Essays in Dynamic Macroeconomic Policy

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Ali Shourideh

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Varadarajan V. Chari
Larry E. Jones

May, 2012
Acknowledgements

I am deeply indebted to Larry E. Jones and V. V. Chari, my advisors, for their constant support, guidance and advice throughout my graduate studies. Over the years, Chari has been my motivation for continuing research. His insights and novel ideas, have always improved my understanding of economics. He has always treated me as a colleague and collaboration with him has been, and I hope it will be, a fantastic experience. I cannot overstate my debt to Larry. Many of the ideas presented here have been formed through countless conversations with Larry over coffee. He has taught me that at the heart of complicated mathematical models, there are simple economic intuitions and how I should turn complex and immature ideas into simple economic models.

I owe special thanks to Patrick Kehoe, Chris Phelan who deepened my understanding of contracts, Narayana Kocherlakota who inspired the work in the third chapter, and Ellen McGrattan from whom I learned a great deal about data and computational work. They created a fantastic environment that makes Minnesota Economics a unique department. I would also like to thank Warren E. Weber, who provided me with the opportunity to spend time at the Minneapolis Fed.

Throughout my graduate studies, I have been lucky enough to work with my fellow classmate Ariel Zetlin-Jones. He has always patiently listened to my mostly incoherent ideas. I have learned a great deal from him and hope to continue working with him. Many thanks goes to Roozbeh Hosseini for collaboration, friendship and encouraging me to come to Minnesota. I have also enjoyed collaborating with Maxim Troshkin and hope to continue this in the future.
Many of my friends have contributed to this work through various discussions and helpful comments. In particular, I would like to thank Alessandro Dovis, Andrew Glover, and Erick Sager. I should also thank my friends Borghan Nezami Narajabad and Ehsan Ebrahimy. Without their inspiration, I would not have chosen economics as a profession.

Special thanks goes to my parents for all their efforts. Their constant and persistent encouragement has been my true motivation throughout my life.

Finally, I would like to thank my friend, colleague and wife, Maryam Saeedi. Maryam has been an amazing partner in all of these years. Her patience with my idiosyncrasies, her intelligent comments, and useful discussions, has made my life in graduate school an unforgettable experience. I dedicate this work to her.
Abstract

In this dissertation, we study optimal macroeconomic policy in dynamic environments.

In Chapter 1, we focus on optimal tax policies in environments with idiosyncratic capital income risk. We develop a model in which entrepreneurs are subject to idiosyncratic shocks to their capital income. Shocks to capital income have two components: 1) a component that is known to the entrepreneur at the time of investment, 2) a residual component that is realized after investment. This creates two types of incentive problems: a hidden type problem and a hidden action problem. We show that, absent private markets for insurance of idiosyncratic risk, entrepreneurial and non-entrepreneurial capital income should be taxed differently. Moreover, the government should subsidize non-entrepreneurial capital income when the known component is at its highest and lowest value. Furthermore, for a wide variety of distributions, the optimal tax schedule is progressive with respect to entrepreneurial capital income. Finally, the results regarding taxation of entrepreneurial income depend on the extent to which incentives and insurance are provided by private contracts. In particular, private contracts can approximately implement the efficient allocation if convertible securities are available. The prevalence of these securities in venture capital contracts suggest that the forces identified here are important in practice.

In Chapter 2, we study optimal intergenerational transmission of consumption and wealth with endogenous fertility. We use an extended Barro-Becker model of endogenous fertility, in which parents are heterogeneous in their labor productivity, to study the efficient degree of consumption inequality in the long run when parents productivity is private information. We show that a feature of the informationally constrained optimal insurance contract is that there is a stationary distribution over per capita continuation utilities there is an efficient amount of long run inequality. This contrasts with much of the earlier literature on dynamic contracting where immiseration occurs. Further, the model has interesting and novel implications for the policies that can be used to implement the efficient allocation. Two examples of this are: 1) estate taxes are positive and 2) there are positive taxes on family size.
In Chapter 3, we focus on optimal design of pension systems in providing incentives for efficient retirement. We study lifecycle environments with active intensive and extensive labor margins. First, we analytically characterize Pareto efficient policies when the main tension is between redistribution and provision of incentives: while it may be more efficient to have highly productive individuals work more and retire older, earlier retirement may be needed to give them incentives to fully realize their productivity when they work. We show that, under plausible conditions, efficient retirement ages increase in lifetime earnings. We also show that this pattern is implemented by pension benefits that not only depend on the age of retirement but are designed to be actuarially unfair. Second, using individual earnings and retirement data for the U.S. as well as intensive and extensive labor elasticities, we calibrate policy models to simulate robust implications: it is efficient for individuals with higher lifetime earning to retire (i) older than they do in the data and (ii) older than their less productive peers, in sharp contrast to the pattern observed in the data.

In Chapter 4, we focus on optimal policies in remedying problems in secondary loan markets. Loan originators often securitize some loans in secondary loan markets and hold on to others. New issuances in such secondary markets collapse abruptly on occasion, typically when collateral values used to secure the underlying loans fall and these collapses are viewed by policymakers as inefficient. We develop a dynamic adverse selection model in which small reductions in that a variety of policies intended to remedy market inefficiencies do not do so.
Contents

Acknowledgements i
Abstract iii
List of Tables ix
List of Figures x
1 Introduction 1
2 A Mirrleesian Approach to Capital Accumulation 5
  2.1 Introduction ................................................. 5
  2.2 A Two Period Example ...................................... 11
    2.2.1 Modified Inverse Euler Equation ....................... 16
    2.2.2 Wedges .................................................. 18
    2.2.3 Shape of the Consumption Schedule .................... 22
    2.2.4 Implementation .......................................... 25
  2.3 Multi-Period Model .......................................... 29
    2.3.1 Modified Inverse Euler Equation ....................... 34
  2.4 Optimal Taxes .............................................. 37
    2.4.1 Intertemporal Wedge .................................... 37
    2.4.2 Progressive Taxes on Entrepreneurial Income .......... 45
    2.4.3 A Tax Implementation .................................. 47
  2.5 Implementation with Private Contracts .................... 51
  2.6 Numerical Simulations ..................................... 56
2.7 Conclusion .................................................. 60

3 Risk Sharing, Inequality, and Fertility ............. 62
  3.1 Introduction .......................................... 62
  3.2 An Example and Intuition .......................... 67
    3.2.1 A Two Period Example ....................... 68
    3.2.2 Resetting – Intuition via Homotheticity .... 71
  3.3 The Infinite Horizon Model ....................... 73
  3.4 Properties of the Model ......................... 77
    3.4.1 The Resetting Property ...................... 78
    3.4.2 Stationary Distributions .................... 80
    3.4.3 Inverse Euler Equation and a Martingale Property .......... 84
  3.5 Extensions and Complementary Results .......... 88
    3.5.1 Implementation: A Two Period Example ........ 89
    3.5.2 Social vs. Private Discounting .............. 92

4 Providing Efficient Incentives to Work:
  Retirement Ages and Pension System ............. 94
  4.1 Introduction .......................................... 94
  4.2 Environment ......................................... 99
  4.3 Distortions and policies ......................... 103
  4.4 Characterization of efficient retirement ....... 105
    4.4.1 Retirement ages in a baseline case ....... 106
    4.4.2 Labor and retirement distortions .......... 110
  4.5 Actuarially unfair pension system ............ 112
  4.6 Quantitative analysis ............................ 115
  4.7 Conclusion .......................................... 126

5 Adverse Selection, Reputation, and Sudden Collapses
  in Secondary Loan Market ......................... 127
  5.1 Introduction .......................................... 127
    5.1.1 Related Literature ........................... 130
  5.2 Evidence on Sudden Collapses ................... 132
5.3 Reputation in a Secondary Loan Market Model .......................... 134
5.3.1 Static Model: A Unique Equilibrium ................................. 136
5.3.2 Two-Period Benchmark Model ......................................... 140
5.3.3 Sudden Collapses and Increased Inefficiency ......................... 145
5.4 Aggregate Shocks and Uniqueness ......................................... 147
5.4.1 Outline of Proof with Improper Priors ............................... 150
5.4.2 Uniqueness Result with Proper Priors ............................... 154
5.5 The Multi-Period Model .................................................... 155
5.6 Fragility ............................................................................ 156
5.7 Policy Exercises .............................................................. 160
5.8 Conclusion .......................................................................... 162

Bibliography ............................................................................. 164

Appendix A. Appendix to Chapter 2 ........................................... 176
A.1 Proofs .............................................................................. 176
A.1.1 Proof of Proposition 2.1. ................................................ 176
A.1.2 Proof of Lemma 2.2 ......................................................... 178
A.1.3 Proof of Proposition 2.3 ................................................... 179
A.1.4 Proof of Theorem 2.9 ....................................................... 180
A.1.5 Proof of Theorem 2.11 ..................................................... 181
A.1.6 Proof of Proposition 2.12 ............................................... 183
A.1.7 Proof of Lemma 2.16 ....................................................... 184
A.2 Sufficient Conditions for FOA ............................................ 186

Appendix B. Appendix to Chapter 3 ........................................... 192
B.1 Proofs .............................................................................. 192
B.1.1 Proof of Proposition 3.6 ................................................ 192
B.2 Implementation ............................................................... 197
B.2.1 Distortions .................................................................. 197
B.2.2 Proof of Remark 3.12 ................................................... 200
Appendix C. Appendix to Chapter 4  
C.1 Proofs ................................................................. 202
  C.1.1 Proof of Lemma 4.2 ........................................... 202
  C.1.2 Proof of Lemma 4.3 ........................................... 203
  C.1.3 Proof of Proposition 4.5 ................................. 203
  C.1.4 Proof of implementation ............................... 204
C.2 Existence of retirement age .................................. 208
C.3 Sufficiency of first-order approach ..................... 209

Appendix D. Appendix to Chapter 5  
D.1 Proofs ................................................................. 211
  D.1.1 Proof of Lemma 5.8 ........................................... 211
  D.1.2 Proof of Lemma 5.9. .......................................... 212
  D.1.3 Proof of Theorem 5.10. ................................. 214
  D.1.4 Proof of Proposition 5.11. ........................... 215
D.2 Full Characterization of Equilibria in Two Period Game .... 218
D.3 Strategic Types ....................................................... 221
List of Tables

4.1 Parameters. ......................................................... 116
4.2 Summary statistics. .............................................. 118
4.3 Regression results. .............................................. 122
4.4 Retirement ages for the U.S. ................................. 124
# List of Figures

2.1 Perturbation ................................................................. 23  
2.2 Intertemporal wedge, full model .......................................... 58  
2.3 Intertemporal wedge, absent residual component ....................... 59  
2.4 Intertemporal wedge, fully persistent ability ............................ 60  
4.1 Empirical weighted average and simulated efficient retirement ages for the U.S., by lifetime earnings decile. Sources: HRS, PSID, and authors’ calculations. .......................................................... 99  
4.2 Estimated productivity profiles by type over lifecycle, \( \varphi(t, \theta) \). .......... 117  
4.3 Unconditional distribution of retirement ages (as defined in the main text) in the HRS. ......................................................... 119  
4.4 Retirement ages vs. average annual labor earnings. ................. 120  
4.5 Retirement ages vs. logarithm of average annual labor earnings. .... 121  
4.6 Empirical weighted average (left panel) and simulated efficient retirement ages (right panel) for the U.S., by lifetime earnings decile. .......... 123  
4.7 Labor distortions (left panel) and retirement distortions (right panel). . 125  
5.1 New Issuance of Asset-Backed Securities (Source: JP Morgan Chase) . 133  
5.2 Change in Stock of Real Estate Bonds, 1920-1930 ........................ 135  
5.3 Cutoff Thresholds for High-Quality Banks. ............................ 158  
5.4 Invariant Distribution of Reputations of High-Quality Banks ........... 158  
5.5 Volume of Trade as a Function of shock to Default Value. .......... 159
Chapter 1

Introduction

The design of optimal government policies is one of the most important issues in macroeconomics and public finance. This dissertation is a theoretical and quantitative investigation of designing optimal policies in dynamic environments. In the following four chapters, we focus on optimal taxation of capital income in presence of capital income, optimal intergenerational transmission of wealth and consumption in presence of fertility motives, optimal design of pension system as an integral part of the tax system in order to provide efficient incentives for retirement, and the design of optimal policies in secondary loan markets.

In Chapter 2, we study optimal design of capital taxes in an economy with capital income risk. The presence of capital income risk can significantly change the policy implications prescribed by the literature on optimal capital taxation. In fact, previous studies have mainly focused on economies with labor income risk. In these economies, the only rational for saving is smoothing of consumption as well as insurance against future labor income risk.

To this end, We develop a model in which owners of capital or entrepreneurs are subject to idiosyncratic shocks to their capital income. Shocks to capital income have two components: 1) a component that is known to the entrepreneur at the time of investment, 2) a residual component that is realized after investment. This creates two types of incentive problems: a hidden type problem and a hidden action problem. We show that, absent private markets for insurance of idiosyncratic risk, entrepreneurial and
non-entrepreneurial capital income should be taxed differently. Moreover, the government should subsidize non-entrepreneurial capital income when the known component is at its highest and lowest value. Furthermore, for a wide variety of distributions, the optimal tax schedule is progressive with respect to entrepreneurial capital income. Finally, the results regarding taxation of entrepreneurial income depend on the extent to which incentives and insurance are provided by private contracts. In particular, private contracts can approximately implement the efficient allocation if convertible securities are available. The prevalence of these securities in venture capital contracts suggest that the forces identified here are important in practice.

In Chapter 3, based on joint work with Larry E. Jones and Roozbeh Hosseini, we study optimal intergenerational transmission of wealth and consumption in presence of fertility decisions. This issue is particularly important in determination of optimal amount inequality in consumption and wealth. The answer to this question can potentially be useful in how much inequality in consumption and wealth should governments allow. A useful framework for such analysis is dynamic models with private information. While a useful framework, they all feature a common result: optimal inequality in the long run is infinity. That is, in the long run a shrinking fraction of households owns a growing share of wealth and this fraction converges to 1. We show that adding fertility motives resolves this issue.

We use an extended Barro-Becker model of endogenous fertility, in which parents are heterogeneous in their labor productivity, to study the efficient degree of consumption inequality in the long run when parents productivity is private information. We show that a feature of the informationally constrained optimal insurance contract is that there is a stationary distribution over per capita continuation utilities there is an efficient amount of long run inequality. This contrasts with much of the earlier literature on dynamic contracting where immiseration occurs. Further, the model has interesting and novel implications for the policies that can be used to implement the efficient allocation. Two examples of this are: 1) estate taxes are positive and 2) there are positive taxes on family size.

In Chapter 4, based on joint work with Maxim Troshkin, we study optimal provision of incentives for efficient retirement. As noted by many studies, there is significant evidence that pension systems (such as United States Social Security system) together
with income taxes provide incentives for earlier retirement.

In this chapter, we focus on a theoretical and quantitative analysis of the efficient pension system as an integral part of the income tax code. We study lifecycle environments with active intensive and extensive labor margins. First, we analytically characterize Pareto efficient policies when the main tension is between redistribution and provision of incentives: while it may be more efficient to have highly productive individuals work more and retire older, earlier retirement may be needed to give them incentives to fully realize their productivity when they work. We show that, under plausible conditions, efficient retirement ages increase in lifetime earnings. We also show that this pattern is implemented by pension benefits that not only depend on the age of retirement but are designed to be actuarially unfair. Second, using individual earnings and retirement data for the U.S. as well as intensive and extensive labor elasticities, we calibrate policy models to simulate robust implications: it is efficient for individuals with higher lifetime earning to retire \( i \) older than they do in the data (at 69.5 vs. at 62.8 in the data, for the most productive workers) and \( ii \) older than their less productive peers (at 69.5 for the most productive workers \( vs. \) at 62.2 for the least productive ones), in sharp contrast to the pattern observed in the data. Finally, we compute welfare gains of between 1 and 5 percent and total output gains of up to 1 percent from implementing efficient work and retirement age patterns. We conclude that distorting the retirement age decision offers a powerful novel policy instrument, capable of overcompensating output losses from standard distortionary redistributive policies.

Finally in Chapter 5, based on joint work with V. V. Chari and Ariel Zetlin-Jones, we study policies that remedy inefficiencies in secondary loan markets. This issue is particularly important given the events that occurred at the onset of the 2007-2009 recession.

We start by studying the determinants of the decision of whether to hold or to sell loans. Secondary loan markets are often argued to suffer from adverse selection problems when originators of loans are better informed than potential purchasers regarding the quality of the loans. We, then, analyze the role of reputation in mitigating such adverse selection problems. We argue that reputation can both be a blessing and curse, in the sense that reputational incentives lead to multiplicity of equilibria. In one of these
equilibria, reputational forces help mitigate the adverse selection problem while in the other reputational forces actually worsen the adverse selection problem. We use a refinement adapted from the global games literature which leads to a unique equilibrium. This equilibrium is fragile in the sense that small fluctuations in fundamentals can lead to large changes in the volume of loans sold in the secondary market. Our model is consistent with the recent collapse in the volume of loans sold in the secondary market in the United States. We analyze a variety of policies that have been proposed to resolve adverse selection problems in the secondary loan market. We find that many such policies do not help resolve this problem and, indeed, worsen the allocative efficiency of the secondary loan market.
Chapter 2

A Mirrleesian Approach to Capital Accumulation

2.1 Introduction

How should wealth be taxed? The answer to this question requires taking a stand on the process of wealth accumulation. Much of economic theory has tackled this question by using models where households are subject to idiosyncratic labor income risk and accumulate wealth as buffer against future income shocks. However, it has been documented that models with idiosyncratic labor income risk fail to generate a concentration of wealth similar to that observed in the data. It has also been argued that models with entrepreneurs who are subject to capital income risk can generate a concentration of wealth similar to that in the data\(^1\). In this paper, motivated by this insight, I study optimal taxation of entrepreneurial income and wealth.

I analyze optimal design of tax schedules by developing a model where entrepreneurs are subject to idiosyncratic capital income risk and private information. The productivity of investment projects stochastically evolves over time. In particular, productivity has two components, a component that is known by the entrepreneurs in advance at the time of investment and a residual component that is realized once investment is

\(^1\) Aiyagari, 1994]’s seminal paper is an example with idiosyncratic labor income risk that fails to capture the concentration of wealth among the wealthy. For successful models with capital income risk, see [Quadrini, 2000], [Cagetti and De Nardi, 2006], and [Benhabib and Bisin, 2009].
made\footnote{This environment nests the models of entrepreneurship in [Evans and Jovanovic, 1989] and [Gentry and Hubbard, 2004].}. The first component of productivity can be interpreted as entrepreneurial ability. I assume that productivity, investment and consumption are all private information to the entrepreneur. In such an environment, a planner would want to insure entrepreneurs against productivity and income risk via redistributive schemes. These redistributive motives together with private information, leads to a trade-off between incentives to invest and insurance as in [Mirrlees, 1971]; hence a Mirrleesian approach to capital accumulation.

In this environment, I ask two sets of questions: First, when entrepreneurs cannot insure themselves against idiosyncratic productivity risk, how should the government design the tax schedule? In particular, how should the government tax capital income of entrepreneurs from their businesses and non-entrepreneurial capital income, i.e., financial wealth. Second, if we do not restrict private agents to a particular set of contracts but rather, allow them to sign insurance contracts, can they achieve efficiency? If so, can a set of standard securities implement the optimal allocation? Do we observe these contracts for entrepreneurs?

Regarding the first set of questions, I have two main theoretical results. First, existence of heterogeneity in entrepreneurial ability introduces forces toward subsidies to non-entrepreneurial capital income, i.e., financial wealth. In particular, in an extreme case where there is no residual component, wealth taxes are negative for entrepreneurs with lowest and highest ability. When both components are significant, the results are mixed. Using a calibrated version of the model, I find that wealth taxes are negative for entrepreneurs with lowest ability and positive for entrepreneurs with highest ability. Moreover, I show that when the residual component of productivity is significant, the tax schedule with respect to business income is progressive for a wide variety of specification for the distribution of shocks.

As for the second set of questions, in this environment, private agents can achieve constrained efficiency when they can sign an unrestricted set of contracts. My main contribution, here, is to show that the optimal allocation can be implemented with a set of standard securities. In fact, one can reinterpret the model as a contract between an entrepreneur and a venture capitalist, with the optimal contract implemented using
equity, convertible debt and a credit line/saving account with a variable interest rate.

To derive these results, I first study the properties of the constrained efficient allocations over time and in the cross-section. Using a first order approach, I derive a modified version of the Inverse Euler Equation (see [Rogerson, 1985a], [Golosov et al., 2003]). I use this equation to characterize the optimal distortions to intertemporal margin of saving, i.e., the intertemporal wedge, and hence the marginal tax rate on wealth. In particular, the intertemporal wedge is the highest when current incentive constraints are very tight relative to future constraints and vice versa. Unfortunately, in this environment, our modified version of the Inverse Euler Equation cannot be used to determine the sign of the intertemporal wedge. However, the recursive formulation of the problem can be used to see that when there is no residual productivity shock, the intertemporal wedge is negative for the highest and lowest realizations of productivity.

To provide an intuition for the negative intertemporal wedge result, I describe how different forces are at play when capital income is risky as opposed to a situation where labor income is risky, as is typical in the dynamic Mirrlees literature. When labor income is risky, an extra unit of saving decreases marginal utility of agents in the future and decreases their labor supply, i.e., it tightens incentive constraints in the future. Hence, a planner wants to discourage agents from saving in order to provide incentives for working in the future. When capital income is risky, it is the opposite. An extra unit of saving causes agents to invest more since they have more resources available for consumption and investment. Hence, saving relaxes future incentive constraints. However, saving is not without cost. In fact, it tightens the incentive constraints in the current period. Since the incentive constraints are not binding for the highest and lowest value of productivity, when current productivity is at its highest and lowest value, an extra unit of saving has no effect on current incentives. Hence, the planner wants to encourage saving for the most and least productive agents.

To prove the progressivity result, I characterize the properties of consumption in the cross-section by deriving a simple equation that relates consumption to income. When the utility function is of the CARA form, this equation implies that the inverse of marginal utility is a linear function of the hazard ratio implied by the distribution of shocks to income. For a large class of distributions for the residual component of productivity it can be shown that the hazard ratio is concave in the income realization.
Concavity of the hazard ratio implies that the consumption schedule is concave in income and thus the tax schedule with respect to business income is progressive\(^3\).

Although one interpretation of the model is of optimal taxation, I show that it is not necessary for the government to tax entrepreneurs in order to achieve efficiency. In particular, there is a private implementation of the optimal contract using standard securities: equity, convertible debt and a credit line/saving account with a variable interest rate. The role of each security can be associated with the properties of the constrained efficient allocation described above. The presence of convertible debt – a security that is similar to debt but can be converted to equity at a pre-specified price – implies that entrepreneur’s consumption is a concave function of income. Hence, this feature can create the relationship between consumption and income in the constrained efficient allocation. The credit line/saving account with a variable interest introduces an intertemporal wedge in the saving margin of the entrepreneur as in the constrained efficient allocation. The significance of this implementation is that it resembles venture capital contracts. In fact, as noted by [Kaplan and Strömberg, 2003], [Sahlman, 1990] and [Gompers, 1999], a major fraction of securities used in venture capital contracts are in the form of convertible securities, i.e., convertible preferred stock, participating preferred stock, etc. Hence, this implementation sheds light on forces behind the widespread use of convertible securities in venture capital contracts. Moreover, it provides a justification for the forces identified in the model.

In deriving the above results, two implicit assumptions have been made. First, the economy is populated only by entrepreneurs. This feature, however, is not critical regarding the distortions implied by taxes. In particular, it is easy to extend this environment to an environment in which workers and entrepreneurs are distinguishable. In that environment, since a planner can distinguish between workers and entrepreneurs, the efficient allocations can be achieved by a lump sum transfer from entrepreneurs to workers along with taxes/private contracts to achieve efficiency within each group. Second, there is no entry into entrepreneurship. Adding this feature would make the model less tractable thereby making the main forces in the model harder to identify. I leave this extension for future work.\(^4\).

\(^3\) In the two period environment, this result is more general. It holds whenever, \(1/u'(c)\), is a convex function of \(c\). In the special case where \(u(c) = c^{1-\sigma}/(1 - \sigma)\), we must have \(\sigma > 1\).

\(^4\) See [Scheuer, 2010] for an analysis of the entry decision in a static economy.
The theoretical results in this paper point to a need for an important empirical question: How successful are credit markets in providing efficient investment incentives for entrepreneurs? As the analysis in this paper shows, the optimal design of nonlinear taxes for entrepreneurs depends on the answer to this question. As I have shown, contracts with features similar to venture capital can achieve efficiency. However, since venture capital is a small portion of private equity market, a more rigorous analysis of the credit market contracts is needed to answer this question.

Related Literature. This paper builds on the literature on optimal dynamic taxation (see [Golosov et al., 2003], [Farhi and Werning, 2010a], [Golosov et al., 2010] among others.) This literature has mainly focused on environments with idiosyncratic labor income risk and their implications about dynamic taxation of various sources of income. In this paper, I study optimal taxation of various sources of income in a model with capital income risk and show that capital income risk overturns some of the main lessons from the literature, namely that the intertemporal wedge can be negative.

This paper is also related to a growing literature on the effect of taxation on entrepreneurial behavior. [Cagetti and De Nardi, 2009] consider the effect of elimination of estate taxes on wealth accumulation. [Kitao, 2008] and [Panousi, 2009] study how changes in the capital income tax rate affects investment by entrepreneurs. However, none of these studies considers the optimal taxation of entrepreneurial income. In developing my model of entrepreneurs, I have relied on their benchmark models while abstracting from some details for higher tractability. [Albanesi, 2006] and [Scheuer, 2010] are early attempts in studying optimal design of tax system for entrepreneurs. [Albanesi, 2006] focuses on specific implementation of optimal contracts and [Scheuer, 2010] focuses on the decision of entry into entrepreneurship and its implication for differential treatment of entrepreneurs and workers.

An important implication of my paper is the emergence of wealth subsidies when entrepreneurs are subject to capital income risk. This result is related to a large literature on optimal capital taxation including [Chamley, 1986], [Judd, 1985], [Kocherlakota, 2005], and [Conesa et al., 2009a], among others. In most of these studies the optimal tax rate on capital income/wealth is positive or zero\(^5\). Exceptions are [Farhi and Werning, 2008]

\(^5\) [Kocherlakota, 2005] actually shows that wealth taxes are zero in expectation and hence some time negative and some time positive. However, that result is specific to a particular implementation and there are other implementations for which capital income tax rate is equal to the investment wedge
and [Farhi and Werning, 2010b] in which negative marginal tax rates emerge either as a result of a higher social discount factor or binding enforcement constraints in the future. In my model, however, subsidies are optimal since they relax future incentive constraints. To my knowledge, this is the first paper that identifies this force.

In deriving optimal progressivity of the tax code with respect to business income, my paper is related to a small number of papers that study optimal progressivity of the tax system ([Varian, 1980], and [Heathcote et al., 2010]). The most related paper is perhaps [Varian, 1980]. In a two period model that shares similar properties to our model, he shows that it is optimal for the government to make the marginal tax rate an increasing function of income. The model in this paper nests his model and extends it to a dynamic environment with productivity risk. Moreover, I show that the progressivity result holds for a large class of distributions.

In this paper, I show that the constrained optimal allocation can be implemented using a set of standard securities that are widely used in venture capital contracts. This result is related to the literature on optimal firm financing and optimal capital structure. [DeMarzo and Fishman, 2007] and [DeMarzo and Sannikov, 2006] show that in a dynamic model with non-verifiable income the optimal contract can be implemented using credit lines, equity and debt. [Biais et al., 2007] show that in the same environment the optimal allocation can be implemented using cash reserves, debt and equity and use this implementation to study its implication for dynamics of security prices. Finally, [Clementi and Hopenhayn, 2006] consider a moral hazard model and show that the optimal allocations can be implemented using short term debt and equity. The implementation in this paper points to the special role of convertible securities, equity buy backs and credit lines in creating the right incentives for the entrepreneur to invest optimally.

Finally, from a technical point of view, the model in this paper contains two main frictions, a hidden action problem and hidden type problem. In general, this makes the problem very hard to analyze. However, I use the first order approach, as in [Pavan et al., 2009], to simplify the set of incentive constraints and we derive conditions under which this first order approach is valid. Since there are two types of private information, this model shares the same structure as the model in [Laffont and Tirole, 1986] and hence positive; see [Werning, 2010].
who study optimal regulation of a monopolist and more recently [Garrett and Pavan, 2010] and [Fong, 2009].

The rest of the paper is organized as follows: section 2 describes a two period version of the model in order to identify the key economic forces at play. In section 3, we develop the multi-period model and derive the modified inverse Euler Equation. In section 4, we study the intertemporal wedge. Section 5, generalizes the shape of the tax function

2.2 A Two Period Example

In this section, we focus on a two period economy in order to identify the key economic forces. We start with a two period example to show one of the main results of the paper – progressivity. As we see, the Modified Inverse Euler Equation – an equation governing time series properties of consumption – proves useful in the analysis of the intertemporal wedge. Hence, we derive a version of it for the two period example and later extend it to the general environment.

Consider a two period economy in which \( t = 0, 1 \). The economy is populated by a continuum of entrepreneurs. Each entrepreneur is the sole owner of an investment technology or project that is subject to idiosyncratic risk. In particular, entrepreneurs draw a productivity shock, \( \theta \in [\underline{\theta}, \bar{\theta}] \), at \( t = 0 \). I assume that \( \theta \) is distributed according to the distribution function \( F(\theta) \). I also assume that \( F(\cdot) \) is differentiable over the interval \( [\underline{\theta}, \bar{\theta}] \) and \( f(\theta) = F'(\theta) \). The value of the shock, \( \theta \), determines the distribution of returns to individual investment. If an entrepreneur with type \( \theta \) invests \( k_1 \) in his private project, the project will yield an output of \( y \in [0, \bar{y}] \) (\( \bar{y} \in \mathbb{R}^+ \cup \{\infty\} \)) that is distributed according to the c.d.f. function \( G(y|k_1, \theta)(g(y|k_1, \theta) = g(y|k_1, \theta)) \) where \( G(\cdot|\cdot, \cdot) \) is \( C^1 \) in all of its argument. Moreover, the mean value of \( y \), given \( \theta, k_1 \) is given by \( (\theta k_1)\alpha \), i.e., \( \int_0^\bar{y} yg(y|k_1, \theta)dy = (\theta k_1)\alpha \) with \( \alpha \in (0, 1) \). In other words, a more productive entrepreneur has a higher total output as well as higher marginal product of capital. This formulation of the production function is similar to [Lucas, 1978] and [Evans and Jovanovic, 1989]. Notice this formulation can stand-in for a more general constant return to scale production function that employs labor, capital and managerial effort with labor being supplied competitively in the labor market and where managerial...
effort is inelastically supplied. The decreasing returns to scale assumption implies that in any socially optimal allocation, there should be investment in projects of all productivities. For tractability, I assume that capital fully depreciates over time.

In addition, in order to make the analysis easier and in accordance with the rest of moral hazard literature – see [Jewitt, 1988] and [Rogerson, 1985b], we assume that \( g(y|k_1, \theta) \) satisfies the Monotone Likelihood Ratio Property (MLRP):

\[
\frac{\partial}{\partial y} g_k(y|k_1, \theta) > 0 \quad (2.1)
\]

This assumption is necessary in order for the validity of the first order approach in characterizing incentive compatible allocations. I further assume that \( G(y|k, \theta) \) has the following property or a function \( \tilde{G}(y, \cdot) \) must exists such that \( G(y|k_1, \theta) = \tilde{G}(y, \theta k_1) \).

Additionally, the distribution \( G(y|k_1, \theta) \) is an increasing a function of \( k_1 \) and \( \theta \) w.r.t. stochastic first order dominance ordering.

Note that the above formulation of entrepreneurial investment technology is compatible with the literature on entrepreneurial behavior as in [Evans and Jovanovic, 1989] and [Gentry and Hubbard, 2004]. In particular, they assume that output is given by

\[
\varepsilon \theta^\alpha k_1^\alpha
\]

where \( \log \varepsilon \sim N(-\frac{1}{2} \sigma^2, \sigma^2) \). This is essentially a special case of the above formulation where \( \bar{y} = \infty \) and \( G(y|k_1, \theta) = \Phi \left( \frac{\log y - \alpha \log \theta k_1 + \frac{1}{2} \sigma^2}{\sigma^\varepsilon} \right) \).

In addition, entrepreneurs preferences are standard and given by

\[
u(c_0) + \beta u(c_1)\]

---

6. Suppose that the production function is given by \( y = \varepsilon^\psi A \theta^\alpha k^\alpha l^{\alpha_2} m^{1-\alpha_1-\alpha_2} \) where \( l \) is labor input and \( m \) is managerial effort and \( \varepsilon \) is a shock realized once capital is put in place. If managers employ labor at \( t = 1 \), and inelastically supply a unit of managerial effort, the profit maximization decision of the firm in \( t = 1 \) is given by

\[
\max_l \varepsilon^\psi A \theta^\alpha k^\alpha l^{\alpha_2} - wl
\]

and therefore, \( \alpha_2 \varepsilon^\psi A \theta^\alpha k^\alpha l^{\alpha_2-1} = w \). Hence, \( \alpha_2, \alpha_1, \psi, \) and \( A \) can be chosen so that \( y = \varepsilon^\psi k^\alpha \).
where $c_0$ and $c_1$ are consumption of the entrepreneur at each period, where $u(\cdot)$ is a strictly concave and smooth function satisfying $u'(0) = \infty$. Entrepreneurs, therefore, consume in each period and invest at $t = 0$. We assume for simplicity that each agent is endowed with $e_0$ at $t = 0$.

For this economy, an allocation is given by $\{c_0(\theta), c_1(\theta, y), k_1(\theta)\}_{\theta = \theta^*}$. An allocation is said to be feasible if it satisfies the following:

\[
\int_{\theta}^{\theta^*} [c_0(\theta) + k_1(\theta)] dF(\theta) \leq e_0 \tag{2.2}
\]

\[
\int_{\theta}^{\theta^*} \int_{0}^{\theta^*} c_1(\theta, y) g(y|k_1(\theta), \theta) dy dF(\theta) \leq \int_{\theta}^{\theta^*} \theta^\alpha k_1(\theta)^\alpha dF(\theta) \tag{2.3}
\]

**Efficient Allocations with Full Information.** It is useful to characterize efficient allocations when a planner can observe entrepreneurs’ project type, $\theta$, as well as their consumption and investment. In such efficient allocations, the planner will equate returns to investment across all types of projects:

\[
\alpha \theta^\alpha k_1(\theta)^\alpha - 1 = \alpha \theta^\alpha k_1(\theta')^\alpha - 1 = \frac{1}{q}
\]

where $q$ is the shadow value of consumption at $t = 1$ in terms of consumption at $t = 0$; formally, $q$ is the lagrange multiplier on (2.3) divided by the one on (2.2). Moreover, if we consider a utilitarian planner that maximizes entrepreneurs’ ex-ante utility before realization of the shock, the efficient allocation must satisfy:

\[
c_0(\theta) = c_0(\theta') = c_0
\]

\[
c_1(\theta, y) = c_1(\theta', y') = c_1
\]

\[
u'(c_0) = \beta q^{-1} u'(c_1) = \beta \alpha \theta^\alpha k_1(\theta)^\alpha - 1 u'(c_1)
\]

The first two equations are implied by full risk sharing across types and the third is an Euler Equation for each individual. Hence, with full information, efficiency implies that the rate of return to individual investment should be equated across individuals. It follows that entrepreneurs with higher productivity should invest more than entrepreneurs with lower productivity. Next, I argue that an important assumption for this result is the observability of investment and consumption.

**Private Information.** Here we assume that agents are privately informed about their productivities. Moreover, the planner cannot observe consumption and investment
by a particular agent at \( t = 0 \). The planner can only observe income \( y \) at \( t = 1 \). By the Revelation Principle, we can focus on direct mechanisms in which each type reports his productivity. We call an allocation \textit{incentive compatible} if it satisfies the following:

\[
 u(c_0(\theta)) + \beta \int_0^y u(c_1(\theta, y)) g(y | k_1(\theta), \theta) dy \\
\geq \max_{\hat{\theta}, \hat{k}} u \left( c_0(\hat{\theta}) + k_1(\hat{\theta}) - \hat{k} \right) + \beta \int_0^\hat{y} u(c_1(\hat{\theta}, y)) g(y | \hat{k}, \theta) dy
\]

The RHS of the above inequality is the utility that a type \( \theta \) receives when he reports \( \hat{\theta} \) and invests \( \hat{k} \). Moreover, I call an allocation \textit{incentive feasible}, if it is incentive compatible and feasible.

The assumption about private information features two type of incentive problems: a hidden type problem and a hidden action problem. The hidden type problem implies that, when facing the full information efficient allocation, agents with higher productivity – \( \theta \), have incentive to lie downward about their productivity type even if they invest "the right" amount. By lying downward and investing \( \hat{k} \), higher productivity agents can enjoy higher consumption in the first period. Moreover, the hidden action problem implies that even if the agents tell the truth, the full insurance in the second period leads to under-investment in the first period.

Given above definitions, a utilitarian planner that maximizes entrepreneurs’ ex-ante utility solves the following problem:

\[
\max_{c_0(\theta), c_1(\theta, y), k_1(\theta)} \int_\theta \left[ u(c_0(\theta)) + \beta \int_0^y u(c_1(\theta, y)) g(y | k_1(\theta), \theta) dy \right] dF(\theta)
\]

subject to (2.2), (2.3), and (2.4).

