ denote a history of abilities up to date $t$, and $\Theta^t$ the set of all histories up to date $t$.

As is standard in the optimal taxation literature, $\theta_t$ determines an agent’s stochastic productivity, which in turn determines an agent’s labor income. In particular, an agent with ability $\theta_t$ has productivity

$$e_t = e(\theta_t),$$

where $e_t$ is strictly increasing and concave. In most NDPF models, $e_t(\theta_t) = \theta_t$ and to simplify the math, we make this assumption, too. Agents supply labor $\ell_t \geq 0$ at date $t$, and hence $\theta_t \ell_t$ units of effective labor. An agent’s total labor income at each date is his total units of effective labor supplied multiplied by his wage per unit of effective labor, $w(\theta_t)$:

$$y_t = w(\theta_t) \theta_t \ell_t.$$

Agents work and consume, $c_t$, until they are hit by a Poisson retirement shock, $R_t$, with intensity $\kappa$. An agent is hit (or not hit) with the shock at the beginning of each date so that agents who are not hit work and consume as usual, while agents who are hit immediately retire and collect a “social security” payment. The retirement shock is observable. An alternative interpretation is that agents are hit with a “disability shock” and transfers take the form of disability payments.\textsuperscript{14} By the weak law of large numbers, exactly a fraction $\kappa$ of agents of each type retire at each date. After being hit by the shock, $R_t = 1$, an agent retires and receives a lump-sum transfer $C_t$. After

\textsuperscript{14}See Golosov and Tsyvinski (2006) for a model optimal disability insurance in which the retirement/disability shock is unobservable.
receiving this transfer, the agent exits the economy. When a new agent is born, his initial ability is drawn from a lognormal distribution, $\log(\theta_0) \sim N(\mu, \nu^2)$.

An important consequence of stochastic retirement is that the economy will have a stationary equilibrium and the state variables will have a stationary distribution.\textsuperscript{15} In fact, several distributions that are exogenously assumed in the literature emerge endogenously in our setting.

**Example 1.1.** Suppose $\kappa = 0$ and

$$d\theta_t = \theta_t \left[ - (1 - p) \left( \log(\theta_t) - \log(\theta^*) \right) + \frac{\sigma^2}{2} \right] dt + \sigma_\theta \theta_t dZ_t.$$

In this case, Itô’s Lemma implies

$$d \left( \log(\theta_t) \right) = - (1 - p) \left( \log(\theta_t) - \log(\theta^*) \right) dt + \sigma_\theta dZ_t.$$

This is called an Ornstein-Uhlenbeck process with persistence $p \in [0, 1]$ and is the continuous time analogue of the AR(1) process in discrete time. As long as $p < 1$ and $\bar{\theta} \to \infty$ then $\theta$ has a lognormal stationary distribution with mean and variance that depend on the parameters.

**Example 1.2.** Suppose

$$d\theta_t = \left( \mu_\theta + \frac{\sigma^2_\theta}{2} \right) \theta_t dt + \sigma_\theta \theta_t dZ_t$$

and $\bar{\theta} \to \infty$. Then Itô’s Lemma implies

$$d \left( \log(\theta_t) \right) = \mu_\theta dt + \sigma_\theta dZ_t,$$

\textsuperscript{15}Stochastic retirement is not always necessary to guarantee the existence of a stationary distribution, but many times it is sufficient.
so that \( \log(\theta_t) \) is a geometric Brownian motion. In this case, the stationary distribution of \( \theta \) is a double Pareto-lognormal distribution with probability density function

\[
g(\theta) = \frac{ab}{a+b} \left[ \theta^{-(1+a)} \exp \left( a\mu + \frac{a^2\nu^2}{2} \right) \Phi \left( \frac{\log(\theta) - \mu - a\nu^2}{\nu} \right) 
+ \theta^{b-1} \exp \left( -b\mu + \frac{a^2\nu^2}{2} \right) \left( 1 - \Phi \left( \frac{\ln(\theta) - \mu + b\nu^2}{\nu} \right) \right) \right],
\]

where \( a \) and \( -b \) (\( a,b > 0 \)) are the roots of the characteristic equation

\[
\frac{\sigma^2}{2} \zeta^2 + \left( \mu_0 - \frac{\sigma^2}{2} \right) \zeta - \kappa = 0.
\]

The double Pareto distribution is a special case when \( \nu \to 0 \), the Pareto lognormal distribution\(^{17}\) is a special case when \( b \to \infty \), and the lognormal distribution is a special case when \( a = b \to \infty \).

Agents have subjective discount factor \( \rho > 0 \). Each agent’s per-date utility while working is separable in consumption and effort:

\[
\tilde{u}(c_t, y_t; \theta_t) \equiv u(c_t) - \phi \left( \frac{y_t}{w(\theta_t)} \theta_t \right).
\]

As usual, \( u \in C^2(\Omega) \) is increasing and concave while \( \phi \in C^2(\Omega) \) is increasing and convex. Let \( v_t^{R}(C_t) \) denote an agent’s utility upon receiving the terminal transfer payment \( C_t \). Agents place weight \( \psi \) on utility received in retirement so that if an agent retires in date \( t \), he receives utility \( \psi u(C_t) \).

An allocation \( x_t : \Theta^t \to \mathbb{R}^3_+ \) is an \( \mathcal{F}_t \)-progressively measurable triple that specifies consumption, terminal consumption, and labor supplied at each date \( t \), conditional on the history \( \theta^t \): \( x_t = \{ x(\theta^t) \}_{\Theta^t} = \{ c(\theta^t), C(\theta^t), \ell(\theta^t) \}_{\Theta^t} \). Then an agent’s expected

\(^{16}\)See Reed and Jorgensen (2004).
\(^{17}\)See Colombi (1990).
lifetime utility from an allocation is given by

$$U(x_t) = \mathbb{E}_0 \left[ \int_0^\tau e^{-\rho t} \tilde{u}(c(\theta^t) , \ell(\theta^t)) \, dt + e^{-\rho \tau} v^R_{\tau} \left( C(\theta^\tau) \right) \right], \quad (1.8)$$

where $\tau = \inf\{t \mid R_t = 1\}$ is the date when the retirement shock hits. Since the shock has a Poisson arrival time, we can rewrite (1.8) as

$$U(x_t) = \mathbb{E}_0 \left[ \int_0^\infty e^{-(\rho+\kappa)t} \left[ \tilde{u}(c(\theta^t) , \ell(\theta^t)) + \kappa v^R_{\tau} \left( C(\theta^\tau) \right) \right] \, dt \right].$$

In particular, the shock acts just like a reduction in the discount factor.

Agents sell their labor to a representative firm and pay taxes on the income they earn. They can also earn net interest rate $r$ on riskfree savings, $a_t$. Let $T : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ denote a (non-stochastic) tax function on labor income and savings. Then the agent solves

$$\max_{\{c_t, C_t, \ell_t\}} U(x_t)$$

subject to the dynamic budget constraint

$$da_t = [w(\theta_t) \theta_t \ell_t + r a_t - T(w_t(\theta_t) \theta_t \ell_t, r a_t) - c_t] \, dt.$$

A representative firm hires all workers and produces output via a constant elasticity of substitution (CES) production function. The firm maximizes output, $Y_t$, minus labor costs. That is, the firm solves

$$\max_{\{L(\theta_t)\}} \left\{ \left( \int_{\Theta} L(\theta_t) \frac{\alpha-1}{\alpha} \, d\theta_t \right)^{\frac{\alpha}{\alpha-1}} - \int_{\Theta} w(\theta_t) L(\theta_t) \, d\theta \right\}$$

for all $t$, where

$$L(\theta_t) = \int_0^1 \theta_t \ell(\theta^t_i) g(\theta^t_i) \, di.$$
is the total units of effective labor supplied by workers with skill \( \theta_t \) and \( g \) is the probability density function over all agents.\(^{18}\) In particular, all the firm cares about is skill at \( t \), not the entire skill history, so that all agents with the same skill level at \( t \) are perfect substitutes while agents with different skill levels are imperfect substitutes. This means that an agent’s wage is a function of only his skill today, not his entire history. The firm does not directly observe skill but posts a menu of wage contracts \( \{ w(\theta_t) \} \) to induce agents to choose the contract meant for them. From the firm’s problem,

\[
w(\theta_t) = \left( \frac{Y_t}{L(\theta_t)} \right)^{\frac{1}{\alpha}}.
\]  

(1.9)

This functional form is as in the so-called “canonical model” in the skill premium literature but with all skill levels entering the production function symmetrically. However, it can be easily modified to include skill bias, as we do later, so that the technology puts more weight on more productive workers, in which case the technology resembles that of Katz and Murphy (1992). Although CES production functions are not widely used in the public finance literature, Heathcote, Storesletten, and Violante (2016) and Ales, Kurnaz, and Sleet (2015) both use variations of this function.

As \( \alpha \to \infty \), then \( w(\theta_t) \to 1 \) and we are back in the familiar NDPF setting in which agents are perfect substitutes across types. Now, however, each agent’s wage depends not only on his own ability and labor supply decision, but on every other agent’s ability and labor supply decision (through \( Y_t \)). That the wage differs across agents means there are skill premia across types: the skill premium between agents \( i \) and \( j \) is

\[
\pi_{ij}^t = \frac{w(\theta_t^i)}{w(\theta_t^j)} = \left( \frac{L(\theta_t^j)}{L(\theta_t^i)} \right)^{\frac{1}{\alpha}}.
\]

Again, as \( \alpha \to \infty \) then \( \pi_{ij}^t \to 1 \) and there are no skill premia.

Our equilibrium concept follows Mirrlees (1971):

\(^{18}\)Integrating over \( i \) is shorthand notation for integrating over histories that terminate at \( \theta_t \).
Definition 1.1 (Tax Equilibrium). A tax equilibrium is a tax function $T : \mathbb{R}^2_+ \rightarrow \mathbb{R}$, an allocation $\{x_t\}$, and wage profiles $\{w(\theta_t)\}$ such that, given public spending $\{G_t\}$:

1. The allocation solves an agent’s problem;

2. For each $\theta$ the wage is given by (1.9);

3. The goods market clears,

$$\int_{\Theta} \left[ c(\theta^t) + \kappa C(\theta^t) \right] g(\theta^t) d\theta^t + G_t \leq Y_t.$$ 

Let $X$ denote the set of equilibrium allocations. To characterize tax equilibria we solve for the optimal allocation via a mechanism design problem, with wages and taxes set to ensure that the allocation is implemented as part of the equilibrium.

1.4.2 Planner’s Problem

Following Mirrlees (1971), while the planner observes each agent’s income, $y_t$, and consumption, $c_t$, $C_t$, he cannot observe ability, $\theta_t$. Therefore, he cannot observe an agent’s labor supply decision, $\ell_t$. The planner relies on agents to report $\theta_t$ at each date.

In this section, we set up the planner’s mechanism design problem. A mechanism specifies, given each agent’s report, how much an agent should produce and consume, i.e., an allocation. We do this in two steps: first, we characterize the dynamics of the optimal contract by formulating a continuous time version of the first-order approach and deriving a set of necessary and sufficient conditions for the optimal contract under the relaxed problem to coincide with the optimal contract of the full problem. This also delivers the relevant state variables. Then, we combine the state variables’ laws of motion derived via the first-order approach with the economy-wide law of motion and the firm’s constraints to derive the planner’s full recursive problem.
**Contracting Problem**

By the Revelation Principle, we can restrict attention to direct revelation mechanisms in which agents truthfully report their types. A report $\sigma_t(\theta^t) \in \Sigma_t$ is an $\mathcal{F}_t$-progressively measurable function that specifies some $\theta_t \in \Theta$ that the agent reports to the planner following the history $\theta^t$, and a reporting strategy $\sigma = \{\sigma_t(\theta^t)\} \in \Sigma$ is a history of reports. Let $\sigma^t \in \Sigma^t$ denote the history of reports generated by reporting strategy $\sigma$. The planner can specify allocations directly as a function of histories of reports, $x(\theta^t) = x(\sigma^t)$. Without loss of generality, we can focus on reporting strategies with $\Sigma^t \subseteq \Theta^t$ since otherwise, the planner immediately detects a lie.

Given a history $\theta^t$ and a reporting strategy $\sigma$, let $v_{\sigma}^t(\theta^t)$ denote an agent’s promised utility/continuation value:

$$
v_{\sigma}^t(\theta^t) \equiv \mathbb{E}_t \left[ \int_t^T e^{-\rho(s-t)} \tilde{u}(\sigma_s(\theta^s)) \, ds + e^{-\rho \tau} v_{R}^\tau(\sigma^\tau(\theta^\tau)) \right].
$$

The truth-telling strategy specifies $\sigma_t(\theta^t) = \theta_t$ after all histories, for all $t$. Let $v_t \in \mathcal{V}$ denote the value of $v_{\sigma}^t$ under truth-telling. We say that an allocation is **incentive-compatible** if truth-telling yields weakly higher continuation utility than any other reporting strategy $\sigma$:

$$
v_t(\theta^t) \geq v_{\sigma}^t(\theta^t) \quad \forall \sigma \in \Sigma, \ \theta^t \in \Theta^t. \tag{1.10}
$$

Let $X^{\text{IC}}$ denote the set of incentive-compatible allocations, i.e., those allocations that satisfy the incentive-compatibility constraint (1.10). Of course, for an agent to participate in the game, his utility from an allocation must exceed that of some outside option:

$$
U(x) \geq U. \tag{1.11}
$$

This is an agent’s participation constraint.
While agents do not have access to a savings market, the planner can save on their behalf, earning net interest rate \( r > 0 \). Then the cost to the planner of delivering an allocation \( x = \{c, C, \ell\} \) is

\[
\Psi(x) = E_0 \left[ \int_0^\infty e^{-rt} \left[ c(\theta^t) - y(\theta^t) + \kappa C(\theta^t) \right] dt \right].
\]

The expectation \( E_0 \) is over the distribution of agents in the economy, \( c - y \) is the planner’s net cost to the \( 1 - \kappa \) agents who do not retire, while \( \kappa C \) is the cost of social security payments to the \( \kappa \) agents who do retire. The planner selects a tax equilibrium, i.e., an allocation \( x \) and tax function \( T \), that is individually rational and incentive compatible for all agents, and is as cheap as possible.

The planner has a utilitarian welfare function so that all agents are equally-weighted. Then the planner’s contracting problem, i.e., ignoring the aggregate law of motion and firm’s problem, written sequentially, is

\[
K(\mathcal{U}) \equiv \min_{x \in \mathcal{X}} \Psi(x)
\]

subject to (1.10), (1.11), and \( c(\theta^t), C(\theta^t), \ell(\theta^t) \geq 0 \). Let \( x^* \) denote an optimal equilibrium allocation and \( T^* \) an optimal equilibrium tax function. The planner has both an insurance and redistributive motive: the planner insures agents against fluctuations over time in their ability and redistributes, within a date, from higher-ability agents to lower-ability ones. At the same time, the tax system must induce agents to supply labor efficiently. Therefore, the planner must balance insurance/redistribution and incentives.

This problem is intractable as currently written because (1.10) is a hopelessly complicated set of constraints because we need to rule out every possible deviation following every possible history. To get around this, we use the first-order approach. Then we will formulate the planner’s relaxed problem recursively. Finally, we have
to go back and check ex post that the solution to the relaxed problem also solves the full problem above. However, we instead invoke a simple sufficient condition under which this holds.

Instead of going through the first-order approach in full detail, we will just present the relevant equations as results with a brief description. The results in this section are special cases of those in Chapter 2 of this dissertation so all of the relevant proofs and more detailed discussion can be found there.

To formulate the planner’s problem recursively, we need each state variable’s law of motion. We already have the law of motion for $\theta_t$, which is simply (1.7). The law of motion for $v_t$ is standard in the continuous time optimal contracting literature.

**Result 1.1.** Fix a contract and a reporting strategy $\{\hat{\theta}_t\}$ with finite expected payoff to the agent. Then the process $\{v_t\}$ corresponds to the agent’s continuation utility if and only if there exists a process $\hat{\Delta}_t \in H^2$ such that

$$
dv_t = (\rho v_t - \tilde{u}_t)\,dt + \sigma_{\theta} \hat{\Delta}_t \,dZ_t + (v^R_t - v_t) (dR_t - \kappa \,dt) \tag{1.12}
$$

and the transversality condition $E_t \left[ e^{-\rho(T\wedge\tau)} v_{T\wedge\tau} \right] \to 0$ as $T \to \infty$ holds.

The drift of (1.12) tracks promise-keeping: $\tilde{u}_t$ is just-delivered utility and $\rho v_t$ is everything owed going forward. The transversality condition ensures that in the long-run, the agent receives everything he is owed. The process $\hat{\Delta}_t$ is the sensitivity of $v_t$ to shocks in the underlying process $\theta_t$.

The third state variable is an agent’s information rent, which arises because if an agent misreports today, he affects the trajectory of the planner’s beliefs into the future. Formally, an agent’s information rent, $\Delta_t$, is the sensitivity of his continuation value to his report, evaluated at truth-telling:

$$
\Delta_t \equiv \left. \frac{\partial v_t}{\partial \hat{\theta}_t} \right|_{\hat{\theta}_t = \theta_t} = \frac{\partial v_t}{\partial \theta_t}. \tag{1.13}
$$
Let $\Delta_t \in \mathcal{D}$. We are now ready to state the first-order condition.

**Result 1.2.** A necessary condition for truth-telling to be optimal is that for all $t$,

$$\hat{\Delta}_t = \Delta_t, \quad (1.14)$$

where $\hat{\Delta}_t$ is the process in Result 1.1.

Let $X^{FOA}$ denote the set of allocations that satisfy (1.14), which necessarily satisfies $X^{IC} \subseteq X^{FOA}$. Using Result 1.2, we can derive the law of motion.

**Result 1.3.** The finite process $\{\Delta_t\}$ is characterized by (1.13) if and only if, for some $\sigma_{\Delta,t} \in H^2$,

$$d\Delta_t = \left(\left(\rho - \mu'(\theta_t)\right) \Delta_t - \bar{u}_{\theta,t} - \sigma^2_{\theta} \sigma_{\Delta,t}\right) dt + \sigma_{\theta} \sigma_{\Delta,t} dZ_t + \left(\Delta_t^R - \Delta_t\right) dR_t - \kappa dt$$

(1.15)

and the transversality condition $E_t \left[e^{-\rho(T\wedge \tau)} \Delta_{T\wedge \tau}\right] \to 0$ as $T \to \infty$ holds.

The planner uses the process $\sigma_{\Delta,t}$ to control the flow of information rent over time. In the drift of (1.15), $-\mu'(\theta_t) \Delta_t$ is the decay in information rent while $\bar{u}_{\theta,t} + \sigma^2_{\theta} \sigma_{\Delta,t}$ is delivered information rent. The term $\Delta_t^R$ is the sensitivity of continuation utility at retirement, where $\Delta_t$ can jump: if it jumps down then the principal must provide more information rent while the agent works.

Using Results 1.1 and 1.3, we can formulate the planner’s relaxed problem:

$$K(U) \equiv \min_x \Psi(x)$$

subject to

$$v_0 = E_0 \left[\int_0^\infty e^{-(\rho + \kappa)t} \left(\bar{u}_t + \kappa v_t^R\right) dt\right],$$

the first-order condition (1.14), and $c, C, \ell \geq 0$. Consider the state variables $(\theta_t, v_t, \Delta_t)$ driven by the controls $\{c_t, C_t, \ell_t, \sigma_{\Delta,t}\}$ according to the above laws of motion and the
appropriate transversality conditions. Then Results 1.1 and 1.3 imply that this control problem (with the above objective function) is equivalent to the relaxed problem.

An important issue is whether or not the solution to the relaxed problem necessarily solves the full problem, too. Kapička (2013) and Pavan, Segal, and Toikka (2014) derive complicated integral monotonicity conditions that, if satisfied, imply the optimal contract derived under the first-order approach is globally optimal. Instead, we invoke the simple condition derived in Chapter 2. We need the following assumption.

**Assumption 1.1.** The drift of an agent’s skill process satisfies $2\mu'(\theta_t) + \sigma^2_\theta < \rho$.

This assumption rules out processes that are “too explosive.” If log($\theta_t$) is an Ornstein-Uhlenbeck process, then this condition is $2\sigma^2_\theta < \rho$ if $p = 1$ and

$$\frac{2\sigma^2_\theta - \rho}{2(1 - p)} + \log(\theta^*) < 1 + \log(\theta_t),$$

if $p < 1$. While the literature usually sets $p = 1$ so that our condition holds under common estimates of $\sigma^2_\theta$, it still holds if $p < 1$ as long as $\theta_t$ is not too small. If log($\theta_t$) is a Brownian motion with drift, then the condition is

$$\mu_\theta + \sigma^2_\theta < \rho,$$

which holds for reasonable values of $(\mu_\theta, \sigma_\theta, \rho)$ (typically, $\rho \gg \sigma^2_\theta > \mu_\theta$).

**Result 1.4.** Suppose agents cannot overreport their types, $\hat{\theta}_t \leq \theta_t$ for almost all $t$. Then under Assumption 1.1 and the transversality condition of Result 1.1, the optimal contract under the first-order approach is globally optimal.

The assumption in Result 1.4 that agents cannot overreport is fairly ubiquitous in the optimal taxation literature. One way to micro-found this restriction is to assume that an agent must be able to show that he has any income that he reports.
Fully Problem

Now we formulate the planner’s full recursive problem. In addition to the contracting
constraints, the planner must account for the firm’s problem and the labor market
clearing conditions. Rearranging (1.9),

\[ w(\theta_t) \alpha L(\theta_t) = Y_t. \]

Combining this with the labor market clearing condition for each \( \theta \) yields

\[
w(\theta_t) \alpha L(\theta_t) = w(\theta_t)^\alpha \int_{\mathcal{V} \times \mathcal{D}} \theta_t \ell(\theta_t, v_t, \Delta_t) g(\theta_t, v_t, \Delta_t) d(v_t, \Delta_t) \\
= \int_{\Theta \times \mathcal{V} \times \mathcal{D}} w(\theta_t) \theta_t \ell(\theta_t, v_t, \Delta_t) g(\theta_t, v_t, \Delta_t) d(v_t, \Delta_t, \theta_t) = Y_t \quad (1.16)
\]

for all \( \theta \). The wage \( w(\theta_t) \) acts like an additional state variable in the planner’s
problem so we need its law of motion in order to apply standard stochastic optimal
control methods. Define

\[ w'(\theta_t) = z(\theta_t). \quad (1.17) \]

This ensures that the wage function is differentiable and the tax equilibrium is well-
behaved. Here, \( z_t \) is a control so the planner controls the shape of the wage function.
Recall that since all the firm cares about is skill at \( t \), the wage is a function of \( \theta_t \)
alone and the planner must respect this measurability constraint. Finally, we need a
law of motion for the economy-wide distribution, \( g(\theta, v, \Delta, t) \), which is given by the
Kolmogorov Forward Equation (KFE)

\[
\frac{\partial g}{\partial t} = \mathcal{A}^* g + \kappa \tilde{f}(\theta_0, v_0, \Delta_0), \quad (1.18)
\]

\[
1 = \int_{\Theta \times \mathcal{V} \times \mathcal{D}} g(\theta, v, \Delta, t) d(\theta, v, \Delta), \quad (1.19)
\]
where $\tilde{f}(\theta_0, v_0, \Delta_0) = \tilde{f}_0$ is the density of reborn agents\(^{19}\) and $A^* g$ is the adjoint of the infinitesimal generator

$$A^* g = -\sum_{i=1}^{n} \frac{\partial}{\partial s_i} [\mu_i(s, m, A) g(s, t)] + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2}{\partial s_i \partial s_j} \left[ \left( \sigma(s) \sigma(s)^T \right)_{i,k} g(s, t) \right] - \kappa g(s, t).$$

Here, $s = (s_1, \ldots, s_n)$ is the vector of state variables, $m$ is the vector of controls, and $A$ is the vector of aggregate variables. In our case, these are, respectively, $(\theta, v, \Delta)$, \{c, C, $\ell$, $\sigma$, $\Delta$, z\}, and the wages $w(\theta)$. We will solve for the stationary equilibrium so that the left side of (1.18) is zero and $g$ does not depend on $t$.

Summarizing, the planner’s objective is to find the cheapest possible allocation subject to the participation constraint (1.12), the incentive constraint (1.15), the market clearing conditions (1.16), the wage law of motion (1.17), and the law of motion of the distribution as captured by the KFE, (1.18) and (1.19). We call an allocation $x$ efficient if it solves this problem.

Let $k = k(v, \Delta, \theta)$ denote the planner’s cost of insuring an agent in state $(v, \Delta, \theta)$, and let $K(\theta) = \int_{V \times D} k(\theta, v, \Delta) g(\theta, v, \Delta) d(v, \Delta)$ denote the social impact function for each $\theta$, i.e., the cost to the planner for all agents with skill $\theta$. The total cost is then $K = \int k(\theta, v, \Delta) g(\theta, v, \Delta) d(\theta, v, \Delta)$. We follow Nuño and Moll (2015) to derive the planner’s Hamilton-Jacobi-Bellman (HJB) equation by writing the planner’s problem as an inner product and taking the appropriate first-order conditions. The planner’s objective function, written as an inner product,\(^{20}\) is

$$K_t = \left\langle e^{-rt} \left( c(\theta_t, v_t, \Delta_t) + \kappa C(\theta_t, v_t, \Delta_t) - w(\theta_t) \theta_t \ell(\theta_t, v_t, \Delta_t) \right) , g(\theta_t, v_t, \Delta_t) \right\rangle_{\Theta \times V \times D \times [0, \infty)}.$$

\(^{19}\)Earlier, we specified that the marginal distribution of $\theta_0$ is a lognormal. If all agents are reborn at a single point $(\theta_0, v_0, \Delta_0)$, this becomes the Dirac delta function evaluated there, $\delta(\theta_0, v_0, \Delta_0)$.

\(^{20}\)For arbitrary functions $u, g \in L^2(\Omega)$, we define the inner product $\langle u, g \rangle = \int_{\Omega} u(\omega) g(\omega) d\omega$, where $\omega \in \Omega$. 

27
Let $\chi(\theta_t)$ denote the Lagrange multiplier on (1.16) and $\eta(\theta_t)$ the multiplier on (1.17). Then the Lagrangian functional for the planner’s problem at each $t$ is

$$L(g_t) = \left\langle e^{-rt} \left( c_t + \kappa C_t - w(\theta_t) \theta_t \ell_t + Ak_t - rk_t + \partial_t k_t + \kappa k_t \bar{f}_0 \right), g(\theta_t, v_t, \Delta_t) \right\rangle_{\Theta \times V \times D \times [0,\infty)}$$

$$+ \left\langle e^{-rt} \chi(\theta_t), w(\theta_t)^a L(\theta_t) - Y_t \right\rangle_{\Theta \times V \times D \times [0,\infty)} + \left\langle e^{-rt} \eta(\theta_t), z(\theta_t) \right\rangle_{\Theta \times V \times D \times [0,\infty)};$$

where $L(\theta_t)$ and $Y_t$ are as in (1.16) and $Ak$ is the infinitesimal generator

$$Ak = \sum_{i=1}^n \frac{\partial k}{\partial s_i} \mu_i(s, m, A) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left( \sigma(s) \sigma(s)^T \right) i,j \frac{\partial^2 k}{\partial s_i \partial s_j} - \kappa k.$$ 

Note that we can write the integral over (1.16) as

$$\int_\Theta \chi(\theta) \left[ w(\theta)^a \theta \int_{\Theta \times D} \ell(\theta, v, \Delta) g(\theta, v, \Delta) d(v, \Delta) - \int_{\Theta \times V \times D} w(\theta') \theta' \ell(\theta', v, \Delta) g(\theta', v, \Delta) d(v, \Delta, \theta') \right] d\theta$$

$$= \int_{\Theta \times V \times D} \chi(\theta) \left[ w(\theta)^a \theta \ell(\theta, v, \Delta) g(\theta, v, \Delta) - \int_{\Theta} w(\theta') \theta' \ell(\theta', v, \Delta) g(\theta', v, \Delta) d\theta' \right] d(v, \Delta, \theta)$$

$$= \int_{\Theta \times V \times D} \left[ \chi(\theta) w(\theta)^a \theta \ell(\theta, v, \Delta) - \left( \int_{\Theta} \chi(\theta') d\theta' \right) w(\theta) \theta \ell(\theta, v, \Delta) \right] g(\theta, v, \Delta) d(v, \Delta, \theta),$$

where the last step follows from integration by parts. Let $\mathcal{X} \equiv \int_{\Theta} \chi(\theta') d\theta'$. Taking the Gateaux derivative of the Lagrangian with respect to $k$ along an arbitrary variation and setting it to zero yields the planner’s HJB equation in the stationary equilibrium given $\theta$:

$$(r + \kappa) K(\theta) = \inf_{\{c,C,k,\ell,\sigma,\Delta,\bar{z}\}} \left\{ \int_{\Theta \times D} \left( c + \kappa C - \theta \ell w(\theta) + k_\theta \mu(\theta) + k_v \left[ (\rho + \kappa) v - \bar{u} - \kappa v^R \right] 

+ k_\Delta \left[ (\rho + \kappa - \mu'(\theta)) \Delta - \bar{u}_\theta - \kappa \Delta^R - \sigma_\theta^2 \sigma_\Delta \right] + \frac{\sigma_\theta^2}{2} \left[ k_\theta \theta^2 \Delta^2 + k_\Delta \Delta \sigma^2 + k_{\theta \theta} \theta \right] 

+ \sigma_\theta^2 \left[ k_\theta \Delta \sigma_\Delta + k_{\theta \theta} \theta^2 \Delta + k_{\theta \theta} \theta \Delta \sigma_\Delta \right] + \kappa k_\bar{f}_0 

+ \chi(\theta) w(\theta)^a \theta \ell - \mathcal{X} w(\theta) \theta \ell g(\theta, v, \Delta) d(v, \Delta) + \eta(\theta) z(\theta) \right\}. \quad (1.20)$$
Since (1.18) and (1.20) are both three-dimensional second-order partial differential equations, we need six boundary conditions for each one to pin down a unique solution; we discuss these boundary conditions in Appendix B. This system plus the boundary conditions, what Lasry and Lions (2007) call a “mean field game,” together characterize the stationary equilibrium. Intuitively, agents care about only the “mean” of all other agents’ actions, not each one individually. The policy functions implied by this system yield the efficient allocation.

This mean field game illustrates a key distinction between familiar models with a linear production functions and our model. Agents’ policy functions depend on $w$ and with linear production function, this is simply one so that $z(\theta) = 0$ for all $\theta$ and $g$ does not matter for solving HJB equation. Therefore, each agent’s optimal policies are independent of the distribution. This means we can solve the system sequentially: derive the optimal policies via (1.20), plug these into the KFE, and then solve for the stationary distribution via (1.18). On the other hand, with a CES function then $w$ depends on $g$ so that the system is coupled and we have to solve the system simultaneously: each agent’s optimal policies depend on the distribution, which itself depends on each agent’s policies. In particular, (1.20) depends on the aggregate quantity $\mathcal{X}$, not just $\chi(\theta)$. This generalizes the framework in Nuño and Moll (2015), where all of the market clearing conditions can be aggregated linearly. In their setting, capital is the mean field and the planner’s HJB equation can be solved state-by-state while here, the mean fields are the wage functions and we need to solve for everything at once.

It turns out that the HJB equation (1.20) is difficult to work with because several derivatives depend on the choice variable $\sigma$. To get around this, we use duality to work in an alternative state space parametrized by the Lagrange multipliers in the above problem. We briefly explore this approach in the next section to characterize optimal tax policies.
We solve our model using a finite difference method to approximate the partial differential equations. We use an “upwind scheme” to avoid dealing with boundary conditions, although it turns out that in the reparametrized state space, we do not need boundary conditions anyway to recover a unique solution; more detail can be found in Appendix D.

1.5 Optimal Tax Policies

This section characterizes the allocations obtained as solutions to the relaxed problem.

Marginal distortions in agents’ choices can be characterized with “wedges” that represent marginal tax rates. Given a history $\theta^t$, the two key wedges are the labor wedge, $\tau_L(\theta^t)$, and the intertemporal, or savings wedge, $\tau_K(\theta^t)$, both of which are standard in the literature:

$$
\tau_L(\theta^t) = 1 + \frac{\bar{u}_{y,t}}{u_{c,t}} = 1 - \frac{\phi'(\ell_t)}{w(\theta_t) \theta_t u_{c,t}},
$$

$$
\tau_K(\theta^t) = r - \rho + \frac{1}{dt} \mathbb{E}_t [du_{c,t}] = \theta_t u_{c,t}.
$$

The labor wedge is simply the gap between the marginal rate of substitution and the marginal rate of transformation between consumption and labor while the savings wedge is the difference between the marginal rate of intertemporal substitution and the return on savings. In the first-best allocation with perfect information, both of these wedges are zero. Also, both of these wedges are pre-retirement because at retirement there is no more asymmetric information. As long as the tax function $T$ is differentiable, $\tau_L$ is the derivative of $T$ with respect to income (and similarly for $\tau_K$).

\footnote{We use “wedge” and “tax rate” interchangeably, sidestepping the issue of how to implement the constrained efficient allocation with a simple tax system. Kocherlakota (2005) and Albanesi and Sleet (2006) show how to use tax systems to implement the optimal allocation when skill is not too persistent, while Golosov, Troshkin, and Tsyvinski (2016a) show that a tax system based on consolidated income accounts can generally implement the optimal allocation.}
Now we will derive and characterize the optimal tax formula. To characterize the wedges we will solve the recursive planner’s problem. The first-order conditions of (1.20) are

\[ c : 1 = k_v u_c, \]
\[ C : 1 = k_v v^R_C + k_\Delta \Delta_C^R, \]
\[ \ell : -\theta w (\theta) = k_v \tilde{u}_\ell + k_\Delta \tilde{u}_{\theta \ell} - \chi (\theta) w (\theta)^\alpha \theta + \lambda w (\theta) \theta, \]
\[ [\sigma] : k_\Delta = k_\Delta \sigma_\Delta + k_v \lambda \Delta + k_\Delta \theta, \]
\[ [z (\theta)] : \eta (\theta) = \int_{V \times D} k_\Delta \tilde{u}_{\theta z} g (\theta, v, \Delta) d (v, \Delta). \]

The first two equations are easy to interpret: increasing consumption by one unit increases the planner’s cost by one unit. However, this lowers an agent’s marginal utility of consumption and makes it easier to satisfy the promise-keeping constraint going forward. The third equation balances the effects of asking a given agent to work more by accounting for the effect on everyone’s wages. The fourth equation weighs the cost of delivering more information rent today against the benefit of relaxing the incentive constraints. The final equation weighs the cost of increasing the slope of the wage function against the benefit of loosening incentive constraints. This last equation also highlights that an individual agent’s policy functions depend on everyone else in the economy.

Instead of working with the state space as given, it is simpler to work with “dual” variables. In particular, following Farhi and Werning (2013), define \( \lambda \equiv k_v \) and \( \gamma \equiv k_\Delta \). This reparametrizes the state space as \( (\lambda, \gamma, \theta) \in \Lambda \times \Gamma \times \Theta \). This new state space has two advantages over the original. First, as we show here, it is quite simple to characterize the optimal allocations, wedges, and their dynamics. The second advantage, which we explore in Appendices C and D, is that it makes numerically solving for the policy functions simpler.
Using the new state space, the first-order conditions are

\[ c : 1 = \lambda u_c, \tag{1.23} \]
\[ C : 1 = \lambda v^R_C + \gamma \Delta^R_C, \tag{1.24} \]
\[ \ell : - \theta w (\theta) = \lambda \tilde{u}_\ell + \gamma \tilde{u}_{\theta \ell} - \chi (\theta) w (\theta)^\alpha \theta + X w (\theta) \theta, \tag{1.25} \]
\[ [\sigma_\Delta] : \gamma = k_{\Delta \Delta} \sigma_\Delta + k_{\Delta \theta} \theta \Delta + k_{\Delta \theta} \theta, \tag{1.26} \]
\[ [z (\theta)] : \eta (\theta) = \int_{\Lambda \times \Gamma} \gamma \tilde{u}_{\theta z} g (\lambda, \gamma, \theta) d (\lambda, \gamma). \tag{1.27} \]

There is also an envelope condition from the state variable \( w (\theta) \),

\[ [\text{Env}] : - \eta' (\theta) = L (\theta) [1 + \alpha \chi (\theta) w (\theta)^{\alpha-1} - \mathcal{X}] + \int_{\Lambda \times \Gamma} \gamma \tilde{u}_{\theta w} g (\lambda, \gamma, \theta) d (\lambda, \gamma). \]

Since the firm’s problem is static, this depends on only \( \theta \) and the other two state variables are integrated out.

Since we are working with new state variables, we need their equations of motion. The next result, which is similar to Proposition 5 in Farhi and Werning (2013), does just that.

**Proposition 1.1.** The solution to the planner’s problem is characterized by the following: there exists a function \( \sigma_{\lambda,t} \) such that the stochastic processes for \( \{\lambda_t\} \) and \( \{\gamma_t\} \) satisfy the following stochastic differential equations:

\[ d\lambda_t = (r - \rho) \lambda_t \, dt + \sigma_\theta \sigma_{\lambda,t} \lambda_t \, dZ_t + \left( \lambda^R_t - \lambda_t \right) dR_t, \tag{1.28} \]
\[ d\gamma_t = \left( (r - \rho + \mu' (\theta_t)) \gamma_t - \sigma^2_{\theta} \theta_t \sigma_{\lambda,t} \lambda_t \right) dt + \sigma_{\theta} \gamma_t \, dZ_t + \left( \gamma^R_t - \gamma_t \right) dR_t \tag{1.29} \]

with \( \gamma_0 \) free.

The law of motion for \( \lambda_t \) is the familiar “inverse Euler equation” and shows that the inverse of the marginal utility of consumption is a submartingale if \( r > \rho \), a
supermartingale if \( r < \rho \), and a martingale if \( r = \rho \). Note that these equations are much simpler and more explicit than those of \( \{v_t\} \) and \( \{\Delta_t\} \). Finally, it turns out that \( \Delta_t^R \), and hence \( \gamma_t^R \), is zero. This is because once the retirement shock hits, there is no more asymmetric information so agents do not receive any information rent at retirement, and since all process are càdlàg information rent must equal zero at the jump. If the retirement shock is private as in Golosov and Tsyvinski (2006) then \( \Delta_t^R \neq 0 \) since agents can use the shock to extract more information rent.\(^{22}\) We set \( \gamma_0 \) so that agents have zero information rent at birth.

Using the equations of motion of the dual variables, we can compute the wedges. To derive an expression for the labor wedge, first note that

\[
\tilde{u}_t = -\phi_t \implies \tilde{u}_\theta t = \left( \frac{1}{\theta} + \frac{w'(\theta)}{w(\theta)} \right) (\phi_t + \ell \phi_{\ell t}) = \left( \frac{1}{\theta} + \frac{w'(\theta)}{w(\theta)} \right) \left( 1 + \frac{1}{\varepsilon} \right) \phi_t,
\]

where \( \varepsilon \equiv \frac{\phi_t}{\ell \phi_{\ell t}} \) is the Frisch elasticity of labor supply.\(^{23}\) Plugging these into (1.25) and recalling the definition of \( \tau_{L,t} \) from (1.21),

\[
\frac{\tau_{L,t}}{1 - \tau_{L,t}} = -\gamma_t \left( 1 + \frac{1}{\varepsilon} \right) \left( \frac{1}{\theta_t} + \frac{w'(\theta_t)}{w(\theta_t)} \right) \phi_{\ell,t} - w(\theta_t) \left[ \chi_t (\theta_t) w(\theta_t)^{\alpha - 1} \theta_t - X_t \theta_t \right] \implies
\]

\[
\frac{\tau_{L,t}}{1 - \tau_{L,t}} = -\gamma_t \frac{1 + 1}{\lambda_t} \left( \frac{1}{\theta_t} + \frac{w'(\theta_t)}{w(\theta_t)} \right) + \frac{w(\theta_t) \theta_t \chi_t (\theta_t) w(\theta_t)^{\alpha - 1} - X_t \theta_t}{\lambda_t \phi_{\ell,t}}.
\]

This expression generalizes the ones in Stiglitz (1982) and Ales, Kurnaz, and Sleet (2015) to include dynamics and the ones in Farhi and Werning (2013) and Golosov, Troshkin, and Tsyvinski (2016b) to include wage compression forces.

The Mirrlees term is standard in all Mirrlees models and captures the planner’s insurance motive against shocks. It is the tax rate the planner would implement if

\[^{22}\text{They show that agents do not receive a final transfer unless their assets are sufficiently low.}\]

\[^{23}\text{More generally,} \frac{1 + \varepsilon^u}{\varepsilon^c}, \text{where} \varepsilon^u, \varepsilon^c \text{are the uncompensated and compensated elasticities of demand; with separable preferences, this reduces to} 1 + \frac{1}{\varepsilon}.\]
he set agents’ wages optimally but ignored any additional effects of the nonlinear production function. The only difference between our expression and the one in Farhi and Werning (2013) is the $w′(\theta_t)/w(\theta_t)$ term, which reflects that wages are not constant across types in our model. Under standard distributions, this ratio is positive for large $\theta$ and negative for small $\theta$. This means that more skilled workers also receive a larger wage, making it easier for them to mimic less skilled workers and tightening their incentive constraint. This is a force for a higher marginal tax rate on more skilled agents and a lower rate on less skilled agents. However, the Mirrlees term is always nonnegative and so on its own is inconsistent with the EITC.