\textbf{First Order Approach.} As can be seen, the set of incentive compatibility constraints is large and this complicates the characterization of optimal allocations. Here, I appeal to the first order approach to simplify the set of incentive compatibility constraints and discuss the validity of this approach in this environment. In particular, let \( U(\theta) \) be the utility of type \( \theta \) from truth-telling. Then we must have

\[
U(\theta) = \max_{\hat{\theta}, \hat{k}} u \left( c_0(\hat{\theta}) + k_1(\hat{\theta}) - \hat{k} \right) + \beta \int_0^\hat{y} u(c_1(\hat{\theta}, y)) g(y | \hat{k}, \theta) dy
\]
If we assume that the allocations are $C^1$ in $\theta$ and $y$, then incentive compatibility yields the following first order conditions and Envelope condition:

$$u'(c_0(\theta)) = \beta \int_0^y u(c_1(\theta, y)) g_k(y|k_1(\theta), \theta) dy$$ (2.5)

$$u'(c_0(\theta)) \left[ c_0'(\theta) + k_1'(\theta) \right] + \beta \int_0^y u'(c_1(\theta, y)) c_{1\theta}(\theta, y) g(y|k_1(\theta), \theta) dy = 0$$ (2.6)

The Envelope condition associated with this problem is given by

$$U'(\theta) = \frac{\partial}{\partial \theta} u \left( c_0(\hat{\theta}) + k_1(\hat{\theta}) - \hat{k} \right) + \beta \int_0^y u(c_1(\hat{\theta}, y)) g(y|\hat{k}, \theta) dy \bigg|_{\hat{\theta} = \theta, \hat{k} = k_1(\theta)}$$

$$= \beta \int_0^y u(c_1(\hat{\theta}, y)) g_\theta(y|\hat{k}, \theta) dy$$

Note that since $g(y|k_1, \theta)$ is a function of $\theta k_1$, I can write $g_\theta(y|k_1, \theta) = \frac{k_1}{\theta} g_k(y|k_1, \theta)$. Hence, the above envelope condition combined with the first order condition simplifies to

$$U'(\theta) = \frac{1}{\theta} k_1(\theta) u'(c_0(\theta))$$ (2.7)

We say an allocation is **locally incentive compatible** if it satisfies (2.5) and (2.7).

The above conditions are necessary for incentive compatibility. However, it is not clear that they are sufficient for incentive compatibility. Our aim, here, is to provide sufficient conditions under which the local incentive compatibility implies incentive compatibility, i.e., the First Order Approach (FOA) is valid. As mentioned before, there are two frictions in this model: an adverse selection problem and a moral hazard problem. As for the moral hazard problem, there is a series of papers giving providing assumption on fundamentals for validity of the FOA – see [Mirrlees, 1999], [Rogerson, 1985b], [Jewitt, 1988]. Regarding the adverse selection problem, there has not been much success in finding general assumptions on primitives that validate the FOA\(^7\). In Appendix A.2, in line with [Pavan et al., 2009], we provide monotonicity conditions on endogenous allocations that can be easily checked and are sufficient to ensure that FOA is valid.

Given the above discussion and conditions provided in Appendix A.2, in what follows, we relax the set of incentive compatible constraints and only impose local incentive

\(^7\) There are special cases for which assumptions on fundamentals exist. For example [Myerson, 1981] and [Guesnerie and Laffont, 1984] show that when principal and agent are both risk neutral, a monotone likelihood ratio assumption on the distribution of types validates the FOA.
compatibility. This further simplifies the analysis of the planning problem and enables us to further characterize the properties of the optimal allocations.

Hence, the relaxed problem becomes the following:

$$\max_{c_{0}(\theta),c_{1}(\theta,y),k_{1}(\theta),U(\theta)} \int_{\theta}^{\theta} U(\theta) dF(\theta)$$ (P1)

subject to

$$\int_{\theta}^{\theta} [c_{0}(\theta) + k_{1}(\theta)] dF(\theta) \leq c_{0} \quad (2.8)$$

$$\int_{\theta}^{\theta} \int_{0}^{\theta} c_{1}(\theta,y) g(y,k_{1}(\theta),\theta) dy dF(\theta) \leq \int_{\theta}^{\theta} \theta \alpha k_{1}(\theta) \alpha dF(\theta) \quad (2.9)$$

$$U(\theta) = u(c_{0}(\theta)) + \beta \int_{0}^{\theta} u(c_{1}(\theta,y)) g(y,k_{1}(\theta),\theta) dy$$

$$U'(\theta) = \frac{1}{\theta} k_{1}(\theta) u'(c_{0}(\theta)) \quad (2.10)$$

$$\beta \int_{0}^{\theta} u'(c_{1}(\theta,y)) g(k_{1}(\theta),\theta) dy = u'(c_{0}(\theta)) \quad (2.11)$$

In what follows, we refer to (2.10) as the adverse selection constraint and to (2.11) as moral hazard constraint.

2.2.1 Modified Inverse Euler Equation

In this section, we provide our version of the inverse Euler Equation that will prove useful in characterizing taxes and wedges. We call this the Modified Inverse Euler Equation. We have the following proposition:

**Proposition 2.1 (Modified Inverse Euler Equation).** Suppose that $c_{t}, k_{1} > 0$, a.e. Then any solution to (P1) must satisfy

$$\frac{q}{\beta} \int_{0}^{\theta} \frac{1}{u'(c_{1}(\theta,y))} g(y,k_{1}(\theta),\theta) dy = \frac{1}{u'(c_{0}(\theta))} + \frac{u''(c_{0}(\theta))}{u'(c_{0}(\theta))} \left[ \frac{1}{\theta} k_{1}(\theta) \mu_{1}(\theta) + \mu_{2}(\theta) \right] \quad (2.12)$$

where $q$ is the relative intertemporal price of consumption, $\mu_{1}$ is the costate associated with (2.10) and $\mu_{2}$ is the lagrange multiplier associated with (2.11). Both $\mu_{1}$ and $\mu_{2}$ are denominated in $t = 0$ consumption.
The proof can be found in the appendix.

This equation extends the results in [Rogerson, 1985a] and [Golosov et al., 2003] to the described environment. A key condition in deriving the IEE in [Golosov et al., 2003] is the fact that marginal utility is observable by the planner. In general optimality of allocations implies that a perturbation of the allocations that keeps utility of all types unchanged must keep the cost unchanged. In particular, any such perturbation at any given period $t$ should imply that

$$MC_t + MC_{t+1} = 0$$

where $MC_t$ is the marginal cost of such perturbation. When marginal utility is observable, i.e., consumption is separable from the source of private information, a perturbation in consumption that keeps utility unchanged along every history does not change incentives $- \beta^t u(c_t(h^t)) + \beta^{t+1} u(c_{t+1}(h^{t+1}))$ is unchanged for all $h_{t+1}$. Since, the source of private information is separate from consumption and the utility from consumption has not changed, the perturbed allocation must be incentive compatible. This implies that the marginal cost of the perturbation at period $t$ is given by $MC_t = \frac{1}{\beta^t u(c_t)}$ while at period $t + 1$, it is given by $MC_{t+1} = -q_{t+1} E \left[ \frac{1}{\beta^{t+1} u(c_{t+1})} | h^t \right]$ with $q_{t+1}$ being the relative shadow value of aggregate consumption. In our environment, however, consumption is non-separable from the source of private information. Hence, a perturbation in consumption alone will induce some agents to lie and breaks the incentive compatibility requirement. Therefore, there are incentive cost associated with such perturbations. The last two terms in (2.12) capture these costs. Here, we give a heuristic derivation of (2.12).

Consider an infinitesimal perturbation of consumption for type $\theta^8$, $\{\varepsilon_{c0}, \varepsilon_{c1}(y)\}$ that preserves type $\theta$’s utility along each history path or

$$u(c_0(\theta) + \varepsilon_{c0}) + \beta u(c_1(\theta, y) + \varepsilon_{c1}(y)) = u(c_0(\theta)) + \beta u(c_1(\theta, y)), \quad \forall y \in [0, \bar{y}] \quad (2.13)$$

There are two types of incentive costs associated with this perturbation. The first is the cost of distorting incentives for truth-telling about $\theta$. By definition, $\mu_1(\theta)$ captures the marginal cost of a unit increase in $U'(\theta)$. The above perturbation increases $U'(\theta)$

---

8. Since this is a heuristic derivation, we suppress the technical details. For example, the perturbation has to be over a positive measure of types. However, a continuity assumption on the allocations with respect to $\theta$, makes the above perturbation plausible.
by $u''(c_0(\theta))\frac{1}{\theta}k_1(\theta)$. Hence, the first type of incentive cost in terms of consumption in the first period is given by

$$u''(c_0(\theta))\frac{1}{\theta}k_1(\theta)\mu_1(\theta)\varepsilon_{c0}$$

The second type of incentive cost is from distortions to the investment decision. Note that the above perturbation leaves the LHS of (2.11) unchanged. This is due to the fact that the above perturbation shifts utility after any realization of shock $y$ by the same amount. This makes the future marginal benefit from investment unchanged. However, due to the perturbation of consumption at $t = 0$, the incentives for investment at $t = 0$ change and the cost of this change in terms of period 0 consumption is captured by

$$\mu_2(\theta)u''(c_0(\theta))\varepsilon_{c0}$$

Hence, the total cost of this perturbation is given by

$$q \int_0^y \varepsilon_{c1}(y) dy + \varepsilon_{c0} + u''(c_0(\theta))\frac{1}{\theta}k_1(\theta)\mu_1(\theta)\varepsilon_{c0} + \mu_2(\theta)u''(c_0(\theta))\varepsilon_{c0}$$

Note that from (2.13), $\varepsilon_{c1}(y) = -\frac{u'c_0(\theta)}{u'(c_1(\theta,y))}\varepsilon_{c0}$. Setting the above cost equal to zero leads to the desired MIEE.

Our version of Modified Inverse Euler Equation implies when consumption is non-separable from the source of private information, what affects the distortions to intertemporal saving margin is the heterogeneity in second period consumption as well as the tightness of the incentive constraints. In particular, the sign of $\frac{1}{\theta}k_1(\theta)\mu_1(\theta) + \mu_2(\theta)$ which captures the tightness of the incentive constraint, is a key determinant of the distortions to intertemporal saving margin. In section 2.2.2, we further discuss how the MIEE is useful in characterizing distortions.

Since the perturbation argument given above is independent of specific welfare weights on different individuals, it is straightforward to show that for social welfare functions other than the utilitarian, i.e., when the planner’s objective is $\int G(U(\theta))dF(\theta)$, the MIEE holds.

### 2.2.2 Wedges

In this section we study the properties of the intertemporal saving wedge implied by the model developed so far. We argue that in this two period model, the intertemporal
wedge is positive. We show this by considering the case where the utility function is exponential. Under this assumption, the model becomes more tractable and we can show that intertemporal wedge is positive. For general utility functions, the model is less tractable. However, we can show that when one source of risk is shut down, i.e., either output is not risky or there is no heterogeneity in productivities, again the intertemporal wedge is positive. Although the main result of the paper regarding negative intertemporal wedges cannot be shown in a two period model, the analysis in this section is useful to see the mechanisms in play in the model. Later in section 2.4.1, we extend the model to more than two periods to show that there are forces toward negative intertemporal wedges when the number of periods increases from two and agents are hit by subsequent productivity, $\theta$, shocks.

In order to show that the intertemporal wedge is positive, we first show that the incentive costs of utility preserving perturbations, $\frac{1}{\theta}k_1(\theta)\mu_1(\theta) + \mu_2(\theta)$, are positive. Then using an argument similar to [Golosov et al., 2003], we can show that the intertemporal wedge is positive.

The following lemma characterizes the multiplier on moral hazard constraint:

**Lemma 2.2** The multiplier on the moral hazard constraint is given by

$$\mu_2(\theta) = \frac{q}{u'(c_0)} \text{Cov}_\theta \left( u(c_1), \frac{1}{u'(c_1)} \right)$$

Now, since $u(c_1)$ and $\frac{1}{u'(c_1)}$ are positively correlated, $\mu_2(\theta)$ is always positive. As we show in the next section, $\mu_2(\theta)$ determines the sensitivity of the consumption schedule $c_1(\theta, y)$ to income realization $y$. Therefore, this result is equivalent to the consumption schedule $c_1(\theta, y)$ being increasing in income realization.

Given the sign of $\mu_2(\theta)$, if we show that the tightness of the adverse selection constraint, $\mu_1(\theta)$, is positive, then I can show that intertemporal wedge is positive. To do so, I use an argument similar to the argument in [Werning, 2000] in the context of a static Mirrlees model. In fact, the result that $\mu_1$ is positive everywhere is reminiscent of the positive marginal tax result in Mirrleesian contexts. That is, to prove that marginal tax rates are positive in a static Mirrlees economy, one only needs to show that the co-state associated with the incentive constraint is positive. I can do this when the utility function has a CARA form since there are no wealth effects. We can also show
it in the case where there is no riskiness in the returns to investment. The positive sign of the co-state, $\mu_1(\theta)$, intuitively means that the relevant local incentive constraints are the downward incentive constraints.

Hence, we have the following proposition:

**Proposition 2.3** Suppose that $u(c) = -\exp(-\psi c)$. Then, $\mu_1(\theta) \geq 0$ for all $\theta \in [\theta, \bar{\theta}]$. Moreover, $\mu_1(\theta) = \mu_1(\bar{\theta}) = 0$ and the above inequality is strict for at least a positive measure of $\theta$'s.

Proof can be found in the Appendix.

Using the same proof, I can also show that for general utility functions, when there is no riskiness in returns, i.e., $G(\cdot|k, \theta)$ puts mass 1 on $(\theta k)^\alpha$, the co-state $\mu_1(\theta) > 0$ is positive – see Appendix for details.

The above discussion on the sign of incentive costs together with the Modified Inverse Euler Equation helps us determine the sign of the intertemporal wedge. That is, since $\frac{1}{q}k_1(\theta)\mu_1(\theta) + \mu_2(\theta) > 0$, then MIEE together with concavity of the utility function implies that

$$\frac{q}{\beta} \int_0^y \frac{1}{u'(c_1(\theta, y))} g(y|k_1(\theta), \theta) dy < \frac{1}{u'(c_0(\theta))}$$

By Jensen’s inequality, we have

$$\frac{q}{\beta} \int u'(c_1(\theta, y)) g(y|k_1(\theta), \theta) dy < \frac{q}{\beta} \int_0^y \frac{1}{u'(c_1(\theta, y))} g(y|k_1(\theta), \theta) dy < \frac{1}{u'(c_0(\theta))}$$

or

$$q^{-1}\beta \int u'(c_1(\theta, y)) g(y|k_1(\theta), \theta) dy > u'(c_0(\theta)) \quad (2.14)$$

Hence, the intertemporal wedge, defined by

$$\tau_s(\theta) = 1 - \frac{u'(c_0)}{q^{-1}\beta \int u'(c_1(\theta, y)) g(y|k_1(\theta), \theta) dy}$$

is positive. One interpretation of positive intertemporal wedge is that in order to provide incentives, the optimal contract encourages consumption early. That is an agent who has access to borrowing and lending at rate $q^{-1}$, facing the efficient allocation, would like to save. To see the intuition for the above inequality, consider decreasing agent $\theta$’s consumption in the first period by $\varepsilon$ and increasing his consumption by $q^{-1}\varepsilon$ after any realization in the second period. In addition to the usual direct cost, $u'(c_1)\varepsilon$, and benefit
\[ \beta q^{-1} \varepsilon \int u'(c_1) g dy \] of such a perturbation, there are two incentive costs associated with it. The first comes from the moral hazard aspect of the model. Since utility function is concave, such perturbation makes investment relatively unattractive, i.e., it decreases \[ \int u(c_1) g dy. \] It also increases the current cost of investment to the individual consumer, \( u'(c) \). Hence, an agent of type \( \theta \) will decrease his investment. The second cost associated with this perturbation is that it increases the slope of the schedule \( U(\theta) \), i.e., \( \frac{1}{q} u'(c_0) k_1 \). Therefore, the entrepreneurs with higher productivity will find optimal to lie downward and work less. Since the marginal cost of such perturbation should be equal to its marginal benefit, we must have the inequality (2.14).

Although, this wedge can be interpreted as a tax on saving, it does not directly translate into a marginal tax rate on saving. In fact, the implementation of the efficient allocation requires tax functions that are non-separable between second period income and saving. I discuss this further in section 2.2.4.

Given the above definition of wedges, it can be shown that a version of [Mirrlees, 1971]-[Saez, 2001] tax formulas holds in this economy as well when the returns are deterministic. In fact, I can derive a formula for saving wedge as a function of the skill distribution, intertemporal elasticity of substitution, investment-consumption ratio and distribution of consumption in the second period. In particular, it can be shown that the following proposition holds:

**Proposition 2.4** Suppose that \( c_1(\theta), k_1(\theta) > 0 \), a.e.-\( F \). Then any solution to (P1) must satisfy

\[
\frac{\tau_s(\theta)}{1 - \tau_s(\theta)} = \frac{1 - F(\theta)}{\theta f(\theta)} \frac{k_1(\theta)}{c_0(\theta)} \left( - \frac{u''(c_0(\theta)) c_0(\theta)}{u'(c_0(\theta))} \right) \int_\theta^1 \times \left[ \frac{u'(c_1(\theta))}{u'(c_1(\hat{\theta}))} - \lambda_0 q^{-1} u'(c_1(\theta)) \right] \frac{dF(\theta)}{1 - F(\theta)}
\]

As we can see, these formulas are very similar to Saez’s formulas since they relate marginal income/saving distortions to tail of skill distribution, \( \frac{1 - F(\theta)}{\theta f(\theta)} \), intertemporal elasticity of substitution, and investment-consumption ratio. Note that in our derivations in the appendix – MIEE and the tax formula, I have not used the fact that the
skill distribution is bounded. In particular, the above formulas hold even in the case that $\bar{\theta} = \infty$. The above formulas are easier to understand for the case where $\bar{\theta} = \infty$ and $\lim_{\theta \to \infty} \frac{1-F(\theta)}{\sigma_f(\theta)} > 0$. In this case, saving wedges at the top are non-zero and can be derived explicitly in terms of fundamentals of the model as in [Diamond, 1998] and [Saez, 2001]. The above analysis implies that the same exercise can be done for our environment.

2.2.3 Shape of the Consumption Schedule

In this section, I provide one of the main results of the paper. That is the possibility of progressive tax schedules. To do so I provide a simple formula for consumption in the second period as function of income realizations in period 2. Using this formula, I can provide conditions under which the consumption schedule is a concave function of income realization. As I argue here and formally show in section 2.2.4, concavity of the consumption schedule with respect to income implies progressivity of the tax schedule.

I first start by providing a simple formula for consumption in the second period:

**Lemma 2.5** Consider any solution to (P1) and assume that the allocations are positive almost surely. Then,

\[
\frac{1}{u'(c_1(\theta,y))} = \int^\hat{y}_{y} \frac{1}{u'(c_1(\theta,\hat{y}))} g(\hat{y}|k_1(\theta),\theta) d\hat{y} + \beta q^{-1} \mu_2(\theta) \frac{g_k(y|k_1(\theta),\theta)}{g(y|k_1(\theta),\theta)}. \tag{2.15}
\]

To intuitively see why this equation holds, consider the following perturbation of the allocation: for $\hat{y} \in [y, y + \varepsilon]$ increase $u(c_1(\theta, \hat{y}))$ by 1 unit and decrease all $u(c_1(\theta, \hat{y}))$ by $\varepsilon g(y|k_1(\theta),\theta)$. Note that this perturbation preserves period 1 utility of type $\theta$. Hence, it does not violate (2.10). It does, however, change investment incentives for type $\theta$. Note that the uniform decrease in utility for all $\hat{y}$’s does not change the marginal return to investment. As a result, the marginal decrease in utility for all $\hat{y}$’s does not change the marginal return to investment. Hence the resource cost of this perturbation is given by $\frac{1}{u'(c_1(\theta,y))} g(\hat{y}|k_1(\theta),\theta) \varepsilon$. While the benefit from lowering consumption and relaxing the incentive constraint is given by

\[
\varepsilon g(y|k_1(\theta),\theta) \int^\hat{y}_{y} \frac{1}{u'(c_1(\theta,\hat{y}))} g(\hat{y}|k_1(\theta),\theta) d\hat{y} + q^{-1} \mu_2(\theta) \beta g_k(y|k_1(\theta),\theta) \varepsilon.
\]
Figure 2.1: Perturbation
Equating the cost and benefit leads to (2.15). This perturbation is depicted in Figure 2.1.

Equation (2.15) implies that the marginal cost of providing utility to income level \( y \), \( \frac{1}{w(c(y))} \), is a linear function of the hazard rate. As in [Holmstrom, 1979], \( \frac{\partial g}{\partial y} \), is the derivative of the likelihood function \( \log g(y|k, \theta) \) where \( k \) can be treated as unobservable from planner’s point of view. Hence, when \( \frac{\partial g}{\partial y} \) is the highest, the planner is the most inclined to infer from \( y \) that the agent took the right action. Hence, the rewards to the agent are the highest in those states. Note, also, that the MLRP assumption implies that \( \frac{\partial g}{\partial y} \) is an increasing function of \( y \). Since \( u(\cdot) \) is concave, \( \frac{1}{u'(c)} \) is an increasing function of \( c \) and therefore from (2.15), we deduce that the consumption schedule \( c_1(\theta, y) \) is an increasing function of \( y \).

Given the above formula on the consumption schedule, it is rather straightforward to provide sufficient condition under which the consumption schedule is concave. In fact, when \( \frac{\partial g}{\partial y} \) is concave in \( y \) and \( \frac{1}{u'(c)} \) is a convex function of \( c \), the schedule is concave in \( y \). Notice that when \( u(c) = \frac{c^{1-\sigma}}{1-\sigma} \), the convexity of \( \frac{1}{u'(c)} \) requires that \( \sigma > 1 \).

That is the intertemporal elasticity of substitution must be bigger than 1. To better understand the requirement on the hazard ratio, \( \frac{\partial g}{\partial y} \), I consider the environment in [Evans and Jovanovic, 1989](EJ economy henceforth). Suppose that

\[
y = \varepsilon \theta^a k^a
\]

where \( \varepsilon \sim H(\varepsilon) \) with density \( h(\varepsilon) \) and \( \varepsilon \in [0, \infty) \). In this case,

\[
g(y|k, \theta) = \frac{1}{(\theta k)^\alpha} h\left( \frac{y}{(\theta k)^\alpha} \right)
\]

and hence

\[
\frac{g_k(y|k, \theta)}{g(y|k, \theta)} = -\alpha k^{-1} \left[ 1 + \frac{\frac{y}{(\theta k)^\alpha} h'\left( \frac{y}{(\theta k)^\alpha} \right)}{h\left( \frac{y}{(\theta k)^\alpha} \right)} \right]
\]

Now consider the following examples for the distribution function \( h \):

1. Log-normal distribution: \( h(\varepsilon) = \kappa \varepsilon^{-1} e^{-\frac{(\log \varepsilon - \mu)^2}{2\sigma^2}} \). In this case

\[
\frac{g_k(y|k, \theta)}{g(y|k, \theta)} = \alpha k^{-1} \log y - \alpha \log(\theta k) - \mu \]

and hence the hazard ratio is concave in \( y \).
2. Gamma distribution: \( h(\varepsilon) = \kappa \varepsilon^{-\zeta} e^{-\varepsilon/\eta} \). In this case,

\[
\frac{g_k(y|k, \theta)}{g(y|k, \theta)} = \alpha k^{-1} \left( \frac{1}{\eta (\theta k)^\alpha} - \zeta \right)
\]

and hence the hazard ratio is linear in \( \varepsilon \).

3. Pareto distribution in the tail: \( h(\varepsilon) = \kappa \varepsilon^{-\zeta-1} \). In this case, \( \frac{g_k}{g} = \zeta \alpha k^{-1} \) and hence the hazard ratio is constant.

Hence, for the EJ economy, MLRP implies that \( \frac{\varepsilon h'(\varepsilon)}{h(\varepsilon)} \) be decreasing and when \( \frac{\varepsilon h'(\varepsilon)}{h(\varepsilon)} \) is convex, the consumption schedule is concave in income.

The concavity of the consumption schedule in realized income has an important interpretation regarding tax system. In fact, the slope of the consumption function determines the marginal tax rate on income. In particular, when this slope is decreasing, i.e., consumption schedule is concave, the marginal tax rate is increasing and hence the income tax schedule is progressive – I will discuss this in detail in section 2.2.4. Here, progressivity of the tax system works as an insurance mechanism against income shocks. Due to moral hazard, only partial insurance is feasible and therefore consumption schedule is not fully flat.

The analysis so far points to ways a planner can resolve the two types of informational asymmetries, the moral hazard and the adverse selection problem. Loosely speaking, the intertemporal wedge induces agents to tell the truth regarding their productivity type. Once the productivity type is revealed, equation (2.15) induces the agent to make the right amount of investment.

### 2.2.4 Implementation

In this section, I discuss ways for a government to implement the optimal allocations discussed above. The construction of the tax function below, demonstrates that the tax function is unique given the market structure imposed. Note that the market structure assumed for the implementation plays a key role in determining government policy. Here, we assume that the entrepreneurs, in addition to the individual investment opportunity, have access to a centralized market for risk free asset in net zero supply. We, then, construct a tax schedule that implements the optimal allocation. Using the
properties of the allocations discussed above, we characterize the properties of such optimal tax system.

A key assumption in the following implementation is that agents are unable to sign contracts before realization of their productivity type, \( \theta \). Otherwise, the results in [Prescott and Townsend, 1984] imply that private contracts are able to achieve the constrained efficient allocation discussed above. This assumption gives rise to a need for redistributive policies by the government. Later, in section 2.5, we show that if ex-ante contracting is available, the optimal allocation can be implemented with a set of contracts that are widely used in financial markets and venture capital contracts. This assumption is in line with the rest of the literature on dynamic public finance.

As mentioned above, we assume that each entrepreneur can invest in his private investment project and can borrow and save from centralized market. The agent may purchase and sell the risk free bond at price \( Q \). Hence, the agent’s budget constraint at \( t = 0 \) is given by:

\[
c_0 + k_1 + Qb_0 \leq e_0
\]

The government observes \( b_0 \) and \( y \) at \( t = 1 \) and can tax agents based on observables according to the tax function \( T(b_0, y) \). Given this tax function, the budget constraint of the agent in the second period is given by

\[
c_1 \leq y + b_0 - T(b_0, y)
\]

Hence, facing a particular tax function \( T(b_0, y) \), an entrepreneur of type \( \theta \) solves the following maximization problem

\[
\max_{c_0, c_1(y), k_1, b_0} u(c_0) + \beta \int_0^y u(c_1(y))g(y|k_1, \theta)dy
\]

subject to

\[
c_0 + k_1 + Qb_0 \leq e_0
\]

\[
c_1(y) \leq y + b_0 - T(b_0, y)
\]

Here, we show that given any incentive compatible allocation \( \{c^*_0(\theta), \{c^*_1(\theta, y)\}, k^*_1(\theta)\} \) together with an intertemporal price of consumption \( q \), there exists a tax system of the
above form that implements it. To do so we need to make the following assumption about the allocation:

**Assumption 2.6** For all $\theta \neq \theta'$, $c^*_0(\theta) + k^*_1(\theta) \neq c^*_0(\theta') + k^*_1(\theta')$ and allocations are $C^1$ in $\theta$.

A sufficient condition for the above assumption is that transfers in the first period are increasing in type. In fact, if the allocations are continuous in $\theta$, the above assumption implies that transfers, $c^*_0(\theta) + k^*_1(\theta)$, are monotone in $\theta$.

Given this assumption, we can show the following:

**Proposition 2.7** Consider an incentive compatible allocation $\{c^*_0(\theta), \{c^*_1(\theta, y)\}, k^*_1(\theta)\}$ together with a risk-free bond price $q$. If Assumption 2.6 holds, there is tax function $T(\cdot, \cdot)$ that implements the allocation. Moreover, the tax function is $C^1$.

**Proof.** We start by constructing the saving level, $b^*_0(\theta)$, for each type

$$b^*_0(\theta) = q^{-1} [c_0 - k^*_1(\theta) - c^*_0(\theta)]$$

Assumption 2.6 implies that $b^*_0(\theta)$ is a one-to-one function of $\theta$. Notice that continuity of the allocations together $b^*_0(\cdot)$ being one-to-one implies that there exists an interval $[\bar{b}, \bar{b}]$ such that $b^*_0([\bar{\theta}, \bar{\theta}]) = [\bar{b}, \bar{b}]$ and $b^*_0$ is a bijection over $[\bar{\theta}, \bar{\theta}]$. Hence, we can define the following tax function $T(\cdot, \cdot)$:

$$T(b, y) = \begin{cases} 
  y + b - c_1((b^*_0)^{-1}(b), y) & b \in [\bar{b}, \bar{b}] \\
  y + b & b \notin [\bar{b}, \bar{b}] 
\end{cases}$$

(2.17)

Here, we show that the above tax function implements the desired allocation when the price risk-free bond at $t = 0$ are given by $q$. First, note that if an agent of type $\theta$, chooses $c^*_0(\theta), \{c^*_1(\theta, y)\}, k^*_1(\theta), b^*_0(\theta)$, the utility he receives is equal to the utility he receives from the allocation, $U(\theta)$. Second, it is easy to see that in (2.16) $b_0 \in [\bar{b}, \bar{b}]$, otherwise consumption following any income realization is zero. At last, consider a possible solution to (2.16), $\{\hat{c}_0, \hat{c}_1(y), \hat{k}_1, \hat{b}_0\}$. Since $b^*_0$ is a bijection, there exists a unique $\hat{\theta} \in [\bar{\theta}, \bar{\theta}]$ such that $b^*_0(\hat{\theta}) = \hat{b}_0$. Then, by definition of $b^*(\cdot)$, $c_0 - qb^*(\hat{\theta}) = c^*_0(\hat{\theta}) + k^*_1(\hat{\theta})$ and given the budget constraint at $t = 0$, $\hat{c}_0 + \hat{k}_1 = c^*_0(\hat{\theta}) + k^*_1(\hat{\theta})$. Moreover,
by definition of $T(\cdot, \cdot)$, $\hat{b}_0 + y - T(\hat{b}_0, y) = c_1(\hat{\theta}, y)$. Hence, the utility that the agent receives from this allocation is given by

$$u(c^*_0(\hat{\theta}) + k^*_1(\hat{\theta}) - \hat{k}_1) + \beta \int u(c^*_1(\hat{\theta}, y))g(y|\hat{k}_1, \theta)dy$$

By incentive compatibility (2.4),

$$U(\theta) \geq u(c^*_0(\hat{\theta}) + k^*_1(\hat{\theta}) - \hat{k}_1) + \beta \int u(c^*_1(\hat{\theta}, y))g(y|\hat{k}_1, \theta)dy$$

Therefore, it is optimal for the agent to choose $\{c^*_0(\theta), k^*_1(\theta), b^*_0(\theta)\}$. Q.E.D.

A point worth noticing is that given $q$, the above implementation is unique. In fact, knowing $q$ and the allocation, one can uniquely pin down saving levels and under Assumption 2.6, $T(\cdot, \cdot)$ is uniquely determined by the allocation.

Given the above tax function, properties of the optimal allocation leads to certain properties of the tax function. As we have shown in 2.2.2, intertemporal wedge is positive. This implies that average value of $T_b$ weighted by marginal utility is positive. To see this, note that the first order condition from (2.16) is given by

$$q^{-1} \beta \int_0^\theta u'(c^*_1(\theta, y))(1 - T_b(b^*_0(\theta), y))g(y|k^*_1(\theta), \theta)dy = u'(c^*_0(\theta))$$

Moreover, since $q^{-1} \beta \int_0^\theta u'(c^*_1(\theta, y))g(y|k^*_1(\theta), \theta)dy > u'(c_0(\theta))$ as shown in section 2.2.2, we must have

$$\int_0^\theta u'(c^*_1(\theta, y))T_b(b^*_0(\theta), y)g(y|k^*_1(\theta), \theta)dy > 0$$

Note that in order to $T_b$ in for each income realization, we need to know the way $c^*_1(\theta, y)$ moves as a function of $\theta$.

As mentioned before, a key result of this paper is that the optimal tax schedule with respect to entrepreneurial income is progressive. To show this, note that in this environment, marginal tax rate on income $T_y$ is given by

$$T_y(b^*_0(\theta), y) = 1 - \frac{\partial}{\partial y}c^*_1(\theta, y)$$

Recall from section 2.2.3 that,

$$\frac{1}{u'(c_1(\theta, y))} = a(\theta) + b(\theta)\frac{g_k(y|\theta, k_1(\theta))}{g(y|\theta, k_1(\theta))}$$
where \( a(\theta), b(\theta) \) are independent of \( y \). Hence, the derivative of the consumption function with respect to \( y \) is given by

\[
\frac{\partial}{\partial y} c_1(\theta, y) = \left( \frac{1}{w'} \right)^{-1} \left( a(\theta) + b(\theta) \frac{g_k(y, \theta, k)}{g(y, \theta, k)} \right)
\]

Based on the above formula, when \( \frac{g_k}{g} \) is concave in \( y \) and \( \frac{1}{w'} \) is convex, \( c_1 \) becomes concave in \( y \) and hence \( T_y \) is increasing, i.e., entrepreneurial income taxes are progressive. We have the following proposition:

**Proposition 2.8** Suppose that \( \frac{1}{w'(c)} \) is convex and \( \frac{g_k}{g} \) is concave in \( y \). Then the tax function \( T(\cdot, \cdot) \) defined by (2.17) is progressive, i.e., \( T_y(b, \cdot) \) is increasing in \( y \).

Note that when \( w(c) = \frac{1-\sigma}{1-\sigma} \), then convexity of \( \frac{1}{w'(c)} \) is equivalent to \( \sigma \geq 1 \). This result is in contrast with regressivity of the income tax schedule at the top in [Mirrlees, 1971] when skill distribution is bounded (see [Farhi and Werning, 2010a] for a dynamic extension). Intuitively, progressivity arises in order to provide the right incentives to invest to the entrepreneur. Hence, the rewards to higher realization, i.e., the slope of consumption schedule, must be higher for low realizations of income. Although, we have shown progressivity in a model of entrepreneurs with capital income risk, this result is not specific to this environment. In particular, it is natural to guess that in a model with risky human capital and private information, same result would hold.

### 2.3 Multi-Period Model

In this section, we extend the analysis to a multi-period environment. In this context, we derive the general version of the MIEE and show how results change. We extend the two period model and derive the MIEE. Using the properties of the model, we study the implications of the model on taxation.

Time is discrete and indexed by \( t = 0, 1, \ldots, T \) where \( T \in \mathbb{N} \cup \{\infty\} \). There is a unit measure of entrepreneurs. The entrepreneurs are endowed with \( e_0 \) units of good at \( t = 0 \). Each entrepreneur has access to a private risky investment technology that evolves as follows: At each date \( t = 0, \ldots, T - 1 \), agent draws a productivity type \( \theta_t \in \Theta = [\underline{\theta}, \bar{\theta}] \) according to a differentiable c.d.f function \( F^t(\theta_t | \theta_{t-1}) \) (with its derivative given by \( f^t(\theta_t | \theta_{t-1}) \)). The initial draw of productivity, \( \theta_0 \), is distributed according
to a differentiable c.d.f. function $F^0(\theta_0)$. At each date $t$, the agent can privately invest $k_{t+1}$ in the project. Given the entrepreneur’s investment, $k_{t+1}$, and productivity type $\theta_t$, his income, $y_{t+1} \in Y = [0, \bar{y}]$, realized at $t + 1$, is a random variable that is distributed according to a differentiable c.d.f. $G^{t+1}(y_{t+1}|\theta_t, k_{t+1})$ (with its derivative given by $g^{t+1}(y_{t+1}|\theta_t, k_{t+1})$). Similar to the two-period example, the function $G^{t+1}$ has the following properties:

1. $G^{t+1}(y_{t+1}|\theta_t, k_{t+1})$ is strictly decreasing in $\theta_t$ and $k_{t+1}$ – hence $y_{t+1}$ is increasing according to first order stochastic dominance ordering.

2. The mean value of $y_{t+1}$ is given by
   \[
   \int_Y y_{t+1} g^{t+1}(y_{t+1}|\theta_t, k_{t+1}) dy_{t+1} = (\theta_t k_{t+1})^\alpha
   \]

3. For any $\theta, \theta', k$,
   \[
   G^{t+1}(y|\theta, k) = G^{t+1}(y|\theta, \frac{\theta k}{\theta'}) , \forall y \in [0, \bar{y}]
   \]

4. $G^{t+1}(y|\theta, k)$ satisfies MLRP:
   \[
   \frac{\partial}{\partial y} \left( \frac{g^{t+1}_k(y|\theta, k)}{g^{t+1}(y|\theta, k)} \right) > 0 \forall y, k, \theta
   \]

5. Given $k_{t+1}$ and $\theta_t$, $y_{t+1}$ is independent from $\theta_{t+1}$.

The 5th assumption above is crucial. It is important that conditional on $k_{t+1}$ and $\theta_t$, $y_{t+1}$ and $\theta_{t+1}$ are not perfectly correlated. In case of perfect correlation, a deviation over $k_{t+1}$ at $t$, must be accompanied by a certain report of $\theta_{t+1}$ at $t + 1$. This further complicates the problem. For analytical tractability, we assume that they are independent. However, the analysis here will go through if we assume that $y_{t+1}$ and $\theta_{t+1}$ are partially correlated.

Given this environment, an allocation is given by
\[
\{c_t(\theta^t, y^t), k_{t+1}(\theta^t, y^t)\}_{t=0}^T
\]
where $\theta^t = (\theta_0, \cdots, \theta_t) \in \Theta^{t+1}$ and $y^t = (y_1, \cdots, y_t) \in Y^t$. When $t = 0$, $y^0$ is the empty history and $\theta^t = \theta^{T-1}$ -- there are no draw of productivity at $T$ and no draw of income.
at 0. For ease of notation, we assume that \( \mu_t(\theta^t, y^t; k^{t-1}) \) is the joint distribution of all possible histories at period \( t \) given a sequence of investments \( k^{t-1} = (k_1, \ldots, k_{t-1}) \). Note that by definition,

\[
\mu_t(A_0 \times \cdots \times A_t, B_1 \times \cdots \times B_t; k^{t-1}) = \int_{A_0} \cdots \int_{A_t} f^0(\theta_0) \prod_{\tau=1}^{t} f^\tau(\theta_\tau|\theta_{\tau-1}) \prod_{\tau=1}^{t} \left( \int_{B_\tau} g^\tau(y|\theta_{\tau-1}, k_{\tau}) dy \right) d\theta_t \cdots d\theta_0
\]

An allocation is feasible if it satisfies

\[
\int_{\Theta} \left[ c_0(\theta_0) + k_{1}(\theta_0) \right] f_0(\theta_0) d\theta_0 \leq e_0 \tag{2.18}
\]

\[
\int_{\Theta^{t+1} \times Y^t} \left[ c_t(\theta^t, y^t) + k_{t+1}(\theta^t, y^t) \right] d\mu_t(\theta^t, y^t; k^{t-1}(\theta^{t-1}, y^{t-1})) \tag{2.19}
\]

\[
\leq \int_{\Theta^{t+1} \times Y^t} \left( \theta_{kt+1}(\theta^t, y^t) \right)^{\alpha} d\mu_t(\theta^t, y^t; k^{t-1}(\theta^{t-1}, y^{t-1}))
\]

As before, we assume that the planner observes the income in each period and productivity type as well as consumption and investment is privately known by the agent. Given this assumption about information structure, the nature of the incentive compatibility constraints depends on whether \( y_{t+1} \) is stochastic or deterministic. As it is clear in the two period example, with deterministic returns, agents are not free to deviate to any investment level where as with risky return these deviations are not detectable. Below, we describe an incentive compatible allocation under each set of assumption.

When returns are risky, at each period, an agent can lie about its productivity type and pick a different level of investment. Hence, a deviation strategy by the agent is a reporting strategy \( \sigma = \{\hat{\sigma}_t(\theta^t, y^t)\} \) as well as an investment strategy \( \hat{k} = \{\hat{k}_t(\theta^t, y^t)\} \). The utility from a deviation strategy given allocation is given by

\[
U(\{c_t, k_{t+1}\}; \hat{\sigma}, \hat{k}) = \sum_{t=0}^{T} \beta^t \int_{\Theta^{t+1} \times Y^t} u(c_t(\hat{\sigma}_t(\theta^t, y^t), y^t) + k_{t+1}(\hat{\sigma}_t(\theta^t, y^t), y^t) - \hat{k}(\hat{\sigma}_t(\theta^t, y^t), y^t))
\]

\[
d\mu_t(\theta^t, y^t; \hat{k}^{t-1}(\theta^{t-1}, y^{t-1}))
\]

Let \( \sigma^* \) be the truth-telling strategy, an allocation is then incentive compatible, if

\[
U(\{c_t, k_{t+1}\}; \sigma^*, k) \geq U(\{c_t, k_{t+1}\}; \hat{\sigma}, \hat{k}) \tag{2.20}
\]
When returns are not risky, agents can freely chose a reporting strategy, $\sigma = \{\hat{\sigma}_t(\theta^t, y^t)\}$. They, however, cannot choose any investment level freely. Given $\hat{\sigma}$, the entrepreneur has to invest $\frac{\partial}{\partial t} k_t(\hat{\sigma}_t, y^t)$. Since income is observable, any other investment level is in discrepancy with the original report and therefore detectable by the planner. With a little abuse of notation, an allocation is said to be incentive compatible with safe returns if

$$U(\{c_t, k_{t+1}\}; \sigma^*, k) \geq U(\{c_t, k_{t+1}\}; \hat{\sigma}, \hat{\sigma}_t(\theta^t, y^t))$$ \hspace{1cm} (2.21)

Hence, a planner solves the following maximization problem

$$\max \sum_{t=0}^T \beta^t \int_{\Theta \times Y} u(c_t(\theta^t, y^t)) d\mu_t(\theta^t, y^t; k^t|\theta^t, y^t)$$ \hspace{1cm} (P)

subject to (2.20) or (2.21), (2.18) and (2.19).

**First Order Approach to Incentive Compatibility Constraints**

The set of incentive compatibility constraints in the program (P) is very large. Using the First Order Approach we greatly simplify the set of incentive constraints. Here, we derive necessary first order conditions that any incentive compatible allocation must satisfy.

To do so, let $U_t(\theta^t, y^t)$ be the utility of agent with history $(\theta^t, y^t)$,

$$U_t(\theta^t, y^t) = \sum_{\tau=t}^T \beta^{\tau-t} \int_{\Theta \times Y} u(c_t(\theta^\tau, y^\tau)) d\mu_t(\theta^\tau, y^\tau; k^\tau|\theta^t, y^t)$$

If we focus on one shot deviations – deviation in one period and telling the truth thereafter – an allocation is *one-shot incentive compatible* if

$$u(c_t(\theta^t, y^t)) + \beta \int_{\Theta \times Y} U_{t+1}(\theta^t, \theta_{t+1}, y^t, y_{t+1}) g^{t+1}(y_{t+1}|\theta_t, k_{t+1}(\theta^t, y^t)) f^{t+1}(\theta_{t+1}|\theta_t) dy_{t+1} d\theta_t$$

$$\geq \max_{k, \hat{\theta}} u(c_t(\theta^{t-1}, \hat{\theta}, y^t)) + k_{t+1}(\theta^{t-1}, \hat{\theta}, y^t) - \hat{k}$$

$$+ \beta \int_{\Theta \times Y} U_{t+1}(\theta^t, \hat{\theta}, y^t, y_{t+1}) g^{t+1}(y_{t+1}|\theta_t, \hat{k}) f^{t+1}(\theta_{t+1}|\theta_t) dy_{t+1} d\theta_t$$ \hspace{1cm} (2.23)

When $T$ is finite, it is easy to see that the above is equivalent to (2.20). When $T = \infty$, if the utility from the allocations stays bounded, then (2.23) is equivalent to (2.20). This incentive constraint implies that any incentive compatible allocation that
is $C^1$, must satisfy the following first order condition with respect to investment and
Envelope Theorem—similar to (2.5)-(2.6):

$$
\frac{\partial}{\partial \theta_t} U_t(\theta^t, y^t) = \beta \int_{\Theta \times Y} U_{t+1}(\theta^t, \theta_{t+1}, y^t, y_{t+1}) g^{t+1}_{\theta}(y_{t+1}|\theta_t, k_{t+1}(\theta^t, y^t))
$$

$$
\quad + \beta \int_{\Theta \times Y} U_{t+1}(\theta^t, \theta_{t+1}, y^t, y_{t+1}) g^{t+1}_{\theta}(y_{t+1}|\theta_t, k_{t+1}(\theta^t, y^t))
$$

$$
\quad \quad f^{t+1}_{\theta}(\theta_{t+1}|\theta_t) dy_{t+1} d\theta_t
$$

$$
\quad + \beta \int_{\Theta \times Y} U_{t+1}(\theta^t, \theta_{t+1}, y^t, y_{t+1}) g^{t+1}_{\theta}(y_{t+1}|\theta_t, k_{t+1}(\theta^t, y^t))
$$

$$
\quad \quad f^{t+1}_{\theta}(\theta_{t+1}|\theta_t) dy_{t+1} d\theta_t
$$

(2.24)

Using property 3 for $G^{t+1}$, we know that

$$
k_{t+1} g^{t+1}_{\theta}(y_{t+1}|\theta_t, k_{t+1}) = \theta_t g^{t+1}_{\theta}(y_{t+1}|\theta_t, k_{t+1})
$$

Hence, we can rewrite the first equation as

$$
\frac{\partial}{\partial \theta_t} U_t(\theta^t, y^t) = \frac{1}{\theta_t} k_{t+1}(\theta^t, y^t) u'(c_t(\theta^t, y^t))
$$

(2.25)

$$
\quad + \beta \int_{\Theta \times Y} U_{t+1}(\theta^t, \theta_{t+1}, y^t, y_{t+1}) g^{t+1}_{\theta}(y_{t+1}|\theta_t, k_{t+1}(\theta^t, y^t))
$$

$$
\quad \quad f^{t+1}_{\theta}(\theta_{t+1}|\theta_t) dy_{t+1} d\theta_t
$$

Therefore, the relaxed planning problem is the following

$$
\max \int_{\Theta} U_0(\theta_0) f^0(\theta_0) d\theta_0
$$

(2.26)

subject to

$$
U_t(\theta^t, y^t) = u(c_t(\theta^t, y^t))
$$

$$
\quad + \beta \int_{\Theta \times Y} U_{t+1}(\theta^t, \theta_{t+1}, y^t) g^{t+1}(y_{t+1}|\theta_t, k_{t+1}) f^{t+1}_{\theta}(\theta_{t+1}|\theta_t) dy_{t+1} d\theta_{t+1}
$$

and (2.18),(2.19),(2.24), and (2.25).
2.3.1 Modified Inverse Euler Equation

In this section, we derive the general version of the MIEE for the model set up above. We start with a recursive formulation of the model. Since, $\theta_t$ is a first order Markov Process, a result from [Fernandes and Phelan, 2000] implies that a sufficient statistic for the history is promised utility that is the continuation utility if the agent tells the truth and threat utility, continuation utility when the agent lies – potentially a function when there is a continuum of types. The FOA, implies that a sufficient statistic for threat utility is

$$\int U_{t+1} f_{\hat{\theta}_{t-1}}(\theta|\theta_{-1}) d\theta.$$ 

Hence, given an allocation, we define the following

$$w_t(\theta_{t-1}, y^t) = \int_\Theta U_t(\theta^t, y^t) f_{\theta_{t-1}}(\theta_t|\theta_{t-1}) d\theta_t$$

$$\Delta_t(\theta_{t-1}, y^t) = \int_\Theta U_t(\theta^t, y^t) f_{\hat{\theta}_{t-1}}(\theta_t|\theta_{t-1}) d\theta_t$$

where $w_t$ is the promise utility to the agent before realization of productivity shock at $t$ and $\Delta_t$ is the sufficient statistic for keeping track of threat utility. We call $\Delta_t$, the marginal promised utility. Given the above definitions, we can rewrite the local incentive constraints as

$$u'(c_t(\theta^t, y^t)) = \beta \int_Y w_{t+1}(\theta^t, y^{t+1}) \phi_{k_{t+1}}(y_{t+1}|\theta_t, k_{t+1}(\theta^t, y^t)) dy_{t+1}$$

$$\frac{\partial}{\partial \theta_t} U_t(\theta^t, y^t) = \frac{1}{\theta_t} k_{t+1}(\theta^t, y^t) u'(c_t(\theta^t, y^t))$$

$$+ \beta \int_Y \Delta_{t+1}(\theta^t, \theta_{t+1}, y^{t+1}) \phi_{k_{t+1}}(y_{t+1}|\theta_t, k_{t+1}(\theta^t, y^t)) dy_{t+1}$$

Now, if we let $Q_t$ be the lagrange multiplier on the feasibility constraint – by Theorem 1, Section 8.3 in [Luenberger, 1969], such multiplier exists. We can interpret these multiplier as price of consumption at period $t$. Conversely, $\frac{Q_t}{Q_{t+1}}$ can be interpreted as a return on a risk free bond at period $t$. Given these prices, we can rewrite the dual of the above planning problem as follows

$$P^t(w, \Delta, \theta_{-1}) = \max_{c,k,w',\Delta',U} \left[ \int_\Theta \left[ \frac{Q_{t+1}}{Q_t} (\theta k(\theta))^\alpha - c(\theta) - k(\theta) \right. \right.$$

$$+ \left. \frac{Q_{t+1}}{Q_t} \int_Y P_{t+1}(w'(\theta, y), \Delta'(\theta, y), \theta) \phi_{k_{t+1}}(y|\theta, k(\theta)) dy \right] f^t(\theta|\theta_{-1}) d\theta$$

(P2)
subject to

\[ w = \int_\Theta U(\theta)f^t(\theta|\theta_{t-1})d\theta \]
\[ \Delta = \int_\Theta U(\theta)f^t_{\theta_{t-1}}(\theta|\theta_{t-1})d\theta \]
\[ U(\theta) = u(c(\theta)) + \beta \int_Y w'(\theta, y)g^{t+1}(y|\theta, k(\theta))dy \]
\[ \frac{d}{d\theta}U(\theta) = \frac{1}{\theta} k(\theta)u'(c(\theta)) + \beta \int_Y \Delta'(\theta, y)g^{t+1}(y|\theta, k(\theta))dy \] (2.27)
\[ u'(c(\theta)) = \beta \int_Y w'(\theta, y)g^{t+1}_k(y|\theta, k(\theta))dy \] (2.28)

Note that, the first term is the aggregate output for an agent of type \( \theta_t = \theta \) in period \( t + 1 \) and hence it is discounted by \( Q_{t+1}/Q_t \) in order to be in terms of consumption at period \( t \).

The following proposition extends (2.12) to the environment described above. Technically, it is a result of marginal cost \( P^t_w \) being an Auto Regressive process with autocorrelation \( \frac{\beta Q_t}{Q_{t+1} \beta} \) and how \( P^t_w \) is related to expected reciprocal of marginal utility.

**Theorem 2.9** Any solution to (P2) must satisfy the following Modified Inverse Euler Equation:

\[ \frac{1}{u'(c_t)} + \frac{u''(c_t)}{u'(c_t)} \left[ \frac{1}{\theta_t} k_{t+1} \mu_1 + \mu_2 \right] = \frac{Q_{t+1}}{\beta Q_t} E_t \left\{ \frac{1}{u'(c_{t+1})} + \frac{u''(c_{t+1})}{u'(c_{t+1})} \left[ \frac{1}{\theta_{t+1}} k_{t+1} \mu_1 + \mu_2 \right] \right\} \]

where \( \mu_1 \) is the costate associated with (2.27) and \( \mu_2 \) is the lagrange multiplier associated with (2.28).

Proof can be found in the appendix.

Notice that \( \mu_1 \) and \( \mu_2 \) represent the tightness of the incentive constraints at period \( t \). Similar to lemma 2.2, we can show that

\[ \mu_{2t} = -\frac{Q_{t+1}}{Q_t} \frac{1}{u'(c_t)} Cov \left( P^{t+1}_w, w_{t+1}|(\theta^t, y^t) \right) \] (2.29)

Hence, when \( P^{t+1}_w \) is decreasing with respect to \( w_{t+1} \) – an example of this is the case where \( P^{t+1}_w \) is concave and \( \theta_t \) is i.i.d., \( \mu_{2t} \) is always positive.
What the above equation implies is that the sign of distortions on saving is affected not only by the heterogeneity of consumption, as shown by [Golosov et al., 2003], but also it depends on relative tightness of incentive constraints across periods. In particular, if this tightness increases or decreases in expectation, it might change the sign of the distortions. Saving distortions are the highest when this difference is the highest. Here we perform a heuristic analysis of the above equation. In particular, suppose that $\mu_{1t}$ is always positive and that project returns are deterministic, i.e., local downward incentive constraints are binding. Note that we always have, $\mu_{1t}(\theta^{t-1}, \bar{\theta}, y^t) = 0$. In this case, (2.29) implies that

$$\frac{1}{u'(c_t)} < \frac{Q_{t+1}}{\beta Q_t} E_t \left\{ \frac{1}{u'(c_{t+1})} \right\}$$

Since $\theta_t = \bar{\theta}$, incentive constraints are relatively tighter in the future. In this case, reciprocal of marginal utility should increase. This creates a force toward decreasing the intertemporal wedge. Later, we show that the intertemporal wedge is in fact negative at the top. On the other hand, when current incentive constraints are tighter relative to future incentive constraints, we must have

$$\frac{1}{u'(c_t)} > \frac{Q_{t+1}}{\beta Q_t} E_t \left\{ \frac{1}{u'(c_{t+1})} \right\}$$

This creates a force toward increasing the intertemporal wedge and therefore the intertemporal wedge is positive. To our knowledge, this feature is new to this model. What it implies is that contrary to previous results as in [Golosov et al., 2003], there is a possibility of saving subsidies in a model with capital income risk. The closest result to the above is perhaps [Albanesi, 2006]. She shows that in an environment with moral hazard, there is a possibility of negative taxes. However, in that environment, since the source of private information is separable from consumption, Inverse Euler Equation is satisfied and intertemporal wedge is always positive. The negative tax result, however, is specific to the particular implementation rather than being a property of the optimal allocation.

---

9 The same approach can help us characterize saving distortions in a model in which period utility function is non-separable in consumption and leisure. I suspect, a similar result holds in that environment.
2.4 Optimal Taxes

In this section, we show the main result of the paper regarding negative intertemporal wedges. Moreover, we show that the progressivity result extends to the dynamic model. Finally, we provide a tax schedule that implements the optimal allocations.

2.4.1 Intertemporal Wedge

In this section we focus on the intertemporal wedge implied by the efficient allocation discussed above. In particular, we derive conditions under which its sign is negative. First, in a model with deterministic returns and i.i.d. shocks, we show that the intertemporal wedge is negative at the top and bottom and positive in the middle of the distribution of returns. Moreover, we show that when ex-ante heterogeneity is shut down, i.e., \( \theta \) is not risky and the only source of risk is in returns to investment, the intertemporal wedge is positive. When both types of risk are present, these two forces act against each other and when ex-ante heterogeneity is sufficiently high, the intertemporal wedge is negative at the top. Throughout, this section, we assume that utility function has the CARA form for which there are no wealth effect.

**Assumption 2.10** The period utility function has the CARA form \( u(c) = -\exp(-\psi c) \).

To prove our main result, negative intertemporal wedge at the top, we start with the model with safe returns and i.i.d. shocks.

**Safe Returns – A Negative Wedge Result**

Here we discuss the case where \( \theta_t \) is i.i.d. over time and the returns to investment is deterministic. That is the return to investment is \((\theta_t k_{t+1})^\alpha\) at \( t + 1 \). In this case, since the income from the project is not risky once \( \theta \) is known, \( \mu_{2t} = 0 \) and therefore

\[
\frac{1}{u'(c_t)} - \psi \frac{1}{\theta_t} k_{t+1} \mu_{1t} = \frac{Q_{t+1}}{\beta Q_t} E_t \left\{ \frac{1}{u'(c_{t+1})} - \psi \frac{1}{\theta_t} k_{t+2} \mu_{1t+1} \right\} \quad (2.30)
\]

Moreover, in this case we can show that

\[
\frac{1}{\theta_t} \mu_{1t} = \left[ \frac{Q_{t+1}}{Q_t} \alpha \theta_t^\alpha k_{t+1}^{\alpha-1} - 1 \right] \frac{1}{u'(c_t)}
\]
That is, $\mu_{1t}$ measure the distortions to productive efficiency – how different is the marginal return of an individual project from the economy-wide rate of return. In particular, using the same argument as in 2.3, we can show that $\mu_{1t} \geq 0$. This implies that the inside return from the project, $\alpha\theta_k k_{t+1}^{\alpha-1}$ is higher than the outside return $\frac{Q_t}{q_{t+1}}$. That is, the entrepreneurs in this model look “borrowing constrained”, i.e., the investment in the project is less than what it would have been without frictions. This result is related to a strand of literature in corporate finance that deals with Modigliani-Miller theorem and its determinants (see [Tirole, 2006].) Hence, the sign of the intertemporal wedge depends on how distortions to productive efficiency evolve over time. In particular, when distortions to productive efficiency are higher relative to future, $\frac{1}{u'(c_t)} > \frac{Q_{t+1}}{\beta E_t u'(c_{t+1})}$ and hence intertemporal wedge is positive. When, the distortion to productive efficiency is lower relative to future, namely at the top(bottom) of the distribution of $\theta$ where it is zero, we must have

$$\frac{1}{u'(c_t)} < \frac{Q_{t+1}}{\beta E_t u'(c_{t+1})}.$$ 

Unfortunately, this inequality cannot be used to determine the sign of the intertemporal wedge. Therefore, we use a direct argument using the recursive formulation of the problem to show that the intertemporal wedge is negative at the top(bottom). In particular, we show the negativity in two steps:

1. The margin between $c_{t-1}(w, \bar{\theta})$ and $w_t(w_{t-1}, \bar{\theta})$ is undistorted, or,

$$u'(c_{t-1}(w, \bar{\theta}))P^t_w(w_t(w_{t-1}, \bar{\theta})) = -\frac{\beta Q_{t-1}}{Q_t}$$

2. The marginal utility of increasing cost $P$ by one unit, $-\frac{1}{P_t}$, is more than the marginal utility from increasing consumption at each state, $E_t u'(c_t)$.

The first step is a natural implication of the no-distortions-at-the-top result. That is, since no other type wants to pretend to be the highest type, the margin between $c_{t-1}$ and $w_t$ is undistorted. The marginal cost of increasing utility in the future by one unit, $-\frac{Q_t}{\beta q_{t-1}}P^t_w$ is equal to the marginal benefit of decreasing utility in the current period by one unit, $\frac{1}{u'(c_{t-1})}$. Step 2 implies that a unit of saving relaxes incentive constraints in the future.
To show step 2, note that $P_t(w) = \frac{A_t}{\psi} \log(-w) + B_t$ where $A_t = \frac{1}{Q_t} \sum_{s=t}^{T} Q_s$ and hence $P'_w(w) = \frac{A_t}{w}$. Moreover, it is easy to see that the margin between $c_t$ and $w_{t+1}$ is distorted downward for all $\theta$ or

$$-\frac{1}{\beta} \frac{Q_{t+1}}{Q_t} \frac{A_{t+1}}{w_{t+1}(w_t, \theta)} \leq \frac{1}{u'(c_t(w_t, \theta))}$$

(2.31)

Consider a perturbation that increases $u(c_t)$ by one unit and decreases $w_{t+1}$ by $\frac{1}{\beta}$, this perturbation relaxes the incentive constraint (2.27) by decreasing marginal utility. The cost of such perturbation is $\frac{1}{w'(c_t)}$ and its benefit is $-\frac{1}{\beta} \frac{Q_{t+1}}{Q_t} P'_w(w_{t+1}(w_t, \theta))$. Hence, the above inequality implies that

$$-\frac{1}{\beta} \frac{Q_{t+1}}{Q_t} \frac{A_{t+1}}{w_{t+1}(w_t, \theta)} \leq -\frac{1}{u'(c_t(w_t, \theta))}$$

(2.32)

or

$$\frac{Q_{t+1}}{Q_t} A_{t+1} u(c_t(w_t, \theta)) \geq \beta w_{t+1}(w_t, \theta)$$

(2.33)

Integrating the above inequality and using promise keeping constraint implies that

$$\left[\frac{Q_{t+1}}{Q_t} A_{t+1} + 1\right] \int u(c_t(w_t, \theta)) f^t(\theta) d\theta > w$$

or

$$-\frac{A_t}{\psi} \int u'(c_t(w_t, \theta)) f^t(\theta) d\theta > w$$

and therefore

$$\int u'(c_t(w_t, \theta)) f^t(\theta) d\theta < -\frac{1}{P'_w(w_t)}$$

(2.34)

By step 1,

$$u'(c_{t-1}(w, \tilde{\theta})) = -\frac{\beta Q_{t-1}}{Q_t} \frac{1}{P'_w(w_t(w_{t-1}, \tilde{\theta}))}$$

and by step 2

$$u'(c_{t-1}(w, \tilde{\theta})) = -\frac{\beta Q_{t-1}}{Q_t} \frac{1}{P'_w(w_t(w_{t-1}, \tilde{\theta}))} > \int u'(c_t(w_t, \theta)) f^t(\theta) d\theta$$

We summarize the above discussion in the following theorem:
Theorem 2.11 Suppose that assumption 2.10 holds. Then any solution to program (P2) satisfies the following

\[
\frac{\beta Q_{t-1}}{Q_t} \int u'(c_t(w_{t-1}, \bar{\theta}), \theta) f_t(\theta) d\theta < u'(c_{t-1}(w_{t-1}, \bar{\theta}))
\]

Proof is given in the appendix.

There are two key ingredients in the above argument for step 2. The first ingredient is inequality (2.33). This inequality implies that between current utility, \(u(c_t)\), and promised utility, \(\beta w'\), the planner allocates more to \(u(c_t)\) relative to their weight in the objective \(Q_{t+1}A_{t+1}\) function. Note that, for a general utility function, inequality (2.32) holds whenever \(P_t\) is concave in \(w_t\). The fact that (2.33) is implied by (2.32) is a direct consequence of Assumption 2.10. The second ingredient is the fact that with CARA utility \(u'(c_t)\) is proportional to \(u(c_t)\) and hence using the promise keeping constraint we can show that (2.34) holds.

As noted before, this result is in contrast with the seminal result of [Golosov et al., 2003] where intertemporal wedge is always positive. In what follows, we illustrate how the two models are different and what leads to negative wedges in this model. To do so, it is useful to switch to a model with finite number of types \(\theta_1 < \cdots < \theta_N\). For illustrative reasons, we also make two other assumptions: 1. only local downward constraints are binding, 2. future promised utility is increasing in \(\theta\). Note that with local downward constraints binding, we must have

\[u(c_t) + \beta w'_i = u(c_{i-1} + k_{i-1}(1 - \frac{\theta_i}{\theta_{i-1}})) + \beta w'_{i-1}\]

Since \(w_i\) is increasing \(i\), the above equality implies that

\[c_{i-1} + k_{i-1}(1 - \frac{\theta_{i-1}}{\theta_i}) < c_i\]

This inequality implies that the current utility of an agent increases when he lies down\(^{10}\). Now consider an \(\varepsilon\) increase in \(c_i\) for all \(i\)'s. Then the LHS of the above inequality goes up by \(u'(c_i)\varepsilon\) while the RHS is increased by \(u'(c_{i-1} + k_{i-1}(1 - \theta_{i-1}/\theta_i))\varepsilon\). Above inequality and concavity of \(u\) imply that the incentive constraints are relaxed by such perturbation. The added cost of this perturbation is \(\varepsilon\) while overall utility is increased by \(\varepsilon E_{t-1} u'(c_t)\).

---

\(^{10}\) In the model where \(\theta \in [\underline{\theta}, \bar{\theta}]\), this inequality becomes \(c'(\theta) < \frac{1}{\theta} k(\theta)\). That is current utility from lying \(u(c(\theta) + k(\bar{\theta})(1 - \theta/\theta))\) is decreasing in \(\theta\) when \(\theta = \bar{\theta}\).
Hence, if we set $\varepsilon = \frac{1}{E_{t-1}w'(c_t)}$, cost increases by $\frac{1}{E_{t-1}w'(c_t)}$ and overall utility increases by 1. Optimality of the allocations then implies that $-P_w < \frac{1}{E_{t-1}w'(c_t)}$. That is the implied increase in cost from a unit increase in promised utility, $-P_w$, must be less than the added cost from a uniform increase in consumption $\frac{1}{E_{t-1}w'(c_t)}$. In other words, saving relaxes incentive constraints.

In contrast, consider the model in [Golosov et al., 2003] with the same assumptions: discrete types, local incentive constraints and increasing promised utility. In this model dis-utility of effort is separable from consumption. Hence, the local downward incentive constraints become the following

$$u(c_i) - v(l_i) + \beta w'_i = u(c_{i-1}) - v\left(\frac{\theta_{i-1}}{\theta_i}l_i\right) + \beta w'_{i-1}$$

Note that in this model, since consumption is separable from the source of private information, the margin between $c$ and $w'$ is undistorted. Hence, the fact that $w_i$ is increasing in $i$ implies that $c_i$ is also increase in $i$. Now consider an $\varepsilon$ increase in $c_i$ for all $i$'s, as before. The RHS of the above constraint increases by $u'(c_i)\varepsilon$ while its LHS is increased by $u'(c_{i-1})\varepsilon$. Concavity of $u$ together with $c_i > c_{i-1}$ then implies that this perturbation tightens the set of incentive constraint. That is saving tightens the incentive constraints and therefore intertemporal wedges are positive.

The above analysis also suggests that when $w'(\theta)$ is increasing in (P2), intertemporal wedges are negative at the top and the bottom. In fact, in the Appendix, we show that this result is true when only downward incentive constraints are binding or $\mu_1(\theta) \geq 0$\textsuperscript{11}. That is, we have the following proposition:

**Proposition 2.12** Suppose that $\theta_t$ is i.i.d. Moreover, suppose that in the solution to (P2), $w'(\theta)$ is increasing in $\theta$ and the co-state $\mu_1(\theta)$ is always positive. Then, the intertemporal wedge is negative at the top, i.e.,

$$\frac{\beta Q_{t-1}}{Q_t} \int u'(c_t(w_t(w_{t-1}\bar{\theta}, \theta)))f'(\theta)d\theta < u'(c_{t-1}(w_{t-1}\bar{\theta}))$$

\textsuperscript{11} One can show that $\mu_1(\theta) \geq 0$ whenever the value function is concave. In the appendix, we provide conditions under which the value function is concave and show how concavity of the value function leads to a positive sign for $\mu_1(\theta)$. Proof can be found in the appendix.
So far, we have assumed that the process for productivity is i.i.d. The case where \( \theta \) is persistent is worth discussing. In this case, we can do the same perturbation as above. We again assume that when an agent lies his current utility increases, i.e., \( \frac{1}{t} k(\theta) > c'(\theta) \), then a uniform increase in consumption in all states relaxes the incentive constraints. However, this perturbation increase the overall utility, \( w \), and it changes the overall marginal promised utility, \( \Delta \). What this implies is that

\[
-P^t_w \int u'(c(\theta)) f^t(\theta|\theta_-) d\theta - P^t_\Delta \int u'(c(\theta)) f^t_{-1}(\theta|\theta_-) d\theta < 1
\]

Hence, in this case, the sign of the intertemporal wedge depends on the sign of \( P^t_\Delta \) and \( \int u'(c(\theta)) f^t_{-1}(\theta|\theta_-) d\theta \). However, it is implied by the local approach that \( P^t_\Delta (w_t, \Delta_t, \bar{\theta}) = 0 \). That is, the threat keeping constraint at \( t \) is slack for the entrepreneur with the highest shock at period \( t + 1 \). This result is an implication of the first order approach. An implication of the first order approach is that no other agent wants to pretend to be the highest type. Hence, the threat keeping constraint is not binding at the top, i.e., \( P^t_\Delta (w, \Delta, \bar{\theta}) = 0 \). This implies that we again have

\[
-P^t_w \int u'(c(\theta)) f^t(\theta|\theta_-) d\theta < 1
\]

and hence the intertemporal wedge is negative. Therefore, we have the following proposition:

**Proposition 2.13** Suppose that in the solution to (P2), \( \frac{d}{d\theta} w'(\theta) > \Delta'(\theta) \) and the co-state \( \mu_1(\theta) \) is always positive. Then, the intertemporal wedge is negative at the top, i.e.,

\[
\frac{\beta Q_t-1}{Q_t} E_{t-1} u'(c_t(\theta^t)) < u'(c_{t-1}(\theta^{t-2}, \bar{\theta}))
\]

**The Role of Residual Component**

So far, we have shown that when there is no residual component to productivity, intertemporal wedge is negative at the top and bottom. Here, we discuss how inclusion of residual shocks affect the sign of the intertemporal wedge. To do so, we start from a special case where productivity is only residual and there is no heterogeneity in \( \theta \), what we call a pure moral hazard economy. In this example and under Assumption 2.10, we can show that the movements in tightness of the incentive constraint only depends on
the movements in \( Q_t \) over time and hence, in steady state MIEE becomes the same as
the Inverse Euler Equation. Therefore, saving wedges are positive.

Consider a version of the model in section 2.3, where \( \theta_t = \theta \) is fixed and known and
\( g' = g \) is time independent. Then the recursive problem becomes

\[
P_t(w) = \max_{Q_t} \frac{Q_{t+1}}{Q_t} \theta^\alpha k^\alpha - c - k + \frac{Q_{t+1}}{Q_t} \int P_{t+1}(w'(y)) g(y|k) dy \quad (P3)
\]

subject to

\[
u(c) + \beta \int w'(y) g(y|k) dy = w
\]

\[
\beta \int w'(y) g_k(y|k) dy = u'(c_0)
\]

In this problem, as before, the value function satisfies

\[
P_t(w) = B_t + A_t \log(-w)
\]

where \( A_t = \frac{1}{\psi Q_t} \sum_{s=t}^{T} Q_s \). Moreover, since there are no wealth effects, the policy func-
tions satisfy the following

\[
c_t(\theta, w) = -\frac{1}{\psi} \log(-w) + c_t^\ast \quad (2.35)
\]

\[
w_t'(w, y) = (-w) \cdot w_{t+1}^\ast(y) \quad (2.36)
\]

where \( c_t^\ast \) and \( w_{t+1}^\ast(y) \) are independent of \( w \) but dependent on time. Recall the Modified
Inverse Euler Equation. In this case, since there is no heterogeneity in \( \theta \), \( \mu_{1t} = 0 \). Moreover, we know that

\[
\mu_{2t} = -\frac{Q_{t+1}}{Q_t} \frac{1}{u'(c_t)} \text{Cov}(P_{t+1}^w, w_{t+1}^\ast | y')
\]

Given the above properties of the policy functions, it is easy to see that \( \text{Cov}(P_{t+1}^w, w_{t+1}^\ast) \)
is independent of individual history, \( w_t \). This is due to the fact that \( P_{t+1}^w \) is proportional to \( \frac{1}{w_t} \) and \( w_{t+1} \) is proportional to \( w_t \). Hence \( \text{Cov}(P_{t+1}^w, w_{t+1}^\ast) \) is independent of \( w_t \). That is

\[
\mu_{2t} = -\frac{Q_{t+1}}{Q_t} \frac{1}{u'(c_t)} \text{Cov}(A_{t+1}^w, w_{t+1}^\ast)
\]

Therefore, the Modified Inverse Euler Equation becomes

\[
\frac{1}{u'(c_t)} \left[ 1 + \frac{Q_{t+1}}{Q_t} \text{Cov}(A_{t+1}^w, w_{t+1}^\ast) \right] = \frac{Q_{t+1}}{\beta Q_t} \left[ 1 + \frac{Q_{t+2}}{Q_{t+1}} \text{Cov}(A_{t+2}^w, w_{t+2}^\ast) \right] E_t \frac{1}{u'(c_{t+1})}
\]
Note that the terms in the brackets are time dependent but independent of the individual history. In particular they depend on the aggregate state of the economy represented by $Q_t$. However, if we assume that $T = \infty$ and the economy is on aggregate in Steady State, i.e., $Q_t/Q_{t+1}$ is constant over time, the term in the bracket becomes constant, since the Bellman equation described above becomes time independent. Hence, in steady state, the usual Inverse Euler Equation emerges and we have

$$\frac{1}{u'(c_t)} = \frac{Q_{t+1}}{\beta Q_t} E_t \frac{1}{u'(c_{t+1})}.$$ 

Therefore, the intertemporal wedge must be positive in steady state. Notice that during transition to steady state, the sign of the wedge might change since the tightness of the incentive constraint depends on the aggregate state of the economy.

A comparison with the model with productivity shocks provides better intuition regarding the differences that cause the change in the sign of the wedge. The best way to describe the negative wedge result in the model with productivity shocks is to consider a small decrease in current consumption by $\varepsilon$ accompanied by a uniform increase in consumption in the next period by $q^{-1}\varepsilon$. Such perturbation has two effects: utility effect and incentive effect. The utility effects are standard: there is a utility benefit $\varepsilon q^{-1}\beta E_t u'(c_{t+1})$ and a utility cost $\varepsilon u'(c_t)$. As for incentive effects, when productivity is currently at the top(bottom), the incentive constraint is not binding currently. Hence, such a perturbation does not have any effect on current incentive constraints. However, due to the reasons discussed above it relaxes incentive constraints in the next period. Hence, the utility benefit of this perturbation, $\varepsilon q^{-1}\beta E_t u'(c_{t+1})$ must be less than its utility cost $\varepsilon u'(c_t)$ and the intertemporal wedge has to be negative.

We can apply the same perturbation in the pure moral hazard economy. The difference is that the incentive effects of such perturbation are more complicated. In fact in the pure moral hazard model, since the incentive constraint is always binding, such perturbation has an effect on current incentives. The perturbation tightens up the incentive constraint since it increases current marginal utility and decreases the slope of promised utility profile in the next period. This perturbation also affects incentives in the next period and relaxes the incentive constraints since it decreases marginal utility. The above analysis shows that in steady state the current costs from tightening the incentive constraints are higher than the future benefit from relaxing the incentive
constraints. Hence, the intertemporal wedges are positive.

So far, we have analyzed two extreme cases: when the residual risk is shot down and when entrepreneurs do not know anything in advance about their future productivity. We have shown that the two extreme cases have different implications on the intertemporal wedge. The novel result of this paper is in fact that intertemporal wedge is negative at the top and bottom when productivity has no residual component. As we have seen, when the known component of productivity is shut down, in Steady State, Inverse Euler Equation emerges and intertemporal wedge is positive. Hence, the result on the general model is indeterminate. As the perturbation argument shows, the sign of the wedge depends on the relationship between current incentive costs of decreasing consumption versus benefits from relaxing the incentive constraints in the future. In fact, in the general model as in the pure moral hazard model, would relax both types of incentive constraints in the future and would tighten the current moral hazard constraint. In section 2.6, we use a reasonably calibrated version of the model and show that intertemporal wedge is negative at the bottom and positive at the top.

2.4.2 Progressive Taxes on Entrepreneurial Income

In this section, we study whether the progressivity of the tax schedule with respect to entrepreneurial income generalizes to the dynamic model. We do so, by characterizing the shape of consumption in each period as a function of income. As, the two period example in section 2.2.3 shows, movements in current consumption as a function of current income, depends on the EIS and the hazard rate $g_k$. In this section, we study the economy with exponential utility function. We show that in the general model described above, when $\theta$ is i.i.d., the inverse of marginal utility is a linear function of the hazard rate. Since, the inverse of marginal utility is a convex function, the shape of the consumption schedule, i.e., the shape of the tax function, is solely determined by the shape of the hazard rate. In particular, when the hazard ratio is concave, the consumption is concave in income realization and tax schedule is progressive. Moreover, when $\theta$ is persistent, the consumption schedule is concave for the highest and lowest value of productivity.