The wage compression term was first identified by Stiglitz (1982) and further analyzed by Ales, Kurnaz, and Sleet (2015), both in static models. In their contexts, the planner should subsidize labor that decreases the wages of skilled agents and tax labor that decreases the wages of unskilled agents. When skill has bounded support, this means the most skilled agents should face a negative marginal tax rate while the least skilled agents should face a positive rate.\(^{24}\) Indeed, the negative rate at the top increases the supply of skilled labor, lowering the wages of the most skilled workers, thus reducing skill premia. The negative rate at the top also loosens incentive constraints and makes skilled agents less likely to mimic unskilled ones. Conversely, a higher rate on unskilled workers reduces the supply of unskilled labor and raises their wages.

To uncover more insight, note that $\chi_t$ is also a dynamic variable with a law of motion. This means that the Mirrlees and wage compression terms have both intra- and intertemporal components.\(^{25}\) The intra- and intertemporal Mirrlees forces have the same interpretations as their analogues in Diamond (1998), Saez (2002), Farhi and Werning (2013), and Golosov, Troshkin, and Tsyvinski (2016b). The intratemporal

\(^{24}\)When skill has unbounded support with a fat-tailed distribution, the rate at the top is no longer negative but it is still lower than what it would be if the wage compression channel was not active.\(^{25}\)One could use Itô’s Lemma on (1.30) to derive the full expressions.
force is what drives the planner to insure agents against $\theta_t$ conditional on the history $\theta^{t-}$. The intertemporal force is the dynamic insurance motive and allows the planner to relax incentive constraints in earlier periods to reduce the cost of insurance against earlier shocks. The main difference is that our expressions are (unsurprisingly) more complicated because other agents’ skill shocks spill over via the wage function.

The most important insight comes from deconstructing the wage compression force into intra- and intertemporal components. In our dynamic model, the planner uses a negative rate at the bottom to deter skilled agents from deviating. While in the short-run a skilled agent benefits from deviating in the future, in the long-run he will receive a lower wage and work more to reach the same income level. This is precisely the static vs. dynamic incentives tradeoff highlighted in Section 1.3 and allows the planner to further front-load incentives. The planner also redistributes more to those states that unskilled agents are more likely to face. In other words, the planner uses negative rates at the bottom to deter skilled agents from mimicking unskilled ones while providing more insurance to unskilled agents. That wages are free to adjust in response to labor supply changes gives the planner an additional channel he can use to affect insurance and incentives.

To see how lowering rates allows the planner to further front-load incentives and provide more insurance, recall that a lower rate at the bottom increases the labor supply and lowers wages, and so raises $\tilde{u}_{\theta,t}$. From (1.15), the drift of $\Delta_t$ declines so $\Delta_{t+dt}$ declines: as in the two-period model, the planner front-loads incentives and provides less information rent in later periods. But from (1.12), this lowers the volatility of $v_t$ and unskilled agents receive more insurance. Since insurance is very valuable for these agents, the intertemporal force dominates the intratemporal one and the total wage compression term is negative. If the Mirrlees term is sufficiently low here then the intertemporal wage compression force makes the overall rate negative, consistent with the EITC. For skilled agents, insurance is less valuable than being
incentivized to tell the truth so the intratemporal force dominates. Thus, in the
right tail the static and dynamic models agree. In the middle and near the top,
the planner sets a (slightly) higher rate because skill premia are less responsive and
incentives matter. We conclude that the EITC accomplishes two goals: it reduces the
cost of insuring unskilled workers while making misreporting costly in the long-run
for skilled agents.

As for the intertemporal wedge, using that \( \tilde{u}_{c,t} = \lambda_t^{-1} \), by Itô’s Lemma,

\[
d\lambda_t^{-1} = \lambda_t^{-1} \left( (\rho - r + \sigma_\theta^2 \sigma_{\lambda t}^2) \, dt - \sigma_\theta \sigma_{\lambda t} \, dZ_t \right) + \left( \lambda_t^{R-1} - \lambda_t^{-1} \right) \, dR_t
\]

It follows from (1.22) that

\[
\tau_{K,t} = \sigma_\theta^2 \sigma_{\lambda t}^2.
\]  

Recall that \( \sigma_{\lambda t} \) is the volatility of the inverse of the marginal utility of consumption.
This is higher when consumption is more volatile so that the planner is providing
more incentives at the cost of less insurance. The intertemporal wedge is larger for
more skilled agents with imperfect substitutability than it is with a linear production
function. To see why, first recall that a positive intertemporal wedge exists in the
first place because agents have an incentive to “double-deviate”: they can report
truthfully today but over-save, and then under-report in the next period and use the
extra savings to over-consume. Since very skilled agents receive higher wages they
have a stronger incentive to double-deviate, so the planner deters this by setting an
even higher tax on savings. When the elasticity of substitution across skills is closer
to one, very skilled agents receive higher wages so the planner imposes a higher tax
on savings.
1.6 Quantitative Analysis

We now parametrize the model and solve it numerically.

1.6.1 Parametrization

Agents have CRRA utility over consumption and isoelastic disutility of labor,

$$u(c_t) = \frac{c_t^{1-\omega} - 1}{1 - \omega} \quad \text{and} \quad \phi(\ell_t) = \frac{\beta}{1 + \frac{1}{\varepsilon}} \ell_t^{1+\frac{1}{\varepsilon}}.$$ 

Agents have logarithmic utility so $\omega = 1$. Chetty (2012) finds that the Frisch elasticity is $\varepsilon = 0.5$ and we set $\beta = 1$. We set $\rho = r = 0.05$ so that the planner and the agents have the same discount rate, 5%, and $\psi = 1$ so that agents value working life and retirement equally.

We set

$$\mu(\theta) = \theta \left[-(1 - p) \left(\log(\theta) - \log(\theta^*)\right) + \frac{\sigma^2}{2}\right],$$

for some constant $p \in [0, 1]$ and $\theta^*$. This implies that $\log(\theta) \sim N\left(\theta^*, \frac{\sigma^2}{2(1-p)}\right)$. Farhi and Werning (2013) and Stantcheva (2015) both use this specification and set $p = 1$ so that $\log(\theta_t)$ follows a random walk, based on the finding in Storesletten, Telmer, and Yaron (2004) that income is very persistent. However, there is no stationary distribution when $p = 1$ unless $\kappa > 0$ and in that case, $\log(\theta)$ has an exponential stationary distribution. We set $\kappa = 0.025$ to ensure an average working life of 40 years, and $p = 0.95$ so that income is still quite persistent but $\log(\theta)$ has a normal distribution. Recall that new agents’ skills are drawn from a lognormal distribution, $\log(\theta_0) \sim N(\mu, \nu^2)$, so we set $\mu = \theta^*$ and $\nu^2 = \frac{\sigma^2}{2(1-p)}$ so that in fact they are drawn from the stationary distribution. We normalize $\theta^* = 1$ and we set $\sigma^2 = 0.0095$ to match the volatility of innovations to skill, as documented by Heathcote, Storesletten, and Violante (2005).
The most important parameter in our model is the elasticity of substitution, $\alpha$. Unfortunately, there are no widely agreed-on estimates. Katz and Murphy (1992) estimate $\alpha = 1.4$ while Acemoglu and Autor (2011) find $\alpha = 2.9$, both in a two-type production function with high school (“low-skill”) and college (“high-skill”) graduates. As far as we know, the only other paper that estimates $\alpha$ in our production function with a continuum of skills is Heathcote, Storesletten, and Violante (2016). In their model, $\text{Var} \left( \log \left( w(\theta) \right) \right) = \frac{1}{\alpha^2}$ and while they do not directly observe $w(\theta)$, they do observe the variance of the logarithm of the “total wage” $\theta w(\theta)$, which depends (additively) on $\frac{1}{\alpha^2}$ and the variances of several other processes. They estimate $\alpha = 3.124$ with a standard error of 0.115. Since the 95% confidence interval is so narrow, we use this point estimate in the first set of numerical experiments. In an alternative calibration, the same authors show that if income has a Pareto tail, then $\alpha$ coincides with the tail index. We repeat the experiments with $\alpha = 2.5$, which is at the upper end of estimates of the tail index to ensure a thinner tail.

1.6.2 Properties of Optimal Tax Policies

We solve for the optimal wedges in the stationary equilibrium, so an agent’s tax rate does not depend on when he was born but it does depend on his history of shocks.

Before exploring the tax rates themselves, it is instructive to first understand how skill affects income in our model versus in models with a linear production function. Figure 1.1 plots the average income of agents with a given skill level for three values of $\alpha$. While in standard Mirrlees models income is approximately linear in skill (dotted line), in our model (solid and dashed lines) income is strictly convex so that income grows more than one-for-one with skill; the lower the elasticity, the more convex is income. Observe that while all agents receive (weakly) higher income, the agents that benefit most relative to in the linear case are those with the most skill, which
means they can more easily mimic less skilled agents than they could with a linear production function.

![Figure 1.1: Average income as a function of $\alpha$.](image)

**Decomposing the Labor Wedge**

We start by showing the mechanisms at work. Let $\tau_j^i(\theta)$ denote the $j \in \{L,K\}$ wedge for an agent with current skill $\theta$, averaged over all histories (of any length) that terminate at $\theta$. Here, $\tau_j^M$ is the Mirrlees tax rate described earlier that accounts for only optimal wage-setting and $\tau_j^*$ is the overall optimal tax rate that accounts for both optimal wages and the wage compression channel.

Figure 1.2 plots $\tau_j^L(\theta)$ (solid line) vs. $\tau_j^M(\theta)$ (dashed line) for $\alpha = 2.5$ (black) and $\alpha = 3.124$ (red). The difference between the solid and dashed lines of the same color is the total wage compression force. In both cases, the Mirrlees rate is approximately hump-shaped, topping out near 80% in both cases, and declining very slowly in the right tail. The overall rate has a more pronounced hump shape, especially in the left tail where it is on average well below the Mirrlees rate. It is also higher than
the Mirrlees rate in the middle and near the top in both cases. However, when $\alpha$ is smaller, the overall rate is far lower in the left tail and higher everywhere else. This is because when $\alpha$ is lower, the wage function is "more convex" and lower in the middle of the distribution. This means the planner wants to raise wages here, which he does by increasing the tax rate and discouraging labor. Further, the gap between the two rates closes more quickly when $\alpha$ is larger, meaning that the overall rate will drop below the Mirrlees rate sooner when $\alpha$ is larger. Again, this is because when $\alpha$ is smaller, the wage function is higher in the right tail so the planner can lower taxes more to compress skill premia. Since these are average rates across all histories, it does not rule out negative rates (and hence the EITC) for some agents.

![Figure 1.2: $\tau^*_L(\theta)$ vs. $\tau^M_L(\theta)$ for $\alpha = 2.5$ and $\alpha = 3.124.$](image)

**EITC is Optimal**

Now we show that the EITC is consistent with optimal policy. Figure 1.3 plots the overall rates of an agent who has been at the 25th percentile for skill for one year (solid line) and 50 years (dashed line); the black line is for when $\alpha = 2.5$ and the
red line for when $\alpha = 3.124$. First, observe that rates are always positive following a short history in both cases, with a much higher minimum rate when $\alpha$ is larger. On the other hand, following a long history, the minimum rate is well below zero in both cases, consistent with the EITC. The reason the EITC does not kick in immediately is that initially, the planner’s problem is essentially static while following long histories, the planner can use negative rates at the bottom to relax more incentive constraints from earlier periods.

A key difference, however, is that the phase-in rate is near 50% when $\alpha = 2.5$ but around only 10% when $\alpha = 3.124$, which suggests that small variations in $\alpha$ can have large effects on the generosity of the EITC. This is consistent with earlier reasoning, as is the observation that the rates in the middle and near the top are higher when $\alpha$ is lower. While at the top the overall rate begins to decline, this occurs so deep in the right tail that our model probably does not accurately describe the labor market for these agents anyway. Therefore, for all intents and purposes, the optimal income tax is progressive in the cross-section and resembles the tax function in Heathcote,
Storesletten, and Violante (2016). Finally, while our tax system is not age-dependent, for this particular agent, since skill is constant over time the length of the history roughly corresponds to age. Then as in Heathcote, Storesletten, and Violante (2016), the EITC in our model is more generous later in life, after the agent has spent many years near the bottom of the income distribution.

Other Properties

Figure 1.3 shows the long-run behavior of the tax system for an agent whose skill never changes. Consider the impact of a one-time shock to skill on the tax rate, i.e., the impulse response function. This allows us to highlight short-run properties of the tax system. We follow Borovička, Hansen, and Scheinkman (2014), who link impulse response functions in continuous time to Malliavin derivatives. We set $\kappa = 0$ and ignore the retirement shock for simplicity. Let $S_t = (\theta_t, \lambda_t, \gamma_t)^T$ denote the vector of state variables with law of motion

$$dS_t = d \begin{bmatrix} \theta_t \\ \lambda_t \\ \gamma_t \end{bmatrix} = \begin{bmatrix} \mu(\theta_t) \\ 0 \\ \mu'(\theta_t)\gamma_t - \sigma^2_\theta \theta_t \sigma_{\lambda,t} \lambda_t \end{bmatrix} dt + \begin{bmatrix} \sigma_\theta \theta_t \\ \sigma_\theta \sigma_{\lambda,t} \lambda_t \\ \sigma_\theta \gamma_t \end{bmatrix} dZ_t$$

and initial condition $S_t = s$. Consider an $\mathcal{F}_T$-adapted function $F(S_T)$. By the Clark-Ocone-Haussman Theorem,

$$F(S_T) = \mathbb{E}[F(S_0) \mid \mathcal{F}_0] + \int_0^T \mathbb{E}[D_t(F(S_T)) \mid \mathcal{F}_t] dZ_t,$$

where $D_t(F(S_T))$ is the Malliavin derivative of $F$. Then by the Malliavin chain rule,

$$D_t(F(S_T)) = F'(S_T)^T H_T H_t^{-1} \sigma(S_t) 1_{t \leq T},$$

42
where \( \{ H_t \} \) is the first variation process defined by 

\[
H_t = \frac{\partial S_t}{\partial s} \quad \text{with}
\]

\[
dH_t = \begin{bmatrix}
\mu'(\theta_t) & 0 & 0 \\
0 & 0 & 0 \\
\mu''(\theta_t) \gamma_t - \sigma^2_\theta \lambda_t \left( \sigma_{\lambda,t} + \theta_t (\sigma_{\lambda,t})_\theta \right) & -\sigma^2_\theta \theta_t \left( \sigma_{\lambda,t} + \lambda_t (\sigma_{\lambda,t})_\lambda \right) & \mu'(\theta_t) - \sigma^2_\theta \theta_t \lambda_t (\sigma_{\lambda,t})_\gamma \\
\sigma_\theta & 0 & 0 \\
\sigma_\theta \lambda_t (\sigma_{\lambda,t})_\theta & \sigma_\theta (\sigma_{\lambda,t} + \lambda_t (\sigma_{\lambda,t})_\lambda) & \sigma_\theta \lambda_t (\sigma_{\lambda,t})_\gamma \\
0 & 0 & \sigma_\theta 
\end{bmatrix} dt \\
+ \begin{bmatrix}
\sigma_\theta \\
\sigma_\theta \lambda_t (\sigma_{\lambda,t})_\theta \\
0 \\
0 \\
\sigma_\theta 
\end{bmatrix} dZ_t
\]

and \( H_0 = I_3 \), where \((\sigma_{\lambda,t})_\theta = \frac{\partial \sigma_{\lambda,t}}{\partial \theta} \), and similarly for the other state variables.

Let \( \varphi_s (S_T) \) denote the impulse response function of \( F(S_T) \) with initial conditions \( S_t = s \) and \( H_0 = I_3 \). Then

\[
\varphi_s (S_T) = \mathbb{E}_s \left[ F' (S_T)^T H_T H^{-1}_t 1_{\{ t \leq T \}} \right] \sigma (s).
\]

Set \( F(S_T) = \tau_L (S_T) \) and consider the response to a unit shock to skill. Figure 1.4 plots the labor wedge’s path relative to where it would be had the shock not occurred and shows it is progressive in the short-run. This stands in sharp contrast to models with linear production, as Farhi and Werning (2013) show that the rate is regressive in the short-run. One can show that with a linear production function, the volatility of the tax function is proportional to \( -\sigma_{\lambda,t} \) while here there are additional positive terms that come from the variation in \( w(\theta) \). This is because a positive shock increases skill premia so the planner undoes this with a higher labor tax.

Finally, Figure 1.5 plots the intertemporal wedge, or the implicit tax on savings, averaged across histories that terminate at \( \theta \). The solid line is the optimal rate with \( \alpha = 2.5 \), the dashed line is when \( \alpha = 3.124 \), and the dotted line is the Mirrlees rate. All three functions are increasing over most of the domain and then decrease to zero in the right tail. The reason for the decline is that the most skilled agents can almost
Figure 1.4: Impulse response function of the labor wedge.

perfectly insure themselves against shocks, which means they have almost smooth consumption. This means that $\lambda_t = u_{c,t}^{-1}$ has almost no volatility, which, by (1.28), means $\sigma_{\lambda,t}$ is close to zero; (1.31) then implies that $\tau_{K,t}$ is close to zero. However, the optimal rate tops out at just under 1.5% when $\alpha = 2.5$, compared to just under 1% when $\alpha = 3.124$ and under 0.3% for the Mirrlees rate. In addition, the optimal rate is higher for more agents when $\alpha$ is smaller. The reason, as per Figure 1.1 again, is that skilled agents not only are more productive but also receive a larger wage, which makes it easier for them to mimic less skilled workers and save too much.

A Word on Welfare

Our mechanism design framework puts an upper bound on economy-wide welfare; what is the welfare loss from moving from an optimal, history-dependent tax system to a simpler one? We compared welfare in our benchmark economy with that in an economy in which agents are unconstrained to borrow and save a riskless asset and face history-independent, linear taxes and found that the welfare gain is essentially a
Figure 1.5: Average intertemporal wedge as a function of $\alpha$.

This finding that the optimal tax schedule is close to linear is well-documented and suggests that the overall welfare gains from the EITC are not exceptionally large. Further, the shape of our tax function closely resembles the one in Heathcote, Storesletten, and Violante (2016),

$$T(y) = y - \tau_0 y^{1-\tau_1}$$

for some parameters $\tau_0, \tau_1$, which they argue is a good approximation to the current U.S. system (and thus already has an embedded EITC). Therefore, the gains from modifying the current U.S. system are unlikely to be large. However, these statements should not be taken too seriously as there are likely other factors (such as moving away from a lognormal skill distribution) that can increase welfare gains. Still, Golosov,

---

26 Following Heathcote and Tsuijyama (2016), extended to our dynamic setting, we first solved for the equilibrium allocations given the tax function and then computed the percentage increase in consumption which, if received every period after all histories, would yield the same gain in lifetime utility.

27 This result dates back to Mirrlees (1971) for static models and Farhi and Werning (2013) for dynamic models.
Troshkin, and Tsyvinski (2016b) show that the welfare gains are unlikely to exceed a few percentage points even with these other factors.

1.7 Discussion

While our previous analysis highlights our theoretical results, we now take a moment to discuss some of our key assumptions and how relaxing them could affect our results.

1.7.1 CES Production Function

First, while a few papers do use CES production functions, its use in the public finance literature is still somewhat limited. The reason, as Salanié (2011) points out, is that, for example, the substitutability between very low- and low-skill workers is probably very different from the substitutability between high- and very high-skill workers, which are both different from the substitutability between low- and high-skill workers. Ales, Kurnaz, and Sleet (2015) get around this by micro-founding the production function with a task assignment framework so that the elasticity of substitution differs across pairs of workers. They show that as skilled workers are increasingly locked into complex tasks, the wage compression force becomes more powerful and is thus a force for even lower rates at the top and higher rates at the bottom. However, since their model is static they neglect the opposing dynamic insurance motive. Since the dynamic force is more powerful near the bottom, it is not clear if the “lock-in” effect generated by their more complicated production function is strong enough to make negative rates suboptimal in a dynamic setting. An alternative is to follow Rothschild and Scheuer (2014) so that workers have a skill vector and allocate time to each task in the vector. This weakens the wage compression channel so it would likely strengthen the EITC here, too.
1.7.2 Skill Distribution

Second, in following the literature, we assumed that log ($\theta_t$) is an Ornstein-Uhlenbeck process. This has two shortcomings: first, the implied stationary distortion is lognormal and hence thin-tailed. It is well-documented that the U.S. income distribution is in fact fat-tailed and Saez (2001) shows that fat tails imply a significantly higher tax rate at the top. However, as the figures above show, convergence in the right tail is so slow that the tax rate near the top as implied by a lognormal stationary distribution is a good enough approximation to the rate implied by a fat-tailed distribution for almost the entire population. On the other hand, a thicker tail would lower the elasticity, $\alpha$, as argued in Section 1.6, and this would only strengthen our results with respect to the EITC.

Second, Guvenen, Ozkan, and Song (2014) and Guvenen, Karahan, Ozkan, and Song (2015) show that the skill process is highly leptokurtic, a property that does not hold for lognormal (or Pareto lognormal) distributions. Gabaix, Lasry, Lions, and Moll (2015) suggest using a skill process with distinct growth regimes,

$$d(\log (\theta_t)) = \mu_{\theta_j} \, dt + \sigma_{\theta_j} \, dZ_t + \text{Injection} - \text{Death},$$

where $\mu_{\theta_j}, \sigma_{\theta_j}$ are the growth rate and volatility, respectively, of skill in regime $j \in \{1, \ldots, J\}$, with agents being born into each regime with some probability and randomly transitioning between regimes during life. Not only does this help with fitting the micro data, they also show that it has the potential to explain the long-run dynamics of inequality.

Golosov, Troshkin, and Tsyvinski (2016b) show that a highly leptokurtic distribution generates a substantially different tax function: instead of being hump-shaped,

---

$^{28}$See Piketty and Saez (2003) and Atkinson, Piketty, and Saez (2011), for example.

$^{29}$Indeed, Golosov, Troshkin, and Tsyvinski (2016b) show that lognormal and Pareto lognormal distributions yield very similar tax functions.
it is U-shaped and larger in magnitude, including in the left tail. This means that for poor agents, the optimal policy is a guaranteed level of income that is taxed at a very high rate, a so-called “negative income tax.” The reason is that the hazard function is U-shaped with leptokurtic distributions, as opposed to monotonic with (Pareto) lognormal ones. However, the hazard function acts through $\gamma_t$ and as (1.30) shows, $\gamma_t$ does not affect the wage compression term. This suggests that while the Mirrlees term will be U-shaped, the wage compression term will have the same pronounced hump-shape as it does with a lognormal distribution.30 Whether it is powerful enough to force the entire rate negative in the left tail is unclear.

1.8 Extensions

1.8.1 Skill Bias in Production

In our baseline model, differences in wages are entirely due to differences in labor supply across skill. To put our model more in line with the skill premium literature, we can add skill bias in the production function so that the production process favors more skilled workers. Following Heathcote, Storesletten, and Violante (2016), let

$$
Y_t = \left( \int_{\Theta} \exp (\varrho \theta_t) L (\theta_t)^{\frac{\alpha - 1}{\alpha}} d\theta_t \right)^{\frac{\alpha}{\alpha - 1}},
$$

which implies that the wage per unit of effective labor is

$$
w (\theta_t) = \exp (\varrho \theta_t) \left( \frac{Y_t}{L (\theta_t)} \right)^{\frac{1}{\alpha}}.
$$

30That said, the hazard functions of the lognormal and the leptokurtic distribution in Golosov, Troshkin, and Tsyvinski (2016b) are very similar in the left tail anyway. Further, Heathcote and Tsuijyama (2016) show that the tax function’s U-shape is actually very sensitive to factors such as exogenous government expenditures.
This implies that the skill premium between two skill levels is

$$\pi_{i,j}^t = \exp \left( \varrho \left( \theta_i^t - \theta_j^t \right) \right) \left( \frac{L(\theta_j^t)}{L(\theta_i^t)} \right)^{\frac{1}{\alpha}}.$$

In particular, the premium includes an amplifying exogenous component. Also, note that now income grows more rapidly as a function of $\theta$.

To compare the model with and without skill bias, we set $\alpha$ and $\varrho$ so as to keep the variance of the total wage the same in both cases; obviously, there are a continuum of possible combinations. Since we do not know of any joint estimates of these two parameters, for each of the cases in the previous figures, we pick two examples. In Figure 1.6, we choose $(\alpha, \varrho)$ to be comparable with the $(2.5, 0)$ case from before and in Figure 1.7, we choose them to be comparable with the $(3.124, 0)$ case. The two example parameter vectors are $(2.8, 0.095)$ and $(3, 0.14)$ in the first case and $(3.5, 0.062)$ and $(4, 0.117)$ in the second case. In both cases, panel (a) reproduces Figure 1.2 while panel (b) reproduces Figure 1.3.

![Average Labor Wedge](image1.png)  
(a) Average labor wedge.  

![History-Dependent Labor Wedge with Skill Bias](image2.png)  
(b) Overall rate for low-skill agent.

Figure 1.6: Labor wedge with skill bias consistent with $(\alpha, \varrho) = (2.5, 0)$.  

49
Figures 1.6 and 1.7 make clear that a strong skill bias calls for a smaller EITC, if at all. Indeed, while optimal policy called for the EITC after a long history of low shocks without skill bias, with skill bias the EITC following a long history is consistent with optimal policy in only one case, \((\alpha, \varrho) = (2.8, 0.095)\). To understand this, recall that an agent’s wage is shaped by both the exogenous component and the endogenous labor supply/demand component. When the skill bias is very strong, the wage is shaped more by the exogenous part than the endogenous one. This effectively makes the production function closer to linear and as argued above, weakens the intertemporal wage compression force and thus the EITC. The figures show that the degree of skill bias has the largest effects in the left tail and very small effects elsewhere in the distribution.

![Average Labor Wedge](image1)

(a) Average labor wedge.

![History-Dependent Labor Wedge with Skill Bias](image2)

(b) Overall rate for low-skill agent.

Figure 1.7: Labor wedge with skill bias consistent with \((\alpha, \varrho) = (3.124, 0)\).

### 1.8.2 Endogenous Human Capital Formation

Recent papers such as Kapička and Neira (2015) and Stantcheva (2015) augment the standard Mirrlees framework to include endogenous human capital formation over the life cycle. Goldin and Katz (2007) find that rising education wage premia explain 60%
to 70% of the increase in U.S. wage inequality between 1980 and 2005 while Heckman, Lochner, and Cossa (2003) argue that the EITC is valuable precisely because it boosts human capital investment. By choosing how much to invest in human capital, such as via schooling or on-the-job training, an agent can directly affect his earning potential, instead of being totally at the mercy of the exogenous process \( \{\theta_t\} \).

This section extends our baseline model à la Stantcheva (2015) to study how our intertemporal wage compression force affects tax policy in the presence of human capital. Let \( h_t \in \mathcal{H} \) denote an agent’s stock of human capital at date \( t \). Agents can augment this human capital at each date by investing money. Specifically, each agent has access to a technology \( H(m_t) \) that allows him to grow his human capital through an observable monetary investment \( m_t \). This technology is best thought of as going to school and exhibits decreasing returns to investment, \( H_m > 0 \) and \( H_{mm} < 0 \).\(^{31}\)

Human capital evolves according to

\[
dh_t = (H(m_t) - \delta h_t) \, dt \equiv \mu_{h,t} \, dt. \tag{1.32}
\]

The \( \delta \)-term captures the fact that human capital depreciates over time if agents do not invest to grow it. One can imagine that a worker’s knowledge and experience in an industry become less useful over time as newer technologies and techniques are developed. Consequently, the worker must continuously invest to maintain the acquired portion of his productivity. Therefore, an agent’s type is now \((\theta_t, h_t) \in \Theta \times \mathcal{H}\).

When new agents are born, \((\theta_0, h_0)\) is drawn from some distribution \( \tilde{f}(\theta_0, h_0) \). To compare with the discrete time model in Stantcheva (2015), to augment human capital \( s_{t-1} \) by an amount \( e_t \), \( s_t = s_{t-1} + e_t \), the agent must spend \( M(e_t) \). Here, \( s_t \) corresponds to \( h_t \), \( m_t \) corresponds to \( M_t \), \( \mu_{h,t} \) corresponds to \( e_t \), and hence \( M'(e_t) \) to \( m_{\mu_{h,t}} \).

\(^{31}\)It is straightforward to modify the function so that more skilled agents accumulate human capital more easily, \( H_{m\theta} > 0 \).
An agent’s productivity now depends on exogenous skill and endogenous human capital,

\[ e_t = e(\theta_t, h_t), \]

where \( e_h > 0 \). Stantcheva (2015) takes

\[ e_t = \left( \theta_t^{1-\rho} + h_t^{1-\rho} \right)^{\frac{1}{1-\rho}}, \]

with \( e_t = \theta_t h_t \) as a special case when \( \rho = 1 \). An agent’s labor income at date \( t \) is then

\[ y_t = w(\theta_t, h_t) e(\theta_t, h_t) \ell_t, \]

where \( w(\theta_t, h_t) \) is still the wage per unit of effective labor, but can now depend on human capital as well. To simplify notation, let \( \bar{w}(\theta_t, h_t) \) denote the total wage component of income, i.e., \( \bar{w}(\theta_t, h_t) \equiv w(\theta_t, h_t) e(\theta_t, h_t) \).

The planner operates the same CES production function as before, the lone difference being that the firm cares about skill and human capital instead of only skill (skill is still unobservable while human capital is observable). Let \( L(\theta_t, h_t) \) denote the total labor supply of agents with profile \((\theta_t, h_t)\) and \( L_x \) the derivative with respect to argument \( x \in \{\theta, h\} \). It follows that the wage per unit of effective labor is

\[ w(\theta_t, h_t) = \left( \frac{Y_t}{L(\theta_t, h_t)} \right)^{\frac{1}{\alpha}}. \]

As in Stantcheva (2015), an important variable here is the Hicksian coefficient of complementarity between ability and human capital, \( \rho_{\theta h, t} \):

\[ \rho_{\theta h, t} \equiv \frac{\bar{w}_{\theta h, t} \bar{w}_t}{\bar{w}_{\theta t} \bar{w}_{h, t}}. \]
There, $\rho_{\theta h} = \rho$ in the CES aggregator above while here that is not the case. For example, if $\rho = 1$ so that $e(\theta, h) = \theta h$, then

$$\rho_{\theta h, t} = 1 - \frac{1}{\alpha} \left( \frac{L_t L_{\theta h, t} - L_{\theta, t} L_{h, t}}{\frac{L_t}{\theta_t} - \frac{L_{\theta, t}}{\theta_t}} \right),$$

which equals 1 as $\alpha \to \infty$ but generally not otherwise. The difference is that here, the derivatives account for both individual and aggregate wage effects that depend on the distribution and each agent’s labor decision, both of which are endogenous. The reason is that while investing in human capital increases an agent’s productivity, it also increases the supply of skilled labor, which then feeds back into agents’ decisions and hence affects the distribution, too. Thus, complementarity is determined by the direct effect via an agent’s productivity and the indirect effect via the wage function. Ability and human capital are complements when $\rho_{\theta h, t} > 1$ and substitutes when $\rho_{\theta h, t} < 1$; see Stantcheva (2015) for a more in-depth discussion of this coefficient.

This highlights the main difference between our model and hers: in her model, although skill has an endogenous component, because the production function is linear, agents’ labor supply and human capital investment decisions still do not affect each other. This means that our wage compression channel is absent and hence labor taxes cannot be negative. As we argued with the baseline model, it is precisely the lower insurance cost driven by these labor supply effects that is responsible for the EITC.

Since human capital investments are observable, the planner’s mechanism design problem is very similar to the one in the baseline model: a reporting strategy is still a report of only $\theta_t$ and the laws of motion for the promise- and threat-keeping constraints are the same. There are a few small differences, however: an agent’s utility $\tilde{u}_t$ depends on $h_t$, the planner’s HJB equation (and KFE) contains an additional term for the human capital law of motion, an allocation now specifies the incremental
human capital change, \( \mu_{h,t} \), there is an additional cost from human capital investment, the resource constraints and their Lagrange multipliers depend on \( h_t \), and there is a constraint for the \( h \)-derivative of the wage function. Thus the planner’s problem is

\[
(r + \kappa) K(\theta, h) = \inf_{\{c, C, \mu_h, \sigma_{\Delta}, z\}} \left\{ \int_{V \times D} \left( c + \kappa C + m(\mu_h) - \ell \bar{w}(\theta, h) + k_\theta \mu(\theta) + k_h \mu_h + k_v \left[ (\rho + \kappa) v - \bar{u} - \kappa v R \right] + k_\Delta \left[ (\rho + \kappa - \mu'(\theta)) \Delta - \bar{u}_\theta - \kappa \Delta R - \sigma^2_\theta \sigma_\Delta \right] + \frac{\sigma^2_\theta}{2} \left[ k_{v \theta} \theta^2 \Delta^2 + k_{\Delta \Delta} \sigma^2_\Delta + k_{\theta \theta} \theta^2 \right] + \sigma^2_\theta \left[ k_{v \Delta} \theta \Delta \sigma_\Delta + k_{v \theta} \theta^2 \Delta + k_{\Delta \theta} \sigma_\Delta \right] + \kappa k f_0 + \chi(\theta, h) w(\theta, h) \ell - X \bar{w}(\theta, h) \ell \right) g(\theta, h, v, \Delta) d(v, \Delta) + \eta_\theta(\theta, h) z_\theta(\theta, h) + \eta_h(\theta, h) z_h(\theta, h) \right\},
\]

where \( z_s \) is the derivative of the wage function with respect to the state variable \( s \in \{\theta, h\} \) and \( \eta_s \) is its Lagrange multiplier.

Since human capital investment is now a choice variable, it has a first-order condition:

\[
[\mu_h] : m_{\mu_h} = \frac{1}{H^{-1}(\mu + \delta h)} = -k_h,
\]

where \( m_{\mu_h} \) is computed using the Inverse Function Theorem. Also, since the HJB equation includes another state (with a first-order derivative), we need one more endogenous boundary condition. Let \( \bar{h} \) denote the value of \( h_t \) at which \( dh_t = 0 \).

Since agents will never invest in human capital above this point, this is a reflecting boundary, \( k_h(\theta, \bar{h}, v, \Delta) = 0 \).

As before, we can characterize marginal distortions using wedges. The labor and intertemporal wedges are defined in the same way as before and have the same interpretations, the lone difference being that \( h_t \) is an additional state variable. Let
$\tau_H(\theta^t)$ denote the human capital wedge defined by

$$
\tau_H(\theta^t) = m_{\mu_h,t} - \frac{1}{\rho + \kappa} \left[ \ell_t \bar{w}_{h,t} (1 - \tau_{L,t}) + \frac{1}{dt} \mathbb{E}_t \left[ d \left( u_{c,t} (m_{\mu_h,t} - \tau_{H,t}) \right) \right] \right]. \quad (1.35)
$$

This is simply the continuous time analogue of $\tau_{S,t}$ in Stantcheva (2015). This wedge represents the implicit subsidy from the planner to an agent for an incremental investment in human capital: an agent receives $\tau_{H,t}$ when human capital increases by $\mu_{h,t}$.

This version of the human capital wedge takes as given labor and savings distortions and so acts only to undo these distortions;\footnote{Jacobs (2012) defines a human capital wedge that ignores labor distortions.} this is the “Siamese twins” result from Bovenberg and Jacobs (2005).

Because the human capital wedge explicitly and implicitly depends on the other two wedges, we need to adjust it to account for these distortions. We define the net subsidy on human capital expenses, $t_{h,t}$, as

$$
t_{h,t} = \frac{\tau_{H,t}^* - \tau_{L,t} m_{\mu_h,t}^* + p_t}{(1 - \tau_{L,t}) \left( m_{\mu_h,t}^* - \tau_{H,t}^* \right)} - \frac{\mathcal{X}_t \bar{w}_{h,t} - \chi (\theta_t, h_t) (w_{e,t}^0 e_t)}{\bar{w}_{h,t} (1 - \tau_{L,t})}, \quad (1.36)
$$

where

$$
\tau_{H,t}^* \equiv \tau_{H,t} - \frac{1}{\rho + \kappa} \frac{1}{dt} \mathbb{E}_t \left[ d(u_{c,t} \tau_{H,t}) \right]
$$

is the dynamic, risk-adjusted subsidy,

$$
m_{\mu_h,t}^* \equiv m_{\mu_h,t} - \frac{1}{\rho + \kappa} \frac{1}{dt} \mathbb{E}_t \left[ d(u_{c,t} m_{\mu_h,t}) \right]
$$

is the dynamic, risk-adjusted cost and

$$
p_t \equiv \frac{1 - \tau_{L,t}}{\rho + \kappa} \left[ m_{\mu_h,t} \tau_{K,t} + \frac{1}{dt} \mathbb{E}_t \left[ du_{c,t} \cdot dm_{\mu_h,t} \right] \right]
$$
is the risk-adjusted savings distortion. This net subsidy ensures that the tax system is neutral with respect to human capital. The second term in (1.36) is new and accounts for the fact that investing in human capital affects other agents’ wages and hence their incentive to invest themselves. Given an agent’s (distorted) labor and savings decisions, this subsidy ensures that human capital investment is efficient. In effect, it captures the redistributive and insurance properties of the human capital subsidy while filtering out anything that acts only to undo other distortions.

We will now derive an equation that links the labor wedge with the net subsidy. From Itô’s Lemma, the Envelope Theorem, and (1.34),

\[
\frac{1}{dt} E_t [dm_{\mu_h,t}] = (r + \kappa) m_{\mu_h,t} - \ell_t \bar{w}_{h,t} - \gamma_t \bar{u}_{\theta h,t} + (\chi(\theta_t, h_t)) (w^\theta_{t} e_t) h_t - \mathcal{X}_{t} \bar{w}_{h,t} \ell_t.
\]

Rearranging and using the definition of \( t_{h,t} \) in (1.36), at the optimum the labor and human capital wedges satisfy

\[
t^*_t h,t = \left[ \tau^*_t L_t - WC_t \right] \frac{\varepsilon}{1 + \varepsilon} (1 - \rho_{\theta h,t}),
\]

(1.37)

where \( WC_t \) is the wage compression term, now accounting for the effects of human capital. The term in brackets is just the Mirrlees term.

With a linear production function, \( t^*_t h,t \) and hence (1.37) reduce to the expression in Stantcheva (2015). The main insight still holds: if \( 1 < \rho_{\theta h,t} \) then skilled agents do not disproportionately benefit from acquiring human capital so the planner subsidizes its acquisition to make up for higher labor taxes. That said, it may not be the case that \( \rho_{\theta h,t} < 1 \) for everyone at the same time because \( L(\theta_t, h_t) \) varies across types.

Because the wage compression term can be negative and hence the Mirrlees term larger than the optimal rate in the two tails (see Figure 1.2), the subsidy is largest
(relative to if the planner ignored the wage compression force) in these regions. However, the reasons for the relatively larger subsidies differ across agents. In the left tail, $τ_{L,t}^*$ is small, if not negative, but $-WC_t \gg 0$ while in the right tail, $τ_{L,t}^*$ is large and $-WC_t > 0$. This means that not only could low-skill agents receive a transfer via the EITC, they could receive another small subsidy specifically for human capital investment.

Consider an agent with large $e(θ, h)$. Then the planner should subsidize human capital investment because doing so increases the supply of skilled effective labor, pushing down their wages and lowering skill premia as in static models. For agents at the other extreme, subsidizing human capital investment widens skill premia as before but it also allows the planner to provide dynamic insurance against wage fluctuations. Once again, this insurance is more valuable than offsetting rising skill premia so the optimal policy combines lower labor taxes with human capital subsidies for these agents. However, it is not ex ante clear to what extent the human capital subsidy crowds out the EITC, or vice versa. In other words, policy calls for lower labor taxes but not necessarily negative ones because the human capital subsidy potentially accomplishes the same thing. It is also not clear how this affects the savings tax. While we have focused on the two extremes of the distribution, agents in the middle still receive a human capital subsidy but it is smaller than what they would receive absent the wage compression channel.

As a point of comparison, Jacobs (2012) takes $e(θ, h) = θh$ in a static model. Though his definition of the human capital wedge is different, he finds that skilled agents should receive positive human capital subsidies (and unskilled agents should pay a tax) for the same reason that they should face negative labor income taxes,

---

33The distinction is crucial: the Mirrlees term is lowest for the lowest-skilled agents so the actual subsidy, if it exists, is very small for them. However, absent the wage compression term, the subsidy would be even smaller.
namely that both reduce skill premia. Once again, adding dynamics changes the results because the missing intertemporal force works in the opposite direction.