Consider the economy in section 2.3. The first order conditions imply that at each
\[ -P_{w}^{t+1} = a_t + b_t \frac{g_k(y_{t+1} \mid k_{t+1}, \theta_t)}{g(y_{t+1} \mid k_{t+1}, \theta_t)} \]  \hspace{1cm} (2.37)

where \( a_t \) and \( b_t \) are independent of \( y_{t+1} \). In fact, \( b_t = \beta \mu \) and \( a_t \) is a function of the tightness of the first incentive constraint. These multipliers, depend solely on the past history of shock as well as the current realization of \( \theta_t \). When \( \theta_t \) is i.i.d., the value function satisfies \( P^t(w) = A_t \log(-w) + B_t \) and consumption policy function satisfies

\[ c_t(w, \theta) = -\frac{1}{\psi} \log(-w) + \hat{c}_t(\theta) \]

This implies that \( u'(c_t(w, \theta)) = (-w)u'(\hat{c}_t(\theta)) \) and that \( P^t_w = \frac{A_t}{w} \). Therefore, (2.37) becomes the following

\[ \frac{A_{t+1} u' (\hat{c}_{t+1}(\theta))}{u'(c_{t+1}(w, \theta))} = a_t + b_t \frac{g_k(y_{t+1} \mid k_{t+1}, \theta_t)}{g(y_{t+1} \mid k_{t+1}, \theta_t)} \]

Hence, \( \hat{a}_t \) and \( \hat{b}_t \) exist that are history dependent and independent of \( y_{t+1} \) such that

\[ \frac{1}{u'(c_{t+1}(w, \theta))} = \hat{a}_t + \hat{b}_t \frac{g_k(y_{t+1} \mid k_{t+1}, \theta_t)}{g(y_{t+1} \mid k_{t+1}, \theta_t)} \]  \hspace{1cm} (2.38)

Therefore, since \( c_{t+1} \) is concave in \( y_{t+1} \) when \( \frac{\partial k}{\partial y} \) is concave in \( y \). In particular, for the examples given in section 2.2.3, the same analysis holds and consumption schedule is concave in \( y \).

When \( \theta_t \) is persistent, we can show that the value function

\[ P^t(w, \Delta, \theta_-) = A_t \log(-w) + B_t \left( \frac{\Delta}{w}, \theta_- \right) \]

Hence, in this case, the shape of \( B_t \) affects \( P_w \) and hence the above analysis does not apply. However, when \( \theta_- = \tilde{\theta} \), we have \( P^t_{\Delta} = 0 \). Therefore, in that case, the above analysis applies and (2.38) determines the shape of the consumption schedule.

An important assumption above is that given history of actions, \( y_t \) and \( \theta_t \) are independent. This implies that the marginal cost of increase utility by a unit, \( P^t_w \), is related to the inverse of marginal utility in a way described above. However, when \( y_t \) and \( \theta_t \) are perfectly correlated, this is not necessarily true. Considering a correlation between \( y_t \) and \( \theta_t \) would further complicate the model and we do not pursue this idea here.

Given the above analysis, an investigation of entrepreneurial income processes is required in order to determine the progressivity of the tax schedule. Moreover, a key
assumption that leads to this result on marginal tax rates is that, markets are unable to provide any insurance. However, as shown by [Kaplan and Strömberg, 2003]'s analysis of Venture Capital contracts and [Bitler et al., 2005]'s analysis of SCF data shows, certain features of observed private equity contracts are consistent with the optimal contracting theory. This evidence suggests that markets are able to provide some insurance. Hence, a natural question is what is the role of government in providing insurance. Although an important question, this question is beyond the scope of this paper. In [Shourideh, 2010], we partially try to address this question by considering an environment where there is a role for government and study its implication for optimal taxation.

### 2.4.3 A Tax Implementation

In this section, we analyze the implementation of efficient allocations discussed above. In particular, we show that the tax functions used in the two period example can be extended to the multi-period model. To do so, we impose that agents have only access to a risk free asset \( b_t \) at each date traded at price \( Q_t \). We also assume that the planner can observe \( b_t \) as well as income at each period \( t \). Thus, our aim is to find a tax schedule \( \{ T_t(y^t, b^t) \}_{t=0}^T \), where \( b^t \) is the history of asset holdings for each agent and \( y^t \) is the history of income. The value \( T_t \) is the tax paid by the agent at period \( t \). Later, we discuss how the properties of the optimal allocations discussed above translate into properties of the tax function.

Facing such schedule, each agent solves the following maximization problem

\[
\max \sum_{t=0}^{T} \beta^t \int_{\Theta^{t+1} \times Y^t} u(c_t(\theta^t, y^t)) d\mu_t(\theta^t, y^t; k^{t-1}, y^{t-1})
\]

subject to

\[
c_t(\theta^t, y^t) + k_{t+1}(\theta^t, y^t) + \frac{Q_{t+1}b_{t+1}(\theta^t, y^t)}{Q_t} \leq b_{t}(\theta^{t-1}, y^{t-1}) + y_t - T_t(y^t, b^{t+1}(\theta^t, y^t))
\]

Now consider the optimal allocation \( \{ c^*_t(\theta^t, y^t), k^*_t(\theta^t, y^t) \}_{t=0}^T \) which is the solution to the planning problem (P). A difficulty, in extending the implementation to a multi-period model is an indeterminacy in the level of risk free asset. Note that in the two period model, the asset level at period 0 is pinned down. For each agent, the level of
non-entrepreneurial wealth is given by \( q^{-1}[e_0 - c_0(\theta) - k_1(\theta)] \). When the number of periods is more than 2 and in an intermediate period, the allocation only pins down the number \( Q_t b_{t+1} + T_t (y^t, b_{t+1}) \). In order to show that the tax system is progressive in income, we construct \( b_{t+1} \) such that its slope as a function of \( y_t \) is the same as the slope of \( -\frac{Q_{t+1}}{Q_t} k_{t+1} \). This assumption helps us later in proving that the tax schedule is progressive.

Formally, the construction of \( b^*_t \) and \( T_t \) given the optimal allocation is as follows:

1. At \( t = 0 \), non-entrepreneurial wealth is defined as
   \[
   b^*_1(\theta_0) = \frac{Q_0}{Q_1} (e_0 - c_0^*(\theta_0) - k_1^*(\theta_0))
   \]

2. At intermediate period \( 1 \leq t \leq T - 1 \), given a history \((\theta^{t-1}, y^t)\) and the optimal allocation:
   \[
   b^*_{t+1}(\theta^t, y^t) = -\frac{Q_t}{Q_{t+1}} k^*_{t+1}(\theta^t, y^t) - \xi_t(\theta^t; \theta^{t-1}, y^{t-1}),
   \]

   The function \( \xi_t \) is arbitrary \( C^1 \) function that makes \( b^*_{t+1}(\theta^t, y^t) \) monotone. Obviously we must have

   \[
   \int_{\Theta^{t+1} \times Y^t} \left[ \xi_t(\theta^t, y^{t-1}) + \frac{Q_t}{Q_{t+1}} k^*_{t+1}(\theta^t, y^t) \right] d\mu_t(\theta^t, y^t) = 0.
   \]

   The above construction of risk-free asset holdings, simply pins down the tax function. However, in order for the tax function to be well-defined, the following assumption on allocations is needed:

**Assumption 2.14** For all histories \((\theta^{t-1}, y^t)\), \( b^*_{t+1}(\theta^{t-1}, \tilde{\theta}_t, y^t) \neq b^*_{t+1}(\theta^{t-1}, \hat{\theta}_t, y^t), \forall \hat{\theta}_t \neq \theta_t \). Moreover, \( b^*_{t+1}(\theta^{t-1}, \theta_t, y^t) \) is continuous in \( \theta_t \) for all \( t \).

The first part of assumption 2.14 is similar to an assumption in [Kocherlakota, 2005] regarding income – Assumption 1, page 1601. However introduction of \( \xi_t \) makes it easier to ensure that the above assumption is satisfied. The second part, ensures that \( b^*_{t+1}(\theta^{t-1}, \cdot, y^t) \) belongs to an interval \( I_t(\theta^{t-1}, y^t) \).

Note that by intermediate value theorem, for any value \( \hat{b} \in I_t \), there is a unique \( \hat{\theta}_t \) such that \( b^*_t(\theta^{t-1}, \hat{\theta}_t, y^t) = \hat{b} \). Given the above assumption, we can define the following
well-defined tax function

\[
T_t(y^t, b^{t+1}) = \begin{cases} 
    b_t^*(\theta^{t-1}, y^{t-1}) + y_t - c_t^*(\theta^t, y^t) & \text{if } b^{t+1} = b^t + (\theta^t, y^t) \\
    -k_{t+1}^*(\theta^t, y^t) - \frac{Q_{t+1}}{Q_t} b_{t+1}^*(\theta^t, y^t) & \text{otherwise}
\end{cases}
\]  

(2.41)

where \(b^{t+1} = (b_1, \ldots, b_{t+1})\). Note that taxes paid in the current period depend on the current level of non-entrepreneurial asset holding \(b_{t+1}\).

In what follows, we show that the above tax function implements the optimal allocation. That is given above \(T_t\), the optimal allocation is a solution to (2.39).

Given assumption 2.14, it is easy to see how the above tax system implements the allocation. First, notice that the above definition of \(T_t(\cdot, \cdot)\) ensures that \(b_{t+1} \in I_t\), otherwise the agent loses all his income. Hence, we only restrict attention to asset choices \(b_{t+1} \in I_t\). In particular, suppose an agent picks a sequence of non-entrepreneurial asset levels \(\{b_{t+1}(\theta^t, y^t)\} \neq \{b_{t+1}(\theta^t, y^t)\} \) where \(b_{t+1}(\theta^t, y^t) \in I_t(\theta^{t-1}, y^t)\). We first show that such sequence of asset holding is equivalent to a reporting strategy \(\hat{\sigma}_t(\theta^t, y^t)\).

Then, incentive compatibility of the optimal allocation implies that such strategy is weakly dominated by truth-telling or \(\{b_t^*(\theta^t, y^t)\}\). Starting from period 0 and the fact that \(\hat{b}_1(\theta_0) \in I_0\), implies that \(\hat{b}_1(\theta_0) = b_1^*(\hat{\sigma}_0(\theta_0))\) for some function \(\sigma_0\). Given \(\sigma_0\) and \(\hat{b}_2(\theta^1, y^1)\), there must exist a \(\hat{\sigma}_1(\theta^1, y^1)\) such that \(\hat{b}_2(\theta^1, y^1) = b_2^*(\hat{\sigma}_0(\theta_0), \hat{\sigma}_1(\theta^1, y^1), y^1)\).

Similarly, we can construct a reporting strategy \(\hat{\sigma}_t(\theta^t, y^t)\) for all possible histories. Hence the choice of asset positions \(\{\hat{b}_{t+1}(\theta^t, y^t)\}\) is equivalent to the choice of a reporting strategy \(\hat{\sigma}^t\).

Note that given \(\hat{b}_{t+1}(\theta^t, y^t)\) the budget constraint for the agent is given by

\[
c_t(\theta^t, y^t) + k_{t+1}^*(\theta^t, y^t) + \frac{Q_{t+1}}{Q_t} b_{t+1}^*(\hat{\sigma}^t(\theta^t, y^t), y^t) \leq b_t^*(\hat{\sigma}^t(\theta^{t-1}, y^{t-1}), y^{t-1}) + y_t - T_t(y^t, b^t + (\theta^t, y^t))
\]

Hence by construction of the tax function in (2.41), the total amount available for consumption and investment is given by

\[
b_t^*(\hat{\sigma}^t(\theta^{t-1}, y^{t-1}), y^{t-1}) + y_t - T_t(y^t, b^t + (\hat{\sigma}^t(\theta^t, y^t), y^t)) - \frac{Q_{t+1}}{Q_t} b_{t+1}^*(\hat{\sigma}^t(\theta^t, y^t), y^t) = c_t^*(\hat{\sigma}^t(\theta^t, y^t), y^t) + k_{t+1}^*(\hat{\sigma}^t(\theta^t, y^t), y^t)
\]
That is given \( \hat{b}^t \) and \( \hat{\sigma}^t \), the agent solves the following problem

\[
\max_{\beta} \sum_{t=0}^{T} \beta^t \int_{\Theta^{t+1} \times Y^t} u(c_t(\theta^t, y^t))d\mu_t(\theta^t, y^t; k^t(\theta^{t-1}, y^{t-1})) \quad (2.42)
\]

subject to

\[
c_t(\theta^t, y^t) + k_{t+1}(\theta^t, y^t) \leq c_t^*(\hat{\sigma}^t(\theta^t, y^t), y^t) + k_{t+1}^*(\hat{\sigma}^t(\theta^t, y^t), y^t) \quad (2.43)
\]

Therefore, the highest utility achievable to the agent is

\[
\max_{\hat{k}} U(\{c_t^*, k_{t+1}^*\}; \hat{\sigma}, \hat{k})
\]

Hence, by incentive compatibility the utility received by the agent is the highest if the agent chooses \( \{b^{t+1}_*(\theta^t, y^t)\} \), i.e., tells the truth, and follows the optimal investment. This implies that the tax function implements the optimal allocation. We summarize this discussion in the following theorem:

**Theorem 2.15** Consider the optimal allocation \( \{c_t^*(\theta^t, y^t), k_{t+1}^*(\theta^t, y^t)\}_{t=0}^{T} \) and suppose that assumption 2.14 holds. Then the tax function as defined by (2.41) implements the optimal allocation.

This implementation is similar to [Kocherlakota, 2005]'s implementation in an economy with labor income risk. In both implementations, the tax function at a period is a function of income and risk free asset holdings in that period. However, as [Werning, 2010] has shown, for the economy with labor income risk there is another implementation that separates the tax paid on asset holding and on labor income. However, in our economy this separation is not possible since allocations are not separable in \( \theta \) and \( y \). Hence this non-separability is necessary for implementation of the optimal allocation.

To see how the properties of the allocations discussed above translate into properties of the tax function. To do so, we assume that the optimal allocations are \( C^1 \) w.r.t. histories. That is \( c_t^*(\theta^t, y^t) \) and \( k_{t+1}^*(\theta^t, y^t) \) are continuously differentiable with respect to each \( \theta_s \) and \( y_s \) for all \( 0 \leq s \leq t \). By definition, \( b_t^* \) and \( T_t \), are also continuously differentiable by definition since they are constructed from \( c_t^* \) and \( k_t^* \) using \( C^1 \) transformations.
Hence, any solution of (2.39) must satisfy the following first order condition:

\[
u'(c_t(\theta^t, y^t)) = \beta \frac{Q_t}{Q_{t+1}} \int_{\theta \times Y} u'(c_{t+1}(\theta^{t+1}, y^{t+1}))(1 - T_{b_{t+1}}(y^{t+1}, b^*_{t+2}))g^{t+1}(y_{t+1}|\theta_t, k_{t+1})
\]

\[f^{t+1}(\theta_{t+1}|\theta_t)dy_{t+1}d\theta_{t+1}\]

Therefore, the intertemporal wedge is given by

\[
\tau^*_t(\theta^t, y^t) = \frac{E_t[u'(c_{t+1})T_{b_{t+1}}(y^{t+1}, b^*_{t+2})]}{E_t u'(c_{t+1})}
\]

and is a weighted average of the derivative of the function \(T(\cdot, \cdot)\) weighted by marginal utility. Hence, our result on negative marginal tax rates imply that on average \(T_b\) must be negative whenever \(b_{t+1} = b^*_{t+1}(\theta^{t-1}, \bar{\theta}, y^t)\).

Next, we turn to the shape of the tax function with respect to income realization. Given the way we have constructed non-entrepreneurial asset holdings, we know that

\[
\frac{\partial}{\partial y_t} b^*_{t+1}(\theta^t, y^t) = -\frac{Q_t}{Q_{t+1}} \left[ \frac{\partial}{\partial y_t} k^*_{t+1}(\theta^t, y^t) \right]
\]

Hence, using (2.41), we must have

\[
\frac{\partial}{\partial y_t} T(b^t, y^t) = 1 - \frac{\partial}{\partial y_t} c^*_t(\theta^t, y^t) - \frac{Q_{t+1}}{Q_t} \frac{\partial}{\partial y_t} b^*_{t+1}(\theta^t, y^t)
\]

That is, the shape of the tax function with respect to income realization is the same as the shape of the consumption schedule as a function of income realization. Hence, whenever consumption is concave in income, the marginal tax rate or \(\frac{\partial}{\partial y_t} T(b^t, y^t)\) is increasing and therefore the tax schedule is progressive.

### 2.5 Implementation with Private Contracts

In this section, we study how private contracts can implement the optimal allocation. We do so by showing that there is an implementation of the optimal allocation that uses the types of contracts used in typical venture capital contracts as documented by [Kaplan and Strömberg, 2003]. We do this in the context of the pure moral hazard model described in section 2.4.1. Moreover, we assume that assumption 2.10 holds and the hazard ratio, \(\frac{g_t}{y}\), is concave in \(y\).
We first describe the types of securities used in our implementation – throughout we assume that the environment is comprised of an outside lender and the entrepreneur:

1. Equity: The equity holders collect dividends paid in each period. At each period, the entrepreneur and the lender own parts of the company and the ownership is evolving over time.

2. Short Term Convertible Debt: this security is risk free debt together with $N$ options. Upon exercising option $i$, $1 \leq i \leq N$, the holder can buy a certain number of shares –fraction $\hat{s}^i$ of total equity, at a pre-specified price, $e^i$. Both sequences $\{\hat{s}^i\}_{i=1}^N$ and $\{e^i\}_{i=1}^N$ are increasing in $i$.

3. Credit Line/Saving Account: A bank account that the entrepreneur can borrow and save with a variable interest rate. The interest rate only depends on the ownership structure of the firm, i.e., the fraction of the equity owned by the entrepreneur.

These securities are very standard and as documented by many authors\footnote{See \cite{Kaplan and Strömberg, 2003}, \cite{Sahlman, 1990}, and \cite{Gompers, 1999}, among others.}, are widely used in venture capital contracts. In what follows, we show that the above securities can approximately implement the optimal allocation – as $N$ tends to $\infty$, the implemented allocation converges to the optimal allocation. Given the above securities, the timing is as follows:

1. At the beginning of the each period and before realization of income, the entrepreneur buys all the shares from the outside lender.

2. Income is realized.

3. The outside lender decides whether to convert the convertible debt.

4. Investment is made by the entrepreneur.

5. Dividends are paid out.

6. The entrepreneur can decide to save or draw funds from the credit line and new convertible debt is issued by the outside lender.
Given the above timing, it is useful to introduce a bit of notation:

- Amount of convertible debt issued by the outside lender: $D$, with price $p$,
- Total equity value of the firm before realization of income in each period: $V_t$,
- Share of the entrepreneur in the company: $s_t$,
- Entrepreneur’s debt level: $B_t$; negative values are associated with saving.
- Interest rate on credit line/saving account: $R(s_t)$,
- Conversion decision at option $i$: $j_t(i) \in \{0, 1\}$. Note that the outside lender can only exercise one of the options and therefore $\sum_{i=1}^N j_t(i) \in \{0, 1\}$.

The securities defined above and the sequence of actions lead to the following budget constraint for the entrepreneur:

$$\sum_{i=1}^N e^i j_t(i) + y_t + B_{t+1} - B_t + pD = R(s_{t-1})B_t + d_t + k_{t+1} + D\left(1 - \sum_{i=1}^N j_t(i)\right) + (1 - s_{t-1})V_t$$

Moreover, $c_t = s_t d_t$. Note, also, that due to the buy-back of the stocks in the beginning of the period $s_t = 1 - \sum_{i=1}^N \hat{s}^i j_t(i)$. The LHS of the budget constraint is revenue available to the firm from various sources: revenue from sale stock in case of conversion, income, credit drawn from the credit line, and money raised through issuance of convertible, respectively. The RHS of the budget constraint is the expenses paid: interest payment on the credit balance, dividends, investment, payments to convertible debt holders in case of no conversion, and the cost of share buy-back.

We need to further describe the conversion decision by the outside lender. Since this debt matures every period, the conversion decision can be easily described by a one time optimality condition. The holder of the debt will convert at option $i$ if and only if the value of converting: $(1 - \hat{s}^i)(d_t + qV_{t+1}) - e^i$ exceeds the face value of the debt $D$.

Moreover, the value of the stock to outside lender is given by

$$V_t = E_{t-1} \sum_{\tau=0}^\infty q^{-\tau} d_{t+\tau}$$

In order to show that the above implementation works, we first show that there exists a fixed interest rate $\hat{R}$ and a transfer function $T(y)$ from the entrepreneur, such
that any stationary efficient solution to (P3) can be implemented where entrepreneur can freely borrow and save at rate $\hat{R}$ and $T(y)$ is taken away from the entrepreneur in each period.

**Lemma 2.16** Consider a solution to (P3) where $Q_{t+1}/Q_t = \hat{q}$ with $\{c_t(y^t), k^*, w_0\}$. Then, there exists a function $T(y)$, interest rate $\hat{R}$ and debt level $B_0$ such that the allocation is the solution to the following optimization problem

$$\max_{k_{t+1}, c_t, B_{t+1}} \sum_{t=0}^{\infty} \beta^t \int_{Y^t} u(c_t(y^t)) d\mu_t(y^t; k^t)$$

subject to

$$c_t + k_{t+1} + (1 + \hat{R})B_t = B_{t+1} + y_t - T(y_t)$$

Proof can be found in the appendix.

The idea behind this lemma is simple. It can be shown that in the pure moral hazard model, due to no wealth effect, the intertemporal wedge is constant in all states. This implies that facing an interest rate $1 + \hat{R} = (1 - \tau_s)q^{-1}$, the agents Euler equation will be satisfied. Moreover, given the policy functions in (2.35)-(2.36), income at period $t$ affects $c_t(y^t)$ in an additively separate way. Moreover, stationarity implies that the tax function is independent of time. Note that the concavity of the hazard ratio, implies that $T(\cdot)$ is a convex function of income.

Lemma 2.16 guides us toward our main implementation result. That is, we show that the above securities can replicate the above transfer function and interest rate. In order to prove our main theorem, we need to make one further assumption and that is:

**Assumption 2.17** The optimal allocation satisfies $\frac{\partial}{\partial y} c_t(y^t) \leq 1$, $\forall y_t \in [0, \bar{y}]$.

The above assumption implies that the in each period, the total payment to the outside lender $y_t - c_t - k_{t+1}$ is increasing in $y_t$. When this assumption is violated, there will be a region such that the slope is bigger than 1. In that case the payment to outside lenders decreases following an increasing in $y_t$. Although it is possible to modify the above implementation in order to implement the optimal allocation, for simplicity we make the above assumption.

The following theorem, contains our main implementation result:
Theorem 2.18 Consider a sequence \( y^1 < \cdots < y^N \) and a solution to (P3) where \( Q_{t+1}/Q_t = \hat{q} \) given by allocations \( \{c_t(y^k), k^*, w_0\} \). Then there exists \( \{e^l\}, \{\hat{s}^l\}, D, p, R(s) \) and \( B_0 \) such that the above security structure exactly implements the allocation for all histories \( y^t \in \{y^1, \cdots, y^N\}^t \).

Proof can be found in the appendix.

The idea behind this implementation can be seen from lemma 2.16. First, we note that by concavity of hazard ratio \( T(y) \) is convex and by assumption 2.17 its slope is always positive. Moreover, the payoff schedule resulting from the convertible debt is a piece-wise linear function with increasing slope. Therefore, the role of the convertible debt is to approximate the function \( T(y) \). However, conversion implies that the outside lenders will be equity holders is in the future. This reduces the incentive for the entrepreneur to invest optimally in the firm. The role of share buy-back is dispose of this problem. Since, the buy-back is done before new investment is made, Finally, since the ownership of the entrepreneur is changing over time the interest rate needs to be changing. In fact, the Euler equation from entrepreneur’s decision problem is given by

\[
 s_t^{-1}u'(c_t) = \beta R(s_t)E_t s_{t+1}^{-1}u'(c_{t+1})
\]

Given the policy functions in (2.35)-(2.36), \( R(s_t) = \frac{1}{\beta s_t E_t s_{t+1}^{-1}[-w^*_t]} \).

The above theorem implies that our security structure can approximately implement the optimal allocation. So, we have the following corollary:

Corollary 2.19 As \( N \to \infty \), the allocation implemented by the above security structure converges to the optimal allocation.

Note that the above implementation is not unique. In fact, we can combine all of the above securities into one security. We can, also, implement the optimal allocation using only debt/saving with a variable interest rate, as used in [Quadrini, 2000]. The value of this implementation is that it points to the role of convertible debt and share buy back\(^{13}\). Moreover, it shows that the allocation can be approximately implemented using securities widely used in venture capital contracts.

\(^{13}\) [Green, 1984] in a two period model shows that a convertible debt with one conversion option does better than non-convertible debt in providing investment incentives to the shareholders. His results have the same flavor as ours. He, however, does not provide optimal security design based on underlying frictions.
The above analysis shows that the implication of the model on taxation is mixed. In fact, we have shown that it is possible for market arrangements that are commonly used in financial contracts to achieve constrained efficiency. Given such arrangements, there is no reason for the government to use taxes to achieve constrained efficiency. In that case, government only crowds out private markets. In [Shourideh, 2010], we try to resolve this issue by allowing for unobservable trades among entrepreneurs. In this case, the price of the risk free bond affects entrepreneur’s incentive for investment and hence private contracts cannot implement constrained efficient allocations. We analyze the optimal policy in a two period model and show that optimal tax policy is linear tax function on income.

2.6 Numerical Simulations

In this section, to fully characterize the properties of optimal allocations, I use a calibrated version of the model in order to calculate optimal intertemporal wedges as well as taxes on entrepreneurial income. To do so, I consider an EJ economy in which \( \theta_t \) is i.i.d and \( \log \theta_t \sim N(\log A - \frac{1}{2} \sigma_{\theta}^2, \sigma_{\theta}^2) \). Moreover, I assume that \( \varepsilon \sim \Gamma(\sigma_{\varepsilon}^{-2}, \sigma_{\varepsilon}^2) \). I keep the assumption that the utility function is exponential, \( u(c) = -e^{\psi c} \). Next, I describe how each parameter is calibrated.

To calibrate the economy, I need to calibrate the parameters \( (\alpha, \beta, \psi, A, \sigma_{\theta}, \sigma_{\varepsilon}) \). I assume that \( \beta = .96 \) so that each period is associated with a year. As we have mentioned before, \( \alpha \) should be thought of as a span of control parameter. Hence, to calibrate \( \alpha \), we consider an entrepreneur with production function \( e(\eta l^{1-\eta})^{\nu} \) who adjusts the labor input, \( l \), once shock \( e \) is realized\(^\text{14} \). This maximization decision implies that the production function can be written as \( e^{k(1-\eta)\nu} \). Hence, the implied share of inputs (other than managerial talent) as a fraction of income is given by \( \frac{\eta l^{1-\eta}}{1-(1-\eta)\nu} \). Further, notice that since we have assumed that capital fully depreciates, we need to adjust \( \alpha \) in order to take that into account. Hence, in this model \( \alpha \) is given by

\[
\alpha = \frac{\text{Payments to factors other than managerial talent} + \text{K-Depreciation}}{\text{Output} + \text{K-Depreciation}}
\]

\(^\text{14} \) See footnote 6.
As we have discussed above,

Payments to factors other than managerial talent = \( \frac{\eta \nu}{1 - (1 - \eta)\nu} \) Output

Given the value \( \alpha \), I pick \( A \) so that output for the average firm is normalized to 1.

Calibrating the variance of productivity is problematic, since there are no precise estimate for this process. [Moskowitz and Vissing-Jørgensen, 2002], study the private-equity returns using the Survey of Consumer Finances but are unable to provide precise estimates for variances of returns at individual level. For their benchmark calculations, they use 0.3 for cross-sectional standard deviation of private-equity firms. [Angeletos, 2007] uses 0.2 in a model where only residual risk in productivity is present. I assume that cross-sectional standard deviation of productivity \( \varepsilon \theta^\alpha \) can take value of \{0.2, 0.3, 0.4, 0.5\}. Moreover, inspired by the estimates of [Evans and Jovanovic, 1989], I assume that \( \sigma_\varepsilon = \alpha \sigma_\theta \). This assumption pins down the value of \( \sigma_\varepsilon \) and \( \sigma_\theta \). For the risk aversion parameter, I use \( \psi = 10^{15} \).

Using these parameter values, I compute the model. In doing so, I use a truncated distribution for \( \theta \). As noted above, I use a first order approach to simplify the set of incentive constraints. I assume that the model is in steady state, i.e., \( \frac{Q_{t+1}}{Q_t} = q \) is constant. Since I have assumed an exponential utility function, the policy functions satisfy the following:

\[
\begin{align*}
    c(w, \theta) &= -\frac{1}{\psi} \log(-w) + \hat{c}(\theta) \\
    w'(w, \theta, y) &= (-w)\hat{w}(\theta, y)
\end{align*}
\]

This implies that the difference in consumption across periods is given by

\[
\int_\Theta \int_Y \frac{1}{\psi} \log(-\hat{w}(\theta, y)) g(y|\theta, k(\theta)) f(\theta) dy d\theta
\]

I find \( q \) so that the above integral is zero. This implies that the total expenditures in the economy do not change over time. Whether total income \( \int \theta^\alpha k^\alpha \) is greater or less than total expenditure \( \int c + k' \), depends on the initial value of promised utility. Because, I am considering exponential utility, the distortions, i.e., intertemporal wedge and the slope of consumption schedule, do not depend on the initial value of the promised utility.

\[15\] I have considered various values for \( \psi \), in the range of 1-20. Surprisingly, the results do not change that much for these parameter values.
Intertemporal Wedge. Figure 2.2, below shows how intertemporal wedge depends on productivity.

In fact, the intertemporal wedge is negative and quite large for low values of $\theta$ and positive for high values. To understand these results, we should note that intertemporal wedge is closely related to the variance of growth rate of consumption. In this model, there are two forces that create variability for consumption. First effect comes from that fact that $\theta$ is private information. This implies that consumption should depend on $\theta$ in order to give incentive for more productive types to invest. The second effect is a result of moral hazard in addition to heterogeneity in $\theta$. Moral hazard implies that the planner, should given incentive to entrepreneurs to invest optimally by creating spread in their consumption in future as well as increasing their consumption in the current period. Moreover, due to decreasing return to scale, the planner wants higher productivity types to invest more in their business. This suggests that the moral hazard problem
is tighter for higher ability types. Hence, it is optimal to have current consumption positively correlated with $\theta$. In order to investigate the contribution of each of these effects, we consider two extreme cases: A case where we shut down the residual risk, $\varepsilon$, and a case where $\varepsilon$ is present and $\theta$ is public information. For the case where there is no residual risk, the intertemporal wedge is shown in Figure 2.3.

We can see that relative to Figure 2.2, the distortions are very small. This implies that the variability in consumption is small in a model with no residual risk. Note that, these number are still large in comparison to [Farhi and Werning, 2010a] or [Golosov et al., 2010], since productivity shocks are i.i.d. Persistence decreases variability of consumption growth rate at an individual level. Figure 2.4 plots the intertemporal wedge for the case that $\theta$ is fully persistent (this case is equivalent to the pure moral hazard model discussed in section 2.4.1). We can see that as shown before, intertemporal wedge is positive and small relative to the case where $\theta$ is i.i.d. This analysis suggests that one should analyze the model where $\theta$ is stochastically evolving over time.
2.7 Conclusion

In this paper, I have studied optimal taxation of entrepreneurial income. I have shown that allowing households to invest in businesses, thereby being subject to idiosyncratic investment risk, changes the standard results on taxation of wealth and personal income. Although the model can be interpreted as one of optimal taxation, I have shown that standard securities commonly used in venture capital contracts can implement efficient allocations.

Although, I have interpreted the agents in the model as entrepreneurs subject to capital income risk, the model can be used a variety of issues. In particular, it can be interpreted as a model with risky human capital and private information. Hence, its implications can be used to draw policy implication for labor income. Moreover, I
have assumed away fixed costs associated with investment. In presence of fixed costs of investment, the model can be interpreted as a model of innovation and it can be used for optimal patent policy evaluations. Hence, the techniques developed in this paper are usefully in analyzing a wide variety of questions.
Chapter 3

Risk Sharing, Inequality, and Fertility

3.1 Introduction

A key question in normative public finance is the extent to which it is socially efficient to insure agents against shocks to their circumstances. The basic trade-off is one between providing incentives for productive agents to work hard – thereby making the pie big – versus transferring more resources to less productive agents to insure them. This is the problem first analyzed in [Mirrlees, 1971] where he characterized the solution to a problem of this form in a static setting. This analysis is at the heart of a deep and important question in economics – What is the optimal amount of inequality in society? Mirrlees provides one answer to this question along with a way of implementing his answer to this question – a tax and transfer scheme based on non-linear income taxes that makes up the basis of the optimal social safety net.

More recently, a series of authors (e.g., [Green, 1987], [Thomas and Worrall, 1990] and [Atkeson and Lucas, 1992]) have extended this analysis to cover dynamic settings – agents are more productive some times and less others. A common result from this literature is that the socially efficient level of insurance (ex ante and under commitment) involves an asymmetry between how good and bad shocks are treated. In particular, when an agent is hit with a bad shock, the decrease in what he can expect in the future is more than the corresponding increase after a good shock – there is a negative drift.
in expected future utility. This feature of the optimal contract has become known as ‘immiseration.’ Immiseration, although interesting on its own, is more important as an indicator of a more severe problem in the models. This that there is not a stationary distribution over continuation utilities – the optimal amount of inequality in society grows without bound over time.$^1$ This weakness means the models cannot be used to answer questions such as: Is there too much inequality in society under the current system?

Two recent papers, [Phelan, 2006] and [Farhi and Werning, 2007] have given a different view of the immiseration result. This is to interpret different periods in the model as different generations. In this case, current period agents care about the future because of dynastic altruism a la [Barro, 1974]. Under this interpretation, social insurance is comprised of two conceptually different components: 1) Insurance against the uncertainty coming from current generation productivity shocks, 2) Insurance against the uncertainty coming from the shock of what family you are born into – what future utility was promised to your parents (e.g., through the intergenerational transmission of wealth). Under this interpretation, immiseration means that optimal insurance for the current generation features a lower utility level for the next generation.

These authors go on to show that optimal contract features a stationary level of inequality if society values the welfare of children strictly more than their parents do. When this is true, the amount of this long run inequality depends on the difference between societal and parental altruism.

In an intergenerational setting with dynastic altruism such as that studied by these authors, a natural question arises: To what extent are these results altered when the size of generations – i.e., fertility – is itself endogenous (e.g., as in [Barro and Becker, 1989] and [Becker and Barro, 1988])? This question is the focus of this paper.

We show that the explicit inclusion of fertility choice in the model alters the qualitative character of the optimal allocations in two important ways. First, we show that even when the planner does not put extra weight on future generations, there is a stationary distribution in per capita variables – there is an optimal amount of long run per capita

---

$^1$ Immiseration and the lack of a meaningful stationary distribution are not equivalent in general. There are other examples in the literature (e.g., see [Khan and Ravikumar, 2001] and [Williams, 2009]) that show that immiseration need not hold. However, in those examples there is no stationary distribution over continuation utilities – variance grows without bound.
inequality and no immiseration in per capita terms. In addition to this, since fertility is explicitly included, the model has implications about the properties of fertility. Because of this, the model also has implications about the best way to design policies relating to family size (e.g., child care deductions, tax credits for children, education subsidies, etc.)

From a mechanical point of view, the inclusion of fertility gives the planner an extra instrument to use to induce current agents to truthfully reveal their productivities. That is, the planner can use both family size and continuation utility of future generations to induce truthful revelation. In the normal case (i.e., without fertility choice), in order to induce truth telling today, the planner (optimally) chooses to ‘spread’ out continuation utilities so as to be able to offer insurance in current consumption. The incentives for the planner to do this are present after every possible history. Because it is cheaper to provide incentives in the future when continuation utilities are lower this outward pressure is asymmetric and has a negative bias. Thus, continuation utilities are pushed to their lower bound – inequality becomes greater and greater over time and immiseration occurs.

In contrast, when fertility is endogenous, this optimal degree of spreading in continuation utility for the parent can be thought of as being provided through two distinct sources – spreading of per child continuation utility and spreading in family sizes. In general both of these instruments are used to provide incentives, but the way that they are used is different. While dynasty size can grow or shrink without bound for different realizations, we find that there is a natural limit to the amount of spreading that occurs through per child continuation utilities. Specifically, per child continuation utilities lie in a bounded set. Thus, even if the promised utility to a parent is very low, the continuation utility for their children is bounded below. Similarly, even if the promised utility to a parent is very high, the continuation utility for their children is bounded above. These results form the basis for showing that a stationary distribution exists.

This property, boundedness of per child continuation utilities, is shown by exploiting an interesting kind of history independence in the model. This is what we call the ‘resetting’ property. There are two versions of this that are important for our results. The first concerns the behavior of continuation utility for children when promised utility for a parent is very low. After some point, reducing promised utility to the parent
no longer reduces per child continuation utility – continuation utility for children is bounded below. An implication of this is that if promised utility to the parent is low enough, continuation utility for his children will automatically be higher, no matter what productivity shock is realized. The second version says something similar for high promised utilities for parents. A particularly striking version of this concerns the behavior of the model whenever a family experiences the highest value of the shock: There is a value of continuation utility, \( w_0 \), such that all children are assigned to this level no matter what promised utility is. That is, continuation utility gets ‘reset’ to \( w_0 \) subsequent to every realization of a high shock. So, even if a family has a very long series of good shocks, the utility of the children does not continue to grow but stays fixed at \( w_0 \).

This reasoning concerning the limits of long run inequality in per capita variables differs from what is found in [Farhi and Werning, 2007] in two ways. The first of these is the basic reason for the breakdown in the immiseration result. In [Farhi and Werning, 2007], it is because of a difference between social and private discounting – society puts more weight on future generations than parents do. Here, immiseration breaks down even when social and private discounting are the same because of resetting at the top and bottom of the promised utility distribution. The second difference concerns the movements of per capita consumption over time. The version of the Inverse Euler Equation that holds in the [Farhi and Werning, 2007] world when society is more patient than individuals implies that consumption has a mean reversion property. In our model, there is a lower bound on continuation utilities for children which is independent of both the promised utility to parents and the current productivity shock. Moreover, as discussed above, the resetting property at the top implies that continuation utility reverts to \( w_0 \) each time a high shock is realized. Thus, the type of intergenerational mobility that is present in the two models are quite different.

In contrast to these results about the limits to spreading through per child continuation utility, we find that there is no upper bound on the amount of spreading that occurs through the choice of family size. If the discount factor is equal to the inverse of the interest rate, we show that along any subset of the family tree, population dies out.\(^2\)

However, this does not necessarily imply that population shrinks, since this property

\(^2\) This does not hold if the discount factor is not the inverse of the interest rate, see the discussion
holds even when mean population is growing. Rather, some strands of the dynasty tree
die out and others expand. Those that are growing are exactly those sub-populations
that have the best ‘luck.’

From a mechanical point of view, the key technical difference between our results
and that from earlier work is that here, bounds on continuation utility arise naturally
from the form of the contracting problem rather than from being exogenously imposed.
Contracting problems in which social and private discount factors are different (such
as those studied in [Phelan, 2006] and [Farhi and Werning, 2007]) can equivalently be
thought of as problems where the social and private discount factors are the same,
but there are lower bounds on the continuation utility levels of future generations. As
such, they are closely linked to the approach followed in [Atkeson and Lucas, 1995] and
[Sleet and Yeltekin, 2006]. Here, the endogenous bounds arise because of the inclusion
and optimal exploitation by the planner of a new choice variable, family size.

There are several other interesting differences between the two approaches. For
example, one of the key ways that [Phelan, 2006] and [Farhi and Werning, 2007], differ
from earlier work is that in the socially efficient scheme, the Inverse Euler Equation need
not hold. Indeed, in those papers, the inter-temporal wedge can even be negative.\(^3\)

This can be interpreted (in some implementations) as requiring a negative estate
tax. This has been interpreted as meaning that, in order to overcome immiseration, a
negative inter-temporal wedge may be necessary. This does not hold for us however,
since a version of the Inverse Euler Equation holds. This implies that there will always
be a positive ‘wedge’ in the FOC determining savings and hence, estate taxes are always
implicitly positive.

Finally, an interesting new feature that emerges is the dependence of taxes on family
size. What we find is that for everyone other than the highest type, there is a positive
tax wedge on the fertility-consumption margin – fertility is discouraged to better provide
incentives for truthful revelation.

Our paper is related to the literature on dynamic contracting including [Green, 1987],
[Thomas and Worrall, 1990], [Atkeson and Lucas, 1992] (and many others). These pa-
pers established the basic way of characterizing the optimal allocation in endowment

\(^3\) [Farhi and Werning, 2010c] in the same environment show that estate taxes have to be progressive.
In economies where there is private information. They also show that, in the long run, inequality increases without bound, i.e. the immiseration result. [Phelan, 1998] shows that this result is robust to many variations in the assumptions of the model. Moreover, [Khan and Ravikumar, 2001] establish numerically that in a production economy, the same result holds and although the economy grows, the detrended distribution of consumption has a negative trend. We contribute to this literature by extending the model to allow for endogenous choice of fertility. We employ the methods developed in the aforementioned papers to analyze this problem.

Finally, our paper has some novel implications about fertility per se. First, the socially efficient allocation is characterized by a negative income-fertility relationship—indpendently on specific assumptions on curvature in utility (see [Jones et al., 2008] for a recent summary). This suggests that intergenerational income risk and intragenerational risk sharing may be important factors to explain the observed negative income-fertility relationship. Second, very few papers have analyzed ability heterogeneity and intergenerational transmission of wealth in dynastic models with fertility choice. Our paper is most related to [Alvarez, 1999] who analyzes intergenerational income risk but assumes that it is uninsurable.

In section 3.2 we present a two period, two shock version of the to show the basic results in a simple setting. Section 3.3 contains the description of the general model with private information. In section 3.4 we study the properties of the model relating to long run inequality. In section 3.5, we discuss some extension and complimentary results.

### 3.2 An Example and Intuition

In this section we illustrate the key idea for our results in two steps. First, we study our basic incentive problem in a two period model and derive a property, which we call ‘resetting.’ This property shows that there is a (high) level of utility that is assigned to all children of workers that have high productivity independent of the level of promised utility to the worker. This provides a strong intuition for why adding fertility to the model has such a large impact on the asymptotic behavior of continuation utility – there is a continual recycling of utility levels up to this ‘resetting value’ each time a high
value of the shock is realized. This by itself does not imply that there is a stationary
distribution over continuation utilities, but it is one important step in the argument.

Second, we show that the reason that this occurs in the [Barro and Becker, 1989]
model of dynastic altruism is because of a homotheticity property of utility in this
model. 4

In sum, the key feature that fertility brings to the contracting environment is a
distinction between total and per capita continuation utility. The implications of this
difference are particularly sharp in the Barro and Becker utility case, but they are not
limited to it.

3.2.1 A Two Period Example

Immiseration concerns the limiting behavior of continuation utility as a history of shocks
is realized. In a stationary environment, this is determined by the properties of the policy
function describing the relationship between future utility, $W'$, as a function of current
promised utility, $W_0$ (and the current shock). Immiseration is then the statement that
the only stationary point of this mapping is for continuation utility to converge to its
lower bound.

To gain some intuition, we will mimic this in a two period example by reinterpreting
$W'$ as second period utility and $W_0$ as ex ante utility. Then, a necessary condition for
immiseration is: when $W_0$ is low, $W'$ is even lower. This is the version that we will
explore in this section.

To this end, consider a two-period economy populated by a continuum of parents
with mass 1 who live for one period. Each parent receives a random productivity $\theta$ in
the set $\Theta = \{\theta_L, \theta_H\}$ with $\theta_H > \theta_L$. At date 1, each parent’s productivity, $\theta$, is realized.
After this, they consume, work and decide about the number of children. The cost of
having a child is in terms of leisure. Every child requires $b$ units of leisure to raise. The
coefficient $b$ can be thought of as time spent raising children (or the market value of
maternity leave for women). The child lives for one period and consumes out of the

---

4 In the Supplementary Appendix we show that even for more general utility functions, a similar
result will hold with fertility in the model as long as a certain combination of elasticities is bounded
away from infinity.
savings done by their parents. The parent’s utility function is:

\[ u(c_1) + h(1 - l - bn) + \beta n^n u(c_2) \]

in which \( l \) is hours worked, \( n \) is the number of children and \( c_i \) is consumption per person in period \( t \). From this, it can be seen that the parent has an altruistic utility function where the degree of altruism is determined by \( \beta \).

We assume that a worker of productivity \( \theta \in \Theta \) who works for \( l \) hours has effective labor supply of \( \theta l \) and that both \( \theta \) and \( l \) are private information of the parent. As is typically assumed, we assume that the planner can observe \( \theta l \). In what follows we denote the aggregate consumption of all children by \( C_2 = n_2c_2 \).

Suppose each parent is promised an ex ante utility \( W_0 \) at date zero and that the planner has access to a saving technology at rate \( R \). Thus, the planner wishes to allocate resources efficiently subject to the constraints that - 1) Ex ante utility to each parent is at least \( W_0 \), 2) Each ‘type’ is willing to reveal itself - IC.

\[
\begin{align*}
\min_{c_1(\theta), n(\theta), l(\theta), c_2(\theta)} & \sum_{\theta_L, \theta_H} \pi(\theta) \left( c_1(\theta) + \frac{1}{R} n(\theta)c_2(\theta) \right) - \sum_{\theta_L, \theta_H} \pi(\theta)\theta l(\theta) \quad (3.1) \\
\text{s.t.} & \quad \sum_{\theta_L, \theta_H} \left[ \pi(\theta) \left( u(c_1(\theta)) + h(1 - l(\theta) - bn(\theta)) + \beta n(\theta)^n u(c_2(\theta)) \right) \right] \geq W_0 \quad (3.2) \\
& \quad u(c_1(\theta_H)) + h(1 - l(\theta_H) - bn(\theta_H)) + \beta n(\theta_H)^n u(c_2(\theta_H)) \geq u(c_1(\theta_L)) + h\left( 1 - \frac{\theta L}{\theta H} l(\theta_L) - bn(\theta_L) \right) + \beta n(\theta_L)^n u(c_2(\theta_L)) \quad (3.3)
\end{align*}
\]

After manipulating the first order conditions for the allocation of the highest type, we obtain:

\[ \eta u(c_2(W_0, \theta_H)) = u'(c_2(W_0, \theta_H))c_2(W_0, \theta_H) + b\theta H Ru'(c_2(W_0, \theta_H)). \quad (3.4) \]

This is an equation in per child consumption for children of parents with the highest shock. Notice that none of the other endogenous variables of the system appear in this

---

5 Both here, and throughout the remainder of the paper, we will assume that only downward incentive constraints bind. Under certain conditions it can be shown focusing on the downward incentive constraints is sufficient. (See [Hosseini et al., 2009].)
equation. Similarly, $W_0$ does not appear in the equation. Because of this, it follows that the level of consumption for these children, $c_2(W_0, \theta_H)$, is independent of $W_0$.

We will call this the ‘resetting’ property – i.e., per capita consumption for the children of parents with the highest shock is reset to a level that is independent of state variables.

There are two key features in the model that are important in deriving this result. The first of these comes from our assumption that in this problem we have assumed that no one pretends to be the highest type $\theta_H$. Because of this, it follows from the usual argument that the allocations of this type are undistorted.

The second important feature comes from the fact that the problem is ‘homogeneous/homothetic’ in aggregate second period consumption, $C_2 = nc_2$, and family size, $n$. Because of this, $c_2/W_0 = c_2$ is independent of $W_0$ in undistorted allocations.\footnote{As it turns out, in the full information analog of this problem, this line of reasoning holds for all types, not just the highest one. I.e., with full information, the consumption of a child of a parent of type $\theta$ is given by $c_2(W_0, \theta_H)$ which is independent of $W_0$.}

In sum, $\lim_{W_0 \to -\infty} C_2(W_0, \theta_H) = 0$, $\lim_{W_0 \to -\infty} n(W_0, \theta_H) = 0$, but $C_2(W_0, \theta_H)/n(W_0, \theta_H)$ is constant.

The next step is to use this result to say something about immiseration. There are two ways to interpret continuation utility in our setting: continuation utility from the point of view of the parent, $\beta n^\eta u(c_2)$, and, continuation utility from the point of each child, $u(c_2)$. When fertility is exogenous these two alternative notions are equivalent.

From our discussion above, the ‘resetting’ property implies that $u(c_2(W_0, \theta_H)) = u(c_2(\theta_H))$ is bounded away from the lower bound of utility and hence there is no immiseration in this sense. However, since $u(c_2(W_0, \theta_H))$ is bounded below and $n(W_0, \theta_H) \to 0$, it follows that $\beta n(W_0, \theta_H)^\eta u(c_2(W_0, \theta_H))$ converges to its lower bound.\footnote{There are two cases of relevance here. The first is $u \geq 0$ and $0 \leq \eta \leq 1$. In this case, $\beta n(W_0, \theta_H)^\eta u(c_2(W_0, \theta_H)) \to 0$. The second case is when $u \leq 0$ and $\eta < 0$. In this case $\beta n(W_0, \theta_H)^\eta u(c_2(W_0, \theta_H)) \to -\infty$.}

We summarize this discussion as a Proposition:

**Proposition 3.1**

1. $c_2(W_0, \theta_H) = C_2(W_0, \theta_H)$ is independent of $W_0$;

2. $u(c_2(W_0, \theta_H))$ is bounded below;

\footnote{We can show that at the full information efficient allocations the downward constraints are binding and upward constraints are slack. We also verify the slackness of upward constraints in the numerical example (in infinite horizon environment). In general we cannot prove that only downward constraints are binding because the preferences do not exhibit single-crossing property.}
3. \( \beta n(W_0, \theta_H)^\eta u(c_2(W_0, \theta_H)) \) converges to its lower bound.

It turns out that similar results also hold in a larger class of environments including settings with a goods cost for children, and/or with taste shock rather than productivity shock (see [Hosseini et al., 2009]).

Thus, there is a sense in which there is no immiseration – from the point of view of the children – and a sense in which there is immiseration – from the point of view of the parents. As can be seen from this discussion, the key feature, when fertility is included as a choice variable, is the difference between aggregate and per capita variables. While it is hard to think about infinite horizon and stationarity in a two period example, (2) and (3) provide partial intuition. For example, (2) implies that, per capita utility of children does not have a downward trend (as a function of \( W_0 \)) – there is no immiseration from the point of view of the children. Interpreting (3) is even more difficult, but it, along with a statement that \( \beta n(W_0, \theta_H)^\eta u(c_2(W_0, \theta_H)) \) is below \( W_0 \) when \( W_0 \) is low enough implies that a form of immiseration does hold from the point of view of the parents.

This discussion is complicated by two additional factors when considering the infinite horizon model. The first of these is that to show that a stationary distribution exists (in per capita variables), it is not enough to show that there is no immiseration for the highest type. It must also be shown that utility is bounded below for other shocks too. This is a key step of the main result in the paper discussed below.

Second, showing that utility is bounded below for the highest type is also not sufficient for another reason. This is that the proportions of the population that are children of the highest type is itself endogenous. I.e., the result in the Proposition would not have much bite if, for example, \( n(W_0, \theta_H) = 0 \) for all \( W_0 \). This is also discussed below.

### 3.2.2 Resetting – Intuition via Homotheticity

Some intuition for the ‘resetting’ property can be obtained by reformulating the planner’s problem.

As a first step, rewrite the problem above by letting \( m = 1 - l - bm \) be parents leisure:

\[
\min_{c_1(\theta), n(\theta), m(\theta), C_2(\theta)} \sum_{\theta \in \Theta} \pi(\theta) \left( c_1(\theta) + \theta m(\theta) \right) + \sum_{\theta \in \Theta} \pi(\theta) \left( b \theta n(\theta) + \frac{1}{R} C_2(\theta) \right)
\]

(3.5)
s.t. \[
\sum_{\theta \in \Theta} \pi(\theta) (u(c_1(\theta)) + h(m(\theta))) + \sum_{\theta \in \Theta} \pi(\theta) \beta n(\theta)^{\eta} u \left( \frac{C_2(\theta)}{n(\theta)} \right) \geq W_0 \tag{3.6}
\]

\[
u(c_1(\theta_H)) + h(m(\theta_H)) + \beta n(\theta_H)^{\eta} u \left( \frac{C_2(\theta_H)}{n(\theta_H)} \right) \geq u(c_1(\theta_L)) + h \left( \frac{\theta_L}{\theta_H} m(\theta_L) + (1 - \frac{\theta_L}{\theta_H})(1 - b n(\theta_L)) \right) + \beta n(\theta_L)^{\eta} u \left( \frac{C_2(\theta_L)}{n(\theta_L)} \right) \tag{3.7}
\]

The first term in the objective function is the planner’s expenditure on parents’ consumption and leisure (denominated in parents’ consumption). The second term is the total expenditure on children: their total consumption and time spent parenting (again, denominated in parents’ consumption).

For an allocation to be the solution to this problem, there should not be a way to adjust \(n(\theta_H)\) and \(C_2(\theta_H)\), while holding the other variables fixed, which lowers cost while still satisfying the constraints.

As can be seen from this problem, there are no interactions between the variables \(n(\theta_H)\) and \(C_2(\theta_H)\) and the other variables. That is, they enter additively in the objective function and together, but separate from all other variables in the constraints (i.e., only through the terms based on \(\beta n(\theta_H)^{\eta} u \left( \frac{C_2(\theta_H)}{n(\theta_H)} \right)\)). Because of this, it follows that, given the other variables in the problem, the optimal choice of \((n(\theta_H), C_2(\theta_H))\) must solve the sub-problem:

\[
\min_{C_2, n} b \theta_H n + \frac{1}{R} C_2 \tag{3.8}
\]

s.t. \(n^{\eta} u \left( \frac{C_2}{n} \right) = W(W_0, \theta_H)\)

The resetting property for high productivity parents can be understood by studying this problem. As can be seen, the objective function in this problem is homogeneous of degree one, while the constraint set is homogeneous of degree \(\eta\). This is analogous to an expenditure minimization problem with a homothetic utility function. One property that problems of this form have is linear income expansion paths. In this case, this means that the ratio \(\frac{C_2(W(W_0, \theta_H), \theta_H)}{n(W(W_0, \theta_H), \theta_H)}\) does not depend on \(W(W_0, \theta_H)\) (and therefore, does not depend on \(W_0\)). Instead, it only depends on technology and preference parameters.
Drawing an analogy with consumer demand theory, relative demand, \( \frac{C_2}{n} \), only depends on relative prices, \( bR\theta \), and not on promised utility. Therefore, the resetting property that we find is an immediate consequence of the homotheticity of the utility function in the Barro and Becker formulation of dynastic altruism.

This same argument does not hold for the low type \( \theta = \theta_L \) because \( n(\theta_L) \) and \( C_2(\theta_L) \) do not separate from the other variables in the maximization problem. Mathematically, this is because of the fact that \( n(\theta_L) \) also enters the leisure term of a high type who pretends to be a low type in the incentive constraint. This effect is absent in the full information version of this problem and hence, in that case, there is ‘resetting’ for all types – with full info, there are values of per capita consumption, \( c_2(\theta) \) such that \( c_2(W_0, \theta) = c_2(\theta) \) for all \( W_0 \).

As the above discussion shows, the resetting property relies on the way parents’ utility from children. Following [Barro and Becker, 1989] and [Becker and Barro, 1988] we have made two assumptions. One, the way parents care about utility of children is multiplicatively separable in the number of children and the consumption of each child. Second, the component of the utility that depends on the number of children is homothetic (\( n^\eta \)). The discussion above indicates that this homotheticity is important in getting the resetting property. If this function is not homothetic, the per capita consumption of each child whose parent receives a high shock may depend on the promised utility of the parent \( (W_0) \). However, as it turns out, it can be shown that under a very general condition per capita consumption, and hence, continuation utility, of each child remains bounded away from zero. In the Supplementary Appendix, we give a general set of conditions under which there is no ‘immiseration’.

### 3.3 The Infinite Horizon Model

In this section, we will extend the model in section 3.2 to an infinite horizon setting. The set of possible types is given by \( \Theta = \{\theta_1 < \cdots < \theta_I\} \). As above, the planner can observe the output for each agent but not hours worked nor productivity. Using the revelation principle, we will only focus on direct mechanisms in which each agent is asked to reveal his true type in each period. As is typical in problems like these, it can be shown that the full information optimal allocation does not satisfy incentive
compatibility. Although the argument is more complex than in the usual case, we show (see [Hosseini et al., 2009]) that under the full information allocation, a higher productivity type would want to pretend to be a lower productivity type.

In addition to this, in Mirrleesian environments with private information where a single crossing property holds, one can show only downward incentive constraints bind. We do not currently have a proof that the only incentive constraints that ever bind are the downward ones. In keeping with what others have done (e.g., [Phelan, 1998] and [Golosov and Tsyvinski, 2007]), we assume that agents can only report a level of productivity that is less than or equal to their true type. In the Supplementary Appendix, we give a sufficient condition for this to be true.

Under this assumption, we can restrict reporting strategies, \( \sigma \), to satisfy

\[
\sigma_t(\theta_t) \leq \theta_t.
\]

(Here, for every history \( \theta_t \), \( \sigma_t(\theta_t) \) is agent’s report of its productivity in period \( t \) and \( \sigma^{t}(\theta^{t}) \) is the history of the reports.) Moreover, because of our assumed restriction on reports, we have \( \sigma_t(\theta_t) \leq \theta_t \). Call the set of restricted reports \( \Sigma \). Then, an allocation is said to be incentive compatible if

\[
\sum_{t, \theta_t} \beta^t \pi_t(\theta^t) N_t(\theta^{t-1}) \eta \left[ u \left( \frac{C_t(\theta^t)}{N_t(\theta^{t-1})} \right) + h \left( 1 - \frac{L_t(\theta^t)}{N_t(\theta^{t-1})} - b \frac{N_{t+1}(\theta^t)}{N_t(\theta^{t-1})} \right) \right] \geq (3.9)
\]

\[
\sum_{t, \theta^t} \beta^t \pi_t(\theta^t) N_t(\sigma^{t-1}(\theta^{t-1})) \eta \left[ u \left( \frac{C_t(\sigma^{t}(\theta^t))}{N_t(\sigma^{t-1}(\theta^{t-1}))} \right) + h \left( 1 - \frac{\sigma_t(\theta^t)L_t(\sigma^{t}(\theta^t))}{\theta_t N_t(\sigma^{t-1}(\theta^{t-1}))} - b \frac{N_{t+1}(\sigma^{t}(\theta^t))}{N_t(\sigma^{t-1}(\theta^{t-1}))} \right) \right] \forall \sigma \in \Sigma
\]

Hence, the planning problem becomes the following:

\[
\sum_{t, \theta^t} \beta^t \pi_t(\theta^t) N_t(\theta^{t-1}) \eta \left[ u \left( \frac{C_t(\theta^t)}{N_t(\theta^{t-1})} \right) + h \left( 1 - \frac{L_t(\theta^t)}{N_t(\theta^{t-1})} - b \frac{N_{t+1}(\theta^t)}{N_t(\theta^{t-1})} \right) \right]
\]

s.t.

\[
\sum_{t, \theta^t} \frac{1}{R_t} \pi_t(\theta^t) \left[ C_t(\theta^t) - \theta_t L_t(\theta^t) \right] \leq K_0
\]

\footnote{In numerically calculated examples, this assumption is redundant.}
\[ \sum_{t, \theta^t} \beta^t \pi(\theta^t)^N, \pi(\theta^t)^{\theta^t-1}^{\eta} \left[ u \left( \frac{C_t(\theta^t)}{N_t(\theta^t-1)} \right) + h \left( 1 - \frac{L_t(\theta^t)}{N_t(\theta^t-1)} - b\left. N_{t+1}(\theta^t) \right/ N_t(\theta^t-1) \right) \right] \geq \]
\[ \sum_{t, \theta^t} \beta^t \pi(\theta^t)^N, \pi(\theta^t)^{\sigma^t-1}(\theta^t)^{\eta} \left[ u \left( \frac{C_t(\sigma^t(\theta^t))}{N_t(\sigma^t-1)(\theta^t-1))} \right) + h \left( 1 - \frac{\sigma_t(\theta^t)L_t(\sigma^t(\theta^t))}{\theta_t N_t(\sigma^t-1)(\theta^t-1))} - b\left. N_{t+1}(\sigma^t(\theta^t)) \right/ N_t(\sigma^t-1)(\theta^t-1)) \right] \forall \sigma \in \Sigma \]

Using standard arguments, we can show that the above problem is equivalent to the following functional equation:

\[ V(N, W) = \min_{C(\theta), L(\theta), N'(\theta)} \sum_{\theta} \pi(\theta) \left[ C(\theta) - \theta L(\theta) + \frac{1}{R} V(N'(\theta), W'(\theta)) \right] \quad (P1) \]

s.t.
\[ \sum_{\theta} \pi(\theta) \left[ \sum_{t} \left( \frac{C(\theta)}{N} + h \left( 1 - \frac{L(\theta)}{N} - b\left. N'(\theta) \right/ N \right) \right) + \beta W'(\theta) \right] \geq W \]
\[ \sum_{\theta} \pi(\theta) \left[ \sum_{t} \left( \frac{C(\theta)}{N} + h \left( 1 - \frac{\theta L(\theta)}{\theta N} - b\left. N'(\theta) \right/ N \right) \right) + \beta W'(\theta) \right] \geq 0 \quad (3.11) \]

As we can see, the problem is homogeneous in \( N \) and therefore as before, if we define \( v(N, w) = \frac{V(N, N^n w)}{N} \), \( v(\cdot, \cdot) \) will not depend on \( N \) and satisfies the following functional equation:

\[ v(w) = \min_{c(\theta), l(\theta), n(\theta), w'(\theta)} \sum_{\theta} \pi(\theta) \left( c(\theta) - \theta l(\theta) + \frac{1}{R} n(\theta) v(w'(\theta)) \right) \quad (P1') \]

s.t.
\[ \sum_{\theta} \pi(\theta) \left[ u(c(\theta)) + h(1 - l(\theta) - b n(\theta)) + \beta n(\theta)^{\eta} w'(\theta) \right] \geq w \]
\[ u(c(\theta)) + h(1 - l(\theta) - b n(\theta)) + \beta n(\theta)^{\eta} w'(\theta) \geq u(c(\theta)) + h(1 - \frac{\hat{\theta} l(\theta)}{\hat{\theta}} - b n(\hat{\theta})) + \beta n(\hat{\theta})^{\eta} w'(\theta), \forall \theta > \hat{\theta} \]

In what follows, we will assume that the solution to the minimization problem, (P1), has several convenient mathematical properties. These include strict convexity and
differentiability of the value function as well as the uniqueness of the policy functions. Normally, these properties can be derived from primitives by showing that $V(N,W)$ is strictly convex, that the constraint set is convex, etc. Because of the presence of the incentive compatibility constraints, the usual lines of argument will not work (due to the non-convexity of the constraint set). In some contracting problems, these issues can be partially resolved. For example, in some cases, a change of variables can be designed so that convexity of the constraint set is guaranteed. Here, because of the way that fertility and labor supply enter the problem, this will no longer work. An alternative way to resolve this issue is by allowing for randomization. Allowing for randomization, makes all the constraints linear in the probability distributions and therefore the constraint correspondence is convex. This is the method used in [Phelan and Townsend, 1991] and [Doepke and Townsend, 2006] (see also [Acemoglu et al., 2008]). This is not quite enough for us since it only implies convexity of $V$, not strict convexity, and hence, uniqueness of the policy function cannot be guaranteed. Because of this, we simply assume that $V$ has the needed convexity properties. Similar considerations hold for the differentiability of $V$. The following lemma on $V$ provides useful later in the paper:

**Lemma 3.2** If $V(N,W)$ is continuously differentiable and strictly convex, then $v(w)$ is continuously differentiable and strictly convex. Moreover, $\eta v'(w)w - v(w)$ is strictly increasing.

See Supplementary Appendix for the proof.

In addition, for the purposes of characterizing the solution, we will want to use the FOC’s from this planning problem in some cases. This requires that the solution is interior. The usual approach to guarantee interiority is to use Inada conditions. We use a version of these here to guarantee that $c, 1 - l - bn$, and $n$ are interior. The version that we use is stronger than usual and necessary because of the inclusion of private information and fertility.

**Assumption 3.3** Assume that both $u$ and $h$ are bounded above by 0, and unbounded below. Note that this implies that $\eta < 0$ must hold for concavity of overall utility and

\[10\] In our numerical examples the value function is convex even without the use of lotteries. In the [Hosseini et al., 2009], we study a special case where we can show that the constraint correspondence is convex.
hence, an Inada condition on \( n \) is automatically satisfied. Finally, we assume that \( h(1) < 0 \).\(^{11}\)

Under this assumption, it follows that consumption, leisure and fertility are all strictly positive. This is not enough to guarantee that the solution is interior however, since hours worked might be zero. Indeed, there is no way to guarantee that \( l > 0 \) in this model. This is because of the way hours spent raising children enter the problem. Because of this feature of the model, it might be true that the marginal value of leisure exceeds the marginal product of an hour of work even when \( l = 0 \). The usual way of handling this problem by assuming that \( h'(1) = 0 \) will not work in this case since we know that \( n > 0 \). Hence, the marginal value of leisure at zero work will always be positive, even if \( h'(1) = 0 \). Because of this, when continuation utility is sufficiently high, it is always optimal for work to be zero.

In addition to this, in some cases, there are types that never work. This will be true when it is more efficient for a type to produce goods through the indirect method of having children and having their children work in the future than through the direct method of working themselves. This will hold for a worker with productivity \( \theta \), if \( \theta < \frac{E(\theta)}{bR} \). That is, \( l(w, \theta) = 0 \) for all \( w \) if \( \theta < \frac{E(\theta)}{bR} \). For this reason, we will rule this situation out by making the following assumption:

**Assumption 3.4** Assume that, for all \( i \), \( \theta_i > \frac{E(\theta)}{bR} \).

This assumption does not guarantee that \( l(w, \theta) > 0 \) for all \( w \), but it can be shown that when continuation utility is low enough, \( l > 0 \). As we will show below, this is sufficient to guarantee that a stationary distribution exists.

In what follows, we will simply assume that \( l > 0 \) for most of the paper. We will return to this issue below when we show that a stationary distribution exists.

### 3.4 Properties of the Model

In this section, we lay out the basic properties of the model. These are:

1. A version of the resetting property for the infinite horizon version of the model,

\(^{11}\) This would hold, for example, if \( h(\ell) = \frac{\ell^{1-\sigma}}{1-\sigma} \) with \( \sigma > 1 \).
2. A result stating that there is a stationary distribution over per capita variables, and

3. A version of the Inverse Euler Equation adapted to include endogenous fertility.

Taken together these imply that, when $\beta R = 1$, there is no immiseration in per capita terms but, there is immiseration in dynasty size. When $\beta R > 1$, this need not hold.

3.4.1 The Resetting Property

We have shown in the context of a two period model, children’s consumption is independent of parent’s promised utility when $\theta = \theta_H$. Here we will show that a similar property holds in the infinite horizon version of the model. This can be derived from the first order conditions of the recursive formulation. Taking first order conditions with respect to $n(\theta_I), l(\theta_I)$ and $w'(\theta_I)$ respectively, gives us the following equations:

$$\pi(\theta_I) \frac{1}{R} v'(w'(\theta_I)) = b \left( \lambda \pi(\theta_I) + \sum_{\hat{\theta} < \theta_I} \mu(\theta_I, \hat{\theta}) \right) \left[ h'(1 - l(\theta_I) - bn(\theta_I)) + \beta \eta (n(\theta_I))^{\eta - 1} w'(\theta_I) \right]$$

$$\pi(\theta_I) n(\theta_I) v'(w'(\theta_I)) = - \left( \lambda \pi(\theta_I) + \sum_{\hat{\theta} < \theta_I} \mu(\theta_I, \hat{\theta}) \right) \beta (n(\theta_I))^{\eta - 1}$$

Combining these gives

$$v(w'(\theta_I)) - \eta w'(\theta_I) v'(w'(\theta_I)) = -b R \theta_I.$$  \hspace{1cm} (3.12)

We can see that $w'(w, \theta_I)$ is independent of promised continuation utility. That is, $w'(w, \theta_I) = w'(\hat{w}, \theta_I)$ for all $w, \hat{w}$. Denote by $w_0$ this level of promised continuation utility – $w_0 = w'(w, \theta_I)$. 

The resetting property means that once a parent receives a high productivity shock, the per capita allocation for her descendants is independent of the parents level of wealth – an extreme version of social mobility holds.

Because of this, it follows that there is no immiseration in this model, under very mild assumptions, in the sense that per capita utility does not converge to its lower bound. To see this, first consider the situation if \( n(w, \theta) \) is independent of \((w, \theta)\). In this case, from any initial position, the fraction of the population that will be assigned to \( w_0 \) next period is at least \( \pi(\theta_I) \). This by itself implies that there is not a.s. convergence to the lower bound of continuation utilities. When \( n(w, \theta) \) is not constant, the argument involves more steps. Assume that \( n \) is bounded above and below – \( 0 < a \leq n(w, \theta) \leq a' \). Then, the fraction of descendants being assigned to \( w_0 \) next period is at least \( \pi(\theta_I)a \frac{\pi(\theta_I)a}{(1-\pi(\theta_I))a} \). Again then, we see that there will not be a.s. immiseration. We summarize this discussion in a Proposition.

**Proposition 3.5** Assume that \( v \) is continuously differentiable and that there is a unique solution to 3.12. Then, continuation utility has a ‘resetting’ property, \( w'(w, \theta_I) = w_0 \) for all \( w \).

Intuitively, the reason that the resetting property holds here mirrors the argument given above in the two period case. That is, since no ‘type’ wants to pretend to have \( \theta = \theta_I \), the allocation for this type is marginally undistorted. Again, due to the homogeneity properties of the problem, per capita variables (i.e., continuation utility) are independent of promised utility.

Next, we argue that a similar property holds when continuation utility is low enough for any type. That is, even as promised utility, \( w \), gets lower and lower, continuation utility, \( w'(w, \theta) \), is bounded away from \(-\infty\).

To this end, we show that as \( w \to -\infty \), the optimal allocation converges to \( c = 0, l = 1, n = 0 \). The interesting thing about this allocation is that no incentive constraints are binding and hence, the optimal allocation has properties similar to those in the full information case. Formally:

**Proposition 3.6** Suppose that \( V \) is continuously differentiable and strictly convex.
Then there exists a $w_i \in \mathbb{R}$, such that

$$\lim_{w \to -\infty} w'(w, \theta_i) = w_i$$

See Appendix B.1.1 for the proof.

A key step in the proof is to show that when promised utility is sufficiently low, incentive problems, as measured by the values of the multipliers on the incentive constraints, converge to zero. An important part of the proof uses the fact that $h$ is unbounded below.

This property is one of the key technical findings in the paper. It can also be shown that, with utility unbounded below, this also holds in models with exogenous fertility. Loosely speaking, as $w$ gets smaller, the allocations look more and more similar to full information allocations, whether fertility is endogenous or exogenous. What makes an endogenous fertility model different from an exogenous one is the properties of full information allocations – continuation utility is bounded below (by shock-specific resetting values for per child continuation utility) when fertility is endogenous.

From this, it follows that as long as $w'(w, \theta)$ is continuous, $w'$ will be bounded below on any closed set bounded away from 0.

**Corollary 3.7** Suppose that $V$ is continuously differentiable and strictly convex. Then for all $\hat{w} < 0$, $w'(w, \theta)$ is bounded below on $(-\infty, \hat{w}]$ – there is a $w(\hat{w})$ such that $w'(w, \theta) \geq w(\hat{w})$ for all $w \leq \hat{w}$ and all $\theta$.

The proof can be found in the Supplementary Appendix.

### 3.4.2 Stationary Distributions

The results from the previous section effectively rule out a.s. immiseration as long as $n$ is bounded away from 0. This is not quite enough to show that a stationary distribution exists however. This is the topic of this section. There are two issues here. First, is there a stationary distribution for continuation utilities and is it non-trivial? Second, because the size of population is endogenous here and could be growing (or shrinking), we must also show that the growth rate of population is also stationary. We deal with this problem in general here.
Consider a measure of continuation utilities over $\mathbb{R}$, $\Psi$. Then, applying the policy functions to the measure $\Psi$, gives rise to a new measure over continuation utilities, $T\Psi$:

$$T(\Psi)(A) = \int_{\mathbb{R}} \sum_{\theta} \pi(\theta) 1_{\{(\theta, w); w'(\theta, w) \in A\}}(w, \theta)n(\theta, w)d\Psi(w)$$  \hspace{1cm} (3.13)

For a given measure of promised value today, $\Psi$, $T(\Psi)(A)$ is the measure of agents with continuation utility in the set $A$ tomorrow. The overall population growth generated by $\Psi$ is given by

$$\gamma(\Psi) = \int_{\mathbb{R}} \sum_{\theta} \pi(\theta)n(\theta, w)d\Psi(w)$$

Now, suppose $\Psi$ is a probability measure over continuation utilities. $\Psi$ is said to be a stationary distribution if:

$$T(\Psi) = \gamma(\Psi) \cdot \Psi$$

This is equivalent to having a constant distribution of per capita continuation utility along a Balanced Growth Path in which population grows at rate $(\gamma(\Psi) - 1) \times 100$ percent per period.

To show that there is a stationary distribution, we will show that the mapping $\Psi \rightarrow \frac{T(\Psi)}{\gamma(\Psi)}$ is a well-defined and continuous function on the set of probability measures on a compact set of possible continuation utilities. To do this, we need to construct a compact set of continuation utilities, $[w, \bar{w}]$, such that:

1. For all $w \in [w, \bar{w}]$, there is a solution to problem $P1'$;
2. For all $w \in [w, \bar{w}]$, $w'(w, \theta) \in [w, \bar{w}]$;
3. $n(w, \theta)$ and $w'(w, \theta)$ are continuous functions of $w$ on $[w, \bar{w}]$;
4. $\gamma(\Psi)$ is bounded away from zero for the probabilities on $[w, \bar{w}]$.

First, we define $w$ and $\bar{w}$.

For any fixed $w < 0$, consider the problem:

$$\max_{n \in [0, 1/6]} h(1 - bn) + \beta n^\eta w.$$
Note that there is a unique solution to this problem for every \( w < 0 \). Moreover, this solution is continuous in \( w \). Let \( g(w) \) denote the maximized value in this problem and note that it is strictly increasing in \( w \). Because of this, \( \lim_{w \to 0} g(w) \) exists. In a slight abuse of notation, let \( g(0) = \lim_{w \to 0} g(w) \). Further, since \( w < 0 \), it follows that \( g(w) < h(1) \) and hence, \( g(0) \leq h(1) \). In fact, \( g(0) = h(1) \). To see this, consider the sequences \( w_k = -1/k \), \( n_k = k^{1/(2\eta)} \). Then, for \( k \) large enough, \( n_k \) is feasible and therefore, \( g(w_k) \geq h(1 - b n_k) + \beta n_k^\eta w_k \). Hence,

\[
    h(1) = \lim_{k \to \infty} h(1 - b n_k) - \beta k^{-1/2} = \lim_{k \to \infty} h(1 - b n_k) + \beta n_k^\eta w_k \leq \lim_{k} g(w_k) = g(0) \leq h(1).
\]

Thus, in a neighborhood of \( w = 0 \), \( g(w) < w \).

Assume that \( b < 1 \) (thus it is physically possible for the population to reproduce itself). Then, it also follows that for \( w \) small enough, \( g(w) > w \).

Hence, there is at least one fixed point for \( g \). Since \( g \) is continuous, the set of fixed points is closed. Given this there is a largest fixed point for \( g \). Let \( \overline{w} \) be this fixed point. Since \( g(w) < w \) in a neighborhood of 0, it follows that \( \overline{w} < 0 \).

Following Corollary 3.7, choose \( w = w(\overline{w}) \).

With these definitions, it follows that, as long as a solution to the functional equation exists for all \( w \in [\overline{w}, \bar{w}] \), \( w' w(\theta) \in [\overline{w}, \bar{w}] \). I.e., 2 above is satisfied.