1.9 Conclusion

We develop a dynamic Mirrlees model of optimal taxation in which agents with different skills are imperfect substitutes in the production process and uncover a novel intertemporal wage compression channel that the planner uses to lower the cost of insurance against shocks. We demonstrate that this insurance motive can drive labor income taxes negative for low-skill agents, helping us to rationalize redistributive programs like the EITC. Since more skilled agents are paid higher wages and can more easily mimic less skilled agents, we further show that the planner imposes a higher savings tax on these agents to deter this behavior. Finally, in contrast to models with linear production functions, the forces that we introduce cause labor taxes to be progressive in the short-run. This occurs because positive skill shocks spill over to other agents and raise skill premia.

We extend our model to include skill bias in the production function and endogenous human capital formation. The former weakens the intertemporal wage compression force and hence the EITC because it weakens labor supply/demand effects. The latter introduces an additional margin through which intertemporal wage compression forces influence savings behavior. In contrast to static models of human capital formation where unskilled workers pay a tax to deter them from acquiring human capital, optimal human capital subsidies in our setting are largest, relative to what they would be in the absence of wage compression forces, for agents at both ends of the skill distribution.

We also make a methodological contribution that addresses an obstacle in the literature, namely solving heterogeneous agent Mirrlees models numerically. By working
in continuous time we can take advantage of techniques from the theory of mean field games to solve our heterogeneous agent model. This should allow the literature to solve Mirrlees models with even richer degrees of heterogeneity and hence analyze tax systems that account for more drivers of inequality.

While our model highlights an important determinant of tax policy relevant for understanding redistributive programs, there are many avenues for future research that would bring our model closer to the policy realm. Adding endogenous retirement decisions, for instance, would allow one to jointly analyze tax and retirement policies with imperfect substitutability. In addition, employing richer production and skill functions, as described in Section 1.7, would allow for a more accurate quantitative assessment of the optimal tax schedule relevant for policy recommendations while adding an extensive labor supply margin would allow us to jointly analyze the most important drivers of the EITC.
1.10 Appendix

A Section 1.5 Proofs

Proof of Proposition 1.1

By the Envelope Theorem with respect to $v$,

$$(r + \kappa) k_v = (\rho + \kappa) k_v + \frac{\mathbb{E}[dk_v]}{dt}$$

while by Itô’s Lemma,

$$d\lambda_t = \mathbb{E}_t[d\lambda_t] + \sigma_\theta (k_{vv,t} \theta_t \Delta_t + k_{v\Delta,t} \sigma_{\Delta,t} + k_{v\theta,t} \theta_t) dZ_t + (\lambda^R_t - \lambda_t) dR_t$$

$$= (r - \rho) \lambda_t dt + \sigma_\theta (k_{vv,t} \theta_t \Delta_t + k_{v\Delta,t} \sigma_{\Delta,t} + k_{v\theta,t} \theta_t) dZ_t + (\lambda^R_t - \lambda_t) dR_t,$$

where the second line uses the envelope condition. From the first-order condition for $\sigma_{\Delta,t}$,

$$\sigma_{\Delta,t} = \frac{k_{\Delta,t} - k_{v\Delta,t} \theta_t \Delta_t - k_{v\theta,t} \theta_t}{k_{\Delta\Delta,t}}.$$

Plugging this in yields the volatility:

$$\text{Vol} (\lambda_t) = \sigma_\theta \left( k_{vv,t} \theta_t \Delta_t + \frac{k_{v\Delta,t}}{k_{\Delta\Delta,t}} (K_{\Delta,t} - k_{v\Delta,t} \theta_t \Delta_t - k_{v\theta,t} \theta_t) + k_{v\theta,t} \theta_t \right)$$

$$= k_{v,t} \cdot \frac{\sigma_\theta}{k_{v,t}} \left( \frac{k_{vv,t} k_{\Delta,t} - k_{v\Delta,t}^2}{k_{\Delta\Delta,t}} \theta_t \Delta_t + \frac{k_{v\Delta,t}}{k_{\Delta\Delta,t}} (k_{\Delta,t} - k_{v\theta,t} \theta_t) + k_{v\theta,t} \theta_t \right)$$

$$\equiv \sigma_\theta \sigma_{\lambda,t} \lambda_t.$$

Next, once again applying the Envelope Theorem to (1.20),

$$(r + \kappa) k_\Delta = (\rho + \kappa - \mu' (\theta)) k_\Delta + \sigma^2_\theta (k_{vv,\theta} \theta \Delta + k_{v\Delta} \sigma_{\Delta} + k_{v\theta} \theta) + \frac{\mathbb{E}[dk_\Delta]}{dt}.$$
Using Itô’s Lemma,

\[ d\gamma_t = \mathbb{E}_t [d\gamma_t] + \sigma_t \left( k_{v\Delta,t} \theta_t \Delta_t + k_{\Delta\Delta,t} \sigma_{\Delta,t} + k_{\Delta \theta,t} \theta_t \right) dZ_t + \left( \gamma_t^R - \gamma_t \right) dR_t \]

\[ = \left( (r - \rho + \mu' (\theta)) \gamma_t - \sigma_t^2 \theta_t \sigma_{\Delta,t} \gamma_t \right) dt + \sigma_t \gamma_t dZ_t + \left( \gamma_t^R - \gamma_t \right) dR_t, \]

where the second line uses (1.26).

We give an alternative proof in Appendix C based on the stochastic maximum principle.

**B  Mean Field Game Boundary Conditions**

**Hamilton-Jacobi-Bellman Equation**

The mean field game in our model comprises two coupled, second-order, three-dimensional partial differential equations, so we need six boundary conditions for each one to pin down a unique solution.

We begin with the HJB equation (1.20). First, both \( \underline{\theta} \) and \( \overline{\theta} \) are reflecting boundaries so

\[ k_{\theta} (v, \Delta, \underline{\theta}) = 0, \quad (B-1) \]

\[ k_{\theta} (v, \Delta, \overline{\theta}) = 0. \quad (B-2) \]

Let \( \underline{v} \) and \( \overline{v} \) denote the lowest and highest possible continuation utilities, respectively, and \( \underline{\Delta} \) and \( \overline{\Delta} \) the lowest and highest possible information rents, respectively. Since an agent can abandon the contract if continuation utility drops below its autarky value, we set \( \underline{v} = v^{\text{aut}} \). For \( v \) to reflect back inward at \( v^{\text{aut}} \), we need

\[ k_{\theta} (v^{\text{aut}}, \Delta, \theta) = 0. \quad (B-3) \]
Since information rent cannot be negative, $\Delta = 0$. At this point, the principal does not ask the agent to produce anything so $y = 0$ and the cost to the planner is simply the cost of the consumption stream needed to achieve promised utility $v$. Now, let
\[
v = \frac{u(c) + \kappa R(C)}{\rho + \kappa}
\]
denote the agent’s promised utility as a function of consumption (since there is full insurance during working life and retirement). Since the retirement shock is observable, the marginal utility of consumption just before and just after retirement must be equal:
\[
c^{-\omega} = u'(c) = vR'(C) = \psi u'(C) = \psi C^{-\omega} \implies C = \psi^{1-\omega} c.
\]
This provides a link between consumption pre- and post-retirement. Then
\[
v = \frac{u(c) + \kappa \psi u\left(\psi^{1-\omega} c\right)}{\rho + \kappa} \implies c = \left(\frac{(1-\omega)(\rho + \kappa) v + 1 + \kappa \psi}{1 + \kappa \psi^{1-\omega}}\right)^{\frac{1}{1-\omega}}.
\]
It follows that\(^{34}\)
\[
k(v, 0, \theta) = \frac{c + \kappa C}{\rho + \kappa} = \left(\frac{(1-\omega)(\rho + \kappa) v + 1 + \kappa \psi}{1 + \kappa \psi^{1-\omega}}\right)^{\frac{1}{1-\omega}} \frac{1 + \kappa \psi^{1-\omega}}{\rho + \kappa}.
\]
For the fifth boundary condition, we impose a reflecting boundary at $\overline{\Delta}$, or
\[
k_\Delta (v, \overline{\Delta}, \theta) = 0.
\]
This upper reflecting boundary is endogenous and comes from setting $\mathbb{E}_t [d\overline{\Delta}_t] = 0$ and using the first-order condition for $\ell$. Finally, since an agent will not be promised more utility than he receives in the first-best allocation, we set $\overline{v} = v^{FB}$. Let $k^*$ denote the planner’s cost function under perfect information, i.e., the solution to the original
\[^{34}\text{With logarithmic utility, the condition is } k(v, 0, \theta) = \exp\left(\frac{(\rho + \kappa) v - \kappa \psi \ln(\psi)}{\rho + \kappa \psi}\right) \frac{1 + \kappa \psi^{\frac{1}{\omega}}}{\rho + \kappa}.
\]
HJB equation while ignoring (1.15). Then the first-best and constrained efficient cost functions must be the same at \( v = v^{FB} \):

\[
k (v^{FB}, \Delta, \theta) = k^* (v^{FB}, \Delta, \theta) .
\]

(B-6)

The six boundary conditions are (B-1)–(B-6).

**Kolmogorov Forward Equation**

Note that the FKE can be written as

\[
0 = - \sum_{j=1}^{3} \frac{\partial}{\partial x_j} (a_i (x) g) + \frac{1}{2} \sum_{i,j=1}^{3} \frac{\partial^2}{\partial x_i \partial x_j} (\sigma_{ij} (x) g) .
\]

It is useful to write this in divergence form,

\[
\nabla \cdot S = 0,
\]

where

\[
S_i = a_i (x) - \frac{1}{2} \sum_{j=1}^{3} \frac{\partial}{\partial x_j} (\sigma_{ij} (x) g)
\]

is the \( i \)th element of \( S \). Here, \( S \) represents the “probability flux” representing the flux of particles through the point \( x = (x_1, x_2, x_3) = (\theta, v, \Delta) \).

Several of the boundary conditions are reflecting. If variable-\( i \) is reflecting, then

\[
S_i \cdot \hat{n}_i = 0,
\]

where \( \hat{n}_i \) is the outward unit normal vector in the \( i \)th direction evaluated at the reflecting boundary. Since \( \theta \) has reflecting boundaries, this implies that \( \hat{n}_i = (0, 0, 1) \)
at $\bar{\theta}$ and $\hat{n}_c = (0,0,-1)$ at $\bar{\theta}$. Thus, the two boundary conditions are

$$
\mu (\bar{\theta}) = \frac{1}{2} \sum_{j=1}^{3} \frac{\partial}{\partial x_j} (\sigma_{3j}(x) g) \bigg|_{\bar{\theta} = \bar{\theta}},
$$

$$
\mu (\bar{\theta}) = \frac{1}{2} \sum_{j=1}^{3} \frac{\partial}{\partial x_j} (\sigma_{3j}(x) g) \bigg|_{\bar{\theta} = \bar{\theta}}.
$$

Similarly, $\{v,v\} = \{v^\text{aut},v^{\text{FB}}\}$ and $\{\Delta,\Delta\} = \{0,\bar{\Delta}\}$ are reflecting. The outward unit normal vector associated with these are $(\mp 1,0,0)$ and $(0,\mp 1,0)$, respectively. The remaining four boundary conditions can then be derived as above by using the appropriate unit normal vectors at each boundary point.

C Contracts via Optimal Stochastic Control

We now construct an equivalent, but simplified planner’s problem. We use the stochastic maximum principle to solve an unconstrained maximization problem in the space of Lagrange multipliers.

To simplify the problem, we work with $\log (\theta_t)$ instead of $\theta_t$. In addition, $v_t$ and $\Delta_t$ now refer to the state variables in terms of $\log (\theta_t)$. Their laws of motion are

$$
d (\log (\theta_t)) = \left( \frac{\mu (\theta_t)}{\theta_t} - \frac{\sigma^2_{\theta}}{2} \right) dt + \sigma_{\theta} dZ_t = \mu^{\log} (\log (\theta_t)) dt + \sigma_{\theta} dZ_t,
$$

$$
dv_t = (\rho v_t - \bar{u}_t) dt + \sigma_{\theta} \Delta_t dZ_t + (v_{R} - v_t) (dR_t - \kappa dt),
$$

$$
d\Delta_t = \left( \rho - \frac{d\mu^{\log} (\log (\theta_t))}{d\log (\theta_t)} \right) \Delta_t - \bar{u}_{\log(\theta),t} dt + \sigma_{\theta}\sigma_{\Delta,t} dZ_t + (\Delta_{R} - \Delta_t) (dR_t - \kappa dt).
$$

To simplify notation, we will relabel $\log (\theta_t)$ with $\theta_t$ and $\tilde{\mu} (\theta) \equiv \mu^{\log} (\log (\theta))$. The planner’s HJB equation (1.20) is very difficult to work with in its current form because the directions of the higher-order derivatives depend on the choice of $\sigma_{\Delta}$. We instead reparametrize the state space in terms of Lagrange multipliers on the state variables.
We will set up and solve the planner’s problem via the stochastic maximum principle. Recall that the controls are \(\{c, C, \ell, \sigma_{\Delta}, z\}\). Let \(\lambda\) and \(\gamma\) denote the Lagrange multipliers on the laws of motion of \(v\) and \(\Delta\), respectively, and \(\sigma_{\theta}\sigma_{\lambda}, \sigma_{\gamma}\) their respective volatilities. Then the Lagrangian is

\[
L(g) = \int_{\Lambda \times \Gamma} \mathcal{L}(\lambda, \gamma, \theta) g(\lambda, \gamma, \theta) \, d(\lambda, \gamma) = \left\langle e^{-rt} \left[ c + \kappa C - w(\theta) e^{\theta} \ell \right] , g \right\rangle + \left\langle e^{-rt} f, A^* g + \kappa \tilde{f}_0 \right\rangle + \left\langle e^{-rt} \chi(\theta), w(\theta)^\alpha L(\theta) - Y \right\rangle + \left\langle e^{-rt} \eta(\theta), z(\theta) \right\rangle.
\]

Integrating by parts,

\[
L(g) = \left\langle e^{-rt} \left[ c + \kappa C - w(\theta) e^{\theta} \ell \right] , g \right\rangle + \left\langle e^{-rt} \left( Af + \kappa \tilde{f}_0 - rf \right) , g \right\rangle + \left\langle e^{-rt} \chi(\theta), w(\theta)^\alpha L(\theta) - Y \right\rangle + \left\langle e^{-rt} \eta(\theta), z(\theta) \right\rangle.
\]

Expanding this out and substituting in \(\lambda\) and \(\gamma\) yields

\[
L(g) = \left\langle e^{-rt} \left[ c + \kappa C - w(\theta) e^{\theta} \ell + \kappa \tilde{f}_0 - (r + \kappa) f \right] , g \right\rangle + \left\langle e^{-rt} \left[ \lambda \left( \dot{v} - ((\rho + \kappa) v - \bar{u}) \right) \right] , g \right\rangle + \left\langle e^{-rt} \left[ \gamma \left( \dot{\Delta} - (\rho + \kappa - \bar{\mu}'(\theta)) \Delta + \bar{u}_{\theta} \right) \right] , g \right\rangle + \left\langle e^{-rt} \left[ f_{\theta} \left( \dot{\theta} - \bar{\mu}'(\theta) \right) \right] , g \right\rangle - \left\langle e^{-rt} \left[ \sigma_{\theta}^2 \sigma_{\lambda} \Delta + \sigma_{\theta} \sigma_{\Delta} \sigma_{\gamma} + \sigma_{\theta} \text{Vol}(f_{\theta}) \right] , g \right\rangle + \left\langle e^{-rt} \chi(\theta), w(\theta)^\alpha L(\theta) - Y \right\rangle + \left\langle e^{-rt} \eta(\theta), z(\theta) \right\rangle.
\]

Integrating by parts one more time,

\[
L(g) = \left\langle e^{-rt} \left[ c + \kappa C - w(\theta) e^{\theta} \ell + \kappa \tilde{f}_0 - (r + \kappa) f \right] , g \right\rangle - \left\langle e^{-rt} \left[ \lambda \left( (\rho + \kappa) v - \bar{u} \right) + v \lambda \right] , g \right\rangle - \left\langle e^{-rt} \left[ \gamma \left( (\rho + \kappa - \bar{\mu}'(\theta)) \Delta - \bar{u}_{\theta} \right) + \dot{\Delta} \gamma \right] , g \right\rangle + \left\langle e^{-rt} \left[ f_{\theta} \left( \dot{\theta} - \bar{\mu}'(\theta) \right) \right] , g \right\rangle - \left\langle e^{-rt} \left[ \sigma_{\theta}^2 \sigma_{\lambda} \Delta + \sigma_{\theta} \sigma_{\Delta} \sigma_{\gamma} + \sigma_{\theta} \text{Vol}(f_{\theta}) \right] , g \right\rangle + \left\langle e^{-rt} \chi(\theta), w(\theta)^\alpha L(\theta) - Y \right\rangle + \left\langle e^{-rt} \eta(\theta), z(\theta) \right\rangle.
\]
Differentiating with respect to $\sigma_\Delta$ yields

$$[\sigma_\Delta] : -\sigma_\gamma^2 \sigma_\theta = 0.$$  

Thus $\gamma_t$ has zero volatility under the optimal contract and $\lambda_t = \tilde{u}_{c,t}^{-1}$. Next, we can see that the drift of $\lambda_t$ is $\mathcal{L}_\nu + (r + \kappa) \lambda_t = 0$ and the drift of $\gamma_t$ is $\mathcal{L}_\Delta + (r + \kappa) \gamma_t = (r - \rho + \tilde{\mu}'(\theta_t)) \gamma_t - \sigma_\theta^2 \sigma_{\lambda,t} = \tilde{\mu}'(\theta_t) \gamma_t - \sigma_\theta^2 \sigma_{\lambda,t}$ since we have assumed $r = \rho$. Thus we can write

$$d\lambda_t = \sigma_\theta \sigma_{\lambda,t} dZ_t + (\lambda_t^R - \lambda_t) dR_t, \quad (C-1)$$

$$d\gamma_t = (\tilde{\mu}'(\theta_t) \gamma_t - \sigma_\theta^2 \sigma_{\lambda,t}) dt + (\gamma_t^R - \gamma_t) dR_t. \quad (C-2)$$

In particular, $\lambda_t$ has no drift and $\gamma_t$ has no volatility, i.e., it is a slow-moving variable.

If we compare these first-order conditions to the ones that we would obtain from the original HJB equation (under the reparametrized state space), then we see that $\lambda_t = k_{v,t}$ and $\gamma_t = k_{\Delta,t}$.

We solve for $v, \Delta$, and the planner’s profit function, $p$, via a system of partial differential equations over the new state space. Let $P(\theta) = \int_{\Lambda \times \Gamma} p(\lambda, \gamma, \theta) g(\lambda, \gamma, \theta) d(\lambda, \gamma)$. Then these functions have a recursive representation:

$$(r + \kappa) v = \tilde{u} + \kappa v^R + v_\gamma (\tilde{\mu}'(\theta) \gamma - \sigma_\theta^2 \sigma_\lambda) + \frac{\sigma_\theta^2}{2} [v_{\lambda\lambda} \sigma_\lambda^2 + v_{\theta\theta}] + v_{\lambda\theta} \sigma_\theta^2 \sigma_\lambda;$$

$$(r + \kappa) \Delta = \tilde{u}_\theta + \Delta_\theta \tilde{\mu}(\theta) + \Delta_\gamma (\tilde{\mu}'(\theta) \gamma - \sigma_\theta^2 \sigma_\lambda) + \frac{\sigma_\theta^2}{2} [\Delta_{\lambda\lambda} \sigma_\lambda^2 + \Delta_{\theta\theta}] + \Delta_{\lambda\theta} \sigma_\theta^2 \sigma_\lambda;$$

$$(r + \kappa) P(\theta) = \int_{\Lambda \times \Gamma} \left\{ w(\theta) e^{\theta \ell} - c - \kappa C + p_{\theta\theta}(\theta) + p_\gamma (\tilde{\mu}'(\theta) \gamma - \sigma_\theta^2 \sigma_\lambda) + \frac{\sigma_\theta^2}{2} [p_{\lambda\lambda} \sigma_\lambda^2 + p_{\theta\theta}] + p_{\lambda\theta} \sigma_\theta^2 \sigma_\lambda + \kappa p_{f_0} + \chi(\theta) w(\theta)^\alpha e^{\theta \ell} - \lambda w(\theta)^e \ell \right\} g(\theta, \lambda, \gamma) d(\lambda, \gamma) + \eta(\theta) z(\theta).$$

Alternatively, taking the Gateaux derivative of the Lagrangian with respect to $f$ along an arbitrary variation, all three equations are characterized by the function
\[ f : (\lambda, \gamma, \theta) \rightarrow \mathbb{R} \text{ that solves} \]

\[
(r + \kappa) F(\theta) = \max_{\{c,C,\ell,z,\sigma\}} \left\{ \int_{\Lambda \times \Gamma} \left[ w(\theta) e^\theta \ell - c - \kappa C + \lambda (\tilde{u} + \kappa v^R) + \gamma \tilde{u}_\theta + f_\theta \tilde{\mu}(\theta) \right. \right.
\]

\[
+ f_\gamma (\tilde{\mu}'(\theta) \gamma - \sigma_\alpha^2 \sigma_\lambda) + \sigma_\theta^2 \left( \frac{f_{\lambda \alpha} \sigma_\lambda^2 + f_{\theta \theta}}{2} + \sigma_\lambda f_{\lambda \theta} \right) + \kappa f \tilde{f}_0
\]

\[
+ \chi(\theta) w(\theta)^\alpha e^\theta \ell - \mathcal{X} w(\theta) e^\theta \ell \left] \right. g(\lambda, \gamma, \theta) d(\lambda, \gamma) + \eta(\theta) z(\theta) \right\}, \quad \text{(C-3)}
\]

where for any triple \((\lambda, \gamma, \theta)\), the optimal allocation solves

\[
c \in \arg\max_{c'} \{ \lambda \tilde{u} - c' \},
\]

\[
C \in \arg\max_{C'} \{ \lambda v^R - C' \},
\]

\[
\ell \in \arg\max_{\ell'} \left\{ w(\theta) e^\theta \ell' + \lambda \tilde{u} + \gamma \tilde{u}_\theta + \chi(\theta) w(\theta)^\alpha e^\theta \ell' - \mathcal{X} w(\theta) e^\theta \ell' \right\},
\]

\[
\sigma_\lambda \in \arg\max_{\sigma'_\lambda} \left\{ \frac{f_{\lambda \lambda} \sigma_\lambda^2}{2} + (f_{\lambda \theta} - f_{\gamma}) \sigma'_{\lambda} \right\},
\]

\[
z \in \arg\max_{z'} \left\{ \int_{\Lambda \times \Gamma} [\lambda \tilde{u} + \gamma \tilde{u}_\theta] g(\lambda, \gamma, \theta) d(\lambda, \gamma) + \eta z' \right\}.
\]

Notice that \(p\) is the objective function in a constrained optimization problem while \(f\) is the objective function of an “unconstrained” problem. Therefore, these four equations can be linked via

\[
f(\lambda, \gamma, \theta) = \max_{\{v, \Delta\}} \{ p(v, \Delta, \theta) + \lambda v + \gamma \Delta \}.
\]

Then the Envelope Theorem implies

\[
v = f_\lambda, \quad \Delta = f_\gamma, \quad p = f - \lambda f_\lambda - \gamma f_\gamma.
\]

Note that the fourth condition implies that the second-order terms in (C-3) can be simplified to \(\frac{\sigma_\lambda^2}{2} (f_{\theta \theta} - f_{\lambda \lambda} \sigma_\lambda^2)\). Crucially, (C-3) is considerably easier to solve than is
The boundary conditions are

\[ f_\theta (\cdot, \cdot, \theta) = \lambda f_{\lambda \theta} (\cdot, \cdot, \theta) + \gamma f_{\gamma \theta} (\cdot, \cdot, \theta), \]
\[ f_\theta (\cdot, \cdot, \bar{\theta}) = \lambda f_{\lambda \theta} (\cdot, \cdot, \bar{\theta}) + \gamma f_{\gamma \theta} (\cdot, \cdot, \bar{\theta}), \]
\[ f_\lambda (\cdot, 0, \cdot) = \frac{u (c) + \kappa v^R (C)}{\rho + \kappa}, \]
\[ f_\gamma (\cdot, 0, \cdot) = 0, \]
\[ p (\cdot, 0, \cdot) = f (\cdot, 0, \cdot) - \lambda f_\lambda (\cdot, 0, \cdot) = -\frac{c + \kappa C}{\rho + \kappa}. \]

However, since \( \gamma_t \) never reaches zero there is a single non-explosive solution to each equation so the boundary conditions are not actually necessary to solve the problem.

The result below provides a set of sufficient conditions for a contract as determined by \((C-1)\) and \((C-2)\) to solve the relaxed problem.

**Proposition C1.** Suppose that the function \( f \) solves \((C-3)\). Then \((C-1)\) and \((C-2)\) define a contract in which \((v_t, \Delta_t, p_t)\) are given by

\[ v_t = f_{\lambda, t}, \quad \Delta_t = f_{\gamma, t}, \quad p_t = f_t - \lambda f_{\lambda, t} - \gamma f_{\gamma, t}. \]

Then the contract is a solution of the relaxed problem if the Hessian \( H (f) \) is positive definite.

**Proof.** The proof follows the one in Sannikov (2014) and has two key steps. First, we establish the mappings between the original state variables and the dual variables. Then, we show that the principal’s profit is bounded above under any alternative contract. Since the market clearing and wage derivative constraints do not affect the proof, we omit them for brevity.

**Step 1: Original/Dual Mapping**
Lemma C1. If $f$ solves (C-3) then $v_t = f_\lambda (\lambda_t, \gamma_t, \theta_t)$, $\Delta_t = f_\gamma (\lambda_t, \gamma_t, \theta_t)$, and the planner’s continuation payoff is $f (\lambda, \gamma_t, \theta_t) - \lambda_t v_t - \gamma_t \Delta_t$ in the contract defined by (C-1) and (C-2).

Proof. Differentiating (C-3) with respect to $\lambda$ and using the Envelope Theorem,

$$(r + \kappa) f_\lambda - (\tilde{u} + \kappa v^R) = f_{\lambda\theta} \tilde{\mu} (\theta) + f_{\lambda\gamma} \tilde{\mu}' (\theta) \gamma + \frac{\sigma^2_\theta}{2} \left[ f_{\lambda\lambda\lambda} \sigma^2_\lambda + f_{\theta\theta\lambda} \right] + \sigma^2_\theta \sigma_\lambda \left( f_{\lambda\theta\lambda} - f_{\lambda\gamma} \right).$$

The right side equals the drift of the process $f_\lambda (\lambda_t, \gamma_t, \theta_t)$ when $(\lambda_t, \gamma_t)$ follow (C-1) and (C-2). Thus as long as the transversality condition holds, $f_\lambda$ is an agent’s continuation value $v$. Similarly, differentiating with respect to $\gamma$,

$$(r + \kappa) f_\gamma - \tilde{u}_\theta = f_{\gamma\theta} \tilde{\mu} (\theta) + f_{\gamma\gamma} \tilde{\mu}' (\theta) \gamma + \frac{\sigma^2_\theta}{2} \left[ f_{\gamma\gamma\gamma} \sigma^2_\lambda + f_{\theta\theta\gamma} \right] + \sigma^2_\theta \sigma_\gamma \left( f_{\gamma\gamma\theta} - f_{\gamma\gamma\gamma} \right)$$

and the right side is the drift of $f_\gamma$ so as long as the transversality condition holds then $f_\gamma = \Delta$. Finally, subtracting $\lambda$ times $E [df_\lambda]$ and $\gamma$ times $E [df_\gamma]$ from (C-3),

$$(r + \kappa) (f - \lambda f_\lambda - \gamma f_\gamma) = w (\theta) e^\theta \ell - c - \kappa C + \left[ f_{\theta} - \lambda f_{\lambda\theta} - \gamma f_{\gamma\theta} \right] \tilde{\mu} (\theta)$$

$$+ \left[ f_{\lambda\gamma} - \lambda f_{\lambda\gamma} - \gamma f_{\gamma\gamma} - f_{\gamma} \right] \tilde{\mu}' (\theta) \gamma$$

$$+ \left[ f_{\gamma\theta} - \gamma f_{\gamma\gamma} - f_{\gamma} \right] \tilde{\mu}' (\theta) \gamma$$

$$+ \frac{\sigma^2_\theta}{2} \left[ \frac{\partial^2 (f - \lambda f_\lambda - \gamma f_\gamma)}{\partial \lambda^2} \sigma^2_\lambda + \frac{\partial^2 (f - \lambda f_\lambda - \gamma f_\gamma)}{\partial \theta^2} \sigma^2_\gamma + \frac{\partial^2 (f - \lambda f_\lambda - \gamma f_\gamma)}{\partial \lambda \partial \theta} \right].$$
Hence, the process
\[
\bar{p}_t = \int_0^t e^{-(r+\kappa)s} \left( w(\theta_s) e^{\theta_s} \ell_s - c_s - \kappa C_s \right) ds + e^{-(r+\kappa)t} (f(\lambda_t, \gamma_t, \theta_t) - \lambda_t v_t - \gamma_t \Delta_t)
\]
is a martingale. Since \(\bar{p}_t = \mathbb{E}_t[\bar{p}_{T\wedge \tau}]\) it follows that \(f(\lambda_t, \gamma_t, \theta_t) - \lambda_t v_t - \gamma_t \Delta_t\) is the planner’s continuation payoff under the contract with a given agent.

**Step 2: Bounding the Value Function**

We will now show that under any alternative contract, the principal’s value function is bounded above. We do this in two steps.

**Lemma C2.** Consider an alternative contract characterized by the controls \((c, C, y, \sigma_\Delta)\) and let \((v, \Delta)\) denote the state variables under these controls. If \(H(f)\) is positive definite, then there exist processes

\[
d\lambda_t = \mu_\lambda^t dt + \sigma_{\lambda,t} dZ_t + J_\lambda^t dR_t \quad \text{and} \quad d\gamma_t = \mu_\gamma^t dt + \sigma_{\gamma,t} dZ_t + J_\gamma^t dR_t
\]
such that \(v_t = f_{\lambda,t}\) and \(\Delta_t = f_{\gamma,t}\).

**Proof.** We need to construct the processes above such that the drifts and volatilities match. To match volatilities, by Itô’s Lemma,

\[
\sigma_{\theta} \begin{bmatrix} f_{\lambda\lambda} & f_{\lambda\gamma} & f_{\lambda\theta} \\ f_{\lambda\gamma} & f_{\gamma\gamma} & f_{\gamma\theta} \end{bmatrix} \begin{bmatrix} \sigma_{\lambda} \\ \sigma_{\gamma} \\ 1 \end{bmatrix} = \sigma_{\theta} \begin{bmatrix} \Delta \\ \sigma_\Delta \end{bmatrix}.
\]

Since \(H(f)\) is positive definite, so are all of its principal submatrices, and positive definite matrices are invertible, so there exists a unique solution to this system of
equations. To match drifts, we set up another system of equations:

\[ M(f) \begin{bmatrix} \mu^\lambda \\ \mu^\gamma \\ \tilde{\mu}(\theta) \end{bmatrix} + \cdots = \begin{bmatrix} (\rho + \kappa) v - \tilde{u} - \kappa v^R \\ (\rho + \kappa - \tilde{\mu}'(\theta)) \Delta - \tilde{u}_\theta \end{bmatrix}. \]

Similarly, since \( H(f) \) is positive definite a solution exists. A similar argument works for the jump terms.

We will now show that \( \bar{p} \) is a supermartingale under alternative contracts. By Itô’s Lemma, the drift of \( f(\lambda_t, \gamma_t, \theta_t) - \lambda_t v_t - \gamma_t \Delta_t \) is

\[
\text{Drift} = f_\lambda \mu^\lambda_t + f_\gamma \mu^\gamma_t + f_\theta \tilde{\mu}(\theta_t) + \frac{\sigma^2}{2} \begin{bmatrix} \sigma_{\lambda,t} & \sigma_{\gamma,t} & 1 \\ \sigma_{\lambda,t} & \sigma_{\gamma,t} & 1 \\ \sigma_{\Delta,t} & \sigma_{\Delta,t} & 1 \end{bmatrix} H(f) \begin{bmatrix} \sigma_{\lambda,t} \\ \sigma_{\gamma,t} \\ 1 \end{bmatrix} - \mu^\lambda_t v_t - \mu^\gamma_t \Delta_t
\]

\[- \sigma_\theta^2 \begin{bmatrix} \sigma_{\lambda,t} & \sigma_{\gamma,t} \\ \sigma_{\lambda,t} & \sigma_{\gamma,t} \end{bmatrix} \begin{bmatrix} \Delta_t \\ \sigma_{\Delta,t} \end{bmatrix} - \lambda_t ((\rho + \kappa) v_t - \tilde{u}_t) - \gamma_t ((\rho + \kappa - \tilde{\mu}'(\theta_t)) \Delta_t - \tilde{u}_{\theta,t})
\]

\[= f_\theta \tilde{\mu}(\theta_t) - \frac{\sigma^2}{2} \begin{bmatrix} \sigma_{\lambda,t} & \sigma_{\gamma,t} & 1 \\ \sigma_{\lambda,t} & \sigma_{\gamma,t} & 1 \end{bmatrix} H(f) \begin{bmatrix} \sigma_{\lambda,t} \\ \sigma_{\gamma,t} \\ 1 \end{bmatrix} + \sigma_\theta^2 (f_\lambda \sigma_{\lambda,t} + f_\gamma \sigma_{\gamma,t} + f_\theta)
\]

\[- \lambda_t ((\rho + \kappa) v_t - \tilde{u}_t) - \gamma_t ((\rho + \kappa - \tilde{\mu}'(\theta_t)) \Delta_t - \tilde{u}_{\theta,t}),\]

where the final equality uses volatility matching. Under (C-1) and (C-2), this is

\[ f_\theta \tilde{\mu}(\theta_t) + \frac{\sigma^2}{2} (f_\theta \sigma_{\lambda,t} + f_\gamma \sigma_{\gamma,t}^2) - \lambda_t ((\rho + \kappa) v_t - \tilde{u}_t) - \gamma_t ((\rho + \kappa - \tilde{\mu}'(\theta_t)) \Delta_t - \tilde{u}_{\theta,t}),\]

where \( \sigma_{\lambda,t}^2 \) is the value under the optimal contract. It follows that the drift of \( \bar{p} \) is zero, i.e., it is a martingale under the optimal contract. On the other hand, with
arbitrary laws of motion, the drift of $\overline{p}$ changes by $(e^{-(r+\kappa)t}$ times)

$$-\frac{\sigma^2}{2} \left[ \begin{array}{cc} \sigma_{\lambda,t} & \sigma_{\gamma,t} \\ \sigma_{\gamma,t} & 1 \end{array} \right] H(f) \left[ \begin{array}{c} \sigma_{\lambda,t} \\ \sigma_{\gamma,t} \end{array} \right] + \sigma^2 \left( f_{\lambda l} \sigma_{\lambda,t} + f_{\gamma l} \sigma_{\gamma,t} + f_{\theta l} \right) - \frac{\sigma^2}{2} \left( f_{\theta\theta} - f_{\lambda\lambda} \sigma_{\lambda,t}^2 \right)$$

$$= -\frac{\sigma^2}{2} \left[ \begin{array}{cc} x_t & \sigma_{\gamma,t} \\ \sigma_{\gamma,t} & 0 \end{array} \right] H(f) \left[ \begin{array}{c} x_t \\ \sigma_{\gamma,t} \end{array} \right] \leq 0$$

for some $x_t$ that depends on the other processes and the inequality holds because $H(f)$ is positive definite. That is, under any alternative contract, $\overline{p}$ is a supermartingale and a martingale under the optimal contract. 

\[ \square \]

\section{D Numerical Appendix}

This section describes our numerical procedure to solve the model. Most of our algorithm follows Nuño and Moll (2015) so we defer the interested reader there for more details on individual steps. The state space is $\Theta \times \Lambda \times \Gamma$.

We use a finite difference method and approximate the planner’s value function, $f$, on a grid of $(\theta, \lambda, \gamma)$. Let $(\theta_i, \lambda_j, \gamma_k)$, with $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $k = 1, \ldots, K$ denote a grid point, and

$$f_{i,j,k} = f(\theta_i, \lambda_j, \gamma_k)$$

the function $f$ evaluated at that point. We approximate the derivatives with either a forward- or backward-difference approximation,

$$f_\theta(\theta_i, \lambda_j, \gamma_k) \approx f^F_{\theta;i,j,k} = \frac{f_{i+1,j,k} - f_{i,j,k}}{\theta_{i+1} - \theta_i},$$

$$f_\theta(\theta_i, \lambda_j, \gamma_k) \approx f^B_{\theta;i,j,k} = \frac{f_{i,j,k} - f_{i-1,j,k}}{\theta_i - \theta_{i-1}},$$

72
and similarly for the other state variables. For the second derivative, we use

\[ f_{\theta \theta; i,j,k} \approx 2 \frac{(\theta_i - \theta_{i-1}) f_{i+1,j,k} - (\theta_{i+1} - \theta_{i-1}) f_{i,j,k} + (\theta_{i+1} - \theta_i) f_{i-1,j,k}}{(\theta_i - \theta_{i-1}) (\theta_{i+1} - \theta_i) (\theta_{i+1} + \theta_{i-1})} \]

and for the mixed partial (recall that \( \gamma \) has zero volatility), we use

\[ f_{\theta \lambda; i,j,k} = \frac{f_{i+1,j,k+1} - f_{i+1,j,k-1} - f_{i-1,j,k+1} + f_{i-1,j,k-1}}{4 (d\theta) (d\lambda)} \]

Along the boundaries,

\[ f_{\theta \lambda; i,j,k} (\theta) = \frac{f_{i+1,j,k+1} - f_{i+1,j,k} - f_{i-1,j,k+1} + f_{i-1,j,k}}{2 (d\theta) (d\lambda)} \]
\[ f_{\theta \lambda; i,j,k} (\bar{\theta}) = \frac{f_{i+1,j,k} - f_{i-1,j,k} - f_{i+1,j,k-1} + f_{i-1,j,k-1}}{2 (d\theta) (d\lambda)} \]
\[ f_{\theta \lambda; i,j,k} (\bar{\lambda}) = \frac{f_{i,j,k+1} - f_{i,j,k-1} - f_{i-1,j,k+1} + f_{i-1,j,k-1}}{2 (d\theta) (d\lambda)} \]
\[ f_{\theta \lambda; i,j,k} (\lambda) = \frac{f_{i+1,j,k+1} - f_{i+1,j,k} - f_{i-1,j,k+1} + f_{i-1,j,k}}{2 (d\theta) (d\lambda)} \]

while at the corners,

\[ f_{\theta \lambda; i,j,k} (\theta, \bar{\lambda}) = \frac{f_{i+1,j,k+1} - f_{i+1,j,k} - f_{i,j,k+1} + f_{i,j,k}}{(d\theta) (d\lambda)} \]
\[ f_{\theta \lambda; i,j,k} (\theta, \bar{\lambda}) = \frac{f_{i,j,k+1} - f_{i,j,k} - f_{i-1,j,k+1} + f_{i-1,j,k}}{(d\theta) (d\lambda)} \]
\[ f_{\theta \lambda; i,j,k} (\bar{\theta}, \lambda) = \frac{f_{i+1,j,k} - f_{i-1,j,k} - f_{i,j,k} + f_{i,j,k-1}}{(d\theta) (d\lambda)} \]
\[ f_{\theta \lambda; i,j,k} (\bar{\theta}, \bar{\lambda}) = \frac{f_{i,j,k} - f_{i,j,k-1} - f_{i,j,k} + f_{i,j,k-1}}{(d\theta) (d\lambda)} \]

We use an “upwind scheme” to determine whether to use the forward or backward approximation: use the forward approximation when the drift of the state variable is positive and a backward approximation when the drift is negative. Then, using an implicit method, the discretized version of the HJB equation (setting \( \kappa = 0 \) for
notational simplicity) is\(^{35}\)

\[
\sum_{j,k} \left[ \frac{f_{i,j,k}^{n+1} - f_{i,j,k}^n}{\Delta} + r f_{i,j,k}^{n+1} \right] g_{i,j,k}^n = \sum_{j,k} \left[ y_{i,j,k}^n - c_{i,j,k}^n + \sum_{j',k'} \xi_{i',j',k'}^{i,j,k} f_{i',j',k'}^{n+1} \right. \\
+ \left. \left( \chi_i^n w_i^{n\alpha-1} - \chi_i^n w_i^n \right) e^{\theta_i} \ell_{i,j,k}^n \right] g_{i,j,k}^n + \eta_i^n z_i^n, \quad (D-1)
\]

where \(\xi_{i',j',k'}\) are the (functional) coefficients on \(f_{i',j',k'}^{n+1}\) when the forward and backward approximations are written out, and depend on the drift of the state variables and whether this drift is positive or negative; \((i',j',k') \in \{i-1, i, i+1\} \times \{j-1, j, j+1\} \times \{k-1, k, k+1\}\). Since (D-1) is a system of \(J \times K\) linear equations for each \(i\), we can rewrite it in matrix form:

\[
\frac{\hat{f}_{i}^{n+1} - \hat{f}_{i}^n}{\Delta} + r \hat{f}_{i}^{n+1} = \hat{u}_{i}^n + A_i^n \hat{f}_{i}^{n+1} + \eta_i^n z_i^n,
\]

where \(A_i^n\) is a discrete approximation of the infinitesimal generator given \(\theta_i\) and must satisfy the properties of a Poisson matrix: (1) all rows sum to zero; (2) diagonal elements are nonpositive; and (3) off-diagonal elements are nonnegative. Also,

\[
\hat{f}_{i}^{n+1} = f_{i}^{n+1} \circ g_i^n = (f_{i,1,1}^{n+1}, \ldots, f_{i,J,K}^{n+1})^T \circ (g_{i,1,1}^n, \ldots, g_{i,J,K}^n)^T, \\
\hat{u}_{i}^n = u_i^n \circ g_i^n = (y_{i,1,1}^n - c_{i,1,1}^n, \ldots, y_{i,J,K}^n - c_{i,J,K}^n)^T \circ (g_{i,1,1}^n, \ldots, g_{i,J,K}^n)^T
\]

are the pointwise products of the value and objective functions, respectively, given \(\theta_i\), with the density function. This can be rearranged into

\[
\left[ \left( \frac{1}{\Delta} + r \right) \mathbf{I} - A_i^n \right] \hat{f}_{i}^{n+1} = \hat{u}_{i}^n + \frac{\hat{f}_{i}^n}{\Delta} \equiv \hat{d}_{i}^n + \eta_i^n z_i^n.
\]

\(^{35}\)Abusing notation, here \(\Delta\) refers to the step size, not information rent.
Nuño and Moll (2015) show that the KFE can then be written

\[ \mathbf{A}_i^T \mathbf{g}_i = \mathbf{h}_i, \]

where \( \mathbf{A}_i = \lim_{n \to \infty} \mathbf{A}_i^n \) and \( \mathbf{h}_i \) is the vector of zeros with \(-1\) as the first entry; we do this for each \( i \).