As noted above, we have no way to guarantee from first principles that the requisite convexity assumptions are satisfied to guarantee that a unique solution to the functional equation exists and is unique (i.e., 1 and 3 above). Thus, we will simply assume that this holds. Given this assumption, 4 can be shown to hold since \( n \) must be bounded away from zero on \([\overline{w}, \bar{w}]\) for the promise keeping constraint to be satisfied.

Now, we are ready to prove our main result about the existence of a stationary distribution. Let \( M([\overline{w}, \bar{w}]) \) be the set of regular probability measures on \([\overline{w}, \bar{w}]\).

**Theorem 3.8** Assume that for all \( w \in [\overline{w}, \bar{w}] \), there is a solution to the functional equation and that it is unique. Then there exists a measure \( \Psi^* \in M([\overline{w}, \bar{w}]) \) such that \( T(\Psi^*) = \gamma(\Psi^*) \cdot \Psi^* \).

**Proof.**
Since \([\overline{w}, \bar{w}]\) is compact in \( \mathbb{R} \), by Riesz Representation Theorem ([Dunford and Schwartz, 1958],
IV.6.3), the space of regular measures is isomorphic to the space $C^*(\bar{w},w)$, the dual of the space of bounded continuous functions over $[\bar{w}, w]$. Moreover, by Banach-Alaoglu Theorem ([Rudin, 1991], Theorem 3.15), the set $\{\Psi \in C^*([\bar{w}, w]); ||\Psi|| \leq k\}$ is a compact set in the weak-* topology for any $k > 0$. Equivalently the set of regular measures, $\Psi$, with $||\Psi|| \leq 1$, is compact. Since non-negativity and full measure on $[\bar{w}, w]$ are closed restrictions, we must have that the set

$$\{\Psi : \Psi \text{ a regular measure on } [\bar{w}, w], \Psi([\bar{w}, w]) = 1, \Psi \geq 0\}$$

is compact in weak-* topology.

By definition,

$$T(\Psi)(A) = \int_{[\bar{w}, w]} \sum_{i=1}^n \pi_i 1 \{w'(w, \theta_i) \in A\} n(w, \theta_i) d\Psi(w).$$

The assumption that the policy function is unique implies that it is continuous by the Theorem of the Maximum. It also follows from this that $n$ is bounded away from 0 on $[\bar{w}, w]$ (since otherwise utility would be $-\infty$). From this, it follows that $T$ is continuous in $\Psi$. Moreover,

$$\gamma(\Psi) = \int_{[\bar{w}, w]} \sum_{i=1}^n \pi_i n(w, \theta_i) d\Psi(w) \geq n > 0.$$ 

is a continuous function of $\Psi$ and is bounded away from zero.

Therefore, the function

$$\hat{T}(\Psi) = \frac{T(\Psi)}{\gamma(\Psi)} : \mathcal{M}([\bar{w}, \bar{w}]) \rightarrow \mathcal{M}([\bar{w}, \bar{w}])$$

is continuous. Therefore, by Schauder-Tychonoff Theorem ([Dunford and Schwartz, 1958], V.10.5), $\hat{T}$ has a fixed point $\Psi^* \in M([\bar{w}, \bar{w}]).$

Q.E.D.

This theorem immediately implies that there is a stationary distribution for per capita consumption, labor supply and fertility. Moreover, since promised utility is fluctuating in a bounded set, per capita consumption has the same property. This is in contrast to the models with exogenous fertility where a shrinking fraction of the population will have an ever growing fraction of aggregate consumption.\textsuperscript{12}

---

\textsuperscript{12} There is a technical difficulty with extending this Theorem to settings with a continuum of types. A sufficient condition for the result to hold is that $w'(w, \theta)$ is increasing in $\theta$. 
The resetting property at the top has important implications about intergenerational social mobility. In fact, it makes sure that any smart parent will have children with a high level of wealth - as proxied by continuation utility. Finally, there is a lower bound on how much of this mobility occurs:

**Remark 3.9** Suppose that $\bar{w} = w_0$. Choose $A > 0$ so that $\frac{n(w,\theta_n)}{n(w,\theta)} \geq A$ for all $w$ and $\theta$. Suppose that $l(w,\theta_n) > 0$, for all $w \in [\underline{w}, \bar{w}]$, then for any $\Psi \in M([\underline{w}, \bar{w}])$, we have:

$$\hat{T}(\Psi) \{\{w_0\}\} \geq \frac{\pi_n A}{1 - \pi_n + \pi_n A}.$$

See Supplementary Appendix for proof.

Theorem 3.8, although the main theorem of the paper, says very little about uniqueness and stability as well as its derivation. The main problem is with endogeneity of population. This feature of the model, makes it very hard to show results regarding uniqueness or stability. In the Supplementary Appendix, we give an example of an environment with two values of shocks two productivity. In this case, under the resetting assumption, we are able to characterize one stationary distribution and show that given the class of distributions considered, the stationary distribution is unique. This procedure, as described in the Supplementary Appendix, can be used to construct at least one stationary distribution. The main idea for the construction is to start from full mass at the resetting value $\underline{w}_I$ and iterate the economy until convergence.

The difficulty in proving uniqueness and stability of the stationary distribution, depends heavily on the fact that fertility is endogenous. Endogenity of fertility, implies that the transiotion function for promised value is not Markov. That is, the set of possible per capita promised values in the next period is not of unit measure. This in turn implies that this transition function can have multiple eigenvalues and eigenvectors each corresponding to a population growth rate $\gamma$ and a stationary distribution $\Psi$. Therefore, we suspect that there are example economies in which stationary distribution is not unique.

### 3.4.3 Inverse Euler Equation and a Martingale Property

An important feature of dynamic Mirrleesian models with private information is the Inverse Euler Equation. [Golosov et al., 2003] extend the original result of [Rogerson, 1985a]
and show that in a dynamic Mirrleesian model with private information, when utility is separable in consumption and leisure, the Inverse Euler equation holds when processes for productivity come from a general class. Here we will show that a version of the Inverse Euler equation holds. To do so, consider problem (P1). Suppose the multiplier on promise keeping is $\lambda$ and the multiplier on $\theta_{11}$ is $\mu(\theta, \hat{\theta})$. Then the first order condition with respect to $W'(\theta)$ is the following:

$$\pi(\theta) \frac{1}{R} V_{W}(N'(\theta), W'(\theta)) + \lambda \pi(\theta) \beta + \beta \sum_{\theta > \hat{\theta}} \mu(\theta, \hat{\theta}) - \beta \sum_{\theta < \hat{\theta}} \mu(\hat{\theta}, \theta) = 0.$$ 

Define $\mu(\theta, \hat{\theta}) = 0$, if $\hat{\theta} \geq \theta$. Summing the above equations over all $\theta$’s, we have

$$\frac{1}{R} \sum_{\theta} \pi(\theta) V_{W}(N'(\theta), W'(\theta)) + \beta \lambda \pi(\theta) + \beta \sum_{\theta} \sum_{\hat{\theta}} \mu(\theta, \hat{\theta}) - \beta \sum_{\theta} \sum_{\hat{\theta}} \mu(\hat{\theta}, \theta) = 0.$$ 

Moreover, from the Envelope Condition:

$$V_{W}(N, W) = -\lambda.$$ 

Therefore, we have

$$\sum_{\theta} \pi(\theta) V_{W}(N'(\theta), W'(\theta)) = \beta RV_{W}(N, W).$$ 

Now consider the first order condition with respect to $C'(\theta)$:

$$\pi(\theta) + \lambda \pi(\beta) N^{\eta-1} u' \left( \frac{C(\theta)}{N} \right) + N^{\eta-1} u' \left( \frac{C(\theta)}{N} \right) \sum_{\hat{\theta}} \mu(\theta, \hat{\theta}) - N^{\eta-1} u' \left( \frac{C(\theta)}{N} \right) \sum_{\hat{\theta}} \mu(\hat{\theta}, \theta) = 0.$$ 

Thus, 

$$V_{W}(N'(\theta), W'(\theta)) = \frac{\beta R}{N^{\eta-1} u' \left( \frac{C(\theta)}{N} \right)}.$$ 

which implies that

$$V_{W}(N_{t+1}(\theta^t), W_{t}^t(\theta^t)) = \frac{\beta R}{N_{t}(\theta^{t-1})^{\eta-1} u' \left( \frac{C_{t}(\theta^t)}{N_{t}(\theta^{t-1})} \right)}.$$
Hence, we can derive the Inverse Euler Equation:

\[
E \left[ \frac{1}{N_{t+1}(\theta^t)\eta - 1} u'(C_{t+1}(\theta^t)) \right] | \theta^t \] = \frac{\beta R}{N_t(\theta^{t-1})\eta - 1} u'(C_t(\theta^t)) \left( \frac{C_{t+1}(\theta^t)}{N_{t+1}(\theta^t)} \right). \tag{3.14}
\]

An intuition for this equation is worth mentioning. Consider decreasing per capita consumption of an agent with history \( \theta^t \) and saving that unit. There will be \( R \) units available the next day that can be distributed among the descendants. We increase consumption of agents of type \( \theta^t+1 \) by \( \epsilon(\theta^t+1) \) such that:

\[
n_t(\theta^t) \sum_{\theta} \pi(\theta) \epsilon(\theta) = R
\]

\[
u'(c_{t+1}(\theta^t, \theta)) \epsilon(\theta) = \nu'(c_{t+1}(\theta^t, \theta)) \epsilon(\theta') = \Delta.
\]

The first is the resource constraint implied by redistributing the available resources. The second one makes sure that the incentives are aligned. In fact it implies that the change in the utility of all types are the same and there is no incentive to lie. The above equations imply that

\[
n_t(\theta^t) \sum_{\theta_{t+1}} \pi(\theta_{t+1}) \frac{\Delta}{\nu'(c_{t+1}(\theta^t, \theta_{t+1}))} = R
\]

Since the change in utility from this perturbation must be zero, we must have \( \beta \Delta = \nu'(c_t(\theta^t)) \). Replacing in the above equation leads to equation (3.14). We summarize this as a Proposition:

**Proposition 3.10** If the optimal allocation is interior and \( V \) is continuously differentiable, the solution satisfies a version of the Inverse Euler Equation:

\[
E \left[ \frac{1}{N_{t+1}(\theta^t)\eta - 1} u'(C_{t+1}(\theta^t)) \right] | \theta^t \] = \frac{\beta R}{N_t(\theta^{t-1})\eta - 1} u'(C_t(\theta^t)) \left( \frac{C_{t+1}(\theta^t)}{N_{t+1}(\theta^t)} \right).
\]

Moreover, \( E_t N_{t+1}^{1-\eta} v'(w_{t+1}) = \beta R N_t^{1-\eta} v'(w_t) \). Hence, if \( \beta R = 1 \),

\[
N_{t+1}^{1-\eta} v'(w_t) = \frac{1}{N_{t+1}(\theta^t)\eta - 1} u'(C_{t+1}(\theta^t)) \left( \frac{C_{t+1}(\theta^t)}{N_{t+1}(\theta^t)} \right).
\]

and \( N_t^{1-\eta} v'(w_t) \) are non-negative Martingales.

If \( \beta R = 1 \), we see from above that \( X_t = N_t^{1-\eta} v'(w_t) \) is a non-negative martingale. Thus, the martingale convergence theorem implies that there exists a non-negative random variable with finite mean, \( X_\infty \), such that \( X_t \to X_\infty \) a.s.
As is standard in this literature, to provide incentives for truthful revelation of types, we must have ‘spreading’ in \( (N'(\theta))^{1-\eta}v'(w'(\theta)) \) (details in [Hosseini et al., 2009]) as long as some incentive constraint is binding.\(^{13}\) [Thomas and Worrall, 1990] have shown that in an environment where incentive constraints are always binding, spreading leads to immiseration. We can show a similar result in our environment, under some restrictions:

**Theorem 3.11** If \( \Psi^*\{w; \exists i \neq j \in \{1, \cdots , I\}, \mu(w; i, j) > 0\} = 1 \), then \( N_t \to 0 \ a.s. \)

The condition above ensures that there is always spreading in \( N_t^{\eta-1}v'(w_t) \) when the economy starts from \( \Psi^* \) as initial distribution for \( w \). In this case, the same proof as in Proposition 3 of [Thomas and Worrall, 1990] goes through and the above theorem holds. In fact, spreading implies that \( X_t \) converges to zero in almost all sample paths. Since \( w_t \) is stationary, \( N_t \) converges to zero almost surely.

The failure of the above condition implies that, there exists a subset of promised utilities \( A \) such that \( \Psi^*(A) > 0 \) and \( \forall w \in A, \forall i, j, \mu(w; i, j) = 0 \). For all \( w \in A \), based on the analysis in the Supplemental Appendix, there is no spreading. The evolution of \( N_t \) in this case depends on the details of the policy function \( w'(w, \theta) \). For example, suppose that for all \( w \in A, \theta \in \Theta, w'(w, \theta) \in A \). Then if at some point in time \( w_t \in A \), then \( w_{t'} \in A \) for all subsequent periods \( t' \) and \( N_{t'} \) will evolve so that \( N_{t'}^{1-\eta}v'(w_{t'}) \) is a fixed number - equal to \( N_1^{1-\eta}v'(w_t) \). Since \( w_{t'} \in [\underline{w}, \bar{w}] \), \( N_{t'} \) will also be a finite number and is bigger than zero. In this case, the population would not be shrinking or growing indefinitely following these sequences of shocks and this happens for a positive measure of long run histories. This case, is similar to an example given in [Phelan, 1998] in which case a positive fraction of agents end up with infinite consumption and a positive fraction of agents end up with zero consumption. [Kocherlakota, 2010] constructs a similar example for a Mirrleesian economy.

Intuitively, the planner is relying heavily on overall dynasty size to provide incentives and less on continuation utilities. This is something that sets this model apart from the more standard approach with exogenous fertility.

\(^{13}\) When promised utility of the parent is very high, it is possible that all types work zero hours. In this case, all types receive the same allocation and none of the incentive constraints are binding. See discussion in section 3.3.
Finally, the fact that $N_{t+1}(\theta^t) \to 0$ a.s. does not mean that fertility converges to zero almost surely, rather, it means that it is less than replacement (i.e., $n < 1$). Indeed, in computed examples, it can be shown that for certain parameter configurations (with $\beta R = 1$), $E(N_{t+1}(\theta^t)) \to \infty$ (i.e., $\gamma_{\Psi^*} > 1$). The reason for this apparent contradiction is that $N_t$ is not bounded – it converges to zero on some sample paths and to $\infty$ on others.

Similarly, it can be shown that when $\beta R > 1$, a stationary distribution over per capita variables still exists (see Theorem 3.8) but it need not follow that $N'(\theta^t) \to 0$ a.s. In fact, numerical examples can be constructed in which $N_t \to \infty$ a.s. (see the Supplementary Appendix). In the numerical example, we solve the optimal contracting problem for an example with two values of shocks. For that example, we can calculate the Markov process for $n_t$ induced by the policy functions $w'(w, \theta)$ and $n(w, \theta)$. If the economy starts from the stationary distribution, it can be shown that this Markov process is irreducible and acyclical and therefore has a unique stationary distribution $\Phi^*$. Moreover, in our example $\int \log nd\Phi^* > 0$. Therefore, by theorem 14.7 in [Stokey and Lucas, 1989], the strong law of large number holds for $\log n_t$:

$$\frac{1}{T} \sum_{t=1}^{T} \log n_t \to \int \log nd\Phi^* > 0, \text{ as } T \to \infty, \text{ a.s.}$$

Notice that $N_{t+1} = N_t \times n_t$ and therefore $N_{t+1} = \log N_t + \log n_t$. This implies that $\frac{1}{T} \log N_t \to \int \log nd\Phi^* \text{ a.s.}$ and since $\int \log nd\Phi^* > 0$, $\log N_t \to \infty \text{ a.s.}$ Therefore, through our numerical example, we can see that when $\beta R > 1$, a per capita stationary distribution exists, population grows at a positive rate, and almost all dynasties survive. This is in contrast with [Atkeson and Lucas, 1992] where consumption inequality grows without bound for any value of $\beta$ and $R$.

### 3.5 Extensions and Complementary Results

In this section, we discuss some complementary results. These are:

1. Implementing the optimal allocation through a tax system, and,

2. Differences between social and private discounting.
3.5.1 Implementation: A Two Period Example

Here, we discuss implementing the efficient allocations described above through decentralized decision making with taxes. To simplify the presentation we restrict attention to a two period example and explicitly characterize how tax implementations are used to alter private fertility choices. Similar results can be shown for the analogous ‘wedges’ in the infinite horizon setting.

As in the example in Section 3.2, we assume that there is a one time shock, realized in the first period.

The constrained efficient allocation $c_{1i}^*, l_i^*, n_i^*, c_{2i}^*$ solves the following problem:

$$
\sum_{i=H,L} \pi_i \left[ u(c_{1i}) + h(1 - l_i - bn_i) + \beta n_i^\eta u(c_{2i}) \right]
$$

s.t.

$$
\sum_{i=H,L} \pi_i \left[ c_{1i} + \frac{1}{R} n_i c_{2i} \right] \leq \sum_{i=H,L} \pi_i \theta_i l_i + Rk_0
$$

$$
u(c_{1H}) + h(1 - l_H - bn_H) + n_H^\eta u(c_{2H}) \geq u(c_{1L}) + h(1 - \frac{\theta_L l_L}{\theta_H} - bn_L) + n_L^\eta u(c_{2L}).$$

Now suppose that we want to implement the above allocation with a tax function of the form $T(y,n,c_2)$. Then the consumer’s problem is the following:

$$
\max_{c_1,n,l,c_2} u(c_{1}) + h(1 - l - bn) + \beta n^\eta u(c_{2})
$$

s.t.

$$
c_1 + k_1 \leq Rk_0 + \theta l - T(\theta l, n, c_2)
$$

$$
nc_2 \leq Rk_1
$$

It can be shown that if $T$ is differentiable and if $y$ is interior for both types

$$
T_n(\theta_H l_H^*, n_H^*, c_{2H}^*) = T_y(\theta_H l_H^*, n_H^*, c_{2H}^*) = T_{c_2}(\theta_H l_H^*, n_H^*, c_{2H}^*) = 0
$$

That is, there are no (marginal) distortions on the decisions of the agent with the high shock. Thus, what we need to do is to characterize the types of distortions that are used to get the low type to choose the correct allocation.

It is well known that when the type space is discrete, the constrained efficient allocation cannot be implemented by a continuously differentiable tax function. (This is also
true in our environment.) However, there exists continuous and piecewise differentiable tax functions which implement the constrained efficient allocation. Next, we construct the analog of this for our environment.

Let \( \bar{u}_L \) (resp. \( \bar{u}_H \)) be the level of utility received at the socially efficient allocation by the low (resp. high) type, and define two versions of the tax function:

\[
\bar{u}_L = u(y - T_L(y, n, c) - \frac{1}{R}nc_2) + h(1 - \frac{y}{\theta_L} - bn) + \beta n^\gamma u(c_2),
\]

\[
\bar{u}_H = u(y - T_H(y, n, c) - \frac{1}{R}nc_2) + h(1 - \frac{y}{\theta_H} - bn) + \beta n^\gamma u(c_2).
\]

\( T_L \), is designed to make sure that the low type always gets utility \( \bar{u}_L \) if they satisfy their budget constraint with equality while \( T_H \), is defined similarly. It can be shown that such \( T_L \) and \( T_H \) always exist, and from the Theorem of the Maximum, they are continuous functions of \((y, n, c_2)\). Moreover, since \( c_1 > 0 \) (i.e., \( y - T - \frac{1}{R}nc_2 > 0 \)) they are each differentiable.

We will build the overall tax code, \( T(y, n, c_2) \), by using \( T_L \) as the effective tax code for the low type and \( T_H \) as the one for the high type. Given this, it follows that the distortions, at the margin, faced by the two types are described by the derivatives of \( T_L \) (\( T_H \)) with respect to \( y \) and \( n \).

**Remark 3.12** If the allocation is interior,

1. The tax function

\[
T(y, n, c_2) = \max\{T_L(y, n, c_2), T_H(y, n, c_2)\}
\]

implements the efficient allocation.

2. If the incentive constraint for the low type is slack, there are no distortions in the decisions of the high type - \( \frac{\partial T}{\partial y}(y^*_H, n^*_H, c^*_L) = \frac{\partial T_H}{\partial y}(y^*_H, n^*_H, c^*_L) = 0 \) and \( \frac{\partial T}{\partial n}(y^*_H, n^*_H, c^*_L) = \frac{\partial T_H}{\partial n}(y^*_H, n^*_H, c^*_L) = 0 \).

3. At the choice of the low type, \((y^*_L, n^*_L, c^*_L), T = T_L \) and (i) \( \frac{\partial T_L}{\partial y}(y^*_L, n^*_L, c^*_L) > 0 \); (ii) \( \frac{\partial T_L}{\partial n}(y^*_L, n^*_L, c^*_L) > 0 \); (iii) \( \frac{\partial T_L}{\partial c_2}(y^*_L, n^*_L, c^*_L) = 0 \).

**Proof.** See Appendix.
The new finding here is that the planner chooses to tax the low type at the margin for having more children – \( \frac{\partial T_L}{\partial n} (y^*_L, n^*_L, c^*_2L) > 0 \). In the Mirrlees model without fertility choice, for incentive reasons, the planner wants to make sure that the low type consumes more leisure (relative to consumption) than he would in a full information world – this makes it easier to get the high type to truthfully admit his type. This is accomplished by having a positive marginal labor tax rate for the low type. Here, there is an additional incentive effect that must be taken care of. This is for the planner to make sure that the low type doesn’t use too much of his time free from work raising children. This would also make it more appealing to the high type to lie. To offset this here, the planner also charges a positive tax rate on children for the low type. These two effects taken together ensure that the low type has low consumption and fertility and high leisure thereby separating from the high type.

**Positive Estate Taxes**

In the above example, since all individual uncertainty is realized in the first period, there is no need for taxation of estate/capital. However, we know from [Golosov et al., 2003] that if there is subsequent realization of individual shocks, the optimal allocation features a positive wedge on saving. The same logic applies here since we have a version of Inverse Euler Equation. Consider the model in section 3.3. Recall, from section 3.4 that we have

\[
\frac{\beta R}{u'(c_t(\theta^t))} = \sum_{\theta_{t+1} \in \Theta} \pi(\theta_{t+1}) \frac{1}{n_t(\theta^t)\eta^{-1}u'(c_{t+1}(\theta^t, \theta_{t+1}))}
\]

When, \( c_{t+1}(\theta^t, \theta_{t+1}) \) varies with realization of \( \theta_{t+1} \), the Jensen’s inequality implies that

\[
\frac{\beta R}{u'(c_t(\theta^t))} = \sum_{\theta_{t+1} \in \Theta} \pi(\theta_{t+1}) \frac{1}{n_t(\theta^t)\eta^{-1}u'(c_{t+1}(\theta^t, \theta_{t+1}))} < \frac{1}{\sum_{\theta \in \Theta} \pi(\theta_{t+1})n_t(\theta^t)\eta^{-1}u'(c_{t+1}(\theta^t, \theta_{t+1}))}
\]

and therefore

\[
\beta R \sum_{\theta \in \Theta} \pi(\theta_{t+1}) n_t(\theta^t)\eta^{-1}u'(c_{t+1}(\theta^t, \theta_{t+1})) < u'(c_t(\theta^t))
\]
The above equation implies that there is a positive wedge on saving. This positive wedge, in turn, translates into positive marginal tax rates on bequests. In fact, we think that it is relatively straightforward to extend the tax system in [Werning, 2010] in order to implement the optimal allocation in our environment and that tax system features positive taxes on bequests. This is in contrast with [Farhi and Werning, 2010c], where they have shown that in order to implement an optimal allocation that has a stationary distribution, the planner must subsidize bequests.

3.5.2 Social vs. Private Discounting

In this section we consider a version of the two period example discussed in section 3.2 motivated by the papers by [Phelan, 2006] and [Farhi and Werning, 2007]. This is to include a difference between the discount rate used by private agents and that used by the planner. Thus, we want to analyze the solution to the following planning problem:

$$\max_{c_1(\theta), n(\theta), l(\theta), c_2(\theta)} \sum_{\theta_L, \theta_H} \left[ \pi(\theta) \left( u(c_1(\theta)) + h(1 - l(\theta) - bn(\theta)) + \beta n(\theta)^\eta u(c_2(\theta)) \right) \right]$$

s.t.

$$\sum_{\theta_L, \theta_H} \pi(\theta) \left( c_1(\theta) + \frac{1}{R} n(\theta) c_2(\theta) \right) \leq \sum_{\theta_L, \theta_H} \pi(\theta) \theta l(\theta) + K_0$$

$$u(c_1(\theta_H)) + h(1 - l(\theta_H) - bn(\theta_H)) + \beta n(\theta_H)^\eta u(c_2(\theta_H)) \geq u(c_1(\theta_L)) + h \left( 1 - \frac{\theta_L}{\theta_H} l(\theta_L) - bn(\theta_L) \right) + \beta n(\theta_L)^\eta u(c_2(\theta_L))$$

This can be thought of as finding an alternative Pareto Optimal allocation when the planner puts more weight on children than individual parents do. As before, we have the following result:

$$\eta u(c_2(\theta_H)) = u'(c_2(\theta_H)) c_2(\theta_H) + b \theta_H R u'(c_2(\theta_H)).$$

From this, we can see that there is still ‘resetting’ at the top – $c_2(\theta_H)$ does not depend on $K_0$. The reasoning behind this is the same as that given in section 3.2.2, i.e., the planner always has an incentive of choosing the mix between $C_2$ and $n$ so as to minimize the cost of providing any given level of utility in the second period. Because of this,
the same homogeneity/homotheticity logic still holds (and it does not depend on \( \hat{\beta} \)). Further, \( c_2(\theta_H) \) does not depend on \( \hat{\beta} \).

This suggests that our argument showing that there is a stationary distribution over continuation utilities will go through in this more general case. The one potential difficulty is proving an analog of theorem 3.8.

In addition, we have:

\[
\left( \hat{\beta} - \beta \right) \frac{u'(c_1(\theta_H))}{\beta \lambda} + 1 = \frac{1}{\beta R} n(\theta_H)^{1-\eta} \frac{u'(c_1(\theta_H))}{u'(c_2(\theta_H))}
\]

where \( \lambda \) is the Lagrange multiplier on the resource constraint.

As can be seen from this, when \( \hat{\beta} = \beta \), this gives the usual Euler equation. When \( \hat{\beta} > \beta \) this equation shows that, in general, there is an extra force to increase \( n \). This is because \( c_2(\theta_H) \) is independent of \( \hat{\beta} \) and the term \( \left( \hat{\beta} - \beta \right) \frac{u'(c_1(\theta_H))}{\beta \lambda} \) is strictly positive. Formally, holding \( c_1(\theta_H) \) fixed, increasing \( \hat{\beta} \) increases \( n(\theta_H) \).

Intuitively, when \( \hat{\beta} > \beta \) the planner wants to increase second period utility (relative to the \( \hat{\beta} = \beta \) case). Since \( u(c_2(\theta_H)) \) does not depend on \( \hat{\beta} \) the only channel available to do this is through increasing \( n(\theta_H) \).

In sum, when the planner is more patient than private agents, he will encourage more investment both by increasing population size and increasing savings. Thus, this approach has important implications for population policy over and above what it implies about long run inequality.
Chapter 4

Providing Efficient Incentives to Work: Retirement Ages and Pension System

4.1 Introduction

Economic efficiency suggests that more productive individuals should work more and retire later than their less productive peers. However, if individuals can work with productivity below their maximum, earlier retirement may be needed to provide them with incentives to fully realize their productivities while they work. We study this tension in a class of lifecycle models. We emphasize active intensive and extensive margins of labor supply in the individual decisions of how much to work when not retired and when to retire. The paper provides a theoretical and quantitative analysis of the efficient distribution of retirement ages and examines how the interaction between the tax code and the pension system should be designed to implement the optimum.

Specifically, this paper studies lifecycle environment where individuals differ in two respects. First, individual workers are heterogeneous in their productivities. A worker’s
productivity changes over lifecycle and follows a privately known idiosyncratic hump-shaped productivity profile.\footnote{At least since \cite{Mincer74}, it has been known that productivity typically increases earlier in life and declines later leading some individuals to leave the labor force entirely, i.e., to retire. These changes in productivity do not happen to everyone at the same age or at the same rate. Some individuals experience significant decreases in their ability to produce rather early in life while others remain productive for many more years.} Second, individuals face privately known heterogeneous fixed costs of work.\footnote{In most of the paper we focus on cases where fixed cost is perfectly correlated with productivity type. We later study an extension that allows partial correlation between fixed cost and productivity profiles.} Fixed utility cost of work introduces non-convexity into disutility of working. Combined with a hump-shaped productivity profile, this makes it optimal for a worker to choose to retire at some age while heterogeneity implies that retirement ages differ among workers. In other words, we study lifecycle environment that features both active intensive and extensive labor margins. A government in this environment reallocates resources across time and households to achieve efficiency and a certain level of redistribution. The government, however, cannot use policies contingent on productivity and fixed cost of work since productivity and fixed cost are private information available only to the individual.

Our first main result is to derive conditions on fundamentals under which efficient retirement ages are increasing in lifetime earnings. More generally, the analysis here clearly identifies factors that determine how efficient retirement ages change as a function of productivity. These factors are \textit{(i)} virtual fixed costs of work, i.e., fixed cost of work plus rents from private information,\footnote{The notion of virtual fixed costs here, or virtual types, is akin to Myerson’s virtual types.} and how they change with productivity, \textit{(ii)} the distribution of productivities, and \textit{(iii)} how redistributive the government is. The intuition behind this result is that the economy with private information is equivalent to an economy without frictions but with modified, or virtual, productivity profiles and fixed costs of work. The virtual types depend on the distribution of productivities and on how redistributive the government is. To provide sharper focus on the underlying mechanisms, our baseline formulation is the case without income effects. A particularly tractable version, that abstracts from risk aversion and discounting, allows to derive closed form characterizations that sharply highlight the forces driving our results. We then reintroduce curvature into the utility function. We conclude that, under plausible
conditions, efficient retirement ages increase in lifetime earnings. That is, individuals with higher lifetime earnings should be given incentives to retire older than less productive individuals.

Our second main result is to show that a policy based on actuarially unfair pension benefits can implement the optimum. That is, the pension side of the policy should be age dependent in a particular way - the present value of lifetime retirement benefits should rise with the age of retirement. We argue that distorting individual retirement age decision offers a powerful policy instrument. In particular, one important role of the retirement distortion is to undo part of the retirement incentives provided by a standard distortion of the consumption-labor margin, i.e., by a labor income tax. To demonstrate this, we provide a partial characterization of the distortions to both intensive and extensive labor margins, i.e., labor distortion and retirement distortion. Note that labor and retirement distortions both affect retirement incentives: income taxes decrease payoffs from an extra year in the labor force, while retirement benefits increase payoffs from staying out of the labor force. We show that, in the optimum, retirement distortion is lower than labor distortion at the time of retirement. Intuitively, this suggests that labor income taxes distort retirement decision too much. An implementation of the optimum thus requires a pension system with present value of retirement benefits increasing in retirement age to create benefits above and beyond income taxes.

Our third contribution is to provide a quantitative study of efficient work and retirement incentives. We use individual earnings and hours data from the U.S. Panel Study of Income Dynamics (PSID) in combination with retirement age data from the Health and Retirement Study (HRS) to calibrate and simulate efficient work and retirement choices and policy. We calibrate to also match the estimates of labor supply elasticity at the extensive margin. A quantitative result robust across calibrations is that individuals with higher lifetime earning in the U.S. should retire older than they do now and, more strikingly, older than less productive workers. We find in our benchmark

---

4 The design of the current pension system in the U.S., Social Security, is meant to be actuarially fair before age 65, i.e., benefits rise by 6.67% each year, but actuarially unfair after 65 with a decreasing present value of benefits, i.e., benefits rise by 5.5%. The actuarial fairness is of course affected by the actual life span.

5 [Chetty et al., 2011] emphasize the importance of calibrating the extensive margin elasticity as well as the intensive one. We follow their review of the existing estimates and calibrate to match the range of estimates of the extensive elasticity from the individual studies they analyze.
calibration that in the optimum, the highest productivity types retire at 69.5, whereas in the data their average retirement age is 62.8. At the same time, individuals with lower lifetime earnings should retire younger than they do now, as well as younger than their more productive peers. In particular, the lowest productivity types retire in the optimum at 62.2 years compared to 69.5 for the highest productivity types. This pattern of retirement ages is in sharp contrast with the one found in the current individual data for the U.S., where average retirement age displays a predominantly decreasing pattern as a function of lifetime earnings. We summarize this contrast in Figure 4.1. The dashed line displays the average retirement ages for earnings deciles in the data while the solid line displays simulated efficient retirement ages.

Our quantitative study allows us to measure and decompose welfare gains and total output gains associated with inducing efficient retirement age distribution. We find that providing efficient incentives for both work and retirement results in large welfare gains. We compute welfare gains that range across calibrations between 1 and 5 percent of annual consumption equivalent. Notably, we also find a small but positive change in total output of up to 1 percent. We show that this increase in total output results from the meaningfully active intensive and extensive distortions of labor supply. Increasing standard distortions along the intensive margin generally leads to output losses in favor of redistribution and welfare gains. The additional policy instrument of distorting the retirement decision proves powerful enough to overcompensate by inducing more productive individuals to work more years and thus produce more.

Distorting individual retirement decisions efficiently is a compelling example of a policy reform that can produce perceivable benefits. A recent surge of research points towards evidence of significant effects of the incentives created by the interaction between tax and pension systems: from providing strong incentives to leave labor force at statutory retirement age (see, e.g., [French, 2005]) and resulting in significant amount of redistribution (see, e.g., [Feldstein and Liebman, 2002b] and [Feldstein and Liebman, 2002a]) to penalizing work after statutory retirement age regardless of how productive a worker is (e.g. [Gruber and Wise, 2007]), to cite just a few recent examples. A unifying theme that emerges from this evidence is a need to address the question of how to design these incentives to reap maximum welfare gains and what that implies about when individuals should retire.
The analysis in this paper contributes to several literatures. Most directly, it provides a new and empirically-based policy application of the tools of a literature (see, e.g., [Prescott et al., 2009] and [Rogerson and Wallenius, 2009]) reconciling macro and micro estimates of labor elasticities with meaningfully active intensive and extensive margins of labor supply. It also extends the literature on optimal distortionary policies with both margins of labor supply. That literature was reinvigorated with the contribution of [Saez, 2002], who studies optimal income transfer programs when labor responses are concentrated along the intensive responses or when labor responses are concentrated along the extensive responses. Numerous recent studies provide further theoretical extensions of that literature (see, e.g., [Chone and Laroque, 2010]). Finally, the analysis in this paper also contributes to the empirically-driven Mirrleesian literature that connects labor distortions to estimable distributions and elasticities, as do [Diamond, 1998]; [Saez, 2001], and in dynamic environments [Golosov et al., 2010]. In particular, we provide elasticity-based expressions for labor distortions in the presence of both margins of labor supply. Our result about the increase in total output is related to analysis in [Golosov and Tsyvinski, 2006]. Most modern studies of efficient redistributive policies largely result in increased distortions improving welfare but generally sacrificing total output (see, e.g., [Fukushima, 2010], [Farhi and Werning, 2010a], [Golosov et al., 2010], [Weinzierl, 2011]). Unlike much of the optimal tax literature, rather than focusing on a specific social welfare function, we characterize Pareto efficient allocations similar to [Werning, 2007].

The questions we address and the policy implications we seek are also related to those in [Conesa et al., 2009b] as well as in [Huggett and Parra, 2010]. Their approach differs from ours as they study policies within a set of parametrically restricted functions. One advantage of that approach is that it is computationally more feasible while allowing to study commonly used in practice policies. This paper examines a larger set of policies that are endogenously restricted by the information structure.

The rest of the paper proceeds as follows. The next section describes a lifecycle environment with active intensive and extensive labor margins. Section 4.3 makes precise the notions of distortions and of the tax and pension system in our environment. Section 4.4 provides analytic characterization of the baseline model, including efficient

---

6 For a review as well as international evidence of the effects of labor taxes see [Rogerson, 2010].
Figure 4.1: Empirical weighted average and simulated efficient retirement ages for the U.S., by lifetime earnings decile. Sources: HRS, PSID, and authors’ calculations.

retirement age patterns. Section 4.5 theoretically examines policies that implement those patterns. Section 4.6 provides quantitative analysis based on the individual level U.S. data and intensive and extensive elasticities. Section 4.7 concludes.

4.2 Environment

This section builds a life-cycle environment where intensive and extensive margins of labor supply are emphasized in the individual decisions to work and retire. We use a baseline version of this environment in Section 4.4 to analytically study the tension
between efficiency and equity and examine the efficient distribution of retirement ages. In Section 4.5, we study how the interaction between the tax policy and the pension system should be designed to implement such distribution. We make precise what we mean by the interaction between the tax policy and the pension system in Section 4.3.

Time is continuous and runs from \( t = 0 \) to \( t = 1 \). The economy is populated by a continuum of individuals who are born at \( t = 0 \) and live until \( t = 1 \), at which point they all die. Each individual consumes and works over their lifecycle. Individuals differ in two respects. First, individual workers are heterogeneous in their productivities. A worker’s productivity changes over lifecycle and follows a privately known idiosyncratic hump-shaped productivity profile. Second, individuals face privately known heterogeneous fixed costs of work. Specifically, at time \( t = 0 \), each individual draws a type, \( \theta \), from a distribution of types, \( F(\theta) \) where \( F'(\theta) = f(\theta) > 0 \). Individual’s type, \( \theta \), determines their productivity over their lifecycle as well as their preferences toward working, i.e., fixed utility cost the individual faces whenever she works. One interpretation of \( \theta \) can be the individual’s lifetime income.

An individual’s type \( \theta \) determines this individual’s productivity profile \( \{ \varphi(t, \theta) \}_{t \in [0,1]} \) over lifecycle. That is, when the individual works \( l \) hours at age \( t \), then her income is \( \varphi(t, \theta)l \). The productivity profile for the individual has the following two properties. First, The productivity profile, \( \varphi(t, \theta) \), is continuous and twice continuously differentiable. Second, the productivity follows a hump-shape over the lifecycle, i.e. the profile exhibits an inverse U-shape. In other words, for any type \( \theta \), there exists an age \( t^* \) such that for all \( t < t^* \), \( \varphi_t(t, \theta) > 0 \) and for all \( t > t^* \), \( \varphi_t(t, \theta) < 0 \).