The relevant first-order conditions, not discretized, are

\[
\begin{align*}
[c] : \lambda u_c - 1 &= 0, \\
[C] : \lambda v_C^R - 1 &= 0, \\
[\ell] : w(\theta) e^\theta + \lambda \bar{u}_\ell + \gamma \bar{u}_{\theta \ell} + \chi(\theta) w(\theta) e^\theta - \mathcal{X} w(\theta) e^\theta &= 0, \\
[z] : \int_{\Lambda \times \Gamma} \gamma \bar{u}_{\theta z} g(\lambda, \gamma, \theta) d(\lambda, \gamma) + \eta(\theta) &= 0,
\end{align*}
\]

along with the envelope condition

\[
[w] : L(\theta) + \int_{\Lambda \times \Gamma} \gamma \bar{u}_{\theta w} g(\lambda, \gamma, \theta) d(\lambda, \gamma) + \alpha \chi(\theta) w(\theta) \alpha^{-1} L(\theta) - \mathcal{X} L(\theta) + \eta'(\theta) = 0.
\]

Combining (D-5) and (D-6), plus the boundary condition \( \eta(\theta) = 0 \),

\[
0 = \int_{\Lambda \times \Gamma} \gamma \bar{u}_{\theta w} g(\lambda, \gamma, \theta) d(\lambda, \gamma) \\
+ \int_{\theta} \left[ \int_{\Lambda \times \Gamma} \gamma \bar{u}_{\theta w} g(\lambda, \gamma, \theta) d(\lambda, \gamma) + \alpha \chi(\theta) w(\theta) \alpha^{-1} L(\theta) - \mathcal{X} L(\theta) \right] d\theta.
\]

Since we work with processes that generate unbounded stationary distributions, we need to truncate the domain of the skill process, \( \Theta = [\underline{\theta}, \bar{\theta}] \), to solve the model numerically.

Our solution algorithm to solve the full mean field game is as follows:
1. Guess the wage function and its derivative \( w^0(\theta_i), z^0(\theta_i) \) for \( i = 1, \ldots, I \).

2. Guess the Lagrange multipliers \( \chi^0(\theta_i) \) and \( \eta^0(\theta_i) \) for \( i = 1, \ldots, I \).

3. Guess the value function \( f^0(\theta_i, \lambda_j, \gamma_k) \) for \( i = 1, \ldots, I, j = 1, \ldots, J, \) and \( k = 1, \ldots, K \).

4. For each \((\theta_i, \lambda_j, \gamma_k)\), use (D-2)–(D-6) and their approximations above to solve for the remaining policy functions \( \{c, C, \ell, \sigma_\Delta\} \); iterate on \( f^n \) until convergence.

5. Solve for the stationary distribution via \( 0 = \mathcal{A}^*g + \kappa \tilde{f}_0 \).

6. Use (D-7) to iterate on \( \chi \) for each \( \theta_i \).

7. Based on the resulting stationary distribution, iterate on the wage function via the market clearing condition.

8. If \( \|w^{n+1}(\theta) - w^n(\theta)\| > \epsilon \), return to Step 1; otherwise, stop.
Chapter 2

The First-Order Approach with Persistent Private Information

2.1 Introduction

There are many environments in which agents have private information that persists over time. For example, an individual’s income is likely to be highly correlated over time, as is his ability to produce goods or manage a firm. Alternatively, an agent could take an action today, such as a CEO instituting a new policy, that affects output far into the future. Despite the prevalence of such situations, outside of certain special cases, it is quite difficult to characterize the optimal contract. In this paper, I work with a model in which agents have private information about a persistent shock and I provide a simple, intuitive condition that can be imposed ex ante under which the so-called first-order approach is sufficient for optimality. Specifically, I show that the first-order approach is sufficient if agents cannot overreport their shocks. This setting is general enough to accommodate many models, including models of optimal insurance and dynamic taxation.
In static mechanism design problems, an agent’s private information lasts for one period and decisions today do not directly affect decisions in future periods. For example, if an agent’s skill is i.i.d. over time, then his skill today carries no information about his skill tomorrow. Similarly, in standard moral hazard models, the decision to exert effort (or shirk) today does not convey any information about future decisions to exert effort. In contrast, in dynamic mechanism design problems, private information persists through time and does affect agents’ decisions in future periods. Continuing with the examples above, an agent’s skill tomorrow is likely to be very highly correlated with his skill today, or his decision to shirk today might have long-run effects on output.

The revelation principle allows the mechanism designer to restrict attention to direct revelation mechanisms in which agents truthfully reveal their type or action to the planner. This implies a set of incentive constraints, dictating that an agent’s utility from truth-telling (weakly) dominates his utility from any other report. However, since the set of incentive constraints is likely to be quite large, one typically appeals to the so-called first-order approach, replacing the entire set of incentive constraints with a single first-order or envelope condition. One then solves this “relaxed” problem and checks ex post whether the implied contract is in fact a solution to the original problem.

The issue is that the first-order condition is only necessary for optimality, not sufficient. In particular, by construction, the first-order approach focuses on small, “local” deviations and says nothing about large ones. Thus, the solution to the relaxed problem might not be globally optimal. In static mechanism design problems, however, there are several well-known conditions that guarantee that the first-order approach is sufficient for optimality. For example, Rogerson (1985) shows that certain monotonicity conditions on the distribution of shocks are sufficient, and Jewitt (1988) generalizes these conditions to a wider class of distributions. Importantly, these
conditions can be imposed ex ante and are easy to check. It is well-known, however, that no such conditions exist for general dynamic mechanism design problems.

This is not to say that there has not been progress, though and in fact several recent papers have derived sufficient conditions in specific cases (see below). However, these conditions are not intuitive and have to be checked numerically ex post. In contrast, working with a continuous time model in which agents’ types are private and persistent, I show that as long as agents cannot overreport their types (plus a simple parametric assumption), the first-order approach is sufficient for optimality. This condition is appealing because it is easy to interpret, is economically-motivated, can be imposed ex ante, and is typically satisfied anyway in the class of models I work with.

Although my model is very similar to existing models, I derive a simpler condition than do others because I use a different method of proof. A key step in the proof is to show that the utility gain from misreporting is negative. While other papers (see below) construct a linear approximation to the gain from deviating, I construct a quadratic approximation, which yields sharper results with fewer restrictions.

The first-order approach consists of a few steps. It is well-known that dynamic mechanism design problems require two contracting state variables: continuation/promised utility to track promise-keeping, and information rent to track threat-keeping. First, I derive the law of motion for continuation utility under an arbitrary contract. Then, I derive the first-order necessary condition for optimality. I then use this necessary condition to derive the law of motion for information rent under the optimal contract. Finally, I use both of these laws of motion under the optimal contract to establish sufficiency. Importantly, the first-order approach does not need to hold for arbitrary contracts, only the optimal one.
2.2 Related Literature

My model of persistent private information is very close to those of Williams (2011) and Kapićka (2013). The latter, working in discrete time, shows that the first-order approach is sufficient under a certain integral monotonicity condition, while the former, working in continuous time, derives a simpler but still difficult-to-interpret condition. Prat and Jovanovic (2014) derive a related condition in a continuous time model where the agent has an unknown quality and takes a private action. The difference between this paper and those two, and hence why I can derive a simpler condition than them, is that our proofs use different techniques. I follow Sannikov (2014) and construct a quadratic approximation to the upper bound for an agent’s utility from misreporting. In contrast, they use only a linear approximation. While my model’s setup is identical to the one in Kapićka (2013), I can derive sharper results than him by taking advantage of continuous time methods such as Itô’s Lemma and the Girsanov Theorem; working in continuous time also imposes certain continuity restrictions on reports. Pavan, Segal, and Toikka (2014) work in an even more general discrete time setting and as a result, their condition, also an integral monotonicity condition, is the most complicated.

Most of the techniques in this paper are fairly standard in the continuous time optimal contracting literature. In particular, I adapt the methodology developed by Sannikov (2014) in a model in which moral hazard has long-run effects, and extended by DeMarzo and Sannikov (2015) to a moral hazard model in which the principal also has access to a signal, to a setting with persistent private information. Their main methodological contribution is a verification method for global incentive compatibility, which I adapt to my own setting. The key step in this verification method is conjecturing an appropriate upper bound on the utility gain from deviating. See the above papers and the references therein for more related models.
The remainder of the paper is organized as follows: Section 2.3 describes the model, Section 2.4 outlines the planner’s mechanism design problem, Section 2.5 develops the necessity and sufficiency of the first-order condition, and Section 2.6 concludes.

## 2.3 Model

Time is continuous and is indexed by \( t \in [0, \infty) \). Let \((\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})\) denote a complete, filtered probability space, where the filtration \( \{\mathcal{F}_t\} \) satisfies the usual conditions.\(^1\) The economy has a unit measure of agents who live until hit with a death shock. Formally, an agent is hit with a Poisson shock, \( R_t \), with intensity \( \kappa > 0 \).\(^2\) When an agent dies, a new one is born to maintain a constant population size. At each date, agents have utility over a vector \( X_t = (x_{1,t}, \ldots, x_{N,t}) \in \mathbb{X} \). Agents receive an idiosyncratic shock \( \theta_t \in \Theta \subseteq \mathbb{R}^{++} \) at each date with the initial value, \( \theta_0 \), drawn from some distribution. Crucially, the shock is private information to each agent. Agents have preferences\(^3\)

\[
\mathbb{E}_0 \left[ \int_0^\tau e^{-\rho t} \tilde{u}(X_t; \theta_t) \, dt \right],
\]

where \( \tau \equiv \inf \{ t \mid R_t > 0 \} \) is when the death shock hits and the expectation is over both the shock process and the death shock. I can rewrite this as

\[
\mathbb{E}_0 \left[ \int_0^\infty e^{-(\rho+\kappa) t} \tilde{u}(X_t; \theta_t) \, dt \right].
\]

\(^1\)A filtration \( \{\mathcal{F}_t\} \) satisfies the usual conditions if \( \mathcal{F}_0 \) contains all \( \mathbb{P} \)-null sets and \( \mathcal{F}_t = \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon} \) for all \( t \geq 0 \).

\(^2\)This shock is not necessary but is included to increase generality. I will assume that agents do not receive any utility once the death shock hits but this can be easily generalized. An alternative is to endogenize the retirement time as part of the contract. The model with finitely-lived agents is similar, too.

\(^3\)Throughout, \( \mathbb{E}_t [\cdot] = \mathbb{E}[\cdot | \mathcal{F}_t] \) denotes the date-\( t \) conditional expectation.
The utility function \( \tilde{u} : X \times \Theta \rightarrow \mathbb{R} \) is continuous, differentiable, affine, concave in \( X \), and satisfies the following property: if \( \tilde{u}_{x_i} \leq 0 \) then \( \tilde{u}_{x_i\theta} \geq 0 \). This last property says that if a particular “good” \( x_i \) makes an agent better off, then having higher \( \theta \) has decreasing returns. This environment nests several widely-used models (possibly after redefining variables), such as standard Mirrlees models of optimal taxation (both static and dynamic), the endowment economy settings of Green (1987), Thomas and Worrall (1990), and Williams (2011) where income is stochastic, and games in which sellers of a product have a private production cost.\(^4\)

The private information \( \theta_t \) evolves according to a diffusion process

\[
d\theta_t = \mu(\theta_t) \, dt + \sigma_{\theta} \theta_t \, dZ_t,
\]

where \( Z_t \) is a standard Wiener process adapted to the filtration \( \{ \mathcal{F}_t \} \) for all \( t \), and \( \mu(\theta) \in C^2(\Omega) \) is a measurable function satisfying the standard Lipschitz and growth conditions:

\[
\| \mu(t, x) - \mu(t, y) \| \leq A |x - y|, \quad \| \mu(t, x) \| \leq B (1 + |x|),
\]

for some constants \( A, B > 0 \). Going forward, let \( \mathcal{F}_t \equiv \sigma(\theta_s \mid s \leq t) \) denote the natural filtration with respect to \( \{ \theta_t \} \).

**Assumption 2.1.** The drift of the skill process satisfies \( 2\mu'(\theta) + \sigma_\theta^2 < \rho \).

This assumption rules out processes that are “too explosive” and thus processes satisfying this condition must revert to the mean, in some sense. Intuitively, if this condition did not hold then \( \theta_t \) would grow so fast that the benefits of deviations would accrue too quickly to be deterred with dynamic incentives. Under the most common

\(^4\)However, it does not include the Atkeson and Lucas (1992) taste shock model or the Battaglini and Lamba (2015) repeated buyer-seller game because in those models, \( \tilde{u}_c > 0 \) and \( \tilde{u}_{c\theta} > 0 \).
specification, \( \log(\theta_t) \) is an Ornstein-Uhlenbeck process, this amounts to

\[
\frac{2\sigma_\theta^2 - \rho}{2 (1 - \rho)} + \log(\theta^*) < 1 + \log(\theta_t)
\]

if \( p < 1 \) and \( 2\sigma_\theta^2 < \rho \) if \( p = 1 \). Thus, given standard estimates such as those in Chapter 1, it holds trivially with perfect persistence and otherwise, it holds unless \( \theta_t \) is too small because there, agents might pool at the same contract.\(^5\)

I also impose that this process reflects back inward at two exogenous points \( \underline{\theta}, \overline{\theta} \) with \( 0 < \underline{\theta} < \overline{\theta} < \infty \) so that \( \theta_t \) resides in a compact set. Compactness ensures that all functions are bounded except possibly on a measure-zero set, and \( \theta > 0 \) ensures that zero is not an absorbing state.\(^6\) The reflecting boundaries insure that the model does not have absorbing states, which doom the first-order approach because multi-period deviations might become appealing. Let \( \theta^t = \{\theta_s\}_{s \in [0,t]} \) denote a history of shocks up to date \( t \), and \( \Theta^t \) the set of all histories up to date \( t \).

While the process (2.1) is quite flexible and can generate several commonly-used stationary distributions for \( \theta \) (possibly after a monotonic transformation),\(^7\) because it is driven by a Wiener process it is already restricted considerably. Indeed, it is a first-order Markov process and the continuous time limit of a binomial tree, which implies its stationary distribution is related to the normal distribution.\(^8\) That said, since most of the literature uses lognormal or Pareto distributions anyway, these assumptions are not too restrictive. But it also does not have jumps so sample paths are almost surely continuous. As I will argue, this assumption is essential because it rules out jumps in agents’ reporting strategies.

\(^5\)This is consistent with Battaglini and Lamba (2015), who find that monotonicity tends to fail when \( \theta \) is small (though they work in discrete time and in a different framework).

\(^6\)The boundedness assumption can probably be relaxed at the cost of additional regularity conditions on the utility function and reporting strategies.

\(^7\)For example, lognormal, double Pareto, and double Pareto lognormal when taking \( \underline{\theta} \rightarrow 0 \) and \( \overline{\theta} \rightarrow \infty \).

\(^8\)Kapička (2013) assumes \( \theta_t \) is a first-order Markov process but new draws come from an arbitrary distribution satisfying some regularity conditions, while Pavan, Segal, and Toikka (2014) make neither assumption.
The social planner promises to deliver lifetime utility $U$ to the agent and his objective is to minimize the expected present value of costs. The planner can freely borrow and save at the exogenous net interest rate $r > 0$, which is also his discount factor. Let $K : \mathbb{X} \to \mathbb{R}$ denote the planner’s cost of allocating $X \in \mathbb{X}$ to an agent, with $K$ continuous, increasing, and convex. For example, in an insurance model $K(X) = K(c, y) = c - y$ is the planner’s claims payout minus his income.

At the beginning of each date, agents observe their shocks (and if they die, they immediately leave the economy). Since the planner cannot observe $\theta_t$, agents then report their shock to the social planner, who determines $X_t$, and agents consume.

### 2.4 Planner’s Problem

At $t = 0$, the planner selects an allocation $x = \{x_t\}_{t=0}^\infty$, which is a sequence of $\mathcal{F}_t$-progressively measurable\footnote{A function is progressively measurable if, for every $t$, the map $[0, t] \times \Omega \to \mathbb{X}$ defined by $(s, \omega) \mapsto x_s(\omega)$ is Borel $[0, t] \otimes \mathcal{F}_t$-measurable.} functions $x_t : \Theta^t \to \mathbb{X}$; let $\mathcal{X}$ denote the set of allocations. Then an agent’s utility from an allocation $x$ is

$$U(x) = \mathbb{E}_0 \left[ \int_0^\infty e^{-(\rho + \kappa)t} \bar{u}(x_t(\theta^t); \theta_t) \, dt \right].$$

At each date agents report their shocks to the planner. A reporting strategy $\sigma = \{\sigma_t\}_{t=0}^\infty \in \Sigma$ is a sequence of $\mathcal{F}_t$-progressively measurable functions $\sigma_t : \Theta^t \to \Theta$; let $\sigma^t \in \Sigma^t$ denote the history of reports up to date $t$.

Let $\{x_t (\sigma^t (\theta^t))\}_{t=0}^\infty$ denote an agent’s allocation given reporting strategy $\sigma \in \Sigma$. By the Revelation Principle, I can restrict attention to direct revelation mechanisms in which agents truthfully reveal their shocks to the planner. The truth-telling strategy specifies $\sigma_t (\theta^t) = \theta_t$ for all $t$. Finally, let

$$\Sigma^* \equiv \{\sigma \in \Sigma \mid \sigma_t (\theta^t) = \hat{\theta}_t \leq \theta_t \ \forall t\}$$
denote the set of reporting strategies in which the agent cannot overreport his shock.

There are two important restrictions that I can impose on strategies. First, without loss of generality, I can focus on reporting strategies with $\Sigma^t \subseteq \Theta^t$ since otherwise, the planner would immediately be able to detect a lie. Second, and more importantly, let $Q$ denote the probability measure induced by an arbitrary reporting strategy (defined formally via the Radon-Nikodym derivative); then $Q$ must be absolutely continuous with respect to the true measure $P$. This is important because it rules out jumps in agents’ reports: since Brownian motions are almost surely continuous, an agent’s skill process is almost surely continuous. Therefore, if an agent reports truthfully until date $t$ but the report suddenly jumps, then the planner can be sure the agent is lying and punish him. Obviously, this restriction would not hold if the shock was driven by a jump-diffusion process.

Given a history $\theta^t \in \Theta^t$ and a reporting strategy $\sigma \in \Sigma$, let $v^\sigma_t (\theta^t)$ denote an agent’s promised/continuation utility:

$$
v^\sigma_t (\theta^t) \equiv \mathbb{E}_t \left[ \int_t^\tau e^{-\rho(s-t)} \tilde{u}(\sigma_s (\theta^s)) \, ds \right].$$

Let $v_t$ denote the value of $v^\sigma_t$ under truth-telling. An allocation is incentive-compatible if truth-telling yields weakly higher continuation utility than any other reporting strategy:

$$v_t (\theta^t) \geq v^\sigma_t (\theta^t) \quad \forall \sigma \in \Sigma, \ \theta^t \in \Theta^t. \quad (2.2)$$

Note that this is not a single constraint, but a doubly-continuum of constraints that rules out deviations following any history of shocks and any history of strategies.\footnote{As mentioned above, an agent will not participate in the game unless his utility from}

\footnote{The measure $Q$ is absolutely continuous with respect to $P$, written $Q \ll P$, if for every measurable set $A \subseteq \Omega$, $P (A) = 0$ implies $Q (A) = 0$.}

\footnote{Discrete time models such as Fernandes and Phelan (2000) and Kapička (2013) often use a nearly equivalent concept called “temporary incentive compatibility” that explicitly rules out only one-period deviations.}
an allocation exceeds $U$, which yields the participation constraint

$$U(x) \geq U.$$  \hfill (2.3)

Let $X^{IC}$ denote the set of incentive-compatible allocations that induce the agent to play the game, i.e., those allocations that satisfy the incentive-compatibility constraint (2.2) and participation constraint (2.3). This set is convex since $\tilde{u}_t$ is affine.

The present value of the cost to the planner of delivering an allocation $x$ is

$$
\Psi(x) = \mathbb{E}_0 \left[ \int_0^\infty e^{-rt} K(x_t(\theta_t)) \, dt \right].
$$

Since the planner’s goal is to find the cheapest allocation $x \in X^{IC}$, his problem, written sequentially, is

$$K(U) \equiv \min_{x \in X^{IC}} \Psi(x).$$

An allocation is efficient if it attains this minimum and by Berge’s Theorem of the Maximum, a solution necessarily exists and is continuous.

### 2.5 Relaxed Problem and First-Order Approach

The planner’s problem as constructed above is hopelessly complicated because of the set of incentive constraints. Instead, I follow the first-order approach and replace (2.2) with a single first-order condition that is both necessary and sufficient for optimality.

All proofs are in the Appendix.

I follow the approach outlined in Fernandes and Phelan (2000), Kapíčka (2013), and Pavan, Segal, and Toikka (2014) in discrete time, and Williams (2011) and Sannikov (2014) in continuous time, by reformulating the planner’s problem recursively. To do so, I need to derive the law of motion for each of the planner’s state variables. The first state variable is the agent’s private information, $\theta_t$, whose law of motion is
(2.1). The second state variable is an agent’s continuation utility, $v_t$, as is standard in mechanism design problems.\footnote{See Thomas and Worrall (1988) and Sannikov (2008), for example.} The planner uses continuation utility in order to track promise-keeping through time. My first result, which is standard in the continuous time optimal contracting literature, establishes the law of motion for $v_t$ under an arbitrary contract.

**Proposition 2.1.** Fix a contract and a reporting strategy $\sigma$ with finite expected payoff to the agent. Then the process \{v_t\} corresponds to the agent’s continuation utility if and only if there exists a process $\Delta_t \in H^2$ such that\footnote{A process \{X_t\} is in the space $H^2[0,T]$ if it is $\mathcal{F}_t$-progressively measurable and is square-integrable, $\mathbb{E} \left[ \int_0^T X_t^2 \, dt \right] < \infty$.}

$$
    dv_t = (\rho v_t - \tilde{u}_t) \, dt + \sigma \theta_t \Delta_t \, dZ_t - v_t (dR_t - \kappa \, dt)
$$

(2.4)

and the transversality condition $\mathbb{E}_t \left[ e^{-\rho (T \wedge \tau)} v_{T \wedge \tau} \right] \to 0$ as $T \to \infty$ holds.\footnote{In a finite-horizon economy, the transversality condition is replaced with a terminal boundary condition $v_T = U(T)$ and (2.4) is a backwards stochastic differential equation with a known terminal condition but unknown initial condition.}

The drift of (2.4) tracks promise-keeping: $\tilde{u}_t$ is just-delivered utility and $\rho v_t$ is everything owed going forward. The process $\Delta_t$ is the sensitivity of $v_t$ to variations in the shock process. This also has a natural interpretation in terms of options. Thinking of the contract between the principal and the agent as a package of call options, then $v_t$ represents the value of the options and $\Delta_t$ is the “delta” of the options, i.e., the sensitivity of the value to changes in the underlying shock process.

Since this is a dynamic mechanism design problem, however, I need a third state variable to track threat-keeping.\footnote{If the shock process was driven by a jump-diffusion process instead, then I would need another state variable to keep track of jumps and my results probably would not hold anymore.} This is because misreporting changes the planner’s beliefs about the distribution of the shock process, which affects what the agent can report to the planner at all future dates. In other words, violating the incentive
constraint today affects future incentive constraints. To deter such deviations, the planner has to provide the agent with some information rent. Formally, following DeMarzo and Sannikov (2015), an agent’s information rent is the sensitivity of his continuation value to his report, evaluated at truth-telling:

$$\Delta_t \equiv \frac{\partial v_t}{\partial \theta_t} \bigg|_{\hat{\theta}_t = \theta_t} = \frac{\partial v_t}{\partial \theta_t}.$$ (2.5)

This state variable is present in Fernandes and Phelan (2000), Williams (2011), Kapička (2013), and Pavan, Segal, and Toikka (2014) in settings with private information, and Sannikov (2014) in a setting in which private actions have long-run effects.\(^{16}\) Pavan, Segal, and Toikka (2014) argue that this variable is like an impulse response function that “describe(s) how a change in the agent’s current type propagates through his type process.” Also, this derivative is a Malliavin derivative instead of an ordinary one. The Malliavin derivative is a way to formalize impulse response functions in continuous time by defining what it means to “differentiate” a Brownian motion, since the standard one-shot deviation principle does not have bite in these settings.\(^{17}\) In fact, as the proof of Proposition 2.1 shows, \(\hat{\Delta}_t\) is the “black box” process in the Martingale Representation Theorem. Meanwhile, the Clark-Ocone-Haussman Theorem shows that this process is precisely the Malliavin derivative, thus explaining the link.

**Proposition 2.2.** A necessary condition for truth-telling to be optimal is that,

$$\hat{\Delta}_t = \Delta_t.$$ (2.6)

\(^{16}\)It is not a state variable, however, in Sannikov (2008) and other settings where hidden actions have one-time effects. There, it is well-known that the first-order approach is valid under standard monotonicity assumptions.

\(^{17}\)See Di Nunno, Øksendal, and Proske (2009) for an introduction to Malliavin calculus and Borovička, Hansen, and Scheinkman (2014) for more on how to use Malliavin derivatives to compute impulse response functions; I use these methods on an optimal taxation problem in Chapter 1 of this dissertation.
for all $t$, where $\hat{\Delta}_t$ is the process in Proposition 2.1.

This is the first-order necessary condition for optimality and is quite intuitive: the sensitivity of continuation utility must equal the sensitivity evaluated at truth-telling. Let $\mathcal{X}^{FOA}$ denote the set of allocations that satisfy (2.6), which necessarily satisfies $\mathcal{X}^{IC} \subseteq \mathcal{X}^{FOA}$. This is precisely the condition identified in discrete time by Pavan, Segal, and Toikka (2014) (see their Theorem 1) and in continuous time by Williams (2011) (see his equation (10)). However, Williams (2011) and I arrive at this result via different methods. He first transforms the state space using the Girsanov Theorem and then uses a version of the stochastic maximum principle to derive this expression directly from an agent’s Hamiltonian. On the other hand, I follow Sannikov (2014) and use the Giransov Theorem to directly maximize an agent’s utility function under small deviations from truth-telling, using Malliavin calculus to compute the derivatives.\footnote{He, Wei, Yu, and Gao (2014) apply this method in a moral hazard model with learning, gaining considerable tractability by working with CARA utility.} It is worth noting that in static mechanism design settings such as Sannikov (2008), the derivation of the first-order condition establishes both necessity and sufficiency. That proof considers strategies that deviate up until date $t$ and play truthfully after. However, in that setting, strategies played before $t$ do not affect strategies played after $t$, which is clearly not the case here. Finally, if I restrict strategies so that $\sigma \in \Sigma^*$, then (2.6) becomes $\hat{\Delta}_t \leq \Delta_t$.

With the first-order necessary condition in hand, I can derive the law of motion of $\Delta_t$ under the first-order approach, i.e., the law of motion of information rent under the optimal contract.

**Proposition 2.3.** The finite process $\{\Delta_t\}$ is characterized by (2.5) if and only if, for some $\Gamma_t \in H^2$,

$$d\Delta_t = (\rho + \mu'(\theta_t))\Delta_t - \tilde{u}_{\theta,t} - \sigma^2_{\theta} \theta_t \Gamma_t) dt + \sigma_{\theta} \theta_t \Gamma_t \, dZ_t - \Delta_t (dR_t - \kappa \, dt) \quad (2.7)$$
and the transversality condition $\mathbb{E}_t \left[ e^{-\rho(T\wedge \tau)} \Delta_{T\wedge \tau} \right] \to 0$ as $T \to \infty$ holds.

The planner uses the process $\Gamma_t$ to control the flow of information rent over time. In the drift of (2.7), $-\mu'(\theta_t) \Delta_t$ is the decay in information rent while $\tilde{u}_{\theta,t} + \sigma^2_{\theta} \Gamma_t$ is just-delivered information rent. Continuing with the options analogy, $\Gamma_t$ is related to the “gamma” of the option, or the sensitivity of the delta to changes in the underlying shock process, i.e., the second derivative of promised utility with respect to the shock.

An agent’s incentive to report truthfully depends on how much downside protection the call options offer: the more protection they offer, the higher is the gamma, which reduces the amount of information rent the planner owes going forward.

Williams (2011) and Farhi and Werning (2013) have a similar result. To see where this expression comes from, consider an agent’s Bellman equation:

$$\rho v(\theta) = \max_{\theta'} \left\{ \tilde{u}(\theta'; \theta) + \frac{1}{dt} \mathbb{E} [dv(\theta'; \theta)] \right\}.$$  

In an individual agent’s problem, his shock $\theta$ is the only state variable. Expanding this via Itô’s Lemma,

$$(\rho + \kappa) v(\theta) = \max_{\theta'} \left\{ \tilde{u}(\theta'; \theta) + \frac{\partial v}{\partial \theta} (\theta'; \theta) \mu(\theta) + \frac{1}{2} \frac{\partial^2 v}{\partial \theta^2} (\theta'; \theta) \sigma^2_{\theta} \theta \right\}.$$  

Applying the Envelope Theorem at the optimal $\theta'$ gives

$$(\rho + \kappa) \frac{\partial v}{\partial \theta} (\theta) = \max_{\theta'} \left\{ \tilde{u}_{\theta}(\theta'; \theta) + \frac{\partial^2 v}{\partial \theta^2} (\theta'; \theta) \mu(\theta) + \frac{\partial v}{\partial \theta} (\theta'; \theta) \mu'(\theta) + \frac{\partial^2 v}{\partial \theta^2} (\theta'; \theta) \sigma^2_{\theta} \theta $$

\[+ \frac{1}{2} \frac{\partial^3 v}{\partial \theta^3} (\theta'; \theta) \sigma^2_{\theta} \theta^2 \right\} \]

while applying Itô’s Lemma directly to $\frac{\partial v}{\partial \theta}$ gives

$$d \left( \frac{\partial v}{\partial \theta} \right) = \left( \frac{\partial^2 v}{\partial \theta^2} (\theta) \mu'(\theta) + \frac{1}{2} \frac{\partial^3 v}{\partial \theta^3} (\theta) \sigma^2_{\theta} \theta^2 \right) dt + \frac{\partial^2 v}{\partial \theta^2} (\theta) \sigma_{\theta} \theta \ dZ_t.$$  

90
at the optimal $\theta'$. Now, substitute the envelope condition into this law of motion and take expectations:

$$\frac{1}{dt} \mathbb{E} \left[ d \left( \frac{\partial v}{\partial \theta} \right) \right] = (\rho - \mu' (\theta)) \frac{\partial v}{\partial \theta} - \tilde{u}_\theta - \sigma^2 \theta \frac{\partial^2 v}{\partial \theta^2}.$$

Using the definition of $\Delta$ and defining $\Gamma \equiv \frac{\partial^2 v}{\partial \theta^2}$, it follows that

$$d\Delta_t = ((\rho - \mu' (\theta_t)) \Delta_t - \tilde{u}_{\theta,t} - \sigma^2 \theta_t \Gamma_t) dt + \sigma \theta_t \Gamma_t dZ_t - \Delta_t (dR_t - \kappa dt),$$

as desired. This proof is not technically correct since it relies on a particular recursive representation of an agent’s problem. However, $v(\theta)$ is an agent’s continuation utility on the equilibrium path and, as argued by Fernandes and Phelan (2000) and Kapíčka (2013), the set of lifetime utilities does not have a recursive structure.\(^1\) I get around this by constructing an alternative process and appealing to the Martingale Representation Theorem.

The planner’s relaxed problem is

$$K (U) \equiv \min_x \Psi (x)$$

subject to

$$v_0 = \mathbb{E}_0 \left[ \int_0^\infty e^{-(\rho + \kappa)t} \tilde{u}_t \, dt \right]$$

and the first-order incentive constraint (2.6). By Propositions 2.1 and 2.3, the relaxed problem is equivalent to maximizing the above objective function subject to (2.4) and (2.7).

Before coming to the main result of this paper, I need one more assumption:

**Assumption 2.2.** Agents cannot overreport their shocks, $\sigma \in \Sigma^*$.

\(^1\)They show that the set of all equilibrium and off-equilibrium lifetime utilities *does* have a recursive structure.
This amounts to assuming that upward incentive constraints are irrelevant. In models with private information such as those mentioned earlier (with and without persistence), this restriction is almost always proved, confirmed numerically, or assumed outright. For example, Ales and Maziero (2009) prove that agents never overreport in a dynamic Mirrlees model with i.i.d. shocks when $\Theta = \{\theta_L, \theta_H\}$, Kapicka (2013) solves a dynamic Mirrlees model (with persistence), with $\Theta$ an interval, in the region where agents do not overreport, and Williams (2011) explicitly assumes it in a model with endowment shocks. This restriction also has a natural economic interpretation: agents must be able to show that they have any income that they report, or that agents cannot pretend to be more skilled than they actually are. Below, I give a simple example, from Williams (2011), of when this assumption is automatically satisfied.

**Example 2.1.** Suppose an agent’s income is private information and evolves according to a random walk

$$d\theta_t = \sigma_\theta \, dZ_t.$$  

This is an Ornstein-Uhlenbeck process with $p = 1$ and so trivially satisfies Assumption 2.1. Let $\hat{\theta}_t$ denote a report and $c(\hat{\theta}_t)$ the consumption transfer from the planner as a function of the report. Agents have utility

$$\tilde{u}(c_t; \theta_t) = -\exp \left( -\lambda \left( \theta_t + c(\hat{\theta}_t) \right) \right) = -\exp \left( -\lambda C(\hat{\theta}_t; \theta_t) \right),$$

which satisfies all of the conditions above. Williams (2011) shows that in this case, the model has a closed-form solution with

$$C(\hat{\theta}_t; \theta_t) = \overline{C}(v_0) + \frac{\sigma_\theta^2 \lambda}{2} t + \sigma_\theta Z_t.$$  

---

\(^{20}\) For the purposes of this example, I will ignore the reflecting boundaries.
It follows that overreporting is never an issue because consumption, and hence utility, does not depend on the report.

Unfortunately, outside of very special cases such as those above, it is difficult (if not impossible) to prove that upward incentive constraints are slack; if one could prove this, then Assumption 2.2 obviously would not be necessary.

**Theorem 2.1.** Under Assumptions 2.1 and 2.2 and the transversality condition of Proposition 2.1, the optimal contract under the first-order approach is globally optimal.

To my knowledge, this is the first sufficiency result for the first-order approach entirely in terms of model primitives and strategies. Pavan, Segal, and Toikka (2014) establish a sufficient condition in Markov environments but their condition is extremely complicated (see their Theorem 3) while Kapićka (2013) derives a simpler, but still endogenous integral monotonicity condition that needs to verified ex post (see his Section 6.3).\(^{21}\) Finally, Williams (2011) shows that the first-order approach is sufficient for optimality if \(\theta_t\Gamma_t\) is appropriately bounded by an expression that depends on the utility function and other endogenous objects. This essentially amounts to limiting the volatility of agents’ reports relative to the volatility of the shock process.\(^{22}\) He shows that his condition can sometimes be simplified but that this simplification (and even his original condition) might be overly stringent and that the first-order approach can be sufficient even though the conditions fail.

So why do I obtain a sharper result than Williams (2011) and Kapićka (2013) even though our settings are essentially identical? Since I work in continuous time, I have a more powerful set of tools than does the latter. Working in continuous time also imposes a lot more structure on reports. As for the former, he constructs

\[^{21}\text{Farhi and Werning (2013) and Golosov, Troshkin, and Tsyvinski (2016b) numerically check that his condition holds in their settings.}\]

\[^{22}\text{I thank Drew Fudenberg for this interpretation.}\]
a linear approximation to the utility gain from deviating, and uses the incentive constraint and his condition to ultimately show that the gain is negative. In contrast, I follow Sannikov (2014) and construct a quadratic approximation to the gain from deviating. Using the law of motion (2.7), I show that Assumptions 2.1 and 2.2 are enough to guarantee that this gain is negative. He, Wei, Yu, and Gao (2014) and Di Tella and Sannikov (2016) use a similar verification method in their settings, and it is important to note that the appropriate approximation is very setting-specific and requires some guesswork.

The results in this section established that the planner’s relaxed problem is equivalent to the original problem. For completeness’ sake, I will complete the characterization of the first-order approach by explicitly writing out the planner’s recursive problem. Recall that $K(x)$ is the cost to the planner of delivering an allocation; rewrite this as $K(\theta, v, \Delta)$ to emphasize the dependence on the state variables. Then the solution to the planner’s recursive problem is obtained by solving his Hamilton–Jacobi–Bellman equation:

$$(r + \kappa) K = \inf_{\{x, \Gamma\}} \left\{ K(x) + K_{\theta} \mu(\theta) + K_v \left[ (\rho + \kappa) v - \tilde{u} \right] + K_\Delta \left[ (\rho + \kappa - \mu'(\theta)) \Delta - \tilde{u}_\theta - \sigma^2 \theta \Gamma \right] + \frac{\sigma^2}{2} \left[ K_{\theta \theta} + K_{\theta v} \theta^2 \Delta^2 + K_{\Delta \Delta} \theta^2 \Gamma^2 \right] + \sigma^2 \theta^2 \left[ K_{\theta \Delta} \Delta + K_{\theta \Gamma} + K_{v \Delta} \Delta \Gamma \right] \right\}. \tag{2.8}$$

In general, this equation is very difficult to solve because the directions of the second derivatives depend on the choice of $\Gamma$. While Williams (2011) gives an example with a closed-form solution, in general one can simplify the problem by using duality to formulate an equivalent problem over the space of Lagrange multipliers. I defer to Sannikov (2014), DeMarzo and Sannikov (2015), and Chapter 1 of this dissertation (where I use duality to solve a dynamic Mirrlees model) for more on this procedure.

---

23He derives a very simple sufficient condition, but it still depends on a term similar to $\Gamma_t$. That said, his model has moral hazard instead of persistent private information.
2.6 Conclusion

This paper studies the first-order approach in a widely-used class of dynamic mechanism design models. My main results characterize the first-order necessary condition for optimality, showing that as long as an agent cannot overreport the value of his shock each period, this first-order condition is sufficient, too. This restriction is very reasonable for many types of dynamic mechanism design problems such as optimal taxation models (see Farhi and Werning (2013) and Golosov, Troshkin, and Tsyvinski (2016b), for example). My condition improves on those derived by Williams (2011) and Kapička (2013) in a nearly identical setting. I do so by utilizing a new set of techniques developed by Sannikov (2014).