The latter property is worth discussing. The fact that wages or earnings are inverse U-shaped is a classic results in labor economics known at least since [Mincer, 1974]. Instead of taking a stand on why this is the case, we simply take this stylized fact as given and study its implications for the efficient distribution of retirement ages and how the interaction between the tax policy and the pension system should be designed to implement such distribution.

In addition to productivity, the type \( \theta \) affects individual preferences. In particular, a household that draws a type \( \theta \), has the following preferences

\[
\int_0^1 e^{-\rho t} \left[ u(c(t)) - v \left( \frac{y(t)}{\varphi(t, \theta)} \right) - \eta(\theta) \mathbf{1}[y(t) > 0] \right] dt
\] (4.1)
over the set of all allocations \( \{c(t), y(t)\}_{t \in [0,1]} \) of consumption and income. Here, \( 1[y(t) > 0] \) is an indicator function of positive output, \( y(t) \). The utility function \( u(\cdot) \) is strictly concave, increasing and satisfies standard Inada conditions. Moreover, \( v(\cdot) \) is a strictly convex function with \( v'(0) = 0 \). These preferences exhibit fixed costs of working. This fixed utility cost of work can represent commute time, fixed costs of setting up jobs, etc.\(^7\) While we do not take a stand on the particular interpretation of this parameter, when we turn to a quantitative analysis of calibrated policy models in Section 4.6, we calibrate the fixed cost function \( \eta(\theta) \) to match the observed patterns of retirement in the individual U.S. data. Intuitively, fixed utility cost of work introduces non-convexity into disutility of working. Combined with a hump-shaped productivity profile, this makes it optimal for a worker to choose to retire at some age while heterogeneity implies that retirement ages differ among workers. In other words, this environment features both active intensive and extensive margins in the decisions to work and retire.

Given individual preferences and productivities, we define feasible allocations. An allocation is defined as \( \left( \{c(t, \theta), y(t, \theta)\}_{\theta \in [\underline{\theta}, \bar{\theta}]}, K_t \right)_{t \in [0,1]} \) where \( K_t \) is the aggregate asset holdings of all households. An Allocation is said to be feasible if

\[
\int_{\underline{\theta}}^{\bar{\theta}} c(t, \theta) \, dF(\theta) + \dot{K}_t + G_t \leq \int_{\underline{\theta}}^{\bar{\theta}} y(t, \theta) \, dF(\theta) + rK_t
\]

given \( K_0 \) where \( G_t \) is government expenditure. Throughout the paper, we will use the above budget constraint and its present value equivalent interchangeably:

\[
\int_{0}^{1} e^{-rt} \left[ \int_{\underline{\theta}}^{\bar{\theta}} c(t, \theta) \, dF(\theta) \right] dt + H \leq \int_{0}^{1} e^{-rt} \left[ \int_{\underline{\theta}}^{\bar{\theta}} y(t, \theta) \, dF(\theta) \right] dt + (1 + r)K_0
\]

where \( H \) is the time zero present value of government spending, i.e., \( \int_{0}^{1} e^{-rt} G_t \, dt \). A change in the order of the integrals leads to the following:

\[
\int_{\underline{\theta}}^{\bar{\theta}} \int_{0}^{1} e^{-rt} c(t, \theta) \, dt \, dF(\theta) + H \leq \int_{\underline{\theta}}^{\bar{\theta}} \int_{0}^{1} e^{-rt} y(t, \theta) \, dt \, dF(\theta) + (1 + r)K_0
\]

It can be shown that there exists a retirement age for each type, i.e., an age above which households work and above which they do not. Specifically, for each type

\(^7\) Note that alternatively, one can assume that firms have to pay fixed costs of setting up jobs and hence the fixed costs are in terms of consumption goods (see, e.g., [Rogerson and Wallenius, 2009]) Our formulation significantly simplifies the analysis.
\(\theta\), there exists \(R(\theta)\) such that \(y(t, \theta) > 0\) if and only if \(t < R(\theta)\). To simplify the analysis, from here we assume that this is the case and provide a proof in the Appendix. When this is the case, an allocation is given by \(\{c(t, \theta)\}_{t \in [0,1], \theta \in [\theta, \bar{\theta}]} \cup \{R(\theta)\}_{\theta \in [\theta, \bar{\theta}]}\), \(\{y(t, \theta)\}_{t \in [0, R(\theta)], \theta \in [\theta, \bar{\theta}]} \cup \{K_t\}_{t \in [0,1]}\). Then, the above feasibility constraint can be written as

\[
\int_{\theta}^{\bar{\theta}} \int_0^1 e^{-rt} c(t, \theta) \, dt \, dF(\theta) \leq \int_{\theta}^{\bar{\theta}} \int_0^{R(\theta)} e^{-rt} y(t, \theta) \, dt \, dF(\theta) + (1 + r) K_0 \quad (4.2)
\]

Throughout the paper, we assume that \(\theta\) is privately observed by the individuals and not the planner or the government. By appealing to the revelation principle we focus on direct mechanisms and emphasize incentive compatibility. An allocation is said to be incentive compatible if it satisfies the following condition

\[
\int_0^1 e^{-\rho t} u(c(t, \theta)) \, dt \geq \int_0^{R(\theta)} e^{-\rho t} \left[ v \left( \frac{y(t, \theta)}{\varphi(t, \theta)} \right) + \eta(\theta) \right] \, dt \quad (4.3)
\]

\[
\int_0^1 e^{-\rho t} u(c(t, \bar{\theta})) \, dt \geq \int_0^{R(\bar{\theta})} e^{-\rho t} \left[ v \left( \frac{y(t, \bar{\theta})}{\varphi(t, \bar{\theta})} \right) + \eta(\theta) \right] \, dt
\]

We assume that the government desires to achieve some degree of redistribution and provides incentive for optimal working and retirement. That is, the planner has the following social welfare function

\[
\int_{\theta}^{\bar{\theta}} U(\theta) \, dG(\theta) \quad (4.4)
\]

where \(U(\theta)\) is the lifetime utility of a household of type \(\theta\) given by expression 4.1. The function \(G(\theta)\) is a cumulative density function, i.e., \(G(\bar{\theta}) = 0\), \(G(\theta) = 1\), and \(G'(\theta) = g(\theta) \geq 0\) and \(G(\theta)\) is differentiable over interval \((\theta, \bar{\theta})\). A redistributive motive for the planner implies that \(G(\theta) \geq F(\theta)\) for all \(\theta \in [\theta, \bar{\theta}]\). The case with \(F(\theta) = G(\theta)\) corresponds to the utilitarian social welfare function for the planner, while the case with \(G(\theta) = 1\), for all \(\theta > \bar{\theta}\) corresponds to the Rawlsian social welfare function.

\[8\] In order to allow for extremes of redistribution, i.e., Rawlsian preferences, we restrict the differentiability to the open interval.
In this environment, an allocation is efficient if it maximizes (4.4) subject to satisfying (4.3) and (4.2). We restate the mechanism design problem for convenience as

$$\max \left\{ \{c(t,\theta)\}_{t \in [0,1], \theta \in [\theta, \bar{\theta}]}, \{R(\theta)\}_{\theta \in [\theta, \bar{\theta}]}, \{y(t,\theta)\}_{t \in [0,R(\theta)], \theta \in [\theta, \bar{\theta}]} \right\} \int_{\theta}^{\hat{\theta}} U(\theta) \, dG(\theta) \quad (4.5)$$

subject to incentive compatibility

$$\int_{0}^{1} e^{-\rho t} u(c(t,\theta)) \, dt - \int_{0}^{R(\theta)} e^{-\rho t} \left[ v \left( \frac{y(t,\theta)}{\varphi(t,\theta)} \right) + \eta(\theta) \right] \, dt \geq \int_{0}^{1} e^{-\rho t} u(c(t,\hat{\theta})) \, dt - \int_{0}^{R(\hat{\theta})} e^{-\rho t} \left[ v \left( \frac{y(t,\hat{\theta})}{\varphi(t,\theta)} \right) + \eta(\theta) \right] \, dt$$

and feasibility

$$\int_{\theta}^{\hat{\theta}} \int_{0}^{1} e^{-r t} c(t,\theta) \, dt dF(\theta) + H \leq \int_{\theta}^{\hat{\theta}} \int_{0}^{R(\theta)} e^{-r t} y(t,\theta) \, dt dF(\theta) + (1 + r) K_0.$$

### 4.3 Distortions and policies

In this section, it is useful to make precise the types of policies we will later focus on. Using these policies, we define here the main margins of labor supply that a policy choice distorts as well as the extent of these distortions. Then, in Section 4.4, we analytically characterize constrained efficient allocations in the environment described above and, in Section 4.5, we examine its policy implications.

Consider a working individual as described above that pays age dependent income tax $T(t,y)$ at age $t$ and on income $y$. Upon retiring, the individual is entitled to the present value of pension benefits that depend on her retirement age as well as are a function of her income profile over working life, $b \left( R, Y \left( \{y(t)\}_{t=0}^{R} \right) \right)$. Here, $Y(\cdot)$ can be thought of as a measure of lifetime income or lifetime labor earnings. Facing this tax and benefit schedule, we think of individuals as solving the following problem

$$\max_{c(t), R, y(t), a(t)} \int_{0}^{1} e^{-\rho t} u(c(t)) \, dt - \int_{0}^{R} e^{-\rho t} \left[ v \left( \frac{y(t)}{\varphi(t,\theta)} \right) + \eta(\theta) \right] \, dt$$

subject to

$$c(t) + \dot{a}(t) = (y(t) - T(t,y(t))) \mathbf{1}[t \leq R] + \mathbf{1}[t > R] \frac{rb \left( R, Y \left( \{y(t)\}_{t=0}^{R} \right) \right)}{e^{-r R} - e^{-r}} + ra(t)$$
where in the above formulation, \( \frac{rb(R,Y(\{y(t)\}_{t=0}^R))}{e^{-rR-e^{-r}}} \) is the level of per period benefit from age \( R \) to 1 that will generate a present value of \( b \left( R, Y \left( \{y(t)\}_{t=0}^R \right) \right) \) and \( a(t) \) is the level of asset holdings by an agent at date \( t \).

Note that the above system of taxes and pension benefits resembles several features of the U.S. tax and social security system. In particular, the present value of pension benefits is a function of a measure of lifetime income analogous to the way social security benefits change with average indexed monthly earnings (AIME) in the U.S. Old Age, Survivors, and Disability Insurance program\(^9\). However, the above system is significantly different from the U.S. tax code and social security system in other ways. In particular, contrary to the U.S. social security benefits formula, the present value of benefits in the system above potentially changes directly with the retirement age and the labor income taxes depend on age.

One can rewrite the date-by-date budget constraints above as the following present value budget constraint:

\[
\int_0^1 e^{-rt} c(t) \, dt = \int_0^R e^{-rt} (y(t) - T(t,y(t))) \, dt + b \left( R, Y \left( \{y(t)\}_{t=0}^R \right) \right)
\]

Using this budget constraint, the individual optimal choice of work and retirement implies the following two optimality conditions:

\[
\varphi(t,\theta) \left[ 1 - T_y(t,y(t)) + e^{rt} \delta_{y(t)} Y \left( \{y(t)\}_{t=0}^R \right) by \right] u'(c(t)) = v \left( \frac{y(t)}{\varphi(t,\theta)} \right)
\]

\[
\left[ y(R) - T(R,y(R)) + e^{rR} b_R + e^{rR} by \delta_{R} Y \left( \{y(t)\}_{t=0}^R \right) \right] u'(c(R)) = v \left( \frac{y(R)}{\varphi(R,\theta)} \right) + \eta(\theta),
\]

where \( \delta_{y(t)} Y \) is the Fréchet derivative of \( Y \) with respect to \( y(t) \) and \( \delta_R Y \) is the Fréchet derivative of \( Y \) with respect to \( R \). Note that the above equations describe how and to what extents the intensive and extensive margins are distorted. To see this, notice that, in particular, in an undistorted allocation these conditions become:

\[
\varphi(t,\theta) u'(c(t)) = v' \left( \frac{y(t)}{\varphi(t,\theta)} \right)
\]

\[
y(R) u'(c(R)) = v \left( \frac{y(R)}{\varphi(R,\theta)} \right) + \eta(\theta)
\]

\(^9\) AIME is consisted of an inflation adjusted average of monthly earnings over the highest 35 years of earnings.
Given the above equations, we define in a natural way the extent to which each labor margin, intensive and extensive, is distorted. For any allocation, the labor distortion (also sometime referred to as labor wedge) is given by

\[ \tau_l(t, \theta) = 1 - \frac{1}{\varphi(t, \theta)} u'(c(t, \theta)) \left( y(t, \theta) \right) \]

That is, the labor wedge is a measure of how the intensive margin of labor supply is distorted. Analogously in a natural way, we let retirement distortion (equivalently referred to as retirement wedge) be defined by

\[ \tau_r(\theta) = 1 - \frac{1}{y\left(R(\theta), \varphi(R(\theta), \theta)\right)} \left[ v\left(y\left(R(\theta), \varphi(R(\theta), \theta)\right)\right) + \eta(\theta) \right] \]

In other words, the retirement wedge measures how distorted the extensive margin of labor supply is, or how distorted the retirement decision is.

Given the above definitions and the notion of a system of taxes and pension benefits, we can relate the distortions of the intensive and extensive labor supply margins to the policy instruments by using the two optimality conditions:

\[ \tau_l(t, \theta) = T_y(t, y(t, \theta)) - e^{\gamma t} \delta_{y(t)} Y \left\{ \{ y(t, \theta) \}_{t=0}^{R} \right\} b_Y \]

\[ \tau_r(\theta) = \frac{T\left(R(\theta), y\left(R(\theta), \theta\right)\right)}{y\left(R(\theta), \theta\right)} - e^{\gamma R(\theta)} \frac{b_R + \delta_{R Y} \cdot b_Y}{y\left(R(\theta), \theta\right)} \]

This implies that characterizing the properties of the distortions of both the intensive and the extensive margins of labor supply can inform us about the properties of the system of policy instruments that we focus on. We show in Section 4.5 that a system of policy instruments we presented here can implement constrained efficient allocations. Before we do that, we turn in the next section to the characterization of the baseline form of our environment.

### 4.4 Characterization of efficient retirement

We now analytically study the tension between efficiency and equity in the baseline version of our lifecycle environment. Our analysis emphasizes active intensive and extensive margins of labor supply that represent themselves as the individual decisions to work and retire. In this section, we focus on a theoretical examination of the efficient
distribution of retirement ages. In the next section, we examine how the interaction between the tax policy and the pension system described above should be designed to implement efficient retirement ages.

To provide sharper focus on the main underlying mechanisms and to build intuition, in the baseline formulation we start by temporarily abstracting from risk aversion. We also abstract for now from time discounting. This allows us to derive closed form characterizations in this section. In Section 4.6 we reintroduce the curvature back into the utility function together with discounting and positive interest rates.

4.4.1 Retirement ages in a baseline case

We start by showing that under plausible conditions efficient retirement ages increase in lifetime earnings. That is, individuals with higher lifetime earnings should be given incentives to retire older than their less productive peers. Then, we show that distorting individual retirement decisions provides a novel and surprisingly powerful policy instrument. In particular, one important role of the retirement distortion is to undo part of the retirement incentives provided by a standard distortion of the consumption-labor margin, i.e., a labor income tax.

To build intuition, we first focus on a baseline case with linear \( u(c) \) and no time discounting, i.e., \( \rho = r = 0 \). Assume that Frisch (intensive) elasticity of labor supply is constant, in particular \( v(l) = \psi l^\gamma \gamma^{-1} \) with \( \gamma > 1 \). We can provide closed form characterizations, in particular, of how the retirement age changes with type, or with lifetime income. Some of these results carry over in a straightforward way to the general case.

Assume in addition that the productivity profiles have the following property.

Assumption 4.1 The productivity profile \( \varphi(t, \theta) \) satisfies \( \varphi_{t,\theta}(t, \theta) \geq 0 \).

The above assumption about the productivity profiles ensures that in our mechanism design problem, individuals optimally retire at a certain age and do not re-enter the labor force. Multiple studies that estimate heterogeneous productivity profiles over lifecycle

\(^{10}\) We also abstract from government spending since with risk neutral households, it does not change our results about wedges.
find similar patterns or at least patterns that do not deviate much from the property described in Assumption 4.1 (see, [Altig et al., 2001] and [Nishiyama and Smetters, 2007] among other). In particular, higher earning individuals tend to have steeper growth in early ages and less steep decline in later years of their lives.

Under these assumptions, individuals are indifferent between the timing of their consumption. Hence, we assume that consumption is constant over their lifecycle. Moreover, throughout this section, we will assume that providing incentives against local deviations is enough, i.e., we use the first order approach. In the Appendix, we provide conditions on fundamentals so that this approach is valid. Under this assumption, the above incentive constraint becomes

\[ U' (\theta) = \int_0^{R(\theta)} \psi \varphi(t, \theta) y(t, \theta)^\gamma \varphi(t, \theta)^\gamma dt - \eta' (\theta) R (\theta) \]  

(4.6)

Hence, the planner’s problem can be rewritten as

\[ \max \int_{\theta}^{\bar{\theta}} U (\theta) dG (\theta) \]  

(4.7)

subject to

\[ c (\theta) - \int_0^{R(\theta)} \left[ \frac{1}{\gamma} \left( \frac{y(t, \theta)}{\varphi(t, \theta)} \right)^\gamma + \eta (\theta) \right] dt = U (\theta) \]

\[ \int_{\theta}^{\bar{\theta}} c (\theta) dF (\theta) \leq \int_{\theta}^{\bar{\theta}} \int_0^{R(\theta)} y(t, \theta) dt dF (\theta) \]

\[ U' (\theta) = \int_0^{R(\theta)} \psi \frac{\varphi(t, \theta)}{\varphi(t, \theta)} y(t, \theta)^\gamma \varphi(t, \theta)^\gamma dt - \eta' (\theta) R (\theta) \]

As noted above, the risk neutrality assumption significantly helps in building intuition and highlight the main economic mechanisms. In particular, we fully characterize income \( y(t, \theta) \) by each individual. Furthermore, under some conditions, we can fully characterize retirement age. The following lemma, characterizes income and labor wedge:

**Lemma 4.2** The solution to the above problem satisfies the following:

1. Income for type \( \theta \) at age \( t \leq R (\theta) \) is given by

\[ y(t, \theta) = \psi \frac{1}{1-\gamma} \left[ 1 + \gamma \frac{G(\theta) - F(\theta)}{f(\theta)} \varphi(t, \theta) \right] \frac{1}{\varphi(t, \theta)^{\frac{\gamma}{\gamma-1}}} \]  

(4.8)
2. The labor wedge is given by

\[ \tau_l(t,\theta) = 1 - \frac{1}{1 + \gamma \frac{G(\theta) - F(\theta)}{f(\theta)} \frac{\varphi(t,\theta)}{\varphi(\theta)}} \]

**Proof.** In the Appendix.

The above lemma is reminiscent of the formula derived by the empirically-driven literature that connects labor distortions to productivity distributions and labor elasticities, as do [Diamond, 1998], [Saez, 2001], and as in [Golosov et al., 2010]. We provide elasticity-based expressions for labor distortions in the presence of both margins of labor supply. In particular, to make this obvious one can rewrite the above formula for labor wedge as

\[ \frac{\tau_l(t,\theta)}{1 - \tau_l(t,\theta)} = \gamma \frac{G(\theta) - F(\theta)}{f(\theta)} \frac{\varphi(t,\theta)}{\varphi(\theta)} \]

which is the version for our baseline environment of the formula provided by the rest of the literature. The formula illustrates that labor wedges are driven by several forces: the elasticity of labor supply, redistributive motives imbedded in the Pareto weights, and by the changes in productivity profiles over lifecycle. In particular, the higher the degree of redistribution the higher the marginal tax rate. Moreover, when agent’s are past their highest productivity level, their labor distortion should increases with age, since \( \frac{\varphi}{\varphi} \) is increasing in \( t \).

A variable of interest for our analysis in this environment is retirement age, \( R(\theta) \), and how it changes with \( \theta \) or life-time income. In what follows, we provide a formula that characterizes retirement age and study examples of productivity profiles and their implications for efficient retirement age patterns. The following lemma shows the formula that characterizes the efficient pattern of retirement ages.

**Lemma 4.3** The retirement age \( R(\theta) \) satisfies the following equation:

\[ \frac{\gamma - 1}{\gamma} y(R(\theta),\theta) = \eta(\theta) - \frac{G(\theta) - F(\theta)}{f(\theta)} \eta'(\theta) \]  \hspace{1cm} (4.9)

where \( y(t,\theta) \) is given by (4.8).
Proof In the Appendix.

Since \( y(t, \theta) \) is known, the above formula pins down the retirement age. Moreover, it helps us characterize whether \( R(\theta) \) is increasing in \( \theta \) or not. In particular, suppose that \( y(t, \theta) \) is increasing in \( \theta \) and that \( \eta(\theta) \) is constant and independent of \( \theta \). Then, we can see that retirement age must be increasing in \( \theta \). This is because \( y(t, \theta) \) is decreasing in \( t \) and \( y(t, \theta) \) is increasing in \( \theta \). Hence an increase in \( \theta \) must be accommodated by an increase in \( R(\theta) \). This would hold when the right hand side (4.9) is decreasing in \( \theta \).

We summarize this discussion in the following proposition.

**Proposition 4.4** Suppose that \( y(t, \theta) \) in (4.8) is weakly increasing in \( \theta \) and that \( \eta(\theta) - \frac{G(\theta) - F(\theta)}{f(\theta)} \eta'(\theta) \) is weakly decreasing in \( \theta \). Then the retirement age \( R(\theta) \) is increasing in \( \theta \). In particular, when \( \eta(\theta) \) is constant, \( R(\theta) \) is increasing in \( \theta \).

Proof. In the Appendix.

To see the intuition for this result, notice that in an economy with without informational frictions – full information model, economic efficiency implies that more productive households should retire at a later age provided that productivity profiles are increasing and fixed cost of work is weakly decreasing in lifetime productivity. With private information, this is not necessarily true. In order to provide incentive for truthful revelation of types, a planner might want to have more productive households retire earlier, i.e., by giving them higher utility through lower working length. However, similar to ***Myerson, it can be shown that the economy with private information is equivalent to a full information economy with modified types; an economy with ‘virtual types’, i.e., types adjusted by their informational rent. Now if the virtual types are so that fixed cost of working is weakly decreasing, then by the same efficiency argument, retirement age should be increasing in productivity.

In the model discussed above, virtual fixed cost of work for an agent of type \( \theta \) is given by \( \eta(\theta) - \frac{G(\theta) - F(\theta)}{f(\theta)} \eta'(\theta) \). To provide a partial intuition for this, consider a small increase – of size \( \varepsilon \) – in retirement age for agents of type \( \theta \). Virtual fixed cost is the
effective utility cost of such a change. Note that this increase requires that the planner changes the utility of all the agents above $\theta$, since it changes the RHS of the incentive constraint (4.6) by $-\eta'(\theta)\varepsilon$. The planner can do so by increasing consumption for all types above $\theta$ by $-\eta'(\theta)\varepsilon$. Hence, the total cost of such a change is given by $\eta'(\theta)\varepsilon[1 - G(\theta) - (1 - F(\theta))] = -\eta'(\theta)\varepsilon[G(\theta) - F(\theta)]$. Therefore, the fixed cost of increasing $R$ per unit of worker of type $\theta$ is given by $\eta'_{\theta}\varepsilon[1 - G(\theta) - (1 - F(\theta))] = -\eta'(\theta)\varepsilon[G(\theta) - F(\theta)]$.

The following example, provides more insight into the implications of the above proposition. That is, an example where we provide sufficient conditions on fundamentals under which $R(\theta)$ is increasing in $\theta$. Suppose that $G(\theta) = F(\theta)^\alpha$ with $0 < \alpha < 1$ and that $\varphi(t, \theta) = \theta\hat{\varphi}(t)$ – parallel productivity profiles. Then, we must have

$$y(t, \theta) = \psi^{\frac{1}{\gamma}} \left[ 1 + \gamma \frac{F(\theta)^\alpha - F(\theta)}{\theta f(\theta)} \right]^{\frac{1}{\gamma}} (\theta\hat{\varphi}(t))^{\frac{\gamma}{\gamma - 1}}$$

as well as

$$\psi^{\frac{1}{\gamma}} \left[ 1 + \gamma \frac{F(\theta)^\alpha - F(\theta)}{\theta f(\theta)} \right]^{\frac{1}{\gamma}} (\theta\hat{\varphi}(R(\theta)))^{\frac{\gamma}{\gamma - 1}} = \frac{\gamma}{\gamma - 1} \left[ \eta(\theta) - \frac{F(\theta)^\alpha - F(\theta)}{f(\theta)} \eta'(\theta) \right]$$

and therefore, $R(\theta)$ is increasing in $\theta$ whenever the following conditions are satisfied:

$$\frac{d}{d\theta} \frac{F(\theta)^\alpha - F(\theta)}{\theta f(\theta)} < 0 \quad (4.10)$$

$$\frac{d}{d\theta} \left[ \eta(\theta) - \frac{F(\theta)^\alpha - F(\theta)}{f(\theta)} \eta'(\theta) \right] < 0 \quad (4.11)$$

The following conditions imply that in the economy with virtual types: 1) virtual productivity profiles are increasing in type, condition (4.10); 2) virtual fixed cost of work are decreasing in type, condition (4.11). Furthermore, it establishes that there are two key determinants of the relationship between retirement age and lifetime earnings, or between $R$ and $\theta$. First, how $y(t, \theta)$ moves with $\theta$ and how (Myerson-like) virtual fixed cost of work $\eta(\theta) - \frac{G(\theta) - F(\theta)}{f(\theta)} \eta'(\theta)$ depends on $\theta$.

### 4.4.2 Labor and retirement distortions

Next, we characterize efficient labor distortions and retirement distortions. In particular, we show how retirement and labor wedges are related to each other. The relationship

---

11 Although this has an effect on total disutility form hours, we ignore that since we are interested in fixed cost of work.
between retirement wedge and labor wedge helps us in characterizing the policy system that implements the efficient retirement age pattern. Our main theoretical result here is to show that the retirement distortion is smaller than the labor distortion. In the appendix, we show this result for the general environment as well.

**Proposition 4.5** Suppose that \( \eta' (\theta) = 0 \). Then the retirement wedge \( \tau_r (\theta) \) is lower than the labor wedge at retirement age, \( \tau_l (R(\theta), \theta) \).

**Proof.** In the Appendix.

While the Appendix contains the proof, the result above follows from the following formula:

\[
\tau_r (\theta) y (R(\theta), \theta) = \frac{1}{\gamma} \tau_l (R(\theta), \theta)y (R(\theta), \theta) - \frac{G(\theta) - F(\theta)}{f(\theta)} \eta' (\theta) \tag{4.12}
\]

The above formula ties labor wedge, retirement wedge and the incentive cost of increasing retirement. For instance, it is clear from 4.12 that when \( \eta' (\theta) = 0 \), retirement wedge is lower than labor wedge.

The intuition for this result can be provided by focusing on the incentive cost of a unit increase in income through an increase in retirement age as opposed to an increase in hours worked. Consider a unit increase in \( y (R(\theta), \theta) \). In addition to the effect that this increase has on resources and the utility of the household of type \( \theta \), it has an effect on the incentive constraint. In particular, it increases by

\[
\gamma \psi \phi (R(\theta), \theta) y (R(\theta), \theta)^{\gamma - 1}
\]

On the other hand, an increase of size \( \frac{1}{y(R(\theta), \theta)} \) in \( R(\theta) \) increases income by a unit\(^{12} \) and increases the RHS of the incentive constraint by \( \psi \phi (R(\theta), \theta) y (R(\theta), \theta)^{\gamma - 1} \). That is the incentive cost of an increase in \( R(\theta) \) is lower than the incentive cost of an increase in \( y (R(\theta), \theta) \) of comparable size. Hence, the distortions to retirement margin should be lower than the distortions to the intensive margin.

---

\(^{12}\) To provide better intuition we use a loose argument here. These perturbations should be interpreted as (1) a change in \( y(t, \theta) \) by 1 unit in an interval \([R(\theta) - \varepsilon, R(\theta)]\) for small \( \varepsilon > 0 \), (2) an increase in \( R(\theta) \) by \( \frac{1}{y(R(\theta), \theta)} \).
Equation (4.12) also implies that when $\eta'(\theta)$ is positive the same equation holds. Moreover, when the slope of $\eta'(\theta)$ is negative and low enough, the retirement wedge is lower than labor wedge.

The above result is helpful in characterizing whether labor income taxes distort retirement decision downward or upward, i.e., whether labor taxes provide additional incentives to retire younger or older. In other words, it helps in showing whether pension benefits should be designed to reward later or earlier retirement above and beyond the labor income tax schedule. As we show in the next section, in plausible cases, the above result would imply that retirement should be rewarded by benefit that increases with age in an actuarially unfair way.

### 4.5 Actuarially unfair pension system

In this section, we analytically study the types of policies we introduced in Section 4.3. Our goal here is the design of a pension system as an integral part of the tax code to implement efficient allocations studied above. We show that pension benefits depend on the age of retirement and, moreover, that the pension system should be designed to be actuarially unfair.

To provide a complete implementation of the constrained optimal allocation, we start from the baseline case studied in the previous section. Here, we show that a tax schedule of the form $\{T(t,y), b(R)\}$ can implement the allocations discussed above, where $T(t,y)$ is the income tax schedule at age $t$ and $b(R)$ is the present value benefits.

We start by constructing the tax and benefits schedule as follows: Consider any incentive compatible allocation \(\{y(t,\theta)\}_{t \leq R(\theta)}, R(\theta), c(\theta)\) with the properties that $y(t, \theta)$ and $R(\theta)$ are both increasing functions of $\theta$ for all $t$. Let $T(t,y)$ be defined as a function that satisfies

$$\theta = \arg \max_{\hat{\theta}} y\left(t, \hat{\theta}\right) - T\left(t, y\left(t, \hat{\theta}\right)\right) - \frac{\psi}{\gamma} \frac{y\left(t, \hat{\theta}\right)}{\varphi(t, \theta)\gamma}\gamma$$ (4.13)

The following lemma shows that this tax function exists and is unique.

---

13 In the appendix, we show that $\eta'(\theta) \leq 0$ is a sufficient condition for the first order approach to work. Hence, when $\eta(\theta)$ is increasing, one should make sure (numerically) that the first order approach is valid.
Lemma 4.6 Suppose that \( y(t, \theta) \) is an increasing function \( \theta \). Then there must exist a function \( T(t, y) \) that satisfies (4.13). Moreover, \( T(t, y) \) is uniquely determined over the interval \([\min_\theta y(t, \theta), y(t, \theta)]\) up to a constant.

Proof. In the Appendix.

The idea for the above lemma is very intuitive. The static incentive compatibility of the allocation \((y(t, \theta) - T(t, y(t, \theta)), y(t, \theta))\) determines the slope of the tax function \( T(t, \cdot) \) with respect to \( y \). Hence, \( T(t, y) \) should be uniquely determined over the mentioned interval up to a constant.

Using the tax function constructed above, we define the benefits. We define the function \( \hat{b}(\theta) \) as

\[
\hat{b}(\theta) = c(\theta) - \int_0^{R(\theta)} [y(t, \theta) - T(t, y(t, \theta))] \, dt \tag{4.14}
\]

Since \( R(\theta) \) is an increasing function of \( \theta \), there must exist an increasing function \( b(R) \) such that \( b(R(\theta)) = \hat{b}(\theta) \). For all \( R \neq R(\theta) \) for some \( \theta \), we set \( b(R) \) equal to big negative number so that agents would not choose those retirement ages. The following proposition shows that facing this tax and pension system, the allocation \( \{y(t, \theta)\}_{t \leq R(\theta)}, R(\theta), c(\theta) \) is a local optimal for a household of type \( \theta \). We relegate the complete proof of optimality to the Appendix.

Proposition 4.7 Consider an incentive compatible allocation

\[
\left( \{y(t, \theta)\}_{t \leq R(\theta)}, R(\theta), c(\theta) \right)_{\theta \in [\underline{\theta}, \bar{\theta}]}
\]

such that \( y(t, \theta) \) and \( R(\theta) \) are both increasing in \( \theta \). Moreover, suppose that \( \eta'(\theta) \leq 0 \). Then the tax function \( T(t, y) \) and the benefit schedule \( b(R) \) constructed in (4.13), and (4.14) locally implements this allocation.

Proof. Given the above tax schedule, a household of type \( \theta \)'s optimization problem is given by

\[
\max_{R, y(t)} \int_0^R [y(t) - T(t, y(t))] \, dt + b(R) - \int_0^R \left[ \frac{\psi}{\gamma} y(t)^\gamma \phi(t, \theta) + \eta(\theta) \right] \, dt
\]
We prove this claim in two steps. First, note that if an agent of \( \theta \) works at age \( t \), he will work to produce an income of \( y(t, \theta) \). This is because of definition of \( T(t, y) \) in (4.13). Now, we show that given this, picking \( R(\theta) \) is locally optimal. Suppose on the contrary that the household chooses \( R(\hat{\theta}) \leq R(\theta) \), then given the definition of \( b \), the utility for the household is given by

\[
\int_0^{R(\hat{\theta})} \left[ y(t, \hat{\theta}) - T(t, y(t, \hat{\theta})) \right] dt - \int_0^{R(\hat{\theta})} \left[ \frac{\psi}{\gamma} \frac{y(t, \hat{\theta})}{\varphi(t, \hat{\theta})^{\gamma}} + \eta(\hat{\theta}) \right] dt + c(\hat{\theta}) - \int_0^{R(\hat{\theta})} \left[ y(t, \hat{\theta}) - T(t, y(t, \hat{\theta})) \right] dt \quad (4.15)
\]

Taking a derivative with respect to \( \hat{\theta} \), we have

\[
\left[ y \left( R(\hat{\theta}), \theta \right) - T \left( R(\hat{\theta}), y(R(\hat{\theta}), \theta) \right) \right] - \frac{\psi y(R(\hat{\theta}), \theta)^\gamma}{\gamma \varphi(R(\hat{\theta}), \theta)^\gamma} - \eta(\hat{\theta}) \right] R'(\hat{\theta})
\]

\[
c'(\hat{\theta}) - \left[ y \left( R(\hat{\theta}), \hat{\theta} \right) - T \left( R(\hat{\theta}), y(R(\hat{\theta}), \hat{\theta}) \right) \right] R'(\hat{\theta})
\]

\[
- \int_0^{R(\hat{\theta})} \frac{\partial}{\partial \theta} y(t, \hat{\theta}) \left[ 1 - \frac{\partial}{\partial y} T(t, y(t, \hat{\theta})) \right] dt
\]

Evaluating the above expression when \( \hat{\theta} = \theta \),

\[
c'(\theta) - \left[ \frac{\psi y(R(\theta), \theta)^\gamma}{\gamma \varphi(R(\theta), \theta)^\gamma} + \eta(\theta) \right] R'(\theta) - \int_0^{R(\theta)} \frac{\partial}{\partial \theta} y(t, \theta) \left[ 1 - \frac{\partial}{\partial y} T(t, y(t, \theta)) \right] dt
\]

and by static incentive compatibility (4.13), the above expression becomes

\[
c'(\theta) - \left[ \frac{\psi y(R(\theta), \theta)^\gamma}{\gamma \varphi(R(\theta), \theta)^\gamma} + \eta(\theta) \right] R'(\theta) - \int_0^{R(\theta)} \psi \frac{y(t, \theta)^{\gamma-1}}{\varphi(t, \theta)^\gamma} \frac{\partial}{\partial \theta} y(t, \theta) dt
\]

which is zero by incentive compatibility of the original allocation. This implies that \( \hat{\theta} = \theta \) is a local extreme point of the function (4.15). In the appendix, we show that the second derivative of (4.15) at \( \hat{\theta} = \theta \) is negative and hence \( \hat{\theta} = \theta \) is the local maximizer of (4.15). Hence, the original allocation locally maximizes the utility of a household of type \( \theta \).

Q.E.D.
Intuitively, the proof of the above proposition shows that the local decision of changing retirement age coincides with the decision whether to lie about one’s productivity type. Since the original allocation is incentive compatible, it is also optimal not to deviate and choose a different allocation of work and retirement ages.

4.6 Quantitative analysis

We now turn to the quantitative study of efficient work and retirement patterns. We use individual earnings and hours data in combination with individual retirement age data to calibrate variants of discrete time models in our general lifecycle environment described in Section 4.2. We calibrate to also match micro estimates of labor supply elasticity at the extensive margin. We simulate efficient work and retirement choices and policies that we analyze analytically above. To assess the importance of any potential differences between simulated efficient retirement patterns and the patterns in the data, we compute resulting welfare gains and total output gains.

Parameters.

For our quantitative study we consider discrete time version of the following functional form of $U(\theta)$:

$$
\int_0^1 e^{-rt} c(t,\theta)^{1-\sigma} \left( \frac{1}{1-\sigma} - 1 \right) dt - \int_0^{R(\theta)} e^{-rt} \left[ \frac{1}{\gamma} \left( \frac{y(t,\theta)}{\varphi(t,\theta)} \right)^\gamma + \eta(\theta) \right] dt
$$

As a benchmark, we set $\sigma = 1$ so that we consider log ($c(t,\theta)$) utility of consumption function. The intensive elasticity parameter, $\gamma$, is set to 3. This implies Frisch elasticity of labor supply equal to $\alpha = 1/(\gamma - 1) = 0.5$, consistent with the evidence in [Chetty, 2011]. We later study how robust the results are by also exploring Frisch elasticity of 0.3 and 3. We also explore risk aversion of 0.5 and 3, or alternatively intertemporal elasticity of substitution equal to 2 and 1/3. Individuals in our quantitative environment are born 25 years old, they experience changes in their productivities over discrete time, and they die all at the same age of 85. Table 4.1 summarizes these parameter choices and the robustness ranges.