It would be interesting to investigate whether the results in this paper are unique to the continuous time setting, or if they in fact do hold in discrete time. On one hand, the model is the same in both cases so the economics should not change but on the other hand, taking the time-step to zero and hence reducing the impact of any single report could have tangible impacts. In addition, in discrete time the process \{θ_t\} can “jump” while here the reporting process must be sufficiently well-behaved. Second, while the class of models covered in this paper is widely-applicable, it does not cover settings such as career concerns models à la Holmström (1999), taste shock models like Atkeson and Lucas (1992), or buyer-seller games like in Battaglini and Lamba (2015) where agents have an incentive to overreport. Future work should look into extending the techniques/proofs in this setting to others and, in particular, determining whether the first-order approach is sufficient as long as agents are restricted to behaving as one would expect them to anyway.
2.7 Appendix: Omitted Proofs

A Proof of Proposition 2.1

Fix a contract and a reporting strategy \( \hat{\theta}_t \), and define the deflated gains process

\[
G_t \equiv \mathbb{E}_t \left[ \int_0^\infty e^{-\rho s} \tilde{u}_s \left( \hat{\theta}_s \right) \, ds \right] = \int_0^t e^{-\rho s} \tilde{u}_s \left( \hat{\theta}_s \right) \, ds + e^{-\rho t} v_t.
\]

Since \( G_t \) is a martingale, by the Martingale Representation Theorem, there exist processes \( \hat{\Delta}_t, Q_t \in H^2 \) such that

\[
dG_t = e^{-\rho t} \hat{\Delta}_t \, dZ_t + Q_t \left( dR_t - \kappa \, dt \right).
\]

Differentiating with respect to \( t \) and rearranging,

\[
dv_t = \left( \rho v_t - \tilde{u}_t \right) \, dt + \sigma \theta_t \hat{\Delta}_t \, dZ_t + Q_t \left( dR_t - \kappa \, dt \right).
\]

Since \( v_t \) jumps to \( v_t^R \) at retirement, it must be that \( Q_t = v_t^R - v_t \). To see that the transversality condition holds, note that

\[
e^{-\rho t} v_t = \mathbb{E}_t \left[ \int_t^{T \wedge \tau} e^{-\rho s} \tilde{u}_s \, ds + e^{-\rho (T \wedge \tau)} v_{T \wedge \tau} \right].
\]

Taking the limit \( T \to \infty \) and applying the Monotone Convergence Theorem delivers the result.

For the other direction, if there is a solution to (2.4) with the desired properties, I can integrate

\[
v_t = \mathbb{E}_t \left[ \int_t^{T \wedge \tau} e^{-\rho (s-t)} \tilde{u}_s \, ds + e^{-\rho (T \wedge \tau)} v_{T \wedge \tau} \right].
\]

Take the limit as \( T \to \infty \) and apply the Monotone Convergence Theorem to the first term and the transversality condition to the second term to complete the proof.
B  Proof of Proposition 2.2

Rather than working with the actual process, \( \{\theta_t\} \), I work with its logarithm, \( \{\log(\theta_t)\} \), which corresponds to the growth rate of \( \theta_t \) instead of its level. Its law of motion is

\[
d(\log(\theta_t)) = M(\theta_t)\,dt + \sigma_{\theta}\,dZ_t,
\]

where \( M(\cdot) \) is a function of \( \mu(\cdot) \).

Consider the process defined by

\[
d\hat{Z}_t \equiv dZ_t + \varepsilon a_t\,dt,
\]

where \( \varepsilon \in \mathbb{R} \) and \( a_t \in H^2 \). The reporting space is bounded so the deviation satisfies Novikov’s condition, \( \mathbb{E} \left[ \exp \left( \frac{1}{2} \int_0^T |a_s|^2 \,ds \right) \right] < \infty \). Then this process is a Brownian motion under another (martingale equivalent) measure \( \mathbb{Q} \sim \mathbb{P} \) by the Girsanov Theorem, and is also induced by some alternative reporting strategy \( \hat{\theta}_t(\varepsilon) \neq \theta_t \). In addition,

\[
\frac{dQ}{dP} = \xi_t(\varepsilon) = \exp \left( - \int_0^t \varepsilon a_s \,dZ_s - \frac{1}{2} \int_0^t \varepsilon^2 |a_s|^2 \,ds \right)
\]

is the Radon-Nikodym derivative. It follows that

\[
\left. \frac{d\xi_t}{d\varepsilon} \right|_{\varepsilon=0} = - \int_0^t a_s \,dZ_s.
\]

This corresponds to the Malliavin derivative in the direction of the deviation.

Let \( \hat{\Delta}_t^\ell \) denote the sensitivity of promised utility to changes in \( \log(\hat{\theta}_t) \), and let \( W_t \) denote promised utility as a function of \( \log(\hat{\theta}_t) \), with equation of motion

\[
dW_t = (\rho W_t - \tilde{u}_t)\,dt + \sigma_{\theta}\hat{\Delta}_t^\ell\,dZ_t + (W_t^R - W_t)\,(dR_t - \kappa\,dt).
\]
Define two processes,
\[ \zeta_t'' \equiv \int_t^{t'} a_s \, dZ_s, \]
\[ \Lambda_t \equiv \mathbb{E}_t \left[ \int_t^\infty e^{-\rho(s-t)} \zeta_t'' \tilde{u}_s \, ds \right]. \]

I need the following lemma.

**Lemma B1.** An equivalent expression for \( \Lambda_t \) is
\[ \Lambda_t = \mathbb{E}_t \left[ \int_t^\infty e^{-\rho(s-t)} \sigma_\theta a_s \tilde{\Delta}_s^t \, ds \right]. \]

**Proof.** Note that
\[ \zeta_t^s = \int_t^s a_\tau \, dZ_\tau = \zeta_t^{t'} + \int_{t'}^s a_\tau \, dZ_\tau \]
when \( t \leq t' \leq s \). Then
\[ \Psi_{t'} \equiv \mathbb{E}_{t'} \left[ \int_t^\infty e^{-\rho(s-t)} \zeta_t^s \tilde{u}_s \, ds \right] = \int_t^{t'} e^{-\rho(s-t)} \zeta_t^s \tilde{u}_s \, ds + e^{-\rho(t'-t)} \zeta_t^{t'} W_{t'} + \Lambda_{t'}^t \]
is a martingale, where \( \Lambda_{t'}^t \) is the contribution to \( \Lambda_t \) of the compensation after \( t' \),
\[ \Lambda_{t'}^t \equiv \mathbb{E}_{t'} \left[ \int_{t'}^\infty e^{-\rho(s-t)} \left( \int_{t'}^s a_\tau \, dZ_\tau \right) \tilde{u}_s \, ds \right]. \]
Applying Itô’s Lemma,
\[ d\Psi_{t'} = e^{-\rho(t'-t)} \left\{ \zeta_t^{t'} \left( \tilde{u}_t - \rho W_{t'} \right) \, dt' + a_t W_{t'} \, dZ_{t'} + \zeta_t^{t'} \left[ (\rho W_{t'} - \tilde{u}_{t'}) \, dt' + \sigma_\theta \tilde{\Delta}_{t'} \, dZ_{t'} \right] \right\} \]
\[ + e^{-\rho(t'-t)} \zeta_t^{t'} \left( W_{t'}^R - W_{t'} \right) (dR_{t'} - \kappa \, dt') + e^{-\rho(t'-t)} \sigma_\theta a_t \tilde{\Delta}_{t'} \, dt' + d\Lambda_{t'}^t \]
\[ = e^{-\rho(t'-t)} \left\{ \sigma_\theta a_t \tilde{\Delta}_{t'} \, dt' + \left( \sigma_\theta \tilde{\Delta}_{t'} \zeta_t^{t'} + a_t W_{t'} \right) \, dZ_{t'} + \zeta_t^{t'} \left( W_{t'}^R - W_{t'} \right) (dR_{t'} - \kappa \, dt') \right\} + d\Lambda_{t'}^t. \]
Integrating over \([t, t']\) and taking the expectation at \(t\),

\[
0 = \mathbb{E}_t \left[ \int_t^{t'} e^{-\rho(t'-t)} \sigma_\theta a_{\ell}' \Delta_{t'} \, dt' \right] + \mathbb{E}_t \left[ \int_t^{t'} d\Lambda_{t'} \right].
\]

Taking the limit as \(s \to \infty\) and assuming the transversality condition \(\mathbb{E}_t [\Lambda_s] \to 0\) delivers the result.

With this lemma in hand, I turn to an agent’s problem. An agent’s objective function is

\[
\mathcal{U}_t(\varepsilon) \equiv \mathbb{E}_t \left[ \int_t^\infty e^{-\rho(s-t)} \xi_s(\varepsilon) \tilde{u}_s(\varepsilon) \, ds \right].
\]

For truth-telling to be optimal, \(\mathcal{U}_t(\varepsilon)\) must be maximized at \(\varepsilon = 0\). Under the first-order approach, this means its derivative is zero at \(\varepsilon = 0\). Then

\[
\frac{d\mathcal{U}_t}{d\varepsilon} \bigg|_{\varepsilon=0} = \mathbb{E}_t \left[ \int_t^\infty e^{-\rho(s-t)} \tilde{u}_s \left(-\int_t^s a_\tau \, dZ_\tau\right) \, ds \right] + \mathbb{E}_t \left[ \int_t^\infty e^{-\rho(s-t)} \partial \tilde{u}_s \bigg|_{\varepsilon=0} \left(\frac{\partial}{\partial \varepsilon}\right) \, ds \right] = 0.
\]

Again, technically \(\partial \tilde{u}_s \bigg|_{\varepsilon=0}\) is a Malliavin derivative. By Lemma B1, I can rewrite the above as

\[
\frac{d\mathcal{U}_t}{d\varepsilon} \bigg|_{\varepsilon=0} = -\mathbb{E}_t \left[ \int_t^\infty e^{-\rho(s-t)} \sigma_\theta a_s \tilde{\Delta}_s \, ds \right] + \mathbb{E}_t \left[ \int_t^\infty e^{-\rho(s-t)} \partial \tilde{u}_s \bigg|_{\varepsilon=0} \left(\frac{\partial}{\partial \varepsilon}\right) \, ds \right] = 0.
\]

From the formula for the Malliavin derivative of a diffusion process, the entire second term equals

\[
\mathbb{E}_t \left[ \int_t^\infty e^{-\rho(s-t)} \sigma_\theta a_s \int_s^\infty e^{-\rho(\tau-s)} \tilde{u}'_\tau D_s \left(\ln (\theta_\tau)\right) \, d\tau \, ds \right] = \mathbb{E}_t \left[ \int_t^\infty e^{-\rho(s-t)} \sigma_\theta a_s \tilde{\Delta}_s \, ds \right],
\]

Given a diffusion process \(dX_t = \mu(t, X_t) \, dt + \sigma(t, X_t) \, dZ_t\), where \(\mu, \sigma\) satisfy the Lipschitz growth conditions in the text, its Malliavin derivative is

\[
D_s (X_t) = \sigma(t, X_t) \exp \left( \int_s^t \left( \mu'(\tau, X_\tau) - \frac{1}{2} \sigma'(\tau, X_\tau)^2 \right) \, d\tau + \int_s^t \sigma'(\tau, X_\tau) \, dZ_\tau \right).
\]
where I used that $\Delta^t_t \equiv \mathbb{E}_t \left[ \int_t^\infty e^{-\rho(s-t)} D_t \left( \tilde{u} \left( \ln (\theta_s) \right) \right) ds \right]$ and the Malliavin chain rule. Putting everything together,

$$
\mathbb{E}_t \left[ \int_0^\infty e^{-\rho(s-t)} \sigma_{\theta} a_s \left\{ \tilde{\Delta}_s^t - \Delta^t_s \right\} ds \right] = 0 \implies \tilde{\theta}_s \tilde{\Delta}_s = \tilde{\Delta}_s^t = \Delta^t_s = \theta_s \Delta_s
$$
at truth-telling. Since $\tilde{\theta}_s = \theta_s$ also holds at truth-telling (by definition), it follows that $\tilde{\Delta}_s = \Delta_s$.

### C Proof of Proposition 2.3

Let $\{\theta_t\}$ denote the true process and $\{\theta_t + \varepsilon_t\}$ the reported process. Define a process

$$
\Phi_t \equiv \int_0^t e^{-\rho s} \tilde{u}_{\theta,s} \varepsilon_s \, ds + e^{-\rho t} \Delta_t \varepsilon_t.
$$

The equation of motion for $\varepsilon_t$ is

$$
d\varepsilon_t = \mu' (\theta_t) \varepsilon_t \, dt + \sigma_{\theta} \varepsilon_t \, dZ_t + \sigma_{\theta} \theta_t \left( dZ_t - d\tilde{Z}_t \right),
$$

where $\tilde{Z}_t$ is a Brownian motion under the probability measure $\mathbb{Q}$ induced by the report. Since $\Phi_t$ is a martingale, its drift is zero:

$$
0 = \frac{1}{dt} \mathbb{E}_t [d\Phi_t] = \tilde{u}_{\theta,t} - \rho \Delta_t + \mu' (\theta_t) \Delta_t + \sigma_{\theta}^2 \theta_t \Gamma_t,
$$

where $\sigma_{\theta} \theta_t \Gamma_t$ is the volatility of $\Delta_t$. Rearranging this expression gives the desired result. The transversality condition holds because $\Delta_0$ is finite.

Conversely, suppose that $\{\Delta_t\}$ is any process that satisfies (2.7). Then the process

$$
\Phi^t_{t'} \equiv \mathbb{E}_{t'} \left[ \int_t^{t' \wedge \tau} e^{-\rho(s-t)} \tilde{u}_{\theta,s} \varepsilon_s \, ds \right] + e^{-\rho(t' \wedge \tau - t)} \Delta_{t'} \varepsilon_{t'}
$$

100
is a martingale. Therefore,

$$\Delta_t = \Phi^t_t = \lim_{t' \to \infty} \mathbb{E}^t_t \left[ \Phi^t_{t'} \right] = \lim_{t' \to \infty} \mathbb{E}^t_t \left[ \int_t^{t' \wedge \tau} e^{-\rho(s-t)} \tilde{u}_{\theta,s} \varepsilon_s \, ds \right] + \mathbb{E}^t_t \left[ e^{-\rho(t' \wedge \tau - t)} \Delta_{t'} \varepsilon_{t'} \right].$$

Then the transversality condition implies that $\Delta_t$ satisfies the definition given.

D Proof of Theorem 2.1

Consider any alternative reporting strategy $\{\widehat{\theta}_t\}$ deviating from truth-telling, $\{\theta_t\}$, where $\widehat{\theta}_t = \theta_t + \varepsilon_t$ and $\varepsilon_t < 0$ for almost all $t$. To prove that the payoff to the agent from this deviation cannot exceed the equilibrium payoff $v_t$, I follow Sannikov (2014) by constructing an upper bound for the utility gain from deviating.

To construct the upper bound, I show that it is sufficient to keep track of two additional deviation state variables that matter to the agent’s potential deviation value. These variables are the deviation itself, $\varepsilon_t < 0$, and a variable to capture accumulated information rent:

$$\Phi_t \equiv \int_0^t e^{-\rho s} \tilde{u}_{\theta,s} \varepsilon_s \, ds + e^{-\rho t} \Delta_t \varepsilon_t = \mathbb{E}^t_t \left[ \int_0^\infty e^{-\rho s} \tilde{u}_{\theta,t} \varepsilon_t \, ds \right].$$

By applying Itô’s Lemma, it follows that

$$d\Phi_t = e^{-\rho t} \left[ (\tilde{u}_{\theta,t} + \theta_t^2 \Gamma_t - (\rho - \mu' (\theta_t)) \Delta_t + \mu^A_t) \varepsilon_t \right] \, dt + e^{-\rho t} \varepsilon_t \sigma_{\theta t} (1 + \theta_t \Gamma_t) \, dZ_t$$

$$+ e^{-\rho t} \Delta_t \sigma_{\theta t} \varepsilon_t \left( dZ_t - \tilde{dZ}_t \right) + e^{-\rho t} \varepsilon_t \left( \Delta_t^R - \Delta_t \right) (dR_t - \kappa \, dt)$$

$$d\Phi_t \cdot d\varepsilon_t = e^{-\rho t} \varepsilon_t^2 \sigma_{\theta t}^2 (1 + \theta_t \Gamma_t) \, dt,$$

where I used results from the proof of Proposition 2.3. By the Girsanov Theorem, there exists a process $\{\widetilde{Y}_t\}$ such that $\tilde{Z}_t \equiv Z_t + \int_0^t \tilde{Y}_s \, ds$ is a Brownian motion under the probability measure induced by a deviation; note that $dZ_t - \tilde{dZ}_t = Y_t \, dt \equiv -\tilde{Y}_t \, dt > 0$ since agents cannot overreport.
Given these variables, I construct a candidate upper bound for an agent’s deviation value:

$$\hat{v}(\varepsilon_t, \Phi_t) \equiv v_t + \Phi_t \varepsilon_t + L \varepsilon_t^2,$$

where $L > 0$ is a constant to be specified later. If $\hat{v}(\varepsilon_t, \Phi_t)$ is indeed an upper bound, then for an agent who has not yet deviated with $\varepsilon_t = \Phi_t = 0$, the upper bound of his deviation value is just $v_t$.

I will now prove that $\hat{v}(\varepsilon_t, \Phi_t)$ is indeed an upper bound on an agent’s deviation utility. To simplify notation, I will drop tildes on the utility function so that $u_t \equiv \tilde{u}(\theta_t | \theta_t)$ and $\hat{u}_t \equiv \tilde{u}(\hat{\theta}_t | \theta_t)$. Define the auxiliary deflated gains process $\hat{G}_t$ associated with any feasible report $\{\hat{\theta}_t\}$:

$$\hat{G}_t \equiv \int_0^{t \wedge \tau} e^{-\rho s} \hat{u}_s \, ds + e^{-\rho (t \wedge \tau)} \hat{v}(\varepsilon_{t \wedge \tau}, \Phi_{t \wedge \tau}).$$

To simplify notation, I will omit the retirement time $\tau$. Clearly,

$$\mathbb{E}_0 \left[ \hat{G}_\infty \right] = \mathbb{E}_0 \left[ \int_0^\infty e^{-\rho t} \hat{u}_t \, dt \right]$$

is the expected payoff under any feasible reporting strategy,\(^{25}\) given the transversality condition in Proposition 2.1. On the other hand, $\hat{G}_0 = \hat{v}(\varepsilon_0, \Phi_0)$ is the proposed upper bound of an agent’s deviation value given the current relevant deviation states $(\varepsilon_0, \Phi_0)$. For the upper bound to be valid, I need $\hat{G}_t$ to be a supermartingale for any deviation strategy from the agent’s perspective, and a martingale under truth-telling.

Differentiating $\hat{G}_t$ with respect to $t$,

$$e^{\rho t} d\hat{G}_t = \hat{u}_t \, dt - \rho \hat{v}_t \, dt + d\hat{v}_t.$$

\(^{25}\)Throughout the proof, $\mathbb{E}$ is the expectation under the measure induced by the deviation strategy.
Applying Itô's Lemma and taking conditional expectations to capture only the drift,

\[
e^{\rho t} \mathbb{E}_t \left[ \frac{d\tilde{G}_t}{dt} \right] = \tilde{u}_t - \rho (v_t + \Phi_t \varepsilon_t + L \varepsilon_t^2) + \rho v_t - u_t + d(\Phi_t \varepsilon_t) + L \mathbb{E}_t \left[ d\varepsilon_t^2 \right]
\]

\[
= \tilde{u}_t - u_t - \rho \Phi_t \varepsilon_t - \rho L \varepsilon_t^2
+ e^{-\rho t} \varepsilon_t \left[ (u_{\theta,t} - \rho \Delta_t + \mu_t^\Delta + \mu'(\theta_t) \Delta_t + \sigma_t^\Delta \Gamma_t) \varepsilon_t + \sigma_t \theta_t \Delta_t Y_t \right]
+ e^{-\rho t} \sigma_t^2 (1 + \theta_t \Gamma_t) \varepsilon_t^2 + 2 L \mu'(\theta_t) \varepsilon_t^2 + 2 L \sigma_t \theta_t Y_t \varepsilon_t + L \sigma_t^2 \varepsilon_t^2
\]

\[
= \tilde{u}_t - u_t + \sigma_t \theta_t Y_t \Phi_t + \varepsilon_t \left( \sigma_t \theta_t Y_t \left( e^{-\rho t} \Delta_t + 2L \right) + (\mu'(\theta_t) - \rho) \Phi_t \right)
+ \varepsilon_t^2 \left( e^{-\rho t} \sigma_t^2 (1 + \theta_t \Gamma_t) + L (2 \mu'(\theta_t) + \sigma_t^2) \right),
\]

where the last step follows from Proposition 2.3: under the first-order approach,

\[
\mu_t^\Delta = (\rho - \mu'(\theta_t)) \Delta_t - u_{\theta,t} - \sigma_t^2 \theta_t \Gamma_t.
\]

Now, since \( \tilde{\theta}_t \leq \theta_t \), then \( \varepsilon_t < 0 \) and \( \Delta_t, Y_t \geq 0 \) so it must be that \( \sigma_t \theta_t Y_t \Phi_t < 0 \), too.

Using the assumption that \( \tilde{u}_{x_t} \leq 0 \) implies \( \tilde{u}_{x_t,\theta} \geq 0 \), the definition of \( \Phi_t \), and taking a first-order Taylor expansion around \( \varepsilon_t = 0 \), the first three terms can be written as

\[
\sigma_t \theta_t Y_t \int_0^t e^{-\rho s} u_{\theta,s} \varepsilon_s \, ds + (e^{-\rho t} \sigma_t \theta_t \Delta_t + u_{\theta,t}) \varepsilon_t + O \left( \varepsilon_t^2 \right).
\]

The first two terms are negative so it follows that

\[
e^{\rho t} \mathbb{E}_t \left[ \frac{d\tilde{G}_t}{dt} \right] < \varepsilon_t \left( \sigma_t \theta_t Y_t \left( e^{-\rho t} \Delta_t + 2L \right) + (\mu'(\theta_t) - \rho) \Phi_t \right)
+ \varepsilon_t^2 \left( e^{-\rho t} \sigma_t^2 (1 + \theta_t \Gamma_t) + L (2 \mu'(\theta_t) + \sigma_t^2) \right) + O \left( \varepsilon_t^2 \right)
\]

\[
< \varepsilon_t \Phi_t (\mu'(\theta_t) - \rho) + \varepsilon_t^2 \left( e^{-\rho t} \sigma_t^2 (1 + \theta_t \Gamma_t) + L (2 \mu'(\theta_t) + \sigma_t^2) \right) + O \left( \varepsilon_t^2 \right)
\]

\[
< \varepsilon_t^2 \left( e^{-\rho t} \sigma_t^2 (1 + \theta_t \Gamma_t) + L (2 \mu'(\theta_t) + \sigma_t^2) \right) + O \left( \varepsilon_t^2 \right).
\]
The second inequality uses that $\varepsilon_t \leq 0 \leq Y_t$ and the third inequality uses that $\mu'(\theta_t) < \rho$ (from Assumption 2.1) and $\varepsilon_t, \Phi_t \leq 0$. Now, since $\theta_t \in \Theta_t = [\underline{\theta}, \bar{\theta}] \subseteq \mathbb{R}_{++}$ this means that $\theta_t \Gamma_t$ is bounded above and below. Since $2\mu'(\theta_t) + \sigma_\theta^2 < \rho$ by Assumption 2.1, by choosing $L$ sufficiently large, I can guarantee that the coefficient on $\varepsilon_t^2$ is negative. Therefore,

$$e^{\rho_t} \mathbb{E}_t \left[ \frac{d\hat{G}_t}{dt} \right] < -\beta_t \varepsilon_t^2$$

for some time-dependent coefficient $\beta_t > 0$. This parabola is always nonpositive and strictly negative if $\varepsilon_t \neq 0$. Therefore, $\hat{G}_t$ has strictly negative drift for any deviation strategy, and zero drift under truth-telling. To ensure that $\hat{G}_t$ is a supermartingale, I need to ensure that all of the Brownian and Poisson terms have zero expectation on any time horizon (see Revuz and Yor (1999)). But since everything is bounded, this necessarily holds on any finite time horizon.

Now, given the fact that $\hat{G}_t$ is a supermartingale,

$$\hat{v}(\varepsilon_0, \Phi_0) = \hat{G}_0 \geq \mathbb{E}_0 \left[ \lim_{t \to \infty} \hat{G}_t \right] = \mathbb{E}_0 \left[ \lim_{t \to \infty} \int_0^{t \land \tau} e^{-\rho s} \hat{u}_s \, ds + \lim_{t \to \infty} e^{-\rho(t \land \tau)} \psi_{t \land \tau} \right]$$

$$= \mathbb{E}_0 \left[ \lim_{t \to \infty} \int_0^{t \land \tau} e^{-\rho s} \hat{u}_s \, ds \right],$$

which is an agent’s deviation payoff. Here, the last equality uses the transversality condition on $v_t$ in Proposition 2.1. This implies that $\hat{v}_t$ is indeed an upper bound for the agent’s deviation value, as claimed.

I have thus showed that $\hat{v}(\varepsilon_t, \Phi_t)$ is an upper bound for the agent’s potential deviation value given the deviation states $(\varepsilon_t, \Phi_t)$. Then, for an agent who has not yet deviated with $\varepsilon_t = \Phi_t = 0$, the upper bound of his deviation value is just $v_t$. Because the equilibrium strategy achieves this upper bound, the equilibrium strategy is indeed globally optimal. Thus, the equilibrium strategy that achieves $v_t$ is optimal.
Chapter 3

Securitization, Lending Standards, and Credit Expansion

3.1 Introduction

The period before the financial crisis of 2007–2009 was characterized by a housing bubble across the United States. Notably, households that were previously too risky to receive mortgages saw unprecedented credit expansion and house price growth during the early- and mid-2000s. At the same time, the shadow banking sector—financial institutions that behave like traditional banks but are not subject to the same regulations—grew larger than the traditional banking sector. I build a model that links the growth of the shadow banking sector to the expansion of credit to previously unserved regions of the housing market.

Pozsar, Adrian, Ashcraft, and Boesky (2010) define shadow banking as any of a variety of credit, maturity, and liquidity transformation activities conducted by institutions similar to commercial banks, but without access to public sector guarantees. While traditional, regulated commercial banks relied on deposits from households to fund their lending activity, since shadow banks were unprotected by these regulations,
they had to find other ways to fund their activities. This process of finding alternative ways to operate outside the purview of regulation is called regulatory arbitrage. One example is that shadow banks funded their operations via securitization. Securitization is the process by which individual loans (or other assets) are pooled together, packaged into a security, and this security is then sold to investors. The investor is now entitled to the payments from the original assets while the originator receives a fixed price.\(^1\)

This paper is motivated by three stylized facts about the housing and financial sectors prior to the crisis: the unprecedented rise of the non-agency mortgage-backed security (MBS) market, higher mortgage default rates among securitized mortgages, and a negative correlation between income growth and credit expansion, and house prices. As I explain below, the second fact is unusual from a theoretical point of view, while the third is an empirical anomaly. However, my model, which is based on regulatory arbitrage, shows how the first fact explains the second and third facts.

**Fact 1.** *The non-agency MBS market greatly expanded in the mid-2000s and subsequently crashed at the onset of the financial crisis.*

Banks can make mortgages and then sell them to government-sponsored enterprises (GSEs) such as Fannie Mae and Freddie Mac. MBSs that are guaranteed by these GSEs are called agency securities. The mortgages that go into agency securities, because they are guaranteed by GSEs, have to meet certain requirements for risk (they cannot be too risky) and size (they cannot be too large). Because agency securities are considered very safe, they are often bought by large institutional investors that are required to hold safe assets.

\(^1\)There are many ways to structure securities. The most common, and what I will use in the model, is a “pass-through security,” which simply transfers the payments from the underlying asset to the owner of the security. More complicated securities, such as collateralized debt obligations (CDOs), “tranche” the underlying mortgages in that the CDO pays out only if enough mortgages are repaid.
MBSs created from loans that do not meet the requirements are called non-agency securities. Since non-agency securities are potentially very risky, they are typically bought by hedge funds and other unconstrained investors in the shadow banking sector. The growth of the non-agency MBS market signals that more risky loans were being made, and there was an increased appetite for securities backed by these risky loans. As Figure 3.1 shows,\(^2\) the non-agency MBS market closely tracked the housing market. The top panel shows that the value of the non-agency MBS market peaked at around the same time that the housing market peaked and declined when the housing market declined, while the bottom panel shows that more and more new mortgages were being funded by the sale of securitized old mortgages. That is, banks were making loans, selling them, and using the proceeds to fund new mortgages.

Figure 3.1: The rise and fall of the non-agency MBS market.

**Fact 2.** Securitized mortgages defaulted more often than those that were retained by the originating bank.

\(^2\)Non-agency MBS data is from Inside Mortgage Finance and the Federal Reserve Flow of Funds, and Case-Shiller data is from the Federal Reserve Bank of St. Louis FRED Economic Data.
A curious fact about the recent crisis, documented by (among others) Mian and Sufi (2009) and Keys, Mukherjee, Seru, and Vig (2010), is that loans that were easier to securitize, and loans in regions of the U.S. that experienced an increase in the fraction sold, defaulted far more often than loans that were more difficult to securitize or from regions that did not see a large increase in securitization.

From a theoretical perspective, this is a bit perplexing for two reasons. First, if agents can write complete contracts, then any written contract should account for potential moral hazard or lemons problems. Second, even if agents cannot write complete contracts, they should be able to internalize that only the worst loans will be sold, and the market for MBSs should collapse. But not only did the market not collapse, it grew while the quality of the underlying loans declined.

One common explanation, which is the route pursued by Chari, Shourideh, and Zetlin-Jones (2014) (see below), is that there was adverse selection in the securitization market. However, the evidence for this theory is, at best, mixed: while Elul (2015) argues that the worse performance of privately-securitized loans relative to non-privately-securitized loans is consistent with adverse selection, Keys, Mukherjee, Seru, and Vig (2010) and Agarwal, Chang, and Yavas (2012) find that there was not a clear pattern of adverse selection in the subprime mortgage market. Another theory is that MBS buyers were unsophisticated and ripped off by the originating banks. Foote, Gerardi, and Willen (2012) declare this an outright myth: they find that MBS investors in fact had lots of information about the products they were buying and mortgage market insiders were actually among the biggest losers, while market outsiders were the biggest winners. In my model regulatory arbitrage is the key to explaining this fact because being regulated creates skin in the game, and therefore an incentive to originate high-quality mortgages.
**Fact 3.** From 2002–2005, household income growth was negatively correlated with credit expansion and house price growth; in every other period this correlation is positive.

Mian and Sufi (2009) document that in every period over the last few decades, except for the middle of the housing bubble, income growth was positively correlated with credit expansion and house price growth. This makes sense, since regions that are becoming more wealthy should see property values rise and more people taking out mortgages to buy homes. The opposite occurred during the midst of the bubble, 2002–2005: house prices and credit expansion increased most in subprime regions with negative relative income growth. In other words, during the bubble, low-growth areas that were previously the most likely to be denied credit actually received an unprecedented expansion of credit. In addition, after 2006 these correlations went back to being positive. Figure 3.2 (taken from Mian and Sufi (2009)), particularly the right-hand panel, shows that income growth and credit growth are statistically significantly negatively correlated from 2002 to 2005.

![Figure 3.2: The relationship between income growth and credit growth.](image)

In my model, wealth-constrained households apply for mortgages from intermediaries. Households are prime or subprime, and this is private information. Intermediaries can accept a mortgage application and extend a mortgage, or decline the application. I solve for the optimal lending contract, including approve/deny decisions. Intermediaries fund the mortgages using a combination of their own internal
equity and funds from outsiders; by borrowing heavily from outsiders, intermediaries increase leverage and make more loans.

There is a unit measure of intermediaries, each of whom can choose to operate as a commercial bank or a shadow bank; both try to raise funds from outsiders. Crucially, intermediaries cannot commit to not steal outsiders’ money: an intermediary can borrow from outsiders, make mortgages, but keep all of the returns for itself instead of repaying the outsiders. However, commercial banks pay a regulatory cost that allows them to commit to repaying, which in turn allows them to raise funds via non-contingent deposits. It also means that they retain the mortgages on their balance sheets as skin in the game. This is the source of regulatory arbitrage: since shadow banks are not bound by regulation, they cannot commit to repaying so they cannot issue deposits. Instead, they sell the loans to outsiders, absolving them of any responsibilities (and skin in the game) since the outsiders now own the loans and the associated cash flows. I call this process securitization and shadow banks use the proceeds from securitization to extend new mortgages.

Although a household’s type is private information, intermediaries can pay a screening cost to learn it. Keys, Mukherjee, Seru, and Vig (2010) document that this screening decision had important consequences: mortgages associated with credit scores below 620 are more difficult to securitize because of a regulation so these mortgages were more likely to be retained. As a result, they show that these mortgages were screened more. Conversely, mortgages just above the threshold were more likely to be securitized and thus less likely to be screened. This created a discrete, upward jump in the default rate at the 620-threshold. In the model, I show that commercial banks screen out subprime borrowers because they retain the mortgages while shadow banks, being able to sell them, do not screen. This explains Fact 2 that securitized

---

3See Hart and Moore (1998) for details of the theory behind this friction and Gertler and Kiyotaki (2010) for a model that uses this friction to generate a credit constraint.
mortgages had a higher default rates and also reproduces the jump in default rate at the point where commercial banks give rise to shadow banks.

In the baseline model, the house price is exogenous. I show that the conditions of the mid-2000s and the housing bubble, namely low interest rates and high price expectations, contribute to the growth of the shadow banking sector. In particular, I show that shadow banking is more likely to displace commercial banking and that securitization, leverage, and credit expansion all rise, in line with Fact 1. I also show that in segments where shadow banks operate, credit expansion is greatest in low-income segments, which is the first half of Fact 3, and that high prices today combined with high expectations for the future further fuels expansion (but high prices today without high expectations are not enough to generate this result). The opposite, traditional relationship holds where commercial banks are active. Simple welfare analysis shows that shadow banking can be efficient or inefficient depending on the parameters: there is a welfare gain from previously unserved households now owning houses but a loss from the higher default rate.

I extend the model to include many types of households and I endogenize the house price and in doing so, obtain several new insights. First, I show that, consistent with the second part of Fact 3, the price is highest in low-income segments where shadow banks are active. In fact, rising prices can draw new wealth into the shadow banking sector, which further boosts subprime credit expansion and raises prices even more. Second, I show that there exists a region with multiple equilibria. While commercial banks are always active in this region, they may or may not coexist with shadow banks. Small changes to parameters can move the economy from a region with a unique equilibrium to the multiplicity region (or vice-versa). It also means that small changes can cause the shadow banking sector to suddenly collapse, and prices and welfare can jump up or down. These large jumps in response to small changes to parameters indicate that the financial system with shadow banks is, in a sense,
unstable. Further, a larger shadow banking sector makes the system more unstable by enlarging the region with multiple equilibria.

3.2 Related Literature

This paper fits into the burgeoning literature on shadow banking and the literature on optimal contracting in banking. There are three broad, complementary perspectives on shadow banking: regulatory arbitrage, neglected risk, and liquidity transformation; this paper fits into the first one. Gorton and Metrick (2010) and Acharya, Schnabl, and Suarez (2013) argue that the rise of shadow banking can be attributed largely to regulatory arbitrage, especially to changes that made securitization and off-balance sheet financing easier.

This paper complements other recent work that stresses regulatory arbitrage. A common theme in this literature is that overly-stringent regulation of the traditional banking sector simply makes the shadow banking sector larger and more unstable. In Huang (2016), traditional banks have to pay a regulatory tax to raise deposits so by establishing off-balance sheet shadow banks that are not subject to the tax, they acquire a cheaper but more unstable funding source. As a result, tighter regulation is not necessarily a good thing since it simply pushes more activity to the shadow banking sector. However, unlike in my model, there is no prime/subprime distinction. Plantin (2014) develops a simple model in which the government cannot monitor the asset markets in which shadow banks operate, creating an incentive problem because traditional banks can push more activity into the unregulated market. He finds that the optimal capital requirement is lower than it otherwise would be if there was perfect monitoring. My model stresses the different methods of financing investment and the quality of investment funded. In Ordoñez (2015), shadow banks’ stability is tied to their reputation, and capital requirements plus cross subsidization of banks
can stabilize the system. As in all of these models, in my model securitization-based shadow banking is potentially unstable. Unlike these models, my model rationalizes shadow banking as a product of both regulatory arbitrage and asymmetric information between borrowers and lenders.

The second perspective on shadow banking emphasizes neglected risk. Shleifer and Vishny (2010) develop a stylized model to show how asset mispricing can lead to large swings in bank health and real investment; however, this model is silent on the sources of mispricing and why the banking sector looks the way it does. In Gennaioli, Shleifer, and Vishny (2013), if agents underestimate the probability of a recession, then banks over-invest in risky projects, take on too much leverage, and over-expose the economy to aggregate risk. Interestingly, this is not the case and shadow banking is efficient if agents have rational expectations. My model does not take a stand on whether or not beliefs are correct.

The third perspective stresses the fact that shadow banks funded themselves by creating “safe,” money-like securities from illiquid and risky investments. In Hanson and Sunderam (2013) and Sunderam (2015), shadow banks emerge in response to increased demand for “money-like” claims, and over-issuance exposes the system to instability. In Moreira and Savov (2016), intermediaries fund themselves with securities that are money-like in normal times but become illiquid in crisis times. Gennaioli, Shleifer, and Vishny (2012) combine the second and third perspectives, showing that when intermediaries engineer securities that cater to the demand for safe assets while neglecting crash risk, the financial system is fragile.

I follow the literature on optimal lending contracts in banking. Diamond (1984) and Greenbaum and Thakor (1987) are early examples of models that derive the optimal lending contract when there is asymmetric information between borrowers and lenders. The former emphasizes that if lenders do not monitor the loans they make, then there should be a lemons problem while the latter rationalizes deposit-
vs. securitization-driven financing. Following Wang and Williamson (1998), in my model intermediaries can pay a screening cost to learn the quality of their borrowers; however, as I explain below, the structure of my game is different from theirs, leading to a different equilibrium outcome. In Fender and Mitchell (2009) and Kiff and Kissner (2010), an intermediary’s screening intensity determines the quality of the mortgage pool, but there is no emphasis on the distinction between commercial and shadow banking, and they do not derive optimal contracts. I draw on Holmström and Tirole (1997, 1998) and Shleifer and Vishny (2010) to model shadow banks’ leverage capacity.\footnote{All of the models above have only two or three periods, which makes deriving the optimal contract significantly easier. The alternative is a fully dynamic general equilibrium model, but these must be solved numerically and even then, they cannot be too complex. Landvoigt (2014) and Chen, Michaux, and Roussanov (2014) are examples of the latter: they do not derive the optimal contracts, house prices are exogenous, they ignore the distinction between traditional and shadow banks, and there are no decisions about which types of households to lend to.}

Chari, Shourideh, and Zetlin-Jones (2014) explain fluctuations in the securitization market with a model of adverse selection and reputation. In their model, lenders signal their quality via the securitization contracts they post and securitization volume is driven by the originator’s reputation: when reputation is high and buyers believe that the loans are high-quality, originators can securitize more loans. Securitized loans are lower quality than retained loans because retaining loans signals to the market that they are high-quality. However, their model takes as given that some lenders have higher quality balance sheets than others while in my model, balance sheet quality is a product of the contracting environment. Also, their model ignores credit expansion. Justiniano, Primiceri, and Tambalotti (2016) build a simple dynamic model to explain Fact 3 but are silent on Facts 1 and 2.

On the empirical front, Keys, Seru, and Vig (2012) explore the 620 credit score threshold and find that loans just above the threshold could be securitized more quickly than those just below, and the former defaulted about 20% more often. Dell’Ariccia, Igan, and Laeven (2012) empirically link lending standards and delin-
quency rates in subprime markets to their rapid expansion, finding that delinquency rates rose more sharply in those areas that saw larger increases in origination. Jiang, Nelson, and Vytlacil (2013) distinguish between the ex ante and ex post probability that a loan will be sold. They find that loans with a higher ex ante probability of being sold had higher delinquency rates, but from that pool, those that were actually sold had a lower delinquency rate. They conclude that the prospect of unloading delinquency risk weakened the incentive to screen. Purnanandam (2011) finds that banks that were primarily funded by non-demandable or market-based wholesale debt were the main originators of low-quality loans, consistent with my model.

The remainder of the paper is organized as follows: Section 3.3 lays out the baseline model in which the house price is exogenous in order to highlight the main results, Section 3.4 solves for the equilibrium, Section 3.5 characterizes some of the properties of the equilibrium, including credit growth, Section 3.6 extends the model by endogenizing the house price and adding more household heterogeneity, and Section 3.7 concludes and offers directions for future research.

3.3 Baseline Model

The economy has two dates, $t = 0, 1$, and a continuum of three classes of agents: households, intermediaries, and savers/investors. Households apply for mortgages from intermediaries to buy a house, while intermediaries raise funds to make these mortgages from savers/investors. There are two types of goods: a consumption good, which is the numéraire, and housing. There are three sources of asymmetric information in the economy: an adverse selection problem between households and intermediaries, and two moral hazard problems between intermediaries and savers/investors. A collection of agents and a housing market define a “segment” of the market. Figure 3.3 depicts
a single segment in the economy, each agent’s balance sheet, and which assets change hands. Unless otherwise noted, all parameter values are common knowledge.

Figure 3.3: A schematic representation of the economy.

### 3.3.1 Households

There is a large measure of households in the economy, each of whom wishes to buy exactly one house at $t = 0$. Letting $c_t$ denote consumption, a household maximizes\(^5\)

\[
U^H = \mathbb{E} \left[ c_0 + h \mathbf{1}_{H_0} + c_1 + h \mathbf{1}_{H_1} \right]. \tag{3.1}
\]

Here, $h$ is a household’s utility (in units of the numéraire) from living in a house and $1_{H_t}$ is an indicator function that is one at date $t$ if the household is living in a house at date $t$. I impose limited liability in each period, $c_t \geq 0$.\(^6\) Each household is endowed with $w > 0$ units of the consumption good at $t = 0$, and can either consume

\(^5\)This specification is different from what is usually used in the housing literature. For example, Landvoigt (2014) and Gao, Sockin, and Xiong (2015) use Cobb-Douglas preferences over consumption and housing and housing is divisible. However, for most of the population, the assumption that households buy a single house is realistic, and using Cobb-Douglas preferences here would not change any of the insights, though it would make the algebra considerably more complicated.

\(^6\)One interpretation is that households have some baseline level of consumption, here normalized to zero, and households cannot consume below the baseline level.
this endowment or use it to buy a house within its segment. At $t = 1$, each household receives income $y_H$ with probability $\rho^i \in [0, 1]$, or $y_L$ with probability $1 - \rho^i$, where $y_L < y_H$. This income can be consumed or used to pay down any debt accrued at $t = 0$. The income realizations are independent across households, so by the law of large numbers, there is no aggregate uncertainty.

There are two types of households in the economy, “prime” and “subprime.” Prime households receive $y_H$ with probability $\rho^p$ while subprime households receive $y_H$ with probability $\rho^s$, where $\rho^s < \rho^p$. Thus the probability of receiving a high income is the only dimension along which agents in a segment differ. Prime borrowers make up a fraction $\pi$ of the population and subprime borrowers make up the remaining $1 - \pi$ fraction, and by the law of large numbers, the total fraction of households that receive $y_H$ is $\pi \rho^p + (1 - \pi) \rho^s$. While this two-type assumption is admittedly extreme, it allows me to highlight the main channels of the model; in later sections, I relax this assumption so that there are many types of households.

Let $p_0$ denote the price of a house at $t = 0$ and assume this price exceeds a household’s initial wealth, $p_0 > w$. This ensures a household must take out a mortgage from an intermediary to buy a house; otherwise, the problem is not interesting. Households take this price as given. Because households have linear utility and do not discount future consumption, households use all of their available wealth as down-payments on the houses so that $c_0 = 0$ and households borrow $p_0 - w$. A household may apply for at most one mortgage. If the household takes out a mortgage, then as per (3.1), it lives in the house at $t = 0$ and receives utility $h$. At $t = 1$, the household must repay the mortgage. Since income at $t = 1$ is random, the household might not have enough income to do so, in which case it defaults. Of course, even if the household receives high income, it might choose to strategically default. When I describe intermediaries, I will discuss what happens if a household defaults; I will also derive the relevant

---

7Later, I will derive a household’s participation constraint ensuring that it actually takes out a mortgage instead of remaining in autarky.
incentive constraint to ensure that a household will always repay its debt when it can.

Most importantly, an individual household’s type is private information.\(^8\) This is the first source of asymmetric information and pins down the form of the optimal mortgage contract. As mentioned above, however, all parameters are common knowledge, so this can be interpreted as “hard information” that intermediaries readily have access to. That is, an intermediary knows the overall quality of the segment in which it operates, but not the quality of each individual household (so-called “soft information”).

As a final note, since I am not modeling belief disagreements or excessive optimism/pessimism, it does not matter whether the probabilities correspond to true probabilities or perceived probabilities, as long as all agents agree. Thus, increasing \(\rho^S\), for example, could correspond to actually increasing the probability that subprime households will be able to repay or merely increasing the economy-wide perception that they will be able to repay.

### 3.3.2 Savers/Investors

Savers/investors supply funds to intermediaries; this funding source is perfectly elastic. Savers/investors lack lending expertise so to invest in mortgages, they must do so indirectly by lending to intermediaries. A saver will not fund an intermediary unless he is promised a riskless return, \(1 + r^f \geq 1\). As I describe below, this demand will be met by commercial banks. However, because all agents are risk neutral, all I need is for the expected return be no less than the riskfree rate, and the actual state-contingent payoffs are irrelevant. Indeed, letting \(x = (x_L, x_H)\) denote a saver’s

\(^8\)This is why all households within a segment have the same initial wealth. If prime and subprime borrowers differed in \(w\), then intermediaries would immediately learn their types based on the size of the loan they applied for. This assumption could be relaxed by allowing for heterogeneity in house prices within a segment (making the algebra more cumbersome without adding much insight) or further dividing prime and subprime households into wealthy and poor (adding more constraints to the problem without adding much insight).
state-contingent payoff, as long as \((1 + r_f) L = \mathbb{E}[x]\), where \(L\) is the amount lent, the saver gets the same utility in each case.

Investors have access to some outside option (the “market”) with gross return 
\(1 + r \geq 1 + r_f\), and so will lend to an intermediary as long as the expected return on the loan weakly dominates the market’s return; investors will fund shadow banks. To keep the algebra simple, I set \(r_f = r.\)\(^9\) This outside rate of return is exogenous.

### 3.3.3 Intermediaries

There exists a unit measure of ex ante identical intermediaries. An intermediary can operate in one of two ways, as a regulated commercial bank or an unregulated shadow bank; which one is determined in equilibrium.\(^{10}\) Commercial banks can be thought of as savings and loans institutions, community banks, and other commercial banks while shadow banks are big banks with both commercial and investment banking arms (Citigroup, J.P. Morgan,...) and mortgage brokers (Countrywide,...). The true shadow banking system was far more complicated than this, with commercial banks and mortgage brokers making the mortgages, investment banks securitizing them via special purpose vehicles, and asset backed commercial paper conduits buying the securities with money borrowed from mutual funds (and several of these steps had sub-steps). Collapsing these players into a single agent, the shadow bank, keeps the analysis simple while focusing on the key institutional details.

Each intermediary operates in a competitive market and is endowed with initial equity \(n_0 > 0\). Let \(n_{0}^{\text{CB}}\) and \(n_{0}^{\text{SB}}\) denote total commercial and shadow banking net worth, respectively. Intermediaries combine their own equity with funds raised from savers/investors to make mortgages to households. This allows them to use lever-

\(^9\)Equivalently, there is a single group of outsiders, each outsider has access to the market, and an outsider will lend to any intermediary as long as he is promised an expected return that dominates the market’s return.

\(^{10}\)In what follows, I will use “operating” and “entering” interchangeably when referring to an intermediary choosing a particular mode of banking.
age to boost the number of mortgages they make. Let $c_t$ denote an intermediary’s consumption at the end of date $t$; then each intermediary maximizes

$$U^i = \mathbb{E} [c_0 + c_1],$$  

(3.2)

where consumption comes from any profits from securitization and returns from lending.

As already stated, intermediaries cannot observe a household’s type when it applies for a mortgage. However, intermediaries have access to a screening technology that, if used, perfectly reveals a household’s type. If an intermediary chooses to screen a household, it incurs a non-pecuniary cost (in units of the $t = 1$ consumption good) $\gamma_s > 0$.\footnote{In Fender and Mitchell (2009) and Kiff and Kisser (2010), intermediaries choose the screening intensity and the loan pool’s quality is continuously increasing in this intensity. This implies that the overall default probability of \textit{originated} mortgages is continuously decreasing in credit quality, a prediction that is rejected in the data (see Keys, Mukherjee, Seru, and Vig (2010)).} This cost can be thought of as capturing the time and effort costs of reviewing a household’s credit report, pay stubs... If an intermediary screens, then that household also incurs a non-pecuniary application cost, $\gamma_a > 0$. This can be thought of as capturing the time and effort cost of putting together the necessary paperwork. An intermediary can screen as many or as few households as it wants to, though it must pay the screening cost for each household it screens. Once the intermediary does (or does not observe) a household’s type, it chooses whether or not to extend a mortgage. An intermediary’s screening decision is unobservable to everyone except the intermediary and household. Let $\sigma^i \in \{0, 1\}$ denote intermediary $i$’s screening decision, i.e., $\sigma^i = 0$ if intermediary $i$ does not screen and $\sigma^i = 1$ if it does. An intermediary’s screening decision will be determined in equilibrium.

Mortgage contracts are non-recourse loans secured by the house. Once an intermediary makes a loan, there are two possibilities at $t = 1$: the household repays or it defaults. If it repays, the intermediary receives whatever payment was specified...
in the mortgage contract. If the household defaults, then since the loan is secured
by the house, the intermediary forecloses on the house and seizes it; since the loan
is non-recourse, the intermediary cannot seize any of the household’s other assets.\textsuperscript{12}
By the households’ incentive constraint, households will repay precisely when they
receive high income and default otherwise.

Let \((f_{s'}, D_{s'}, d, \sigma)\) denote a generic mortgage contract; I restrict attention to deter-
ministic mechanisms. Here, \(f_{s'}\) is an intermediary’s foreclosure decision if a household
reports income \(y_{s'} \in \{y_L, y_H\}\) (in particular, \(f_{s'} = 0\) if a household reports income \(y_{s'}\)
and the intermediary forecloses on the house, and \(f_{s'} = 1\) if the intermediary does not
foreclose on the house), \(D_{s'}\) is the payment from a household to an intermediary when
the household reports income \(y_{s'}\), \(d\) is the down-payment on the mortgage (which I al-
ready normalized to \(w\)), and \(\sigma\) is the screening decision. Since I will prove that in the
optimal mortgage contract, when a household repays, its payment is not contingent
on anything else that occurs at \(t = 1\) (for example, how much wealth the household
has), from now on, I will write \((D, \sigma)\), where \(D\) is what the household owes at \(t = 1\).
Although this is not a simple debt contract in the true sense (since the payment in
the default state is the house instead of the borrower’s full income), I will still refer
to the mortgage as a debt contract because the payment in the no-default state is
non-contingent. Intermediaries can commit to the contracts they post, so that, for
example, an intermediary that posts a contract with \(\sigma = 1\) will in fact always screen
the applicant.

The second source of asymmetric information in the model is between interme-
diaries and savers/investors. Specifically, once an intermediary makes a mortgage
and collects the payment, it can run away with the funds without repaying the
savers/investors. At this point, no one would lend to an intermediary because its

\textsuperscript{12}According to Ellis (2008), while “on paper” the U.S. mortgage system resembles the U.K. system,
in which lenders can come after borrower income as well as the house, in practice this is uncommon
because the process is time-consuming and expensive. Thus, assuming non-recourse mortgages is a
reasonable approximation to what actually appears in the U.S.
promise to repay is not credible. Therefore, an intermediary’s lending capacity is completely determined by its own internal equity. This is the same friction used by Gertler and Kiyotaki (2010). However, this friction serves a different purpose here than it does in that paper: in that paper, it puts an upper bound on an intermediary’s leverage while here, it puts restrictions on how intermediaries fund themselves. In addition, a saver/investor might want an intermediary to screen out subprime borrowers (to improve the quality of the mortgage pool, and hence payment stream) but since screening decisions are not observable (and hence cannot be contracted on) and an intermediary can run away with funds, there is a moral hazard problem between intermediaries and savers/investors.

The key distinction between commercial banks and shadow banks is that commercial banks are regulated while shadow banks are not. Formally, for each unit a commercial bank borrows from savers, which I will call a “deposit,” it must pay a fraction $\tau \in (0, 1)$ to the government (that is, it pays $\tau L$ on $L$ assets). This proxies regulatory costs. In exchange, the commercial bank can credibly promise to not divert funds so that when a saver deposits money in a commercial bank, the saver knows the bank will pay him back with interest.\footnote{In Huang (2016), “traditional banks” have to pay a tax on raised funds while shadow banks do not. However, in his model, the tax serves no other purpose and is immediately redistributed back while here, the tax revenue is not returned to intermediaries and the tax solves an asymmetric information problem.} Another consequence is that since commercial banks are on the hook for a fixed payment, they have stronger incentives to screen households, potentially alleviating the moral hazard problem. Nevertheless, now, commercial banks can augment their internal equity with outside funds to expand credit. As I will show, a commercial bank’s leverage is closely tied to the regulatory cost.

On the other hand, shadow banks are not subject to this costly regulation. Since promises to repay are not credible, savers will not lend to shadow banks so any borrowed funds must come from investors. It is clear that if shadow banks raise
money from investors via some contract, this contract cannot specify any payments at $t = 1$ because the shadow bank can simply run away with the funds. Therefore, all transfers/payments must occur at $t = 0$. In particular, once a shadow bank makes a loan, it can immediately sell it to an investor; I call this process “securitization” and the loan a “mortgage-backed security” (MBS) once it is sold. These MBSs, which are not backed by any government guarantees and are issued by private shadow banks, therefore represent non-agency securities. Fluctuations in securitization volume thus represent fluctuations in the non-agency MBS market. Now, the shadow bank receives a payment from the investor. Since the investor holds the loan, he is entitled to its payments at $t = 1$ without having to worry about the shadow banking diverting them. Crucially, since shadow banks are not subject to regulation, they have a lower cost of funding than do commercial banks. However, because shadow banks might not retain all of their loans, they might not have an incentive to screen, creating a misalignment in incentives between shadow banks and investors. Following Shleifer and Vishny (2010), shadow banks earn profits from a securitization fee $f \geq 0$, which will be determined in equilibrium.

Finally, intermediaries cannot costlessly monitor households to ensure repayment. Following Holmström and Tirole (1997, 1998), if an intermediary monitors a particular household at non-pecuniary cost $m > 0$, the household’s probability of receiving high income is its true probability $\rho^i$, while if the intermediary does not monitor, the household “shirks” and receives low income, and hence defaults, with probability one.$^{14}$ While commercial banks already have skin in the game and hence are willing to monitor, now, shadow banks, too, must retain some fraction of the loans they make as skin in the game, and this fraction depends on economic fundamentals. As a result, shadow bank leverage and mortgage volume are explicitly tied to fundamentals. When banks put together an MBS, I assume they randomly sample from the entire

---

$^{14}$This stark assumption is for algebraic simplicity; as long as the default probability is higher without monitoring, the analysis goes through.
pool, and a given mortgage is either sold or retained. In other words, the MBS is a pass-through security that is like an equity stake in the mortgages on the shadow bank’s balance sheet.\textsuperscript{15}

Instead of interpreting $m$ as the cost of literally monitoring households, it can be viewed as the cost of writing covenants into mortgage contracts to ensure that households do not misbehave. Sufi (2007), Drucker and Puri (2009), and Gange and Saunders (2012) find that when banks stepped up securitization, they wrote stricter covenants into the loans they made. Changes to $m$ can also be interpreted as financial innovation since lower $m$ makes securitization cheaper.

### 3.3.4 Housing Market

There is a market for houses at $t = 0$, each with price $p_0$. In the baseline model, this price is exogenous. At $t = 1$ there is an unmodeled market for houses (perhaps being sold to speculators with perfectly elastic demand), at which time their price is $p_1$; this price is exogenous and known for certain. If an intermediary forecloses on a house, then it resells the house at $t = 1$. However, as Campbell, Giglio, and Pathak (2011) show, banks incur considerable costs when foreclosing on houses. Therefore, an intermediary receives only $\kappa p_1$ from foreclosing on and reselling a house, where $\kappa \in (0, 1)$.

I impose three conditions on parameters. First,

**Assumption 3.1.** The parameters satisfy \(\rho^S y_H + (1 - \rho^S) \kappa p_1 < (1 + r) (p_0 - w)\).

This condition ensures that subprime borrowers are unprofitable in expectation: the left side is the highest possible expected revenue an intermediary can get from a subprime mortgage while the right side is what intermediaries owe to savers/investors. It also implies that intermediaries strictly prefer being repaid to foreclosing on houses.

\textsuperscript{15}When a pass-through security is created, it typically contains either all of the payments from a loan, or no payments, i.e., individual loans are not divided up.
Assumption 3.2. The parameters satisfy $y_L + w < \kappa p_1$.

This assumption guarantees that foreclosing on the house yields higher revenue than coming after a household’s income in the low-income state.

Assumption 3.3. The parameters satisfy $\gamma_a + w < h$.

This guarantees that the benefit of owning a house outweighs the cost of applying for one so that the lending market does not automatically shut down.

Let $\tilde{\rho} \equiv \pi \rho^P + (1 - \pi) \rho^S$ denote the fraction of borrowers that repay in a pool of both prime and subprime households.

### 3.3.5 Timing of the Game

The timing of the game is as follows: at $t = 0$, households observe their types. Intermediaries post mortgage contracts and then households decide whether or not to apply for a contract. Once households apply, intermediaries decide which applications to accept and which to reject. Intermediaries then raise funds from savers/investors, extend mortgages, and everyone consumes. At $t = 1$, households receive low or high income. They decide whether to repay or default on their mortgages, at which point intermediaries collect the payments or foreclose; at the end of the period everyone consumes. This is summarized in Figure 3.4.

![Figure 3.4: The timeline of the game.](image-url)
3.4 Equilibrium

This section solves for the equilibrium of the game by analyzing the strategies of each set of agents. The first step is to define an equilibrium.

**Definition 3.1.** An equilibrium of the game is a strategy profile and system of beliefs for each household, intermediary, and saver/investor such that each agent’s strategy is sequentially rational and each player’s beliefs are consistent.

An equilibrium contract specifies which mortgage contracts to post, screening decisions, which contracts to apply for,... and it must satisfy the requirements of a Perfect Bayesian Equilibrium.

All proofs are in the Appendix.

3.4.1 Household’s Problem

To characterize the equilibrium of the game, I begin with the households. All of the analysis that follows assumes that a household that applies for a mortgage receives that mortgage; when I solve the intermediaries’ problems, I will derive which applications will and will not be approved, and therefore, which households will and will not apply. Let \( \rho \in \{ \rho^S, \rho^P \} \) denote the generic probability of receiving high income.

First, a household will not apply for a mortgage unless accepting the mortgage and living in the house yields higher utility than remaining in autarky and consuming its endowment. Since the household uses its entire \( t=0 \) income as a down payment, its expected utility from accepting contract \((D^i, \sigma^i)\) is

\[
h + \rho \left( y_H + h - D^i \right) + (1 - \rho) y_L - \sigma^i \gamma_a.
\]

The first term is utility at \( t = 0 \) from living in the house, the second and third terms are expected utility at \( t = 1 \) from either repaying and staying in the house or
defaulting and consuming income, and the final term is the utility cost from being screened (which is zero if the household is not screened). On the other hand, by consuming its endowment each period, the household’s expected utility is

\[ w + \rho y_H + (1 - \rho) y_L. \]

Therefore, a household’s participation constraint requires that the former expression exceed the latter, or

\[ D^i \leq h + \frac{h - w - \sigma^i \gamma_a}{\rho}. \]  

(3.3)

In deriving the participation constraint, I assumed that a household would automatically repay if it had sufficient income to do so. Of course, for this to be the case, the household’s incentive constraint must be satisfied. Consider a household that has already accepted a mortgage and owes \( D^i \) at \( t = 1 \). Since utility at \( t = 0 \) is independent of whether or not the household repays, all I need to worry about is utility at \( t = 1 \). Its expected utility from repaying is

\[ \rho \left( y_H + h - D^i \right) + (1 - \rho) y_L. \]

With probability \( \rho \), the household receives a high income, repays the debt, consumes what is left, and stays in the house, while with probability \( 1 - \rho \), the household does not have enough income to repay so it defaults and the intermediary seizes the house. However, since the mortgage is non-recourse, the household gets to keep and consume its income. On the other hand, if the household strategically defaults, its utility is

\[ \rho y_H + (1 - \rho) y_L. \]
This is just expected income because in either case, the intermediary seizes the house. Therefore, the incentive constraint is

\[ D^i \leq h. \] (3.4)

Households are protected by limited liability, so the face-value of the mortgage cannot exceed households’ maximum income, or

\[ D^i \leq y_H \] (3.5)

### 3.4.2 Intermediaries’ Problems

The most important step in solving for the equilibrium is characterizing intermediaries’ lending decisions. Let \((D^i, \sigma^i)\) denote the mortgage contract offered by intermediary \(i\).

First, by Assumption 3.1, subprime households are unprofitable on their own, so intermediaries cater to prime households. That is, intermediaries offer contracts that are most favorable to prime households. Since the intermediary sector is competitive, Bertrand competition forces each type of intermediary to offer a contract with the lowest possible \(D\) such that they are indifferent between offering the contract and exiting the market.

The first step is to characterize an intermediary’s screening decision.

**Lemma 3.1.** *Shadow banks do not screen in equilibrium: \(\sigma^{SB} = 0\).*

Since there is some probability that a shadow bank will not be a residual claimant on a mortgage (if it is securitized), screening entails a cost with no benefit for the bank.

**Lemma 3.2.** *Commercial banks screen in equilibrium: \(\sigma^{CB} = 1\).*
Because commercial banks are regulated, they have a higher cost of funding than do shadow banks. Therefore, if they do not screen, since Lemma 3.1 shows that shadow banks do not screen, commercial banks will always be undercut. So the only way a commercial bank can remain active is if it does something that shadow banks do not do, namely screen out subprime households to improve the quality of its loans.

Together, Lemmas 3.1 and 3.2 establish that regulated commercial banks serve only prime households (subprime households are screened out) while unregulated shadow banks serve both types of households. In particular, it makes an important prediction: securitization and lax screening go hand-in-hand. This link was observed by Keys, Mukherjee, Seru, and Vig (2010), who find that loans that were more likely to be securitized were screened less diligently (if at all) than loans that were to be retained by the originator. In addition, Lemma 3.2 implies that subprime households will not apply to commercial banks for mortgages since they will be immediately screened and rejected, while still paying the application cost.

From initial equity \( n_{0}^{\text{CB}} \), a commercial bank can fund \( \frac{n_{0}^{\text{CB}}}{p_{0} - w} \) mortgages. Each of these mortgages can back the payments due on a deposit, allowing the bank to raise \( \frac{n_{0}^{\text{CB}}}{p_{0} - w} \times (p_{0} - w) = n_{0}^{\text{CB}} \) in new funds (the second term is the amount needed to issue another mortgage). However, because the commercial bank is regulated, it must pay \( \tau n_{0}^{\text{CB}} \) to the government, leaving only \( (1 - \tau)n_{0}^{\text{CB}} \) to invest in new mortgages. This means the bank makes \( \frac{(1 - \tau)n_{0}^{\text{CB}}}{p_{0} - w} \) new mortgages, each of which can back a new deposit contract. Doing so, the bank raises \( (1 - \tau)n_{0}^{\text{CB}} \) in fresh funds, and the process repeats.

In total, each commercial bank funds

\[
\ell^{\text{CB}} = \frac{n_{0}^{\text{CB}}}{p_{0} - w} + \frac{(1 - \tau)n_{0}^{\text{CB}}}{p_{0} - w} + \frac{(1 - \tau)^{2}n_{0}^{\text{CB}}}{p_{0} - w} + \cdots = \frac{n_{0}^{\text{CB}}}{\tau (p_{0} - w)} \tag{3.6}
\]
mortgages, all of which go to prime borrowers.\footnote{Since commercial banks have skin-in-the-game, their monitoring constraint does not bind.} The total cost associated with lending is the interest paid on deposits plus regulatory payments to the government:

Deposits: \( n_0^{CB} (1 + r) (p_0 - w) \left( 1 + (1 - \tau) + (1 - \tau)^2 + \cdots \right) = \frac{n_0^{CB} (1 + r) (p_0 - w)}{\tau} \)

Regulation: \( \tau n_0^{CB} (p_0 - w) \left( 1 + (1 - \tau) + (1 - \tau)^2 + \cdots \right) = n_0^{CB} (p_0 - w) \)

Total: \( \frac{n_0^{CB} (1 + r) (p_0 - w)}{\tau} + n_0^{CB} (p_0 - w) = \frac{n_0^{CB} (1 + r + \tau) (p_0 - w)}{\tau} \).

Therefore, a commercial bank’s expected profit on a given mortgage is

\[
\rho^P D^{CB} + (1 - \rho^P) \kappa p_1 - m - \gamma_s - (1 + r + \tau) (p_0 - w) .
\]

The first two terms constitute the expected return of the loan, the third and fourth terms are the monitoring and screening costs, respectively, and the fifth term is the monetary cost just computed. A commercial bank will be indifferent between lending and investing in the outside option if the two options yield the same profit, \( r \tau (p_0 - w) \) per loan: this break-even condition implies

\[
D^{CB} = \frac{(1 + r) (p_0 - w) + m + \gamma_s - (1 - \rho^P) \kappa p_1}{\rho^P} . \tag{3.7}
\]

Now I turn to shadow banks. Investors are pushed to their break-even point so the expected return they receive on an MBS just equals the market’s return. Accounting for the securitization fee, this break-even condition is

\[
\tilde{\rho} D^{SB} + (1 - \tilde{\rho}) \kappa p_1 = (1 + r) (p_0 - w) + f ,
\]

which implies

\[
D^{SB} = \frac{(1 + r) (p_0 - w) + f - (1 - \tilde{\rho}) \kappa p_1}{\tilde{\rho}} . \tag{3.8}
\]
As an aside, it is easy to see that Assumption 3.1 implies that both types of interme-
diaries prefer being repaid to foreclosing on a house.

Now I have to determine $\ell_{SB}$, the number of mortgages a shadow bank makes. To do this, suppose a shadow bank monitors, as investors would like it to. Since the monitoring cost is non-pecuniary, it does not affect the cost of a mortgage. A shadow bank’s utility from monitoring (excluding the securitization fee $f$) is

$$\frac{n_{0}^{\text{SB}} [\tilde{\rho} D + (1 - \tilde{\rho}) \kappa p_1 - m]}{s(D)(p_0 - w)},$$

where $s(D)$ is the fraction of loans the shadow bank retains on its balance sheet when offering a mortgage with face-value $D$. On the other hand, if it does not monitor, the shadow bank’s profit is

$$\frac{n_{0}^{\text{SB}} [(1 - s(D)) (\tilde{\rho} D + (1 - \tilde{\rho}) \kappa p_1) + s(D) \kappa p_1]}{s(D)(p_0 - w)}.$$

The bank sells a fraction $1 - s(D)$ of the loans in the MBS, which must have the same price as before (or else the investors would detect a deviation). The retained fraction $s(D)$ automatically default since the bank does not monitor. Incentive compatibility requires the former to exceed the latter, which, when written out, implies

$$s(D) \geq \frac{m}{\tilde{\rho} (D - \kappa p_1)}.$$

Since shadow banks are risk neutral they want to retain as few loans as possible, meaning this holds with equality. Also, note that $s(D)$ is decreasing in the face-value of the mortgage contract. This fact, which is crucial for the analysis, is because a higher $D$ increases the expect payoff from a mortgage and hence MBS.\footnote{This is because the household can always repay if it has high income so increasing $D$ does not increase the probability of default. This result continues to hold if the support of $t = 1$ income is larger (including a continuum) and there exists some cutoff income such that the household repays if and only if income is above the cutoff. In these cases, even though increasing $D$ shrinks the...}
of risk neutrality, this relaxes investors’ participation constraints, meaning investors require that shadow banks have less skin in the game. Therefore, factors that drive up $D$ allow shadow banks to take on more leverage and make more loans.

As another aside, it is already clear that equilibria with shadow banking are constrained inefficient. Indeed, it is straightforward to show that if shadow banks were credible and the only asymmetric information between shadow banks and investors was the screening moral hazard problem, then shadow banks would tranche the loans so as to retain a small residual claim in the repayment state and receive nothing in the default state. Here, they receive the full payment in each state on retained loans because trancheing requires contingent payouts and these are not credible.

Plugging in $D_{SB}$ from (3.8) yields

$$ s = \frac{m}{(1 + r)(p_0 - w) + f - \kappa p_1}. $$

(3.9)

I will reserve $s$ for this particular fraction and write $s(D)$ for an arbitrary mortgage $D$. Putting everything together, the total number of mortgages originated is\(^{18}\)

$$ \ell_{SB} = \frac{n_0}{s(p_0 - w)} = \frac{n_0[(1 + r)(p_0 - w) + f - \kappa p_1]}{m(p_0 - w)}, $$

(3.10)

of which a fraction $\pi$ are to prime households and fraction $1 - \pi$ are to subprime households.

Finally, a shadow bank’s profit comes from the securitization fee. From Bertrand competition, a shadow bank will be indifferent between a mortgage or investing in the outside option, implying

$$ \tilde{\rho} D_{SB} + (1 - \tilde{\rho}) \kappa p_1 - (1 - s)(1 + r)(p_0 - w) - m = s(1 + r)(p_0 - w) \Rightarrow f = m. $$

\(^{18}\)See Brunnermeier and Pedersen (2009) for another model in which binding credit constraints give rise to an upward-sloping demand curve.

repayment region, because the mortgage contract is non-recourse, the default probability rises by just enough such that the shadow bank’s expected return on the mortgage remains the same.
Note that a shadow bank’s total profit is then \( \ell^{SB} f = \frac{\nu^{SB}_0 m}{s(p_0 - w)} \).

3.4.3 Equilibrium Contracts

So far, I have worked under the assumption that the optimal mortgage contract is a debt contract. Proposition 3.1 states this formally.

Proposition 3.1. The optimal mortgage contract is a simple debt contract: the intermediary receives a fixed payment if the borrower has high income, and forecloses on the house if the household has low income.

This result echoes that of Wang and Williamson (1998), who show that in an environment with screening when income has continuous support, a fixed-payment debt contract is the optimal lending contract.\(^\text{19}\)

The next result gives a condition such that if an equilibrium exists, then both prime and subprime households pool at the same contract.

Proposition 3.2. Suppose \( w \) is sufficiently small; then both prime and subprime households pool at the same mortgage contract.

This result is interesting in its own right because it is the opposite of the well-known result from Rothschild and Stiglitz (1976) that if an equilibrium exists, it must be separating. The reason is that in Rothschild and Stiglitz (1976), the principal posts a menu of contracts, each designed to attract a certain type of agent. Once an agent selects a contract, he gets it. That is, agents reveal their types via the contracts they select and the principal does not reject anyone once a contract is picked.\(^\text{20}\)

On the other hand, here, I follow the three-stage structure introduced by Hellwig

\(^{19}\)However, this result does not appear to be robust to dynamic models. In particular, Piskorski and Tchistyi (2010) show, in a continuous time setting, that the optimal mortgage contract is an option adjustable rate mortgage.

\(^{20}\)Wang and Williamson (1998) show, in their model, that if an equilibrium exists, it must be separating. This is because their equilibrium concept and game structure are similar to those in Rothschild and Stiglitz (1976).
(1987) in which intermediaries can reject applicants even after they select a contract. Hellwig (1987) first pointed out, and Dubey and Geanakoplos (2002), Martin (2007), and Martin (2008) later elaborated on, the implications of this crucial distinction. Specifically, Hellwig (1987) argues that in the three-stage game, the optimal pooling contract is the only one that survives the Kohlberg and Mertens (1986) stability criterion, while Martin (2007) shows that if there are pooling contracts that Pareto dominate separating contracts, then as long as subprime households prefer the pooling contract, it will emerge as an equilibrium contract (indeed, as I show in the proof of Proposition 3.2, subprime households must want to pool at the prime contract).

Now I can characterize the equilibrium contract.

**Theorem 3.1.** There exists a unique equilibrium contract, which is as follows:

1. Suppose
   \[
   D^{SB} \leq \min \left( D^{CB} + \frac{\gamma_a}{\rho_P}, h + \frac{h - w}{\rho_P}, y_H, h \right); \quad (3.11)
   \]
   then shadow banks originate all mortgages and securitize a fraction of them, \( n_0^{SB} = n_0 \), there is no screening, and both prime and subprime households receive mortgages.

2. Suppose
   
   \[
   D^{CB} \leq \min \left( h + \frac{h - w - \gamma_a}{\rho_P}, y_H, h \right); \quad (3.12)
   \]
   \[
   D^{SB} \geq \min \left( D^{CB} + \frac{\gamma_a}{\rho_P}, y_H, h \right); \quad (3.13)
   \]
   then commercial banks originate and retain all mortgages, \( n_0^{CB} = n_0 \), all applicants are screened, and only prime borrowers apply for, and hence receive, mortgages.

3. Suppose
   \[
   D^{CB} \geq \min \left( h + \frac{h - w - \gamma_a}{\rho_P}, y_H, h \right); \quad (3.14)
   \]
\[ D^{SB} \geq \min \left( D^{CB} + \frac{\gamma_a}{\rho^P}, y_H, h \right); \quad (3.15) \]

then no borrowers apply for, and hence receive, mortgages.

While not difficult, the proof of this result is fairly long since in each case, to establish existence I have to rule out all possible deviations by any agent, while to establish uniqueness I need to rule out any other scenarios as an equilibrium outcome. It is important to remember what this result does and does not say: it says that given the house price that emerges when some constellation of intermediaries is active, there exists a unique equilibrium contract; it does not say anything about the price itself.

This result is intuitive. In Case 1, a shadow bank’s mortgage is not only cheap enough to satisfy prime households’ participation and incentive constraints, but it also makes up for the application cost associated with commercial banks’ mortgages. When the utility from home-ownership, the screening cost, or the application cost is high, the constraint is relaxed and shadow banking is more likely. When the latter two costs are high, commercial banks lose any informational advantages they have because the cost of guaranteeing a strong loan pool outweighs the benefit of lower risk. As a result, all intermediaries choose to operate as shadow banks. In Case 2, the opposite is true: the commercial bank contract is cheap enough to induce participation and repayment, and the benefits of a low-risk loan pool outweigh the costs of guaranteeing the low risk. Therefore, all intermediaries operate as commercial banks. In Case 3, no matter how an intermediary operates, each contract is so expensive that households have no reason to apply for one, no reason to repay, or both; thus the lending market collapses.

Finally, in each case in the Theorem, there is at most one type of intermediary active. This is because there are only two types of households, prime and subprime, and neither will be simultaneously served by both types of intermediaries. When I extend the model to include a continuum of types, some intermediaries choose
to operate as commercial banks and serve only high-quality borrowers while other intermediaries operate as shadow banks and lend to low-quality borrowers.

### 3.5 Properties of the Equilibrium

#### 3.5.1 Comparative Statics

Now that I have solved for the equilibrium of the game, I can analyze how changes to certain parameters affect not only how other quantities change, but also which form of intermediation (if any) emerges and how intermediary balance sheets are affected.

**Mode of Intermediation**

I begin by characterizing which parameter changes make intermediation as a whole (commercial banking or shadow banking) more likely.

**Proposition 3.3.** *Intermediation is more likely when $y_H$, $h$, $p_1$, $\rho^P$, $\rho^S$, $\pi$, and $p_1$ are higher and when $p_0$, $r$, $\tau$, $\gamma_s$, $\gamma_a$, and $m$ are lower.*

This result says that comparing two segments that are equal in every way except for $y_H$ (for example), the segment with higher $y_H$ is more likely to have an active mortgage market. All of the changes above relax at least one of the constraints in Cases 1 and 2 of Theorem 3.1. Since some type of intermediary is active in each of these cases, intermediation as a whole is more likely. However, this result does not distinguish between which type of intermediary will become active.

First, consider two segments, one with a higher $y_H$ than the other (but all other parameters are the same). Since $D^{CB}$ and $D^{SB}$ do not depend on $y_H$ the size of the mortgage does not change; however, the limited liability constraint is relaxed which makes it easier for an intermediary to make a profit. It is clear that $h$ behaves exactly like $y_H$ (though in this case, households now get so much utility from living in a house that they are more likely to accept a contract and less likely to default).
Next, a segment with a higher $\rho^P$ (or $\rho^S$) has households that are less likely to default. Clearly, this lowers both $D^{CB}$ and $D^{SB}$, thus expanding the intermediation regions. That is, a mortgage market is more likely to be active in segments where default is less likely. A similar result holds for $\pi$: if a segment has a high fraction of prime households, intermediaries are more likely to operate there.

It is also straightforward to understand the effects of changes to prices. It is easy to see that both $D^{CB}$ and $D^{SB}$ are increasing in $p_0$ so that intermediation is more likely when $p_0$ is low. On the other hand, both $D^{CB}$ and $D^{SB}$ are decreasing in $p_1$ since the expected repayment is larger. This makes intermediation more likely when prices are expected to grow in the future. Finally, both $D^{CB}$ and $D^{SB}$ are increasing in $r$, making intermediation more likely when $r$ is low because borrowing costs are smaller. Intuitively, for each of these variables, intermediation is most likely when borrowing from savers/investors is cheapest, which in turn means borrowing for households is cheapest.

On the other hand, when $\tau$ and $\gamma_s$ are low, it is cheap for commercial banks to remain active. A low $\tau$ means it is cheap for commercial banks to raise deposits and thus lend out more, while a low $\gamma_s$ makes it easier to learn a household’s type; both lower $D^{CB}$. Similarly, a low monitoring cost $m$ lowers $D^{SB}$.

The next step is to determine which forms of intermediation are more likely to appear in response to changing parameters.

**Proposition 3.4.** Commercial banking is more likely when $\tau$ and $\gamma_s$ are lower and when $\rho^P$ and $h$ are higher.

When $\tau$ and $\gamma_s$ are small, commercial banking becomes cheaper, allowing commercial banks to offer cheaper mortgages. As for shadow banks, Proposition 3.5 highlights which parameters make shadow banking more likely.

**Proposition 3.5.** Shadow banking is more likely when $y_H$, $\pi$, $\rho^S$, and $p_1$ are higher and when $m$ and $r$ are lower.
Of course, the opposite conditions as those in Proposition 3.4 also increase the likelihood of shadow banking. More \( t = 1 \) income makes shadow banking more likely because it makes households more likely to apply for, and then repay mortgages. Since commercial banks screen, they do not lend to subprime households, so these households have no impact on a commercial bank’s lending terms. Since shadow banks do serve subprime households, if they make up a small fraction of the market (high \( \pi \)), shadow bank contracts become cheaper, thus closing the door on commercial banks. When subprime households are more likely to repay, the shadow bank loan pool improves in quality, driving down shadow bank mortgages. Since commercial banks do not serve subprime households, they are unaffected by this change and shadow banks are the only ones that benefit. A lower value of \( m \) can be interpreted as investors “trusting” banks more (for example, less strict covenant requirements), or innovation that makes securitization cheaper; either way, while both banking contracts become cheaper, shadow banks’ fall more. Note that \( p_0 \) is not covered in either of the last two results; this is because \( \rho^P > \tilde{\rho} \) but \( \tau > 0 \) so that the effect is ambiguous.

Proposition 3.5 shows that the conditions of the early 2000s, namely rising expectations and low interest rates as controlled by the Federal Reserve, do indeed encourage more securitization and more subprime lending. Lowering \( m \), lowering \( r \), and increasing \( p_1 \) all slacken shadow banks’ incentive constraints for securitizing, i.e., they all spur credit supply. This is consistent with evidence in Mian and Sufi (2009), Nadauld and Sherlund (2013), and Justiniano, Primiceri, and Tambalotti (2015), all of whom argue that the housing boom was driven by positive credit supply shocks to intermediaries, not a sudden increase in demand for credit.

The final major result of this subsection shows that for most parameters, there exist regions such that there is shadow banking in one region, commercial banking in another region, and no lending otherwise.

**Theorem 3.2.** Fix all parameters except the one indicated. Then:
1. There exist thresholds \( y, \bar{y} \) such that there is no mortgage lending if \( y_H \leq y \), there is commercial banking if \( y_H \in (y, \bar{y}) \), and there is shadow banking if \( y_H \geq \bar{y} \). The same result holds for \( \rho^S \).

2. There exist thresholds \( r, \bar{r} \) such that there is no mortgage lending if \( r \geq \bar{r} \), there is commercial banking if \( r \in (r, \bar{r}) \), and there is shadow banking if \( r \leq r \).

3. There exist thresholds \( p_1, \bar{p}_1 \) such that there is no mortgage lending if \( p_1 \leq p_1 \), there is commercial banking if \( p_1 \in (p_1, \bar{p}_1) \), and there is shadow banking if \( p_1 \geq \bar{p}_1 \).

4. There exists a threshold \( \pi^* \) such that there is shadow banking if \( \pi \geq \pi^* \), and either commercial banking or no lending if \( \pi < \pi^* \). The signs are reversed for \( \rho^P \).

5. There exists a threshold \( \gamma^*_s \) such that there is commercial banking if \( \gamma_s \leq \gamma^*_s \) and either shadow banking or no lending if \( \gamma_s > \gamma^*_s \); the same is true for \( m \) and shadow banking. The signs are reversed for \( h \).

The main conclusion of this result is that for many parameters, as these parameters gradually increase (or decrease), the economy gradually transitions from no lending, to traditional commercial banking, to modern shadow banking. In particular, for several of the parameters, commercial banking occurs for intermediate values, while lending collapses or moves to the shadow banking sector only for more extreme values. For example, Case 1 shows that shadow banking does not emerge unless household wealth is sufficiently high, which is consistent with the general boom in economic conditions from the mid-1980s through early 2000s. Cases 2 and 3 show that low interest rates and high house price expectations spur securitization and subprime lending via the growth of shadow banking.
Recall that $\tilde{\rho} = \pi \rho^P + (1 - \pi) \rho^S$ is the fraction of all households that are able to repay a mortgage, i.e., the fraction receiving high income at $t = 1$. This number is an intermediary’s ex ante assessment of a borrower’s credit worthiness, i.e., an intermediary’s “soft information” on a borrower. In real-life, an example of such information is a borrower’s FICO credit score. While this number does not exactly correspond to a credit score, it captures an intermediary’s perception of a borrower’s credit worthiness before screening. Theorem 3.2 sheds some light on how intermediation depends on credit worthiness, and justifies a particular “rule of thumb” that intermediaries used to screen mortgages. Figure 3.5 plots how higher credit worthiness, obtained by increasing $\rho^P$, affects what type of intermediary is active in equilibrium, and by extension, the overall default rate on originated mortgages.

When credit worthiness is sufficiently low, the probability of default is so high that no intermediary expects to remain profitable so there is no lending at all; this is the left-most region. When credit worthiness is in the intermediate region, Theorem 3.2 implies that commercial banks are active. Since commercial banks lend to only prime households, the overall default rate is simply the default rate of prime households, $1 - \rho^P$, which declines as $\rho^P$ increases. Finally, once credit worthiness passes the second threshold, non-screening shadow banks become active. Now, subprime households

\[ \text{Figure 3.5: The effect of higher credit worthiness on default probability.} \]

\[ \text{When credit worthiness is sufficiently low, the probability of default is so high that no intermediary expects to remain profitable so there is no lending at all; this is the left-most region. When credit worthiness is in the intermediate region, Theorem 3.2 implies that commercial banks are active. Since commercial banks lend to only prime households, the overall default rate is simply the default rate of prime households, } 1 - \rho^P, \text{ which declines as } \rho^P \text{ increases. Finally, once credit worthiness passes the second threshold, non-screening shadow banks become active. Now, subprime households} \]

\[ \text{21“Hard information,” obtained via screening, is a borrower’s true probability of being able to repay.} \]
enter the pool, immediately raising the default rate: this is the discrete, discontinuous, upward jump in the figure. The overall default rate is $1 - \tilde{\rho}$ and this declines as the quality of the whole pool further increase. This pattern was documented by Keys, Mukherjee, Seru, and Vig (2010) and Keys, Seru, and Vig (2012), who show that mortgages extended to households with credit scores just above 620 were less likely to be screened than those with scores just below 620. They argue that because it is more difficult to securitize loans below the 620-threshold (as per a regulatory “rule of thumb”), originating banks were more likely to retain these loans on their balance sheets, giving them greater incentive to screen the borrowers; loans above the threshold were more likely to be securitized and thus less likely to be screened. In the model, the second threshold is like a 620 credit score: loans below the threshold are screened because they are retained by originating commercial banks, while loans above the threshold are not screened because they are securitized by shadow banks.

The figure also shows that the default rate for mortgages that are ex ante more likely to be securitized is higher than the default rate for mortgages that are to be retained. This is consistent with the empirical evidence of Mian and Sufi (2009), Dell’Ariccia, Igan, and Laeven (2012), and Jiang, Nelson, and Vytlacil (2013), all of whom show that default rates were higher in regions that saw increased securitization. All of these results are summarized in Corollary 3.1.

**Corollary 3.1.** A household’s probability of receiving a mortgage increases and probability of defaulting on that mortgage decreases when $\pi$, $\rho^i$, and/or $y_H$ increase (except at the threshold where commercial banking transitions to shadow banking).

**Balance Sheet Effects**

Now, I will examine how intermediaries’ balance sheets respond to changes in parameters. All of the analysis assumes the changes are small enough that the economy does not switch between cases in Theorem 3.1.
I begin with commercial banks. First, since none of the probabilities affect $\ell^{\text{CB}}$, commercial banks’ lending capacity is unaffected by households’ abilities to repay. More interesting, however, is the effect of a change to a household’s initial wealth, $w$. All else equal, a segment with a lower $w$ is, in a sense, “more subprime” because the households in that segment have a larger debt-to-income ratio. Mian and Sufi (2011) document that household debt-to-income ratios exploded before the crisis, with credit expansion and house prices increasing the most in those areas with high debt-to-income ratios and low income growth. So consider a change in $w$, which corresponds to comparing a low-income segment with a high-income segment; does commercial banking generate the observed pattern for credit expansion?\footnote{Strictly speaking, I am looking at variations in income, not income growth. However, if two segments were totally identical at date $t = -1$ and totally identical at $t = 0$ except for $w$, then the segment with lower $w$ had lower income growth so the computation is applicable.}

Differentiating (3.6),\footnote{This comparative static (and the similar ones that follow) is a cross-sectional comparison, i.e., across otherwise identical segments at the same point in time, not a time-series one.}

$$\frac{\partial \ell^{\text{CB}}}{\partial w} = \frac{n_0^{\text{CB}}}{\tau (p_0 - w)^2} > 0.$$ 

Thus higher wealth drives credit expansion under commercial banking. The reason is that higher income lowers the face-value of the mortgage, which allows commercial banks to lend more. As per Fact 3, this correlation is consistent with every period in the last few decades, except for the housing bubble from 2002-2005.

Next, since (3.6) shows that $\ell^{\text{CB}}$ is independent of the interest rate, $r$, and the $t = 1$ price, $p_1$, neither of these parameters affect commercial bank credit expansion. This is because intermediaries rely on leverage to expand their balance sheets, and a commercial bank’s leverage is simply

$$\mathcal{L}^{\text{CB}} = \frac{1}{\tau},$$
which is independent of fundamentals such as $r$ and $p_1$. This means that optimism about house prices and low interest rates do not translate into a massive credit expansion if commercial banks are active.\footnote{There is no clear consensus on the cyclicality of commercial bank leverage, or even if it has cyclicality: Nuño and Thomas (2014) find that commercial bank leverage has been largely acyclical since 1984.}

I now move on to what happens when shadow banks are active. Like with commercial banks, none of the probabilities affect $\ell^{SB}$. However, differentiating (3.10) with respect to $w$,

$$\frac{\partial \ell^{SB}}{\partial w} = -\frac{n_0^{SB} (\kappa p_1 - m)}{m (p_0 - w)^2}.$$  

Since $m$ is small, this expression is negative. This means that credit expansion and income growth are negatively correlated under shadow banking, as was the case from 2002–2005. The reason is that lower income $w$ increases the face-value of a mortgage since the household has to borrow more. As explained earlier, a higher face-value allows shadow banks to sell more loans because investors’ participation constraints become easier to satisfy, reducing the skin-in-the-game requirement. Since shadow banks can sell more loans, they can boost leverage and make more mortgages. This feedback is precisely the trend illustrated in the bottom panel of Figure 3.1. In the next section, I extend the model by endogenizing $p_0$ and show that changes to $w$ have further indirect effects through the price, amplifying this result.

Now, suppose that $p_1$ increases so that intermediaries expect house prices to be higher at $t = 1$. Mian and Sufi (2009) and Foote, Gerardi, and Willen (2012) identify excessive optimism about house prices as the driving force behind the housing boom. It is easy to see that this lowers leverage and hence $\ell^{SB}$. The reason is that when prices are expected to be high at $t = 1$, intermediaries receive a higher expected payoff on mortgages. This makes them less willing to sell mortgages and thus less able to expand their balance sheets. So on the surface, it appears that when prices are due to rise shadow banks actually securitize less and shrink their balance sheets.
However, differentiating (3.10) yields

$$\frac{\partial \ell^{SB}}{\partial p_0} = \frac{n_0^{SB} (\kappa p_1 - m)}{m (p_0 - w)^2} \quad \Rightarrow \quad \frac{\partial^2 \ell^{SB}}{\partial p_1 \partial p_0} = \frac{n_0^{SB} \kappa}{m (p_0 - w)^2} > 0.$$

Thus there is a complementarity between $t = 0$ and $t = 1$ house prices in boosting leverage and credit expansion. That is, when prices are high at $t = 0$ and expected to continue growing, credit expansion increases. The model predicts that housing markets in the middle of a boom should be accompanied by rapidly expanding balance sheets and this is indeed consistent with the housing bubble\(^{25}\) in the mid-2000s. Since shadow bank leverage is

$$\mathcal{L}^{SB} = \frac{n_0^{SB} [(1 + r) (p_0 - w) + m - \kappa p_1]}{m},$$

the same comparative statics hold for leverage. Since rising house price expectations are generally consistent with economic booms, the model predicts that shadow bank leverage is procyclical, as documented by Adrian and Shin (2014).

Combining the results of the last two subsections, the model provides an explanation for the housing boom and bust: improving economic conditions, low interest rates, and rising expectations caused mortgage lending to shift, in many regions of the country, from the commercial banking sector to the shadow banking sector. Once this happened, subprime areas received the most credit. When the people started revising their expectations downward (in around 2006), commercial banking returned to dominance, house prices fell, and credit contracted. With a large enough downward swing in some segments, the lending market collapsed completely.

\(^{25}\)Since there is no fundamental value for housing and I did not take a stand on true probabilities versus beliefs, here “bubble” simply refers to high expectations for the future.
3.5.2 Welfare Analysis

Having analyzed how intermediaries’ balance sheets fluctuate with market conditions, it is natural to ask whether securitization actually improves social welfare. Unfortunately, it is difficult to make sharp statements because there are several interacting quantities, though I can say in very broad terms when one form of intermediation is optimal or not.

Total welfare in the economy is the sum of households’, intermediaries’, and savers’/investors’ welfare. Savers/investors make zero expected profits, while commercial and shadow banks earn \( n_0^{CB} r \) and \( \frac{n_0^{SB} f}{s(p_0-w)} \), respectively. When commercial banks are active, only prime households receive loans. Using \( \ell^{CB} \) and (3.1), the total welfare in the economy is

\[
W^{CB} = n_0 \left[ r + h + \rho^P \left( y_H + h - D^{CB} \right) + \left( 1 - \rho^P \right) y_L - \gamma_a \right] \frac{\tau}{\left( p_0 - w \right)}, \tag{3.16}
\]

where \( D^{CB} \) is as in (3.7). Unsurprisingly, welfare is higher when \( p_0 \) is lower because more prime households get mortgages, and they consume more at \( t = 1 \). On the other hand, when shadow banks are active, both prime and subprime households receive mortgages, so total welfare is

\[
W^{SB} = \frac{n_0 \hat{\rho} \left( D^{SB} - \kappa p_1 \right)}{m \left( p_0 - w \right)} \left[ m + h + \hat{\rho} \left( y_H + h - D^{SB} \right) + \left( 1 - \hat{\rho} \right) y_L \right], \tag{3.17}
\]

where \( D^{SB} \) is as in (3.8). Welfare is not monotonic in \( D^{SB} \) because while larger \( D^{SB} \) lowers household utility, it also increases shadow banks’ credit expansion, bringing in more fee revenue, and more households obtain houses. In particular, welfare is increasing when

\[
\hat{\rho} D^{SB} < \frac{\hat{\rho} \kappa p_1 + m + h + \hat{\rho} (y_H + h) + \left( 1 - \hat{\rho} \right) y_L}{2}
\]
because for these values, the mortgage is cheap enough that the welfare gain from additional households being served outweighs the cost of repayment.

By subtracting the former from the latter, notice that there are several discrepancies between (3.16) and (3.17):

\[
W_{SB} - W_{CB} = \frac{n_0}{p_0 - w} \left\{ \tilde{\rho} \left( \frac{D_{SB} - \kappa p_1}{m} \right) \left[ m + h + \tilde{\rho} \left( y_H + h - D_{SB} \right) + (1 - \tilde{\rho}) y_L \right] 
- \frac{r + h + \rho^P \left( y_H + h - D_{CB} \right) + (1 - \rho^P) y_L - \gamma_a}{\tau} \right\}. \tag{3.18}
\]

The key discrepancy is the welfare gain that comes from serving subprime households in tandem with prime households, which commercial banks ignore because they screen out these households.

The above expression is increasing in \( D_{CB} \) and nonmonotonic in \( D_{SB} \) as before. At one extreme, \( D_{CB} - D_{SB} > 0 \) is very large, in which case \( W_{SB} - W_{CB} > 0 \): shadow banks are active, and this is the optimal form of intermediation. At the other extreme, \( D_{SB} - D_{CB} > 0 \) is very large, in which case \( W_{CB} - W_{SB} > 0 \): commercial banks are active, and this is the optimal form of intermediation. In the middle region, \( W_{SB} \) may or may not exceed \( W_{CB} \) and \( D_{SB} \) may or may not exceed \( D_{CB} \). This means that either shadow banking is the optimal form of intermediation but commercial banks are active instead, or commercial banking is the optimal form of intermediation yet shadow banks are active. This is summarized in Proposition 3.6:

**Proposition 3.6.** The following cases determine whether or not an operating intermediary is optimal:

1. If \( D_{SB} \ll D_{CB} \), then all intermediaries operate as shadow banks and this is the optimal form of intermediation;

2. If \( D_{SB} < D_{CB} \) but \( W_{SB} < W_{CB} \), then all intermediaries operate as shadow banks but welfare would be higher if they operated as commercial banks;
3. If $D^{CB} < D^{SB}$ but $W^{CB} < W^{SB}$, then all intermediaries operate as commercial banks but welfare would be higher if they operated as shadow banks;

4. If $D^{CB} \ll D^{SB}$, then all intermediaries operate as commercial banks and this is the optimal form of intermediation.

It is not necessarily the case that each of these scenarios appears; for some of them, the set of parameters for which that case occurs might be degenerate (see Figure 3.6 below). Case 1 is possible when, for example, the screening cost is exceptionally high, in which case shadow banking is optimal because screening is too costly, and securitization is sufficiently cheap that additional welfare gains from newly-served subprime borrowers outweigh the costs of lending to them. In Case 2, securitization is sufficiently expensive relative to screening that the welfare gain from serving more households is outweighed by the cost of expansion: welfare would be higher if commercial banks were active, but their contracts are still too expensive relative to shadow bank contracts. In this case, lowering the regulatory tax (weakly) improves welfare by expanding the commercial banking region. In Case 3, the opposite situation occurs and raising the regulatory cost is welfare-improving. Finally, in Case 4, shadow banking is suboptimal because securitization is so low (because it is costly) that the welfare gain coming from the newly-served subprime borrowers is outweighed by the cost of lending to them.

Figure 3.6 illustrates this result in two cases: it plots $\Delta W \equiv W^{SB} - W^{CB}$ as a function of $D^{SB}$, given some $D^{CB} \in [0, y_H]$ (the vertical, dashed line). In the rendering in panel (a), the third case doesn’t occur. In the orange region, $D^{SB} < D^{CB}$ and $\Delta W > 0$: this is Case 1. In the green region, $D^{SB} < D^{CB}$ but $\Delta W > 0$: this is Case 2. Finally, in the blue region, $D^{CB} < D^{SB}$ and $\Delta W < 0$: this is Case 4. In the rendering in panel (b), the second case doesn’t occur. This means that unless
$D^{CB}$ is very large relative to $D^{SB}$ shadow banking is inefficient relative to commercial banking.

Finally, it is important to note that (3.18) is continuous in the parameters, even on the boundary between commercial banking and shadow banking, because $D^{SB}$ and $D^{CB}$ are both functions of the (same) exogenous price, $p_0$. This means that the economy is stable in a sense since small changes to parameters do not have disproportionately large effects on welfare. When I endogenize the house price, this will no longer be the case.

### 3.6 Extension

In this section, I allow for a continuum of household types and endogenize the house price at $t = 0$ with a simple model of the housing market. To begin, rather than restricting $\rho^i \in \{\rho^S, \rho^P\}$, suppose that $\rho \sim f(\rho)$ with support $[0, 1]$. As in the two-type model, commercial banks will screen and lend to sufficiently profitable households ("prime"). Since commercial banks will learn households’ types through screening, they can offer mortgage contracts that are contingent on $\rho$. Bertrand competition ensures that commercial banks earn the same rate of return on each prime borrower
so that
\[ D_{\text{CB}}(\rho) = \frac{(1 + r)(1 + \tau)(p_0 - w) + \gamma_s + m - (1 - \rho) \kappa p_1}{\rho}. \]

Obviously, a commercial bank will not lend to a household on whom it expects to lose money; let \( \rho^* \) denote the household to whom a commercial bank is just indifferent between lending and not lending. Since households are protected by limited liability, commercial banks will not lend unless
\[ \rho \geq \rho^* \equiv \frac{(1 + r)(1 + \tau)(p_0 - w) + \gamma_s + m}{\min(h, y_H)} - \kappa p_1. \]

Therefore, if commercial banks are active, they will lend to households with \( \rho \in [\rho^*, 1] \).

I will assume that such a \( \rho^* \) necessarily exists so there is always some profitable household. There also exists a household \( \hat{\rho} \) who is indifferent between accepting a commercial bank’s screening contract and a shadow bank’s non-screening contract.\(^{26}\)

I will assume \( \hat{\rho} \leq \rho^* \), which is natural since otherwise commercial banks would lose out on profitable households to shadow banks.

As for intermediaries that choose to become shadow banks, as before, they do not screen and pool together all households to whom they lend. Call a household “subprime” if \( \rho \leq \rho^* \). Let \( \tilde{\rho} = \mathbb{E}[\rho | \rho \leq \rho^*] \) denote the repayment rate from subprime households. Then as in the two-type model, the subprime mortgage contract is
\[ D_{\text{SB}} = \frac{(1 + r)(p_0 - w) + f - (1 - \tilde{\rho}) \kappa p_1}{\tilde{\rho}} \]
and all of the other expressions are the same as before. This highlights a key role of shadow banks: by pooling together very low-\( \rho \) (and hence unprofitable) households

\(^{26}\)This threshold is the fixed point of
\[ \hat{\rho} = \mathbb{E}[\rho | \rho \leq \max(\hat{\rho}, \rho^*)]\left(1 + \frac{\tau(1 + r)(p_0 - w) + \gamma_s + \gamma_a}{(1 + r)(p_0 - w) + m - \kappa p_1}\right). \]
with medium-$\rho$ (and potentially profitable) ones, low-$\rho$ households, who would otherwise be shut out of the market, get an expensive contract that is outweighed by the utility gain from owning a house. In other words, shadow banks subsidize very subprime households to make previously impossible home-ownership possible. An immediate implication is that the distribution of offered mortgage contracts is much tighter when shadow banks have a larger presence; this is consistent with evidence in Purnanandam (2011).

The next step is to determine which banks are active at a given time. Like before, in extreme cases the entire lending market breaks down and no households receive mortgages. In another case, both banks coexist, with shadow banks serving $\rho \leq \rho^*$ and commercial banks serving $\rho \geq \rho^*$; let $p_0^{SB}$ denote the house price in this case. Potentially, commercial banks can be the only ones active and only households with $\rho \geq \rho^*$ receive mortgages. Conversely, consider two households indexed by $\rho', \rho''$ with $\rho' < \rho^* < \rho''$; can $\rho'$ receive a loan but not $\rho''$? For this to be the case, a shadow bank would have to make the loan and for the household to accept, I need

$$D_{SB} \leq \min \left( h, y_H, h + \frac{h - w}{\rho'} \right).$$

The household with $\rho''$ will not accept this contract because $\hat{\rho} \leq \rho^* < \rho''$; therefore, for him to not receive credit he must not want the commercial bank’s contract either, i.e.,

$$D_{CB}(\rho'') \geq \min \left( h, y_H, h + \frac{h - w - \gamma_a}{\rho'} \right).$$

But by Assumption 3.3,

$$D_{SB} \leq \min \left( h, y_H, h + \frac{h - w}{\rho'} \right) = \min \left( h, y_H, h + \frac{h - w - \gamma_a}{\rho'} \right) \leq D_{CB}(\rho'),$$

150
contradicting the fact that the $\rho''$-household prefers the commercial bank’s contract. Therefore, the only way for only shadow banks to be active is that $\rho^* = 1$ and everyone receives an unscreened mortgage.

Next, following Gao, Sockin, and Xiong (2015), at $t = 0$ there is an outside sector of home builders. Let $H$ denote the existing stock of houses on the market, $S$ the total supply of houses, and $\varepsilon \geq 0$ the supply elasticity. The cost of construction is quadratic,

$$C(S) = \frac{1}{2\varepsilon} (S - H)^2.$$  

Builders choose how many new houses to construct so as to maximize their profits,

$$\Pi(S) = \max_{S} \left\{ p_0 S - \frac{1}{2\varepsilon} (S - H)^2 \right\}.$$  

Thus the supply of houses for sale at $t = 0$ is

$$S(p_0) = H + \varepsilon p_0.$$ (3.19)

For simplicity, the price at $t = 1$ is still exogenous and known for certain.

When the price was exogenous, the cases in Theorem 3.1 were mutually exclusive. That is no longer the case here because different prices, and hence different mortgage contracts, emerge depending on which modes of banking are prevalent. Indeed, suppose both commercial and shadow banks are active so the price is $p_{0}^{SB}$. This is an equilibrium as long as subprime households are happy with their contracts and prime households are happy with theirs. As argued above, prime households will not deviate to a shadow bank’s contract, while a subprime household will not deviate (and will continue to accept a shadow bank’s contract) as long as

$$D_{SB}(p_{0}^{SB}) \leq \min \left( D_{CB}^{SB} \left( p_{0}^{SB}, \rho \right) + \frac{\gamma_a}{\rho}, h + \frac{h - w}{\rho}, y_H, h \right).$$
In the second case, only commercial banks are active. For this to be an equilibrium, intermediaries must not have an incentive to operate as shadow banks. But if they do then subprime households will apply to them (and be approved) for mortgages so the price must reflect the new influx of borrowers. Therefore, the commercial bank contract must anticipate what would happen if shadow banks entered. The relevant condition is

\[
D_{CB}(p_{0 SB}^0; \rho) \leq \min \left( \frac{h + \frac{h - w - \gamma a}{\rho}}{\rho}, y_h, h \right);
\]

\[
D_{SB}(p_{0 SB}^0) \geq \min \left( D_{CB}(p_{0 SB}^0; \rho) + \frac{\gamma a}{\rho}, y_h, h \right).
\]

Finally, the lending market collapses if both contracts, under their respective prices, are unappealing:

\[
D_{CB}(p_{0 CB}^0; \rho) \geq \min \left( \frac{h + \frac{h - w - \gamma a}{\rho}}{\rho}, y_h, h \right);
\]

\[
D_{SB}(p_{0 SB}^0) \geq \min \left( D_{CB}(p_{0 CB}^0; \rho) + \frac{\gamma a}{\rho}, y_h, h \right).
\]

Now, there potentially exists a set of parameters for which the first two cases above are simultaneously possible. That is, the string of inequalities

\[
D_{CB}(p_{0 CB}^0; \rho) + \frac{\gamma a}{\rho} < D_{SB}(p_{0 SB}^0) < D_{CB}(p_{0 SB}^0; \rho) + \frac{\gamma a}{\rho}
\]

holds. This is important because, as I discuss later, the potential for multiple equilibria can lead to instability and large welfare jumps.

Now I will actually solve for \( p_0 \) in each case. There are three cases: no lending, only commercial banks, and both commercial and shadow banks. First, when there is no mortgage lending there is no market-clearing house price. Next, when only commercial banks are active, the demand for housing is simply the number of households
served by commercial banks (since each household buys one house). This demand is given by $\ell_{CB}$ in (3.6) while the supply is given by (3.19). Since all intermediaries choose to operate as commercial banks, $n_{0CB} = n_0$. Equating supply and demand,

$$\frac{n_0}{\tau(p_0 - w)} = H + \varepsilon p_0 \implies \varepsilon \tau p_0^2 + \tau (H - \varepsilon w)p_0 - (n_0 + \tau Hw) = 0 \implies$$

$$p_{0CB}^\ell = \begin{cases} \frac{\tau(\varepsilon w - H) \pm \sqrt{(\varepsilon w - H)^2 + 4\varepsilon \tau (n_0 + \tau Hw)}}{2\varepsilon \tau} & \text{if } \varepsilon > 0 \\ w + \frac{n_0}{\tau H} & \text{if } \varepsilon = 0 \end{cases} \quad (3.21)$$

However, since the expression under the square-root is always positive, the smaller root is negative so I can ignore it. That is, there exists a unique house price, $p_{0CB}^\ell$, given by the positive root of (3.21).

Finally, when both commercial and shadow banks are active the demand for housing is the number of households served by each type of bank, $\ell = \ell_{CB} + \ell_{SB}$, while the supply is still given by (3.19). Equating the two,

$$H + \varepsilon p_0 = \frac{n_{0CB}}{\tau(p_0 - w)} + \frac{n_{0SB}}{s(p_0 - w)} \implies$$

$$p_{0SB}^\ell = \begin{cases} \frac{\tau(n_{0SB}(1 + r) - m(H - \varepsilon w)) \pm \sqrt{\tau^2(n_{0SB}(1 + r) - m(H + \varepsilon w))^2 - 4\varepsilon \tau mn_{0SB}(\kappa p_1 - m) - mn_{0CB}^2}}{2\varepsilon \tau m} & \text{if } \varepsilon > 0 \\ w + \frac{\tau n_{0SB}(\kappa p_1 - m) - mn_{0CB}^2}{\tau(n_{0SB}(1 + r) - Hm)} & \text{if } \varepsilon = 0 \end{cases} \quad (3.22)$$

Although I will not explicitly write the bounds, I need to restrict the parameters so that the expression under the square-root is nonnegative.\footnote{One way to write the restriction is $\varepsilon \notin (\varepsilon, \bar{\varepsilon}) \subset (0, \infty)$, where $\varepsilon, \bar{\varepsilon}$ are the $\varepsilon$-roots of the expression under the square-root.} As long as $m(\tau(n_{0SB}^2 + Hw) + n_{0CB}^2) < \tau n_{0SB}^2((1 + r)w + \kappa p_1)$, which is likely since $m$ is small, then both roots are positive. This means there might be multiple equilibria in the housing market when shadow banks are active, in sharp contrast to when only commercial banks are active. I will call the equilibria associated with the smaller and
larger roots the “low-price” and “high-price” equilibria, respectively. However, only one of these equilibria is stable.

**Proposition 3.7.** The high-price equilibrium is a stable equilibrium while the low-price equilibrium is an unstable equilibrium.

Thus, going forward, I can restrict attention to the high-price equilibrium (henceforth, “equilibrium”). Intuitively, suppose the economy is in the high-price equilibrium and the price rises. Then there is excess supply, which drives down the price to reduce the excess supply; this process repeats until the price falls back to its equilibrium value and the excess supply disappears. Conversely, if the economy is in the low-price equilibrium and the price rises, there is excess demand, which drives up the price, away from the low-price equilibrium value and to the high-price one.

With prices in hand, I can once again analyze how changes to certain parameters affect the likelihood of each type of intermediary operating. In the analysis that follows, I will assume that if the economy is in the multiple-equilibria region, then changing parameters does not change the equilibrium mode of intermediation. Since the threshold \( \rho^* \) is increasing in \( p_0 \), decreasing in \( w \), increasing in \( r \), and generally decreasing in \( p_1 \), shadow banks lend to a larger share of the market in these cases, *if already active*. Otherwise, the results from Propositions 3.4 and 3.5 continue to hold (for relevant parameters). The only difference is that now, changes in certain parameters have both direct and indirect (through prices) effects. To see this, note that since \( p_{0CB} \) doesn’t depend on \( p_1 \) or \( r \), changing these parameters has only a direct effect on the commercial banking region, i.e., directly through \( D^{CB} \). However,

\[
\frac{\partial D^{SB}}{\partial p_1} = \frac{1 + r}{\hat{\rho}} \frac{\partial p_{0SB}}{\partial p_1} - \frac{(1 - \hat{\rho}) \kappa}{\hat{\rho}} < -\frac{(1 - \rho^P) \kappa}{\rho^P} = \frac{\partial D^{CB}}{\partial p_1}.
\]

That is, higher expectations drive down the face-value of shadow banks’ mortgages more than it drives down the face-value of commercial banks’ mortgages. Thus
shadow banking is more likely. Similarly,

$$\frac{\partial D_{SB}}{\partial r} = \frac{p_{SB}^0}{\rho} + 1 + r \frac{\partial p_{SB}^0}{\partial r} > \frac{p_{CB}^0 - w}{\rho^P} = \frac{\partial D_{CB}}{\partial r}$$

so that lower interest rates continue to spur shadow banking. Again, this says that lower rates (for example) make households more likely to take a shadow bank’s contract, not that commercial banks want to lend to fewer households.

I can now dive deeper into the credit expansion analysis from the baseline model. In particular, while the baseline model showed that shadow banking can explain the negative correlation between income growth and credit expansion, it cannot speak to how house prices are determined. To start, consider the effect of household wealth when commercial banks are active:

$$\frac{\partial p_{CB}^0}{\partial w} = \begin{cases} \frac{1}{2} \left[ 1 + \frac{\tau (\varepsilon w + H)}{\sqrt{\tau^2 (\varepsilon w - H)^2 + 4 \varepsilon \tau (n_0 + H \tau w)}} \right] & \text{if } \varepsilon > 0 \\ 1 & \text{if } \varepsilon = 0 \end{cases}.$$  

In both cases, the derivative is positive, which means house prices are higher in segments with higher income/lower debt-to-income ratios. This correlation is consistent with every recent period except the housing bubble. Further, differentiating (3.6),

$$\frac{\partial \ell_{CB}}{\partial w} = \frac{n_0}{\tau (p_0 - w)^2} \left( 1 - \frac{\partial p_{CB}^0}{\partial w} \right).$$

When $\varepsilon = 0$ this derivative is zero and $w$ has no effect on credit expansion because the higher price precisely balances the higher $w$; otherwise, it is easy to check that the total derivative is positive so higher $w$ implies higher $\ell_{CB}$. This increases the demand and hence increases the price. Although this increases the face-value of the mortgage, sending the feedback loop in the opposite direction, the first loop outweighs the second loop (and the third loop outweighs the fourth loop,...) and credit expands.
Now, consider the case when both types of intermediaries are active. The derivative of (3.22) with respect to \( w \) is

\[
\frac{\partial p_0^{SB}}{\partial w} = \begin{cases} 
\frac{1}{2} \left[ 1 + \frac{\tau \left( n_0^{SB} (1+r) - m(H+\varepsilon w) \right)}{\sqrt{\tau^2 \left( n_0^{SB} (1+r) - m(H+\varepsilon w) \right)^2 - 4\varepsilon m \left( \tau n_0^{SB} (\kappa p_1 - m) - m n_0^{CB} \right)}} \right] & \text{if } \varepsilon > 0 \\
1 & \text{if } \varepsilon = 0
\end{cases} \tag{3.23}
\]

As long as

\[
\frac{n_0^{CB}}{n_0^{SB}} < \frac{\tau (\kappa p_1 - m)}{m}, \tag{3.24}
\]

the above derivative is negative. This condition is likely to hold if (1) shadow banking is cheap relative to commercial banking (\( m \) is small relative to \( \tau \)), (2) prices are expected to be high in the future (\( p_1 \) is high), or (3) the shadow banking sector is already large relative to the commercial banking sector (\( n_0^{CB} \ll n_0^{SB} \)), all three of which are consistent with the mid-2000s. If the condition holds, then segments with lower \( w \) should have higher house prices, which is consistent with Fact 3. It is also consistent with the evidence that securitization played an important role since these patterns cannot be generated if commercial banks are the only intermediaries operating.

Moving to credit expansion, it is straightforward to check that

\[
\frac{\partial (\ell^{CB} + \ell^{SB})}{\partial w} < 0
\]

so that credit expansion and income growth are negatively correlated, as in the baseline model. Naturally, this correlation becomes more negative when the shadow banking sector is larger and prices are expected to be higher in the future.

Figure 3.7 depicts the feedback mechanism responsible for these results. First, lower income \( w \) increases the face-value of a mortgage since the household has to borrow more. Higher \( D \) allows shadow banks to sell more loans because investors’
participation constraints become easier to satisfy, reducing the skin-in-the-game requirement. Since shadow banks can sell more loans, they can boost leverage and make more mortgages. Since more households get mortgages, demand and hence the price rise. This price increase further drives up the face-value of the mortgage since the higher price means households have to borrow even more, and the cycle repeats. That is, shadow banks lend because other active shadow banks lend, and this mechanism is strongest when (3.24) is very easily satisfied.

![Credit expansion/price feedback loop](image)

Figure 3.7: Credit expansion/price feedback loop.

There is potentially yet another force at work, too. Suppose \( n_{0}^{CB} \) and \( n_{0}^{SB} \) are determined by the fraction of households they serve. That is, \( n_{0}^{SB} \) is an increasing function of \( \rho^{*} \) since higher \( \rho^{*} \) means shadow banks serve more types of households, so more intermediaries choose to operate as shadow banks. But lower \( w \) means higher \( \rho^{*} \) and this means \( n_{0}^{SB} \) increases, which in turn means \( \ell \) increases even more, which feeds back into \( p_{0}^{SB} \). Finally, \( \rho^{*} \) is also increasing in \( p_{0}^{SB} \) and the cycle repeats. This entire cycle is depicted with the dashed lines and shows that rising prices can bring more shadow banks into the market, which further raises prices. Summarizing, the model predicts that when shadow banks are active and securitization is prevalent, low-income segments experience more credit expansion and higher house prices, which in turn attracts more shadow banks and amplifies the process.

In the baseline model I showed that high prices today and tomorrow boost credit expansion. This complementarity still exists when \( p_{0} \) is endogenous. From (3.23), increasing \( p_{1} \) makes the derivative more negative, or \( \frac{\partial^{2} p_{0}^{SB}}{\partial p_{1} \partial w} < 0 \). This means that
low $t = 0$ income and high $t = 1$ price are complements in boosting the $t = 0$ price. Thus the model predicts that prices should be highest in segments with low income (growth) and high expectations for future prices, consistent with the evidence.

When the price is endogenous, if only commercial banks or no intermediaries are active, then the model is qualitatively identical to the baseline model with exogenous prices. Thus, the most interesting case is the one in which commercial and/or shadow banks can be active, i.e., parameter values for which (3.20) holds. Because there are multiple equilibria in this region, I can compare welfare across equilibria. In addition, the endogenous price introduces a pecuniary externality because one intermediary’s decision to extend credit affects the price at which other intermediaries lend.

Assume the economy is in the region with multiple equilibria. Then letting $W^{SB}$ denote welfare in the equilibrium with shadow banks and $W^{CB}$ welfare in the equilibrium with only commercial banks, $\Delta W = W^{SB} - W^{CB}$ is the difference in welfare. The actual expression is very complicated (it depends on all $\rho \in [0, 1]$ and on both $p^{CB}_0$ and $p^{SB}_0$) so it is nearly impossible to quantify welfare gains/losses across equilibria. While many of the conclusions from the model with $p_0$ exogenous go through, there is an important new one: with endogenous prices the economy is now fragile in the sense that small changes to certain parameters can have very large, discontinuous effects on the economy. This is because $\Delta W$ is not continuous in the parameters since $D^{SB}$ is a function of $p^{SB}_0$ while $D^{CB}$ is a function of $p^{CB}_0$. Then by slightly perturbing certain parameters, $D^{SB}$ and/or $D^{CB}$ can change in such a way that the economy moves to the multiple equilibria region. In one case, commercial banks are still the only ones active and not much changes while in other cases, shadow banks enter and welfare jumps depending on the gain from previously unserved subprime households. Conversely, the economy could be at the equilibrium where commercial and shadow banks coexist, but then jump to the one without shadow banks and credit could quickly dry up for subprime households. Even changes to “good” parameters (for example, increasing
$\pi$) can make things substantially worse if the economy selects the inferior equilibrium. Furthermore, as the wealth of shadow banks, $n_{0}^{\text{SB}}$, rises, the inequalities (3.20) hold more easily. In this sense, a larger shadow banking sector is also more unstable in that it expands the region with multiple equilibria. While equilibrium selection is beyond the scope of this paper, the model with endogenous prices nonetheless shows how a large shadow banking sector can contribute to financial instability.

### 3.7 Conclusion

I have presented a simple model of the housing market in which asymmetric information between intermediaries and outsiders leads to regulatory arbitrage and the expansion of the shadow banking system, which in turn drives credit expansion and house price growth in subprime markets. Outside investors do not trust intermediaries to not run away with their funds, but because commercial banks are regulated, they can raise funds via deposit contracts. This also gives them an incentive to monitor the quality of their balance sheets: they screen out subprime households and lend to only prime ones. On the other hand, shadow banks are unregulated, which means they cannot use any method of funding that makes a payment when their assets pay out: their only hope is to sell (securitize) the loan as soon as they make it. Because they are no longer residual claimants, shadow banks do not have an incentive to screen so they lend to both prime and subprime households.

The model makes a number of positive predictions that are borne out in the data. First, mortgages from segments in which a large fraction of mortgages are securitized have a higher default rate, which is consistent with Fact 2. Second, rising house prices go hand in hand with rising securitization, consistent with Fact 1. Third, when commercial banks operate, low-income segments should see less credit expansion while when shadow banks operate and securitization is prevalent, the opposite is true: low-
income segments see more credit expansion, providing a theoretical justification for Fact 3. Fourth, the model predicts that while the mortgage default rate declines as credit quality improves, it also shows that there exists a threshold at which the default rate jumps up, just as Keys, Mukherjee, Seru, and Vig (2010) document when they plot default rate against FICO score: below the threshold, non-securitizing commercial banks screen out subprime households because they have to keep all of the mortgages on their balance sheets, while above the threshold securitizing shadow banks do not screen because they do not have to hold all of the mortgages they make. The model also predicts that shadow banks are more likely to emerge when interest rates are low, prices are expected to grow, and securitization is easier, as was the case in the early- and mid-2000s.

Interestingly, and perhaps contrary to popular opinion, the model does not say that securitization is always excessive or inefficient. Instead, there are times when shadow banking is indeed the efficient form of intermediation (similarly, there are times when commercial banking is the efficient form of intermediation) and in fact, there are times when commercial banks are active but welfare would be higher if shadow banks were active instead.

When I extend the model to include many types of households and an endogenous house price, I uncover several new results. I show that when shadow banks are active, house prices increase most in low-income segments. Further, rising prices draw more wealth into the shadow banking sector, which further boosts credit expansion and house prices. I also show that there exists a region where multiple equilibria exist and while commercial banks are always active in this region, they may or may not coexist with shadow banking. This implies that shadow banking can lead to financial instability: small changes can shift the economy from a region with a unique equilibrium to another region with multiple equilibria, at which point prices and welfare jump (potentially significantly downward).
There are several potential model extensions and directions for future research based on asymmetric information. First, to maintain tractability, my model of the housing market is rather stylized and focuses on only the most salient features. In the model, households receive a one-period mortgage, are locked into that contract, and then repay or default in the next period. In reality, mortgages are long-term contracts that borrowers can renegotiate, default on early, or repay early. The first point is especially relevant for the recent build-up since low interest rates allowed households to refinance at lower rates, while Agarwal, Chang, and Yavas (2012) show that mortgage prepayment risk was as big of a factor, if not a bigger one, than was default risk for banks that originated and securitized mortgages. Thus it would be interesting to add more time periods to understand the roles of pre-payment risk and renegotiation in the face of seemingly more favorable borrowing conditions.

Second, the model focuses on the implications of the mortgage boom for the financial sector and pays little attention to the household side. Mian and Sufi (2011, 2014) show that during the housing bubble, households increasingly used the value of their houses to lever up and boost consumption, especially those in low-income ZIP codes. They then argue that this debt accumulation played a large role in the slow recovery once the housing bubble burst. Chen, Michaux, and Roussanov (2014) capture these facts in a dynamic model, but in their model all households are ex ante identical and house prices are exogenous. It would therefore to be interesting to see if my model’s mechanism can account for these phenomena, in addition to explaining why house prices moved as they did.

Finally, while the model makes a number of positive predictions, it does not make any normative ones in that it does not take a stand on optimal regulatory policies. It would be interesting to derive the optimal regulatory policies to deter excessive shadow banking. In particular, should regulatory authorities impose explicit regulatory costs as in the model (and how large should they be?), or something more
implicit such as capital requirements? Such policies should not only account for the excessive growth of the shadow banking sector, but factors that drive banking into the shadows to begin with.
3.8 Appendix: Omitted Proofs

A Section 3.4 Proofs

Proof of Lemma 3.1

As already argued, the only way a shadow bank can raise funds is by constructing a pass-through security. If a shadow bank screens and lends to a household, the household must be prime, in which case the face-value of the mortgage is

\[ D' = \frac{(1 + r) (p_0 - w) + f + \gamma_s - (1 - \rho P) \kappa p_1}{\rho P}; \]

this is just a shadow bank’s break-even condition (see (3.8)). If it does not screen, the household could be prime or subprime, in which case the face-value of the mortgage is

\[ D'' = \frac{(1 + r) (p_0 - w) + f - (1 - \bar{\rho}) \kappa p_1}{\bar{\rho}}. \]

By assumption, loans are placed randomly into an MBS, say with probability \( s > 0 \). Then shadow banks screen as long as the expected payoff from screening exceeds that of not screening. With probability \( 1 - s \) the shadow bank keeps the loan and gets its payoff and with probability \( s \) it gets the payoff from the MBS, which is always \((1 + r) (p_0 - w)\):

\[ (1 - s) (\bar{\rho} D'' + (1 - \bar{\rho}) \kappa p_1) \leq (1 - s) \left( \rho P D' + (1 - \rho P) \kappa p_1 \right) - \gamma_s. \]

Using the expressions for \( D' \) and \( D'' \) and rearranging yields \( s \leq 0 \), a contradiction.

Proof of Proposition 3.1

Let \((f_{s'}, D_{s'}, d)\) denote a mechanism. By the Revelation Principle, I can restrict attention to truth-telling mechanisms.
Now, a household’s utility is

\[
U^H = \begin{cases} 
  y_s + w + hf_s - D_s & \text{if } D_s \leq \min(y_s + w, h) \\
  y_s + w & \text{if } D_s > \min(y_s + w, h)
\end{cases}.
\]

The incentive constraint for a household to truthfully report \( s = H \) is

\[
hf_H - D_H \geq hf_L - D_L.
\]

By Assumption 3.2, \( y_L + w < \kappa p_1 \), and by Assumption 3.1, intermediaries prefer being repaid to foreclosing on the house. Together, these imply \( \kappa p_1 < D_H \), so that \( y_L + w < D_H \). But since \( y_L + w \) is the household’s maximum wealth if it receives low income, it will never be able to repay \( D_H \) and hence, it has no incentive to lie and report \( s = H \); thus the \( s = L \) incentive constraint automatically holds.

The two participation constraints for \( s = H, L \) are

\[
y_H + w + hf_H - D_H \geq y_H + w \implies hf_H - D_H \geq 0,
\]

\[
y_L + w + hf_L - D_L \geq y_L + w \implies hf_L - D_L \geq 0.
\]

Recall that by limited liability, \( D_L \leq y_L + w < D_H \). Now, suppose that \( f_H = 0 \), i.e., that an intermediary forecloses following a high report. Then

\[
D_L \geq hf_L + D_H > hf_L + y_L + w \geq y_L + w,
\]

a contradiction, so \( f_H = 1 \). If \( f_L = 1 \), i.e., that an intermediary does not foreclose following a low report, then \( h - D_H \geq h - D_L \implies D_L \geq D_H \), a contradiction; then the only way to satisfy the participation constraints is to set \( D_L = 0 \) and \( D_H \leq \min(y_H + w, h) \). This is a simple debt contract.
Proof of Proposition 3.2

Suppose that there exists a separating equilibrium in which subprime households are offered their own contract that will be accepted. Consider the following deviation by an intermediary: slightly lower the face-value of the mortgage designed for the prime borrowers. This intermediary will now capture the entire prime market, forcing other intermediaries to compete for the subprime borrowers. But since the subprime contract will be accepted and subprime borrowers are unprofitable, the other intermediaries would never offer this contract in the first place. Thus, for a separating equilibrium to exist, the contract meant for subprime borrowers must be designed to be rejected.

Let \((t, D^S, D^P)\) denote such a separating equilibrium, where \(t \leq w\) is the downpayment on the mortgage and \(D^j\) is the face-value of the mortgage meant for type-\(j\) borrowers, \(j \in \{S, P\}\). A pooling equilibrium will exist as long as such a separating equilibrium is not possible, namely when prime borrowers prefer accepting their contract to remaining in autarky and subprime borrowers prefer the contract meant for prime borrowers to remaining in autarky (which already dominates accepting the contract meant for them). The former condition implies

\[
w - t + h + \rho^P(y_H + h - D^P) + (1 - \rho^P)y_L \geq w + \rho^P y_H + (1 - \rho^P)y_L \implies h(1 + \rho^P) \geq t + \rho^P D^P.
\]

Meanwhile, subprime borrowers would rather accept the prime contract than remain in autarky if

\[
w - t + h + \rho^S(y_H + h - D^P) + (1 - \rho^S)y_L \geq w + \rho^S y_H + (1 - \rho^S)y_L \implies h(1 + \rho^S) \geq t + \rho^S D^P.
\]
If this latter inequality is satisfied for \( t = w \) then it is satisfied for all \( t \leq w \). Thus a sufficient condition for a pooling equilibrium to exist is \( w \leq h(1 + \rho^S) - \rho^S D^P \) and under Assumption 3.3, this holds as long as \( w \leq \frac{\rho^S}{1 - \rho^S} \gamma_a \).

**Proof of Theorem 3.1**

The proof has five steps. Step 1 proves some lemmas that characterize intermediaries’ profits under arbitrary contracts. Step 2 characterizes prime and subprime households’ strategies. In Step 3, I prove, case-by-case, that the outcome in the Theorem is indeed an equilibrium. To establish uniqueness, in Step 4 I prove that several outcomes cannot constitute an equilibrium and then, in Step 5, I go by case-by-case to rule out any remaining possibilities. Throughout the proof, I will normalize \( n_0 = 1 \) since its actual value is irrelevant for the results.

**Step 1: Preliminary Lemmas**

I begin with some preliminary results that I will use repeatedly in the proofs of the main results, especially Theorem 3.1. For \( \theta \in \{0, 1\} \), \( D \in \mathbb{R} \), \( T \in \{0, \tau\} \), \( F \in \{0, f\} \) define an intermediary’s per-loan expected profit conditional on the borrower’s type:

\[
\Pi(\theta, D, T, F) \equiv \theta \rho^P D + (1 - \rho^P) \kappa p_1 + (1 - \theta) \rho^S D + (1 - \rho^S) \kappa p_1 + F - (1 + r)(1 + T)(p_0 - w) - m.
\]  

(A-1)

When \( \theta = 0 \), this is the expected profit from a subprime borrower and when \( \theta = 1 \), this is the expected profit from a prime borrower.

**Lemma A1.** The profit function satisfies \( \pi \Pi(1, D^{SB}, 0, f) + (1 - \pi) \Pi(0, D^{SB}, 0, f) = f \) and \( \Pi(1, D^{CB}, \tau, 0) = \gamma_s \).

**Proof.** Since

\[
\Pi(1, D^{SB}, 0, f) = \rho^P D^{SB} + (1 - \rho^P) \kappa p_1 + f - (1 + r)(p_0 - w) - m
\]
\[ \Pi(0, D^{SB}, 0, f) = \rho^S D^{SB} + (1 - \rho^S) \kappa p_1 + f - (1 + r)(p_0 - w) - m, \]

plugging in \( D^{SB} \) from (3.8) and using \( f = m \) yields

\[
\pi \Pi(1, D^{SB}, 0, f) + (1 - \pi) \Pi(0, D^{SB}, 0, f) = \tilde{\rho} D^{SB} + (1 - \tilde{\rho}) \kappa p_1 + f - (1 + r)(p_0 - w) - m = f.
\]

Similarly, using \( D^{CB} \) from (3.7),

\[
\Pi(1, D^{CB}, \tau, 0) = \rho^P D^{CB} + (1 - \rho^P) \kappa p_1 - (1 + r)(1 + \tau)(p_0 - w) - m = \gamma_s.
\]

**Lemma A2.** The profit function \( \Pi \) is strictly increasing in \( D \) for \( \theta \in \{0, 1\} \).

**Proof.** Differentiating,

\[
\frac{\partial \Pi}{\partial D} = \theta \rho^P + (1 - \theta) \rho^S > 0.
\]

**Lemma A3.** The profit function \( \Pi \) is strictly decreasing in \( T \).

**Proof.** Differentiating,

\[
\frac{\partial \Pi}{\partial T} = -(1 + r)(p_0 - w) < 0.
\]

**Lemma A4.** The profit function satisfies \( \Pi(0, D, T, F) < F \) for all \( D \leq y_H \).
Proof. Under Assumption 3.1,

\[
\Pi(0, D, T, F) = \rho^S D + (1 - \rho^S) \kappa p_1 + F - (1 + r)(1 + T)(p_0 - w) - m
\]

\[
\leq \rho^S D + (1 - \rho^S) \kappa p_1 + F - (1 + r)(p_0 - w)
\]

\[
\leq \rho^S y_H + (1 - \rho^S) \kappa p_1 + F - (1 + r)(p_0 - w)
\]

\[
< F,
\]

where the second line uses Lemma A3.

Lemma A5. The profit function satisfies \( \Pi(1, D, \tau, F) \leq \gamma_s + F - m \) for all \( D \leq D^{CB} \).

Proof. By plugging in the definition of \( D^{CB} \),

\[
\Pi(1, D, \tau, F) \leq \Pi(1, D^{CB}, \tau, F) = \gamma_s + F - m,
\]

where the inequality follows from Lemma A2.

Step 2: Households’ Strategies

Consider a prime household; its utility from accepting loan contract \((D, \sigma)\) is

\[
U^{H,P} = h(1 + \rho^P) + \rho^P y_H + (1 - \rho^P) y_L - (\rho^P D + \sigma \gamma_a) \equiv h(1 + \rho^P) + \mathbb{E}^P[y] - \Gamma(D, \sigma),
\]

where \( \Gamma(D, \sigma) \equiv \rho^P D + \sigma \gamma_a \) is the cost associated with the contract. Since \( U^{H} \) is strictly decreasing in \( \Gamma(D, \sigma) \), the household chooses the contract with the lowest \( \Gamma(D, \sigma) \) such that his participation constraint (3.3) holds; if the participation constraint was violated, the prime household would exit the market, leaving only unprofitable subprime borrowers (in which case all of the intermediaries would exit the market). However, it is not necessarily the case that individual rationality implies,
or is it implied by, incentive compatibility so going forward, I still need to consider both constraints.

As for subprime households, they will never apply for a contract with $\sigma = 1$ since they incur the application cost $\gamma_a$ with no benefit (they will be immediately rejected). So restricting attention to contracts with $\sigma = 0$, a subprime borrower’s utility is

$$U_{H,S}^H = h (1 + \rho^S) + \rho^S y_H + (1 - \rho^S) y_L - \rho^S D = h (1 + \rho^S) + E^S[y] - \Gamma(D, 0).$$

Subprime borrowers will choose the contract with the lowest $\Gamma$, and hence the lowest $D$. Note that a subprime borrower’s participation constraint is

$$\frac{h - w}{\rho^S} + h \geq D,$$

which, by Assumption 3.3, holds automatically if the prime participation constraint (3.3) holds.

**Step 3: Existence**

**Case 1.** Consider the outcome posited by the Theorem. I will show that there are no strictly profitable deviations by any agent.

Start with households. Under the given contract, a prime household’s utility is

$$h (1 + \rho^P) + E^P[y] - \rho^P D^{SB} \geq h (1 + \rho^P) + E^P[y] - \rho^P h - h + w = E^P[y] + w,$$

which is the household’s utility from autarky, where the inequality uses the condition of Case 1. Thus the prime households apply for the contract. As argued above, if the prime participation constraint holds then so does the subprime one so subprime households apply for the contract, too.

Next, consider a deviation by any inactive intermediary, to the existing contract. A inactive shadow banks earns net profit $r$ while active shadow banks earn net profit
per loan in fees. Since, by Lemma A1, investors earn zero profit, an inactive shadow bank does not strictly gain by entering the market, matching the existing contract, and competing with active shadow banks. Suppose now that an inactive commercial bank enters by matching the existing contract. Then a commercial bank’s profit is

\[
\frac{\pi \Pi (1, D_{SB}, \tau, 0) + (1 - \pi) \Pi (0, D_{SB}, \tau, 0)}{\tau (p_0 - w)} = \frac{\tilde{\rho} D_{SB} + (1 - \tilde{\rho}) \kappa p_1 - (1 + r) (1 + \tau) (p_0 - w) - m}{\tau (p_0 - w)}
\]

\[
= - (1 + r) \tau < 0.
\]

Since the commercial bank can do better by remaining inactive, it will not choose to enter the market at the existing contract.

It remains to show that no active intermediary will deviate from the posited contract. Consider a deviation to \((D, 0)\), where \(D < D_{SB}\). Since the face-value of the contract is lower and there is no screening, this contract attracts all borrowers. Therefore, the shadow bank’s profit is

\[
\frac{\pi \Pi (1, D, 0, f) + (1 - \pi) \Pi (0, D, 0, f)}{s(D) (p_0 - w)} < \frac{\pi \Pi (1, D_{SB}, 0, f) + (1 - \pi) \Pi (0, D_{SB}, 0, f)}{s (p_0 - w)}
\]

\[
= \frac{\tilde{\rho} D_{SB} + (1 - \tilde{\rho}) \kappa p_1 + f - (1 + r) (p_0 - w) - m}{s (p_0 - w)}
\]

\[
= \frac{f}{s (p_0 - w)},
\]

by Lemma A2 and then Lemma A1. Thus no shadow bank will deviate to this contract. Obviously, no shadow bank will deviate to a contract with \(D > D_{SB}\) since this will not attract any borrowers.

Suppose a commercial bank enters the market with a contract \((D, 1)\), where \(D < D_{SB}\). Then it attracts either no households (because of the screening cost), or only prime borrowers. In the latter case, the commercial bank’s profit is

\[
\frac{\Pi (1, D, \tau, 0) - \gamma_s}{\tau (p_0 - w)} < \frac{\Pi (1, D_{CB} + \gamma_a / \rho P, \tau, 0) - \gamma_s}{\tau (p_0 - w)} = \frac{\gamma_s + \gamma_a - \gamma_s}{\tau (p_0 - w)} = - \frac{\gamma_a}{\tau (p_0 - w)} < 0,
\]
where the first inequality uses that $D < D^{SB} \leq D^{CB} + \gamma_a/\rho^P$ by assumption, and the equality uses Lemma A1. Thus undercutting shadow banks is unprofitable.

Finally, playing a mixed strategy that randomizes over contracts is unprofitable since households observe intermediaries’ entry decisions and contracts before applying. Then randomizing puts positive probability on pure strategies that yield (weakly) lower utility than the pure strategy equilibrium contract. Since there are no profitable deviations, the posited outcome is indeed an equilibrium.

**Case 2.** Consider the outcome posited by the Theorem. I will show that there are no strictly positive deviations by any agent.

Start with households. Under the given contract, a prime household’s utility is

$$h \left(1 + \rho^P\right) + \mathbb{E}^P[y] - \rho^P D^{CB} - \gamma_a \geq h \left(1 + \rho^P\right) + \mathbb{E}^P[y] - \rho^P h - h + w = \mathbb{E}^P[y] + w,$$

which is the household’s utility from autarky, where the inequality uses (3.11). Thus the prime households apply for the contract. As argued above, if the prime participation constraint holds then so does the subprime one.

I have to show that no commercial bank deviation yields positive profit, and no shadow bank deviation yields profit higher than $\ell^{SB} f$. First, I claim that $D^{CB} \leq D^{SB}$. To see this, recall that by assumption, either $D^{SB} \geq D^{CB} + \gamma_a/\rho^P$ or $D^{SB} \geq \min(y_H, h)$. In the former case, the claim holds trivially. In the latter case,

$$D^{SB} \geq \min(y_H, h) \geq \min \left(h + \frac{h - w - \gamma_a}{\rho^P}, y_H, h\right) \geq D^{CB}.$$

Now, suppose an inactive commercial bank enters the market by matching the existing contract. Then it earns the same profit so it is indifferent between entering or not. If an inactive shadow bank enters with the existing contract, its profit is

$$\Pi \left(1, D^{CB}, \tau, f, s, (D^{CB})(p_0 - w)\right) = \frac{f + \gamma_s - \gamma_s}{s (D^{CB})(p_0 - w)} < \frac{f}{s (p_0 - w)},$$
where I used Lemma A5 and that $s(D)$ is decreasing in $D$. Thus an inactive shadow bank will not enter at the existing contract.

Next, consider a deviation by an intermediary (inactive or not) to $(D, 1)$, where $D > D^{CB}$. This contract does not attract any borrowers so the offering intermediary does not generate any profit. A deviation to $(D, 0)$, where $D > D^{CB}$, attracts only subprime borrowers. By Lemma A4, an intermediary’s profit will be less than the fee it charges. Therefore, this contract is strictly dominated by the current contract (commercial banks) or by remaining inactive (shadow banks). Thus no intermediary will deviate to this contract.

If a shadow bank enters the market with the contract $(D, 0)$, where $D \leq D^{CB} \leq D^{SB}$, then its expected profit is, using Lemmas A1 and A2,

$$\frac{\pi \Pi(1, D, 0, f) + (1 - \pi) \Pi(0, D, 0, f)}{s(D)(p_0 - w)} \leq \frac{\pi \Pi(1, D^{SB}, 0, f) + (1 - \pi) \Pi(0, D^{SB}, 0, f)}{s(p_0 - w)}$$

$$= \frac{f}{s(p_0 - w)}.$$

Thus this deviation is not strictly profitable. Deviating to $(D, 1)$, where $D \leq D^{CB}$, yields expected profit

$$\frac{\Pi(1, D, \tau, f) - \gamma_s}{\tau(p_0 - w)} < \frac{\Pi(1, D^{CB}, \tau, f) - \gamma_s}{\tau(p_0 - w)} = 0,$$

again by Lemmas A1 and A2. Thus this deviation is not profitable either. Since there are no profitable deviations, the posited outcome is an equilibrium in this case.

**Case 3.** Under the posited outcome, no intermediary offers a contract and thus both are pushed to their outside options. I will show that there are no strictly positive deviations by any agent.
As a first step,

\[ D_{SB} \geq \min \left( D^{CB} + \frac{\gamma a}{\rho^P}, y_H, h \right) \geq \min \left( \frac{h - w}{\rho^P} + h, y_H + \frac{\gamma a}{\rho^P}, h + \frac{\gamma a}{\rho^P}, y_H, h \right) \geq \min (y_H, h). \]

Now, suppose an inactive intermediary enters as a shadow bank enters with the contract \((D, 0)\); if \(D > \min (y_H, h)\) then prime borrowers’ incentive constraint (3.4) is violated so the intermediary loses money since the only borrowers are subprime. So assume \(D \leq \min (y_H, h)\); then everyone accepts the contract and the shadow bank’s profit is

\[
\frac{\pi \Pi(1, D, 0, f) + (1 - \pi)\Pi(0, D, 0, f)}{s(D)(p_0 - w)} \leq \frac{\pi \Pi \left( 1, D_{SB}^{CB}, 0, f \right) + (1 - \pi)\Pi \left( 0, D_{SB}^{CB}, 0, f \right)}{s(p_0 - w)} = \frac{f}{s(p_0 - w)},
\]

where the first line uses that \(D \leq \min(y_H, h) \leq D_{SB}\) and Lemmas A2. Since I have argued that no matter what contract a shadow bank offers, it loses money, there is no strictly profitable deviation for inactive shadow bankers.

Meanwhile, if an inactive intermediary enters as a commercial bank offers \((D, 1)\) then by the same argument as the one above, it generates a negative profit unless \(D \leq \min (y_H, h)\). Thus to satisfy the prime participation constraint (3.3) and the upper bound on \(D\), I need

\[ D \leq \min \left( \frac{h - w - \gamma a}{\rho^P} + h, y_H, h \right) = D^{CB}, \]

where the last step is by assumption. But then the commercial bank’s profit is

\[
\frac{\Pi(1, D, \tau, f) - \gamma_s}{\tau(p_0 - w)} \leq \frac{\Pi(1, D^{CB}, \tau, f) - \gamma_s}{\tau(p_0 - w)} = 0.
\]
It follows that entering is (weakly) unprofitable for a commercial bank. Combined with the fact that shadow banks will not enter either, the posited outcome is an equilibrium.

**Step 4: Uniqueness**

I will establish a series of results that together place a large set of restrictions on possible equilibrium outcomes. I then complete the proof by ruling out any other outcomes for each case individually.

**Lemma A6.** *All active intermediaries impose the same expected cost on prime borrowers, \( \Gamma^i = \Gamma \equiv \rho P D + \sigma \gamma_a \), where \( \Gamma \) is the total expected cost for prime borrowers in equilibrium.*

*Proof.* Order the active intermediaries \( i \) by the cost they impose on prime borrowers, so that \( \Gamma^i < \Gamma^j \) if and only if \( i < j \). Now, suppose \( \Gamma^i < \Gamma^j \) for two active intermediaries, \( i < j \). Then clearly intermediaries \( i' > i \) will not serve any prime households, so, since they are active, they must be serving only subprime households. But then they expect to lose money so leaving the market is a strictly profitable deviation, contradicting the assumption that they are active in equilibrium. It follows that \( \Gamma^i = \Gamma^j \) for all active intermediaries \( i, j \). \( \square \)

**Lemma A7.** *There does not exist an equilibrium in which non-screening commercial banks are active.*

*Proof.* Let \((D, 0)\) denote the contract offered by an active, non-screening bank. By Lemma A6, this intermediary imposes the same cost on prime borrowers, \( \Gamma \), as do any other active intermediaries. Therefore, all active, non-screening intermediaries must offer \((D, 0)\). For this to be an equilibrium, there cannot be a deviation by either agent that yields higher utility than what they would get under the existing contract.
(and investors must break even):

\[
\frac{f}{s (p_0 - w)} \leq \frac{\pi \Pi (1, D, \tau, 0) + (1 - \pi) \Pi (0, D, \tau, 0)}{\tau (p_0 - w)} + \frac{\pi \Pi (1, D, 0, f) + (1 - \pi) \Pi (0, D, 0, f)}{s (p_0 - w)}
\]

\[
= \frac{\bar{\rho} D + (1 - \bar{\rho}) \kappa p_1 - (1 + r)(1 + \tau)(p_0 - w)}{\tau (p_0 - w)} + \frac{\bar{\rho} D + (1 - \bar{\rho}) \kappa p_1 + f - (1 + r)(p_0 - w) - m}{s (p_0 - w)}.
\]

The right side of the first line and the second line is the total expected profit of commercial banks and shadow banks under the contract \((D, 0)\). Evaluating the last line at \(D = D_{SB}\) yields, by Lemma A1,

\[
\frac{f - (1 + r) \tau (p_0 - w) - m}{\tau (p_0 - w)} + \frac{2f - m}{s (p_0 - w)} = -\frac{(1 + r) \tau (p_0 - w)}{\tau (p_0 - w)} + \frac{f}{s (p_0 - w)} < \frac{f}{s (p_0 - w)},
\]

where the final step follows from Assumption 3.1. Since the last line of the expression for the aggregate profit is increasing in \(D\), there exists \(D > D_{SB}\) such that the expression equals \(r\) at \(D = D\). Therefore, the offered contract must satisfy \(D > D\) to be profitable for intermediaries. Consider the following deviation: an inactive shadow banker can offer \(D' \in (D, D)\). This contract imposes a lower cost on prime borrowers so it attracts all of them. At best, these are the only borrowers it attracts and it is profitable; at worst, the contract also attracts all subprime borrowers, in which case the shadow bank’s profit is

\[
\frac{\pi \Pi (1, D', 0, f) + (1 - \pi) \Pi (0, D', 0, f)}{s(D')(p_0 - w)} > \frac{\pi \Pi (1, D_{SB}, 0, f) + (1 - \pi) \Pi (0, D_{SB}, 0, f)}{s(p_0 - w)}
\]

\[
= \frac{f}{s (p_0 - w)}.
\]

Thus a shadow bank can always undercut a non-screening banker, so the latter are never active in equilibrium.
Lemma A8. Non-screening shadow banks and screening commercial banks are simultaneously active in equilibrium with probability zero.

Proof. Let \((D^*, 1)\) denote the contract offered by commercial banks. Since commercial banks do not lend to subprime borrowers, they must serve some prime borrowers. Let \(D\) denote the face-value of the mortgage under the shadow bank contract. If \(D < D^{SB}\), then as proved above, a shadow bank’s profit is less than \(n_0 r\) so this contract isn’t profitable; thus it isn’t offered. If \(D > D^{SB}\) then another shadow bank could deviate to \(D' \in (D^{SB}, D)\), thereby imposing a lower cost on prime borrowers, and hence stealing the whole market while still earning a profit larger than \(f\) on each loan; the argument is the same as in Lemma A7. Thus \(D = D^{SB}\).

There are three cases to consider:

1. If \(D^* + \gamma_a / \rho^P = D^{SB}\), then prime borrowers are just indifferent between the commercial and shadow bank contracts, in which case both types of intermediaries can be active in equilibrium (but this is a probability-zero case).

2. If \(D^* + \gamma_a / \rho^P < D^{SB}\), then prime borrowers strictly prefer the commercial bank contract, so shadow banks serve only subprime borrowers, which is unprofitable. To see this, note that

\[
\frac{\Pi(0, D^{SB}, 0, f)}{s (p_0 - w)} = \frac{\rho^S D^{SB} + (1 - \rho^S) \kappa p_1 + f - (1 + r) (p_0 - w) - m}{s (p_0 - w)} < \frac{\rho^S y_H + (1 - \rho^S) \kappa p_1 + f - (1 + r) (p_0 - w) - m}{s (p_0 - w)} < \frac{f - m}{s (p_0 - w)} = 0.
\]

Thus the shadow bank will deviate and exit the market.

3. If \(D^* + \gamma_a / \rho^P > D^{SB}\), then prime borrowers strictly prefer the shadow bank contract, so commercial banks end up serving no one.
Thus, in the zero-probability first case, both intermediaries are active, in the second case only commercial banks are active, and in the third case, only shadow banks are active.

Lemma A9. If $\Gamma^{CB} > \Gamma^{SB}$ then commercial banks are not active in equilibrium; if $\Gamma^{SB} > \Gamma^{CB}$ then shadow banks are not active in equilibrium.

Proof. Assume $\Gamma^{CB} > \Gamma^{SB}$ and suppose there is an active commercial bank charging $D < D^{CB}$. Then Lemma A5 shows that this bank will receive a negative profit, so this cannot be an equilibrium contract since the bank could withdraw its contract and revert to autarky. Suppose instead that $D \geq D^{CB}$. Then $\Gamma \geq \Gamma^{CB}$, so an inactive shadow bank could enter with $D' \in \left(D^{CB}, \Gamma^{SB}/\rho^P\right)$. For this deviation, $\Gamma = \rho^P D' < \Gamma^{SB}$, so the shadow bank steals all prime borrowers and at worst, subprime borrowers as well in which case, as in Lemmas A7 and A8, the shadow bank still earns a higher profit. Thus, whenever a commercial bank is active, some intermediary has a profitable deviation. The argument for when $\Gamma^{SB} > \Gamma^{CB}$ is similar.

Step 5: Final Steps

Case 1. I just proved that in any equilibrium, either there are only non-screening shadow banks, or no contracts are offered at all. By assumption, $D^{SB} < D^{CB} + \gamma_a/\rho^P$ so that

$$\Gamma^{SB} = \rho^P D^{SB} < \rho^P D^{SB} + \gamma_a = \Gamma^{CB}.$$ 

By Lemma A9, this implies that commercial banks will not be active.

On the other hand, if there are no contracts offered at all, a shadow bank could deviate by entering the market by offering a contract with $D \in \left(D^{SB}, \min(y_H, h)\right)$, which would be accepted by all prime borrowers. Since $D < \min(y_H, h)$, prime households’ incentive constraint is satisfied so, as proved in Lemma A7, since $D > D^{SB}$, this deviation is profitable. Therefore, the no-contract equilibrium is impossible so the only possible equilibrium in this case features only non-screening shadow banks.
Case 2. The proof of this case is the similar to the proof of Case 1, and is omitted.

Case 3. I just proved that in any equilibrium, either there are only non-screening shadow banks, or no contracts are offered at all. In the existence proof for Case 3, I showed that no deviation is profitable, so no alternative can be an equilibrium. Therefore, no contract will be offered in equilibrium.

B Section 3.5 Proofs

Proof of Proposition 3.4

Since $\tau$ and $\gamma_s$ do not appear in any shadow bank quantities, decreasing them does nothing but lower $D_{CB}$, thus expanding the commercial banking region. Increasing $h$ expands the commercial banking region more than it expands the shadow banking region. When the prime repayment probability increases, then for $m$ sufficiently small,

$$
\frac{\partial D_{SB}}{\partial \rho^P} = -\pi [(1 + r)(p_0 - w) + m - \kappa p_1] > -\frac{(1 + r)(p_0 - w) + m - \kappa p_1}{\rho^{P^2}}
\frac{\partial D_{CB}}{\partial \rho^P}.
$$

Thus increasing $\rho^P$ causes $D_{CB}$ to fall more than $D_{SB}$ falls, which expands the commercial banking region and makes it more likely.

Proof of Proposition 3.5

From the conditions in Theorem 3.1, increasing $y_H$ expands the set of contracts shadow banks can offer while remaining profitable. When $\pi$ increases, this has no effect on $D_{CB}$ while

$$
\frac{\partial D_{SB}}{\partial \pi} = -\frac{(\rho^P - \rho^S)[(1 + r)(p_0 - w) + m - \kappa p_1]}{(\pi \rho^P + (1 - \pi) \rho^S)^2} < 0.
$$
Since increasing $\pi$ expands the shadow banking region while having no effect on the commercial banking region, shadow banking is more likely when $\pi$ is higher. When subprime households are more likely to repay, this has no effect on commercial banks because they do not lend to subprime households, while for shadow banks,

$$\frac{\partial D_{SB}}{\partial \rho^S} = \frac{-(1 - \pi) \pi [(1 + r) (p_0 - w) + m - \kappa p_1]}{(\pi \rho^p + (1 - \pi) \rho^S)^2} < 0.$$ 

Thus shadow bank mortgages are cheaper, making shadow banking more likely. Lowering $m$ directly lowers $D_{SB}$ while not affecting $D_{CB}$. As for the prices,

$$\frac{\partial D_{SB}}{\partial p_1} = -\frac{(1 - \tilde{\rho}) \kappa}{\tilde{\rho}} < -\frac{(1 - \rho^p) \kappa}{\rho^p}, \quad \frac{\partial D_{SB}}{\partial r} = \frac{p_0 - w}{\tilde{\rho}} > \frac{p_0 - w}{\rho^p},$$

That is, higher $p_1$ and lower $r$ drive down $D_{SB}$ more than they drive down $D_{CB}$. Since this expands the shadow banking region more than the commercial banking region, shadow banking is more likely.

**Proof of Theorem 3.2**

Before proving the main result, I will prove a preliminary lemma.

**Lemma B1.** The face-values of the mortgage contracts satisfy $\min(D_{CB}, D_{SB}) > 0$.

**Proof.** Since $m > 0$ and $\tilde{\rho} > 0$,

$$D_{SB} = \frac{(1 + r) (p_0 - w)}{\tilde{\rho}} + f - (1 - \tilde{\rho}) \kappa p_1 > \frac{(1 + r) (p_0 - w) - (1 - \tilde{\rho}) \kappa p_1}{\tilde{\rho}} > \frac{(1 + r) (p_0 - w) - \kappa p_1}{\tilde{\rho}} > \frac{(1 + r) (p_0 - w) - (\rho^S y_H + (1 - \rho^S) \kappa p_1)}{\tilde{\rho}} > 0,$$

179
where the last line holds by Assumption 3.1 since $\kappa p_1 < \rho^S y_H + (1 - \rho^S) \kappa p_1$ (which itself holds because $\kappa p_1 < y_H$). Also, since $\tau, \gamma_s > 0$,

$$D^{CB} = \frac{(1 + r) (1 + \tau) (p_0 - w) + \gamma_s + m - (1 - \rho^P) \kappa p_1}{\rho^P} \geq \frac{(1 + r) (p_0 - w) - (1 - \rho^P) \kappa p_1}{\rho^P} \geq 0,$$

as desired. \[\Box\]

I will prove each case for one parameter since the proofs for the other parameters are very similar.

**Case 1.** I will first show that there exists some $y$ at which there is no lending. Note that $D^{SB}$ does not depend on $y_H$ and by Lemma B1, $D^{SB} > 0$. By continuity, there exists $y > 0$ such that $D^{SB} > y \geq \min (y, h)$. But this violates (3.11) so if $y_H = y$, there is no shadow banking in equilibrium. Similarly, since $D^{CB} > 0$ by Lemma B1, the same argument shows that there is no commercial banking in equilibrium if $y_H = y$. Combining the two cases, there is no lending in equilibrium if $y_H = y$.

The next step is to show that there is no lending (of any kind) below this threshold. Since $D^{SB}$ does not depend on $y_H$ and $\min (y_H, h)$ is non-decreasing in $y_H$, lowering $y_H$ simply makes the no-shadow banking condition easier to satisfy, i.e., more slack. Therefore there is no shadow banking if $y_H \leq \underline{y}$. Since $D^{CB}$ does not depend on $y_H$ either and there is no commercial banking at $\underline{y}$, the same logic implies there is no commercial banking if $y_H \leq \underline{y}$. Therefore there is no lending at all if $y_H \in [0, \underline{y}]$.

Now, suppose there is some value of $y_H$ at which there is shadow banking in equilibrium, $\overline{y}$. Let $y_H > \overline{y}$. As per (3.11), $D^{SB} \leq y_H$. Since this becomes easier to satisfy as $y_H$ increases, there must be shadow banking if $y_H$ is large enough. Therefore, there is shadow banking in equilibrium if $y_H \geq \overline{y}$. By construction, there is no commercial banking if $y_H \geq \overline{y}$. Summarizing, there is no lending if $y_H \in [0, \underline{y}]$, commercial banking if $y_H \in (\underline{y}, \overline{y})$, and shadow banking if $y_H \geq \overline{y}$. 180
**Case 2.** As shown in the proof of Proposition 3.5, \( \frac{\partial D_{SB}}{\partial r} > 0 \). As \( r \to \infty \), then \( p_0^{SB} \to \infty \) so there exists \( \overline{r} \) such that \( D_{SB}(\overline{r}) \) is so large that it never induces participation, repayment or both (and since the proof also shows that \( \frac{\partial D_{SB}}{\partial r} > \frac{\partial D_{CB}}{\partial r} \), prime households will prefer the commercial bank mortgage anyway). Therefore, there is no shadow banking if \( r = \overline{r}_H \). Since \( p_0^{CB} \) is independent of \( r \), as \( r \to \infty \) then \( D_{CB} \to \infty \) as well, so for \( \overline{r} \) sufficiently large, there is no commercial banking either (because prime households will not participate, repay, or both).

If \( r > \overline{r} \) then \( D_{SB}(r) > D_{SB}(\overline{r}) \) so there is no shadow banking if \( r \geq \overline{r} \). Similarly, there is no commercial banking if \( r \geq \overline{r} \). That is, there is no mortgage lending if \( r \geq \overline{r} \).

Finally, let \( \underline{r}_H \) be such that there is shadow banking, so that \( D_{SB}(r) \) is low enough that it induces participation, repayment, and is cheaper than commercial banks’ mortgages. Then \( D_{SB}(r) \leq D_{SB}(\underline{r}) \) if \( r \leq \underline{r} \) so there is shadow banking if \( r \leq \underline{r} \). If \( r \in (\underline{r}, \overline{r}) \) then the commercial banks’ mortgages are preferred to the shadow banks’, so commercial banking emerges in equilibrium.

**Case 3.** As shown in the proof of Proposition 3.5, \( \frac{\partial D_{SB}}{\partial p_1} < \frac{\partial D_{CB}}{\partial p_1} < 0 \) so that rising expectations cause shadow banks’ mortgages to fall more quickly than do commercial banks’ mortgages. Therefore, as \( p_1 \) increases, \( D_{SB} \) hits zero first. Therefore, there exists \( \overline{p}_1 \) at which there is shadow banking. Since \( D_{SB}(p_1) < D_{SB}(\overline{p}_1) \) if \( p_1 > \overline{p}_1 \), there is shadow banking whenever \( p_1 \geq \overline{p}_1 \). By taking limits in the opposite direction, there exists \( \underline{p}_1 \) at which shadow banking is not profitable, and hence for \( p_1 \leq \underline{p}_1 \) as well. Similar logic applies to commercial banking, so there is no lending at all if \( p_1 \leq \underline{p}_1 \), commercial banking if \( p_1 \in (\underline{p}_1, \overline{p}_1) \), and shadow banking otherwise.

**Case 4.** Let \( \pi^* \) be such that there is shadow banking. Then the right side of (3.11) weakly dominates \( D_{SB}(\pi^*) \) and since the proof of Proposition 3.5 establishes that \( D_{SB} \) is decreasing in \( \pi \), then \( D_{SB}(\pi) < D_{SB}(\pi^*) \leq \min (\cdot, \cdot, y_H, h) \) if \( \pi > \pi^* \). Thus there is shadow banking if \( \pi > \pi^* \).
Now, suppose there is no shadow banking at some $\pi'$, so $D^{SB} (\pi') > \min (\cdot, \cdot, y_H, h)$. Then there is no shadow banking if $\pi < \pi'$ since then $D^{SB} (\pi) > D^{SB} (\pi')$. Whether or not a contract is offered at all, i.e., whether commercial banks are active or not, depends on (3.12), and none of the terms on the right side of that expression depend on $\pi$.

**Case 5.** It is clear that $D^{CB} \to \infty$ as $\gamma_s \to \infty$ so there exists $\gamma^*_s$ above which commercial banking is never profitable. Therefore, there’s never commercial banking in equilibrium at this point, or at any $\gamma_s \geq \gamma^*_s$.

Now, suppose there is commercial banking in equilibrium for some $\gamma^*_s$. Then (3.12) must hold so that

$$D^{CB} (\gamma^*_s) \leq \min \left( h + \frac{h - w - \gamma_a}{\rho^P}, y_H, h \right).$$

By monotonicity, $D^{CB} (\gamma_s) < D^{CB} (\gamma^*_s)$ if $\gamma_s < \gamma^*_s$, so there is commercial banking if $\gamma_s \in [0, \gamma^*_s]$. For $\gamma_s > \gamma^*_s$, whether or not a contract is offered at all, i.e., whether shadow banks are active or not, depends on (3.11).

The final step is to show, for one parameter (the reader can check others), that a commercial banking region actually exists between the no-lending and shadow banking regions. I will focus on $r$ and show that there exists a non-degenerate interval $(\underline{r}_H, \overline{r}_H)$ in which commercial banks are active; assume $y_H$ and $h$ are large enough that neither minimizes (3.11). Define $\underline{r}_H$ and $\overline{r}_H$, respectively, by

$$D^{SB} (\underline{r}_H) = D^{CB} (\underline{r}_H) + \frac{\gamma_a}{\rho^P},$$
$$D^{CB} (\overline{r}_H) = \min \left( h + \frac{h - w - \gamma_a}{\rho^P}, y_H, h \right).$$

By construction, prime households are just indifferent between commercial and shadow bank contracts at $r = \underline{r}_H$, so since $D^{SB}$ falls faster than does $D^{CB}$, shadow
banking is strictly preferred if \( r < \underline{r}_H \) and commercial banking is strictly preferred if \( r > \underline{r}_H \). By assumption,

\[
D^{CB}(\underline{r}_H) < \min \left( h + \frac{h - w - \gamma a}{\rho P}, y_H, h \right) = D^{CB}(\overline{r}_H)
\]

so \( \underline{r}_H < \overline{r}_H \), i.e., the interval is non-degenerate. Also, by construction, if \( r > \overline{r}_H \) then

\[
D^{CB}(r) + \frac{\gamma a}{\rho P} > D^{CB}(\overline{r}_H) + \frac{\gamma a}{\rho P} = \min \left( h + \frac{h - w}{\rho P}, y_H, h \right)
\]

so commercial banks will not be active. Since commercial banking already dominates shadow banking in this region, there is no lending.

**Proof of Corollary 3.1**

From Case 1 of Theorem 3.2, increasing \( y_H \) (while holding other parameters constant) moves the economy from a no-lending equilibrium, to a commercial banking one, to a shadow banking one if \( y_H \) is large enough. The proof also argued that at least one of these regions is non-empty. In the no-lending region, the probability of receiving a mortgage is zero for all households. In the commercial banking region, it is one of prime households and zero for subprime households, and in the shadow banking region it is one for all households. Thus the probability is non-decreasing for all households. Other cases in Theorem 3.2 take care of the other two parameters.

As for default probability, if \( y_H \) is sufficiently small then the default probability is undefined because no household receives a mortgage. In the commercial banking region, the probability remains undefined for subprime households while for prime households, the probability of default is \( 1 - \rho^P \). Finally, in the shadow banking region, prime households have the same default probability while for subprime households it is \( 1 - \rho^S \). In addition, the aggregate default rate goes from undefined, to \( 1 - \rho^P \), to 183.
\[ \pi (1 - \rho^P) + (1 - \pi) (1 - \rho^S) \]. These rates are decreasing, except at the threshold. Again, other cases in Theorem 3.2 take care of the other two parameters.

C Section 3.6 Proofs

Proof of Proposition 3.7

Let \( Z(p) \) denote the excess demand function when the house price is \( p \); the two roots occur at the prices in (3.22). Let \( p^* \) be such that \( Z'(p^*) = 0 \). Then

\[ p^* \in (P_{0\text{SB}}^-, P_{0\text{SB}}^+) \]

with \( Z'(p) > 0 \) for \( p < p^* \) and \( Z'(p) < 0 \) for \( p > p^* \). This means that the supply curve intersects the demand curve from below at the low-price equilibrium and from above at the high-price equilibrium. Thus the low-price equilibrium is unstable and the high-price equilibrium is stable.
Bibliography