Empirical strategy.

Our main sources of individual level data are individual earnings and hours data from the U.S. Panel Study of Income Dynamics (PSID) and individual retirement age data.
Table 4.1: Parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Benchmark Value</th>
<th>Robustness Range</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$ (relative risk aversion)</td>
<td>1</td>
<td>0.5, 1, 3</td>
<td>intertemporal elasticity of substitution: $1/\sigma$</td>
</tr>
<tr>
<td>$\alpha$ (intensive Frisch elasticity)</td>
<td>0.5</td>
<td>0.3, 0.5, 3</td>
<td>$\alpha = 1/(\gamma - 1)$</td>
</tr>
<tr>
<td>$\beta$ (discount factor)</td>
<td>0.9804</td>
<td></td>
<td>discrete time</td>
</tr>
<tr>
<td>$\delta$ (marginal rate of transformation)</td>
<td>1.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>age at $t = 0$</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>age at $t = 1$</td>
<td>85</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

from the RAND files of the Health and Retirement Study (RAND HRS). Our general empirical approach is to treat individuals in the data as optimizing given the existing policies. That is, we observe individual decisions about how much to work and when to retire that are individually optimal given the existing income taxes and pension benefits those individuals faced when they made their decisions. In particular, we take individual retirement age decisions in the data as being individually optimal through the lens of our environment given the existing policies and the estimated productivity profiles. Taking that approach, we back out the unobservable fixed costs of work from the individual optimality conditions and calibrate to match the extensive elasticity estimates.

We start by estimating productivity profiles over lifecycle, $\varphi(t, \theta)$. For the benchmark quantitative case, we group individuals into ten equal sized groups (types) by their average annual labor earnings. Later, we use these types assigned to individual observations as proxies for lifetime earnings deciles. Our main data source for the productivity profiles is individual total labor earnings and total hours data from the PSID. We use the PSID data collection waves from 1990 onward to the latest currently available data wave of 2007 (containing data from 2006). The labor earnings are obtained directly from the PSID waves and are converted to constant 1990 dollars. We consider total labor earnings, which is a sum of a list of variables in the PSID that contain data on salaries and wages, separate bonuses, the labor portion of business income, overtime pay, tips, commissions, professional practice or trade payments, market gardening, additional job
Figure 4.2: Estimated productivity profiles by type over lifecycle, \( \phi(t, \theta) \).
Table 4.2: Summary statistics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Std.Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>retirement age</td>
<td>64.53118</td>
<td>63</td>
<td>6.391109</td>
<td>47</td>
<td>94</td>
</tr>
<tr>
<td>average annual earnings</td>
<td>56,470.39</td>
<td>44,832.22</td>
<td>70,291.19</td>
<td>119,1287</td>
<td>1,994,460</td>
</tr>
</tbody>
</table>

Note: RAND HRS benchmark sample
Number of observations: 2895

income, and other miscellaneous labor income. When using PSID waves, we treat heads of households and their spouses or long-term cohabitants as separate individuals. We restrict the sample to include only individuals with the total labor income of at least $1,000 in 1990 dollars and with at least 250 total hours worked in a year resulting in a sample of 50,624 individuals total from all waves.

We follow a large part of the literature (see, e.g., [Nishiyama and Smetters, 2007] and [Altig et al., 2001]) in using labor income per hour (computed hourly wage as a ratio of total labor earnings and total hours) as a proxy for working ability or productivity assuming that this measure is a sufficient proxy for \( \phi \), which measures the return to effort. In the future versions of the paper we will construct productivities that are directly implied by the data and the individual first-order conditions. The main challenge is then to correctly account for private assets that appear in the individual optimality conditions since our preferences are not without income effects.

We continue assuming that productivity profiles follow a potentially fanning out parametric form described in Section 4.2. In particular, we think of productivity profiles as given by

\[
\varphi(t, \theta) = \theta \varphi(t) t^{\xi \theta}
\]

To provide an interpretation of this functional form, we take logarithm of both sides to obtain

\[
\log \varphi(t, \theta) = \log \theta + \log \varphi(t) + \xi \theta \log t
\]

Here, the first term can be interpreted as lifetime earnings, while the second term represents an age component and the third term is an interaction term between age and lifetime earnings. Consequently, we regress log productivities on lifetime earnings, age, age squared, and interaction terms to estimate empirical productivity profiles. In this
context, fanning out means that for high enough ages $t$, $\varphi(t)$ decreases faster than $t^{\xi\theta}$ increases. Indeed, we observe just that - Figure 4.2 depicts the ten estimated productivity profiles, one for each lifetime earnings decile.\(^{14}\) Productivities of higher types are higher and generally increase faster for younger ages. The declines in productivities in later years of the lifecycle are not as pronounced, especially for higher lifetime earnings deciles.

We also check our results for robustness by instead following closely the approach of [Nishiyama and Smetters, 2007] and grouping individual observations by type and place them into one of seven bins each for a ten year interval of ages - 25-35 years old, 34-45 years old, ..., 74-85 years old (the few remaining individuals older than 85 we put in the last group) - and extrapolate by using shape preserving cubic splines to obtain the

\(^{14}\) The overall shape of these profiles is similar to those obtained in the literature, see, e.g., [Altig et al., 2001].
productivity profiles. Another important check is to supplement our sample from the PSID with the individual observations from the HRS to increase the number of older age observations. The overall patterns of productivity profiles we find stay similar to the ones displayed in Figure 4.2.

The next important piece of empirical evidence is individual retirement ages provided by the HRS. In addition to labor earnings the HRS provides data on individual retirement decisions. Through the lens of our environment, retirement age means zero hours worked (excluding unemployment) from a given age onward. Figure 4.3 presents an unconditional distribution of retirement ages by this definition in our benchmark sample of 2,895 from all waves that we later expand for robustness. Table 4.2 provides summary statistics. We also check that our results are not changed dramatically when we allow for unretirement or re-entry into the labor force.

Figure 4.4: Retirement ages vs. average annual labor earnings.
Figure 4.5: Retirement ages vs. logarithm of average annual labor earnings.
As with the PSID observations above, we classify individuals by their average annual labor earnings deciles. A simple scatter plot reveals the relationship between the retirement age and the labor earnings presented in Figure 4.4. An alternative look at the retirement ages versus log earnings is displayed in Figure 4.5. Using our benchmark definitions of retirement and earnings, the relationship exhibits negative correlation coefficient of $-0.158$ and a regression coefficient $-0.019$ (see the regression line in Figure 4.4). To account for apparent heteroscedasticity, Table 4.3 also reports robust standard error of $0.00669$. This negative (or, at most, flat) relationship appears robust to various changes from how the retirement age is defined (e.g., allowing for coming out of retirement), to how the labor earnings are computed (e.g. considering only individuals who worked full time), and even to including women in the sample. We connect this evidence with the productivity profiles by grouping individuals into deciles by labor earnings. This produces the pattern of retirement ages, for each of the ten types in our benchmark case, shown on the left panel of Figure 4.6.

**Calibration of the fixed costs.**

Given our preferences and the two pieces of empirical evidence above, a productivity profile and a retirement age for a given type, we back out the unobserved fixed cost of work from the individual optimality condition. To see this intuitively, notice that if allocations were undistorted for simplicity, individual optimality would pin down fixed costs of work for a given curvature in the utility of consumption:

$$
 y(R) - \varphi(R) \frac{v(y(R)/\varphi(R))}{v'(y(R)/\varphi(R))} = \frac{\eta}{u'(c)},
$$

which follows directly from the two optimality conditions discussed in Section 4.4. Since the optimality only jointly identifies fixed costs of work and the curvature of utility (on

---

**Table 4.3: Regression results.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std.Err.</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean earnings / 1000</td>
<td>-0.0188595**</td>
<td>0.0066942</td>
<td>-2.82</td>
</tr>
<tr>
<td>constant</td>
<td>67.01545</td>
<td>0.3436527</td>
<td>195.01</td>
</tr>
</tbody>
</table>

Note: ** indicates significance at the 1% level

Number of observations: 2895

F-statistic: 7.94

R-squared: 0.025
Figure 4.6: Empirical weighted average (left panel) and simulated efficient retirement ages (right panel) for the U.S., by lifetime earnings decile.

the right hand side), we calibrate the pattern of fixed costs, $\eta(\theta)$, that so that simulated extensive elasticity of labor supply falls in the range $0.13 - 0.43$ of estimates from the individual studies analyzed in [Chetty et al., 2011]. We follow [Chetty et al., 2011] in emphasizing calibrating the extensive margin elasticity, as well as the intensive one, in environments with meaningfully active both intensive and extensive margins of labor supply. As described above, we explore robustness with CRRA utility of consumption by varying intertemporal elasticity of substitution.

**Results.**

We use the discussed above estimated productivity profiles, calibrated fixed costs, and the calibrated parameters to solve numerically the planning problem. We compute efficient allocations and analyze efficient retirement ages and their effects on welfare and total output.

The computed allocation that results in the benchmark case implies the retirement age pattern displayed on the right panel in Figure 4.6 (as well as summarized earlier in Figure 4.1). Figure 4.7 displays the labor distortions and retirement distortions associated with this allocation. The left panel displays labor distortions that are always positive, lower for low productive types, and generally increase through life. The right panel displays the difference between retirement distortion and labor distortion at the
Table 4.4: Retirement ages for the U.S.

<table>
<thead>
<tr>
<th>Earnings Decile</th>
<th>Retirement Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Empirical Weighted Average</td>
</tr>
<tr>
<td>1st</td>
<td>70.7</td>
</tr>
<tr>
<td>2nd</td>
<td>67.4</td>
</tr>
<tr>
<td>3rd</td>
<td>65.2</td>
</tr>
<tr>
<td>4th</td>
<td>65.1</td>
</tr>
<tr>
<td>5th</td>
<td>64.6</td>
</tr>
<tr>
<td>6th</td>
<td>64.3</td>
</tr>
<tr>
<td>7th</td>
<td>63.0</td>
</tr>
<tr>
<td>8th</td>
<td>62.8</td>
</tr>
<tr>
<td>9th</td>
<td>62.7</td>
</tr>
<tr>
<td>10th</td>
<td>62.8</td>
</tr>
</tbody>
</table>

Note: Empirical ages are computed from RAND HRS. Simulated ages are from the benchmark calibration.

efficient age of retirement. The difference is everywhere negative implying that retirement distortion needs to undo part of the retirement incentives imbedded in the labor distortion, as discussed in Section 4.5.

Figure 4.6 displays a quantitative result largely robust across calibrations - individuals with higher lifetime earning in the U.S. should retire older than they do now (represented by the dashed line) and, importantly, older than less productive workers (represented by a positively sloped solid line). We find in our benchmark calibration that in the optimum, the highest productivity types retire at 69.5, whereas in the data their average retirement age is 62.8. Individuals with lower lifetime earnings retire younger than they do now, as well as younger than their more productive peers. In particular, the lowest productivity types retire in the optimum at 62.2 years compared to 69.5 for the highest productivity types. This pattern of retirement ages is in sharp contrast with the one found in the current individual data for the U.S., where average retirement age displays a predominantly decreasing pattern as a function of lifetime earnings. Table 4.4 summarizes these differences for all earnings deciles.

Our quantitative study also allows us to measure and decompose welfare gains and total output gains associated with inducing efficient retirement age distribution. We find that compared to the allocations with the existing system, providing efficient
Figure 4.7: Labor distortions (left panel) and retirement distortions (right panel).

incentives for both work and retirement results in large welfare gains across calibrations of between 1 and 5 percent in annual consumption equivalent. Perhaps more surprisingly, it also results in a small but positive change in total output across calibrations of up to 1 percent. The result about the increase in total output is in line with the analysis in [Golosov and Tsyvinski, 2006] as well as follows in spirit earlier contributions of [Diamond and Mirrlees, ] and even [Diamond and Mirrlees, 1986], although taking a more empirically driven approach. Note that at the same time, most modern studies of efficient redistributive policies largely result in increased distortions improving welfare but generally sacrificing total output (see, e.g., [Fukushima, 2010], [Farhi and Werning, 2010a], [Golosov et al., 2010], [Weinzierl, 2011]).

We also find that the increase in total output results from the meaningfully active intensive and extensive margins of labor supply. That is, even though increasing standard distortions of the intensive margin leads to output losses in favor of redistribution and welfare gains, the additional policy instrument of distorting the retirement decision proves powerful enough to overcompensate by inducing more productive individuals to work more years and thus produce more.
4.7 Conclusion

This paper theoretically and quantitatively studies the efficient design of the pension system as an integral part of the tax code when both intensive and extensive labor margins are active. We analytically characterize Pareto efficient policies and derive efficient work and retirement age patterns and show that, under plausible conditions, efficient retirement age increases with lifetime earnings. We show that this pattern is implemented by pension benefits that depend on the age of retirement. Moreover, we show that this requires a pension system that is designed to be actuarially unfair. Using individual earnings and retirement data for the U.S. and, importantly, intensive and extensive labor elasticities, we calibrate and simulate the policy models to generate robust implications: first, it is efficient for individuals with higher lifetime earning to retire older than they do in the data, and second, older than less productive workers do. This is in sharp contrast with what is currently observed in the data. To assess the importance of this disparity, we quantify welfare and total output gains from implementing efficient work and retirement patterns. The main economic message of the paper is perhaps that distorting individual retirement decisions provides a novel and powerful policy tool, capable of overcompensating output losses from standard distortionary redistributive policies.
Chapter 5

Adverse Selection, Reputation, and Sudden Collapses in Secondary Loan Market

5.1 Introduction

Following the sharp decline in the volume of new issuances in the U.S. secondary loan market in the fall of 2007, policymakers argued that the market was not functioning normally and proposed and carried out a variety of policy interventions intended to restore the normal functioning of this market. Here we present evidence on sudden collapses and motivated by that evidence, construct a model in which new issuances in the secondary loan market abruptly collapse. This collapse, in our model, is associated with an increase in inefficiency. We also argue that reductions in the value of the collateral used to secure the underlying loans are particularly likely to trigger sudden collapses associated with increased inefficiency. Since sudden collapses are associated with increased inefficiency, our model is consistent with policymakers’ views that the market was functioning poorly. We use this model to analyze proposed and actual policy interventions and argue that these interventions typically do not remedy the inefficiency associated with the market collapse.

In our model, the main economic function of the secondary loan market is to allocate
originated loans to institutions that have a *comparative advantage* in holding and managing the loans. This economic function is disrupted by informational frictions. In our model, loan originators differ in their ability to originate high-quality loans. The originators are better informed about their ability to generate high-quality loans than are potential purchasers. This informational friction creates an *adverse selection* problem. The focus of our analysis is to examine the extent to which *reputational* considerations ameliorate or intensify the adverse selection problem in these markets. In order to analyze these reputational considerations, we develop a dynamic adverse selection model of the secondary loan market.

Our main finding is that our model has fragile outcomes in which sudden collapses in the volume of new issuances in secondary loan markets are associated with increased inefficiency. We say that outcomes are fragile if the model has multiple equilibria or if a large number of originators change their decisions in response to small changes in aggregate fundamentals.

In terms of fragility as multiplicity, we show that our baseline dynamic adverse selection model with reputation has multiple equilibria for a range of reputation levels. In one of these equilibria, labeled the *positive reputational equilibrium*, high-quality loan originators have incentives to sell at a current loss in order to improve their reputations and command higher prices for future loans. In the other equilibrium, labeled the *negative reputational equilibrium*, loan originators who sell their loans are perceived by future buyers to have low-quality loans. These perceptions induce high-quality loan originators to hold on to their loans. Since low-quality originators always sell their loans, the volume of new issuances is larger in the positive reputational equilibrium than in the negative reputational equilibrium. Clearly, with multiple equilibria sunspot like shocks can generate sudden collapses. We show that the positive equilibrium Pareto dominates the negative equilibrium for a range of reputation levels. In this sense, sudden collapses are associated with increased inefficiency.

Although the multiplicity of equilibria has the attractive feature that it implies that the model can be consistent with observations of sudden collapses, such multiplicity makes it difficult to conduct policy analysis. We propose a refinement adapted from the global games literature (see [Carlsson and Van Damme, 1993] and [Morris and Shin, 2003]). Our refinement is also motivated by the idea that sudden collapses in the volume of new...
issuances in loan markets are associated with falls in the value of the collateral that supports the underlying loans. These considerations lead us to add aggregate shocks to collateral values and to assume that the collateral value is observed with an arbitrarily small error.

We show that shocks to collateral values make the outcomes of our model consistent with our second notion of fragility, namely, a large fraction of loan originators choose to change their decisions on whether to sell or hold their loans in response to small changes in collateral values. In this sense, reductions in collateral values can induce sudden collapses in the volume of new issuances for the market as a whole.

Both adverse selection and the dynamics induced by reputation acquisition play central roles in generating sudden collapses from small changes in collateral values. A simple way of seeing the role of adverse selection is to note that the version of our model with symmetrically informed originators and buyers does not produce sudden collapses in new issuances. With asymmetrically informed agents, originators with high reputations receive higher prices for their loans and are therefore more willing to sell their loans. We show that a fall in collateral values makes high-quality originators who were close to being indifferent about selling versus holding to hold. Small changes in collateral values can induce a large number of originators to switch to holding from selling only if they are all close to the point of indifference. In a static model, we have no reason to expect that the distribution of originators by reputation levels will be concentrated close to the indifference point.

In a dynamic model with learning by market participants, we argue that originators' reputations are likely to be clustered. The reason is that in models like ours, the reputation levels of high-quality originators have an upward trend over time, resulting in the reputation levels of many high-quality originators tending to become similar in the long run. We show that in an infinitely repeated version of our model, the long run or invariant distribution of reputation levels displays significant clustering. This clustering in turn implies that small changes in fundamentals can lead a large number of originators to change their decisions when the fundamentals are close to the point of indifference. A related result is that small changes in collateral values, when these values are far away from the point of indifference, do not lead to large changes in the volume of new issuances.
We have argued that our model is consistent with abrupt collapses in secondary loan markets and with the widespread view among policymakers that such abrupt collapses were associated with sharp increases in the inefficiency of the operation of such markets. In the wake of the 2007 collapse of secondary loan markets, policymakers proposed a variety of programs intended to remedy inefficiencies in the market for securitized assets. Some of these programs, such as the proposed Public-Private Partnership and TALF, were implemented at least in part. The TALF program allows participants to purchase securitized assets by borrowing from the Federal Reserve and using the assets as collateral. We use our model to evaluate the effects of various policies. In terms of purchase policies, we show that if the purchase price is set at or below the level that prevails in the positive reputational equilibrium, the equilibrium outcomes do not change and in this sense the policy is ineffective. If the purchase price is set at a sufficiently high level, the policy implies that the government makes negative profits.

We also analyze policies that change the time path of interest rates. We show that temporary decreases in interest rates worsen the adverse selection problem. Interestingly, anticipated decreases in interest rates in the future can have beneficial current effects by reducing the range of reputations over which the economy has multiple equilibria.

5.1.1 Related Literature

Our work here is related to an extensive literature on adverse selection in asset markets, such as [Myers and Majluf, 1984], [Glosten and Milgrom, 1985], [Kyle, 1985], and [Garleanu and Pedersen, 2004] as well as to the related securitization literature, specifically, the work of [DeMarzo and Duffie, 1999] and [DeMarzo, 2005]. See also [Eisfeldt, 2004], [Kurlat, 2009], [Guerrieri et al., 2010] and [Guerrieri and Shimer, 2011] for analyses of adverse selection in dynamic environments. We add to this literature by analyzing how reputational incentives affect adverse selection problems.

Our assumption that buyers have less information concerning the loan quality of a bank is in line with a descriptive literature that argues that secondary loan markets feature adverse selection (see, for example, the work of [Dewatripont and Tirole, 1994],...
Also, a growing literature provides data on the presence of adverse selection in asset markets. For example, [Ivashina, 2009] finds evidence of adverse selection in the market for syndicated loans. [Downing et al., 2009] find that loans that banks held on their balance sheets yielded more on average relative to similar loans which they securitized and sold. [Drucker and Mayer, 2008] argue that underwriters of prime mortgage-backed securities are better informed than buyers and present evidence that these underwriters exploit their superior information when trading in the secondary market. Specifically, the tranches that such underwriters avoid bidding on exhibit much worse than average ex post performance than the tranches that they do bid on.

A recent paper by [Elul, 2011] presents evidence that is consistent with our model. [Elul, 2011] shows that returns on securitized loans and loans held by originators were similar before 2006 and that returns on securitized loans were lower than returns on comparable loans after 2006. This evidence is consistent with our model in the following sense. Our model implies that when collateral values underlying loans are relatively high, most high-quality banks with high costs of managing the loans choose to sell their loans; but when collateral values are relatively low, such banks choose to hold their loans. Before 2006, land values were rising, so it seems reasonable to suppose that collateral values were relatively high. After 2006, land values stopped rising and in some cases fell, so it seems reasonable to suppose that collateral values were lower than they had been.

Finally, [Mian and Sufi, 2009] present evidence that securitized loans were more likely to default than nonsecuritized loans. This evidence is consistent with our model in the sense that for all realizations of the aggregate shock, the default rate of securitized loans is at least as high as that of held loans, and for some realizations the default rate of securitized loans is higher than that of held loans.

Our work is also related to the literature on reputation. [Kreps and Wilson, 1982] and [Milgrom and Roberts, 1982] argue that equilibrium outcomes are better in models with reputational incentives than in models without them. In the banking literature, [Diamond, 1989] develops this argument. More recently, [Mailath and Samuelson, 2001] analyze the role of reputational incentives in infinite horizon economies and provide conditions under which they can improve outcomes. In contrast, [Ely and Välimäki, 2003]
and [Ely et al., 2008] describe models in which reputational incentives can worsen outcomes. Our work here combines the results in this literature by showing that reputational models can have multiple equilibria. In some of these equilibria, reputational incentives can generate better outcomes; in others, they can generate worse. Furthermore, using techniques from the global games literature, we develop a refinement that produces a unique, fragile equilibrium. Perhaps the work most closely related to ours is that of [Ordoñez, 2008]. An important difference between our work and his is that our model has equilibria that are worse than the static equilibrium, so that reputational incentives can lead to outcomes that are ex post less efficient than those in a model without these incentives.

Our analysis of policy is closely related to recent work by [Philippon and Skreta, 2011] who analyze a variety of policies in a model with adverse selection. The main difference with our work is that we focus on the incentives induced by reputation, whereas they analyze a static model.

5.2 Evidence on Sudden Collapses

Here we present evidence on sudden collapses in the market for new issuances of asset-backed securities. Figure 1 displays the volume of new issuances of asset-backed securities for various categories from the first quarter of 2000 to the first quarter of 2009. The figure shows that the total volume of new issuances of asset-backed securities rose from roughly $50 billion in the first quarter of 2000 to roughly $300 billion in the fourth quarter of 2006. The volume of new issuances fell abruptly to roughly $100 billion in the third quarter of 2007 and then fell again to near zero in roughly the fourth quarter of 2008. The figure also shows similar large fluctuations in the volume of new issuances for each category.

[Ivashina and Scharfstein, 2008] document a similar pattern for new issues of syndicated loans. Figure 1, Panel A of their paper shows that syndicated lending rose from roughly $300 billion in the first quarter of 2000 to roughly $700 billion in the second quarter of 2007. This lending declined sharply thereafter and fell to roughly $100 billion by the third quarter of 2008.

The reduction in the volume of new issuances in the secondary market roughly
Figure 5.1: New Issuance of Asset-Backed Securities (Source: JP Morgan Chase)
coincided with a reduction in collateral values. One way of seeing this coincidence is to consider the Case-Shiller home price index\(^1\). This index stopped growing in late 2006 and declined through 2007. The coincidence of the reduction in the volume of new issuances and the reduction in collateral values is consistent with our model.

[White, 2009] has argued that in the 1920s, the United States experienced a boom-bust cycle in securitization of real estate assets that was similar to its recent experience. Figure 2 displays the change in the outstanding stock in real estate bonds in the 1920s based on data in [Carter and Sutch, 2006]. Such bonds were issued against single large commercial mortgages or pools of commercial or real estate mortgages and were publicly traded. To make this data comparable to more recent data, we scale the data from the 1920s by nominal GDP in 2009. Specifically, we multiply the change in the nominal stock of outstanding debt in each year by the ratio of the nominal GDP in 2009 to that in the relevant year. This figure shows that the changes in the stock rose dramatically from essentially 0 in 1919 to an average of $145 billion in the period from 1925 to 1928. The market then collapsed sharply, and changes in the stock fell to roughly $50 billion in 1929. Such large changes in the stock are likely to have been associated with similar large changes in the volume of new issuances.

### 5.3 Reputation in a Secondary Loan Market Model

We develop a finite horizon model of the secondary loan market and use the model to demonstrate how adverse selection and reputation interact to yield abrupt collapses with increased inefficiency. We show that for every history, the last period of the model has a unique equilibrium which we use to construct equilibria in previous periods. We show that equilibria of the multi period model typically exhibit dynamic coordination problems in the sense that for a wide range of parameters, the game has multiple equilibria. Although reputation is always valued, loan originators choose different actions across the different equilibria based on the different inferences future buyers draw from the current actions of originators.

\(^1\) Available at http://www.standardandpoors.com/indices
**Figure 5.2: Change in Stock of Real Estate Bonds, 1920-1930**

Note: Data is annual change in real estate bonds divided by Nominal GDP at relevant year multiplied by Nominal GDP 2009.

5.3.1 Static Model: A Unique Equilibrium

We start with a static model which should be interpreted as describing the last period of a finite horizon model. We show that the static model has a unique equilibrium in which the equilibrium outcomes depend on the informed originator’s reputation.

**Agents.** The model has three types of agents: a loan originator referred to as a bank, a continuum of buyers, and a continuum of lenders. All agents are risk neutral.

The bank is endowed with a risky loan indexed by $\pi$. The loan can also be thought of more generally as an investment opportunity such as a project, a mortgage, or an asset-backed security. Each loan requires $q$ units of inputs, which represents the loan’s size. A loan of type $\pi$ yields a return of $v = \bar{v}$ if the borrower does not default and a return of $v = v$ if the borrower does default. We refer to $v$ as the **collateral value** of the loan. The probability that the borrower does not default is denoted by $\pi$. For the analysis in this section, we normalize $\bar{v}$ to 0. Later, when we allow for aggregate shocks and introduce our refinement, we will allow $v$ to be a random variable, possibly different from zero. We assume that $\pi \in \{\pi, \bar{\pi}\}$ with $\pi < \bar{\pi}$. We refer to a bank that has a loan of type $\bar{\pi}$ as a **high-quality bank** and one with a loan of type $\pi$ as a **low-quality bank**.

We assume that $\pi \bar{v} \geq q$ so that each loan has positive net present value if sold.

The bank either can sell the loan in a secondary market or can hold the loan. Selling the loan at a price $p$ yields a payoff to the bank of $p - q$. The purchaser of the loan is entitled to the resulting return. If the bank chooses to hold the loan, it must borrow $q$ from lenders to finance the loan and repay $q(1 + r)$ at the end of the period, where $r$ is the within-period interest rate paid to lenders. We allow $r$ to be positive or negative in order to examine the effects of various policy experiments described below. If the bank holds the loan, it is entitled to the return from its projects; however, the bank then incurs a cost of holding the loan, $c$, in addition to the cost of repaying its debt, $q(1 + r)$.

Besides the quality of its loan, the bank is indexed by a cost type, which represents the costs, relative to the marketplace, that the bank incurs when it holds the loan to maturity. We intend the cost of the loan to represent funding liquidity costs, servicing costs, renegotiation costs in the event of a loan default, and costs associated with holding a loan that may be correlated in a particular way with the rest of the bank’s portfolio, among other potential factors. We assume that $c \in \{\xi, \bar{c}\}$ with $\xi < -qr < 0 < \bar{c}$. We refer to a bank of type $\bar{c}$ as a **high-cost bank** and a bank of type $\xi$ as a **low-cost bank**.
We normalize the cost of holding and managing the loan for the market to be zero.

Hence, the model has four types of banks: \((\pi, c) \in \{\bar{\pi}, \pi\} \times \{\bar{c}, c\}\). We refer to the different types of banks, \((\bar{\pi}, \bar{c}), (\bar{\pi}, c), (\pi, \bar{c}), (\pi, c)\), as HHI, HL, LH, LL banks, respectively.

**Timing of the Static Game.** We formalize the interactions in this economy as an extensive form game with the following timing. First, nature draws the quality and cost types of the bank. Then, buyers simultaneously offer a price, \(p\), to purchase the loan. Finally, the bank sells the loan to one of the buyers or holds the loan to maturity.

We assume that, as perceived by buyers and lenders, the bank has quality type \(\bar{\pi}\) with probability \(\mu_2\) and quality type \(\pi\) with probability \(1 - \mu_2\). (The subscript 2 on the probability is meant to indicate that these are the beliefs of lenders associated with the second period of our two-period model described below.) Following the work of [Kreps and Wilson, 1982] and [Milgrom and Roberts, 1982], we refer to \(\mu_2\) as the bank’s reputation. Also, buyers believe that the bank has cost type \(c\) with probability \(\alpha\) and cost type \(\bar{c}\) with probability \(1 - \alpha\). The cost and quality types are independently drawn.

**Strategy and Equilibrium.** A strategy for the bank consists of a decision of whether to sell or hold its loan as a function of prices offered by buyers, and which buyer to sell to if the bank chooses to sell. Clearly, the bank will choose the buyer offering the highest price if the bank decides to sell, so we suppress this aspect of the bank’s strategy. Let \(a = 1\) denote the decision of the bank to sell the loan, and let \(a = 0\) denote the decision to hold the loan. A strategy for the bank is a function \(a(\cdot)\) that maps the highest offered price, \(p\), into a decision of whether to sell or hold the loan. The payoffs to a type \((\pi, c)\) bank are given by

\[
w_2(a|p, \pi, c) = a(p - q) + (1 - a)[\pi \bar{v} - q(1 + r) - c].
\]

A strategy for a buyer consists of the choice of a price to offer a bank for its loan. The payoffs to a buyer with an accepted price \(p\) and a strategy \(a_2(\cdot|\pi, c)\) for each type of bank is

\[
u_2(p|a_2) = E_{\pi,c}[v|a_2(p|\pi, c) = 1] - p.
\]

Since buyers move simultaneously, they engage in a form of Bertrand competition, so that the price is equal to the expected return on the loan.
A (pure strategy) Perfect Bayesian Equilibrium is a price $p_2$ and a strategy for each bank type, $a_2(\cdot|\pi,c)$, such that for all $p$, each bank type chooses the optimal loan decision and buyers offer the highest price that yields a payoff of 0; i.e., $p_2 = \max\{p|u_2(p|a_2) = 0\}$.

With full information, when the bank’s type is known by buyers, under the assumption that $c < -qr$, it is easy to show that the high cost bank sells its loan and a low cost bank holds its loan. In particular, the decision of whether to sell or hold the loan does not depend on the quality type of the bank. The reason is that the return on the loan, ignoring the holding cost, is the same for both the bank and the buyers. Notice that the equilibrium allocation under full information is ex post efficient. Low-cost banks have a comparative advantage (over buyers) in holding loans to maturity, while buyers have a comparative advantage over high-cost banks. The full information equilibrium allocates loans to agents with a comparative advantage in holding and managing the loan.

Next, we show that the private information model has a unique equilibrium. For expositional simplicity, we focus on the decisions of the high-quality, high-cost bank (HH) and restrict the strategy sets of the low-cost bank as well as the low-quality, high-cost bank (LH). Specifically, we assume that HL and LL banks hold their loans and the LH bank sells its loan. In Appendix D.2, we show that if $c$ is sufficiently negative, the assumed strategies for these three types of banks are indeed optimal.

To show uniqueness of equilibrium, we show that the HH bank sells its loan for reputation levels higher than a critical threshold, $\mu^*_2$, and holds its loan otherwise. To see this result, note that facing price $p$, the HH bank sells its loan if and only if

$$p - q \geq \bar{\pi} \bar{v} - q(1 + r) - \bar{c}. \quad (5.1)$$

Bertrand competition among buyers implies that buyers must make zero profits so that any candidate equilibrium price at which the HH bank sells must satisfy the following equality:

$$\hat{p}(\mu_2) := [\mu_2 \bar{\pi} + (1 - \mu_2) \bar{\pi}] \bar{v}. \quad (5.2)$$

To determine the threshold, $\mu^*_2$, above which the equilibrium involves the HH selling its loan, substitute from (5.2) into (5.1) and find the thresholds for $\mu_2$ at which (5.1) holds with equality. We obtain

$$\mu^*_2 = 1 - \frac{qr + \bar{c}}{(\bar{\pi} - \pi) \bar{v}}. \quad (5.3)$$
To see that when $\mu_2 \geq \mu_2^*$ the equilibrium must have the HH bank selling, note that if $\mu_2 \geq \mu_2^*$ and the offered price is below $\hat{p}(\mu_2)$, one of the buyers can deviate and offer a price just below $\hat{p}(\mu_2)$ and induce the HH bank to sell. This deviation yields strictly positive profits. For reputation levels below $\mu_2^*$, the HH bank holds even if offered $\hat{p}(\mu_2)$. Thus the equilibrium must have the HH bank holding at these reputation levels and only the low-quality bank selling, so that the equilibrium price must satisfy

$$p = \bar{\pi} \bar{v}.$$  

(5.4)

We use this characterization of the static equilibrium to calculate the payoffs associated with a given level of reputation $\mu_2$ at the beginning of the period before a bank’s cost type is realized. These payoff calculations play a crucial role in our dynamic game. They are given by

$$V_2(\mu_2) = \begin{cases} \bar{\pi} \bar{v} - q(1 + r) - Ec, & \mu_2 < \mu_2^* \\ (1 - \alpha) \{[\mu_2 \bar{\pi} + (1 - \mu_2)\bar{\pi}]\bar{v} - q\} + \alpha[\bar{\pi} \bar{v} - q(1 + r) - c], & \mu_2 \geq \mu_2^*. \end{cases}$$

(5.5)

Similarly, we can define the value of the equilibrium for a low-quality bank:

$$W_2(\mu_2) = \begin{cases} (1 - \alpha) \{\bar{\pi} \bar{v} - q\} + \alpha[\bar{\pi} \bar{v} - q(1 + r) - c], & \mu_2 < \mu_2^* \\ (1 - \alpha) \{[\mu_2 \bar{\pi} + (1 - \mu_2)\bar{\pi}]\bar{v} - q\} + \alpha[\bar{\pi} \bar{v} - q(1 + r) - c], & \mu_2 \geq \mu_2^*. \end{cases}$$

It is clear that $V_2$ is weakly increasing and convex in $\mu_2$. We have proved the following proposition.

**Proposition 5.1** If $\bar{\pi} \bar{v} > q$ and $qr + c > 0$, then for any $\mu \in [0, 1]$, the static model has a unique equilibrium. Let $\mu_2^*$ be defined by (5.3). For $\mu_2 < \mu_2^*$, the HH bank holds its loan and for $\mu_2 \geq \mu_2^*$, the HH bank sells its loan.

Note that we have modeled buyers as behaving strategically. This modeling choice plays an important role in ensuring that the static game has a unique equilibrium. Suppose that rather than modeling buyers as behaving strategically, we had instead simply required that market prices satisfy a zero profit condition. One rationale for this requirement is that buyers take prices as given and choose how many loans to buy as in a competitive equilibrium. It is easy to show that with this requirement the economy has multiple equilibria in the static game if $\mu_2 \geq \mu_2^*$. One of these equilibria corresponds
to the unique equilibrium of our game. In the other equilibrium, the buyers offer a price of $\pi \bar{v}$. At this offered price, the HH bank holds its loan and only the low-quality, high-cost bank sells its loan. We find multiplicity of this kind unattractive in our model because obvious bilateral gains to trade are not being exploited. Each of the buyers has a strong incentive to offer a price slightly below $[\mu_2 \pi + (1 - \mu_2) \bar{v}] \bar{v}$. At this offered price, the HH bank strictly prefers to sell, and the buyer making such an offer makes strictly positive profits. In our formulation, with strategic behavior by the buyers, this low price outcome cannot be an equilibrium.

Although we prefer our strategic formulation, we emphasize that our results that reputational incentives induce multiplicity do not rely on the static game having a unique equilibrium. We chose a formulation in which the static game has a unique equilibrium in order to argue that reputational incentives by themselves can induce multiplicity.

5.3.2 Two-Period Benchmark Model

Consider now a two-period repetition of our static game in which the bank’s quality type is the same in both periods. We assume that the bank’s second-period payoffs are discounted at rate $\beta$. In period 1, a continuum of buyers who are present in the market for only one period choose to offer prices for loans sold in that period. In period 2, a new set of buyers each offer prices for loans sold in that period. This new set of buyers observes whether the bank sold or held its loan in the previous period, and, if the bank sold its loan, buyers observe the realized value of the loan. If the loan is held, we assume that period 2 buyers do not observe the realized value of the loan.

The assumption that period 2 buyers receive no information about the realized value of the loan is convenient but not essential in generating multiplicity of equilibria. Our multiplicity results go through if period 2 buyers receive a sufficiently noisy signal of the realized value of the loan. The critical assumption in generating multiplicity is that the market receives more precise information about the value of the loan if it is sold than if it is held. We think this assumption is natural in that market participants typically receive information only about aggregate returns to bank portfolios and do not receive information on the returns to specific assets. Banks typically hold a variety of assets in their portfolios, some of which can be securitized and others which cannot. In such a setting, the information investors receive about returns on specific assets is typically
not as precise if a bank holds an asset as it would be if the bank sold the asset.

The timing of the game is an extension of that described in the static game. As in that game, at the beginning of period 1, nature draws the bank’s quality and cost type. We assume that the bank’s quality type is fixed for both periods. At the beginning of period 2, nature draws a new cost type for the bank. In any period, the bank’s quality and cost types are unknown to buyers. The timing within each period is the same as in the static game. We also assume that the returns to successful loans, \( v = \bar{v} \), and to unsuccessful loans, \( v = 0 \), are the same in both periods.

In order to define an equilibrium in this repeated game, we must develop language that will allow us to describe how second-period buyers update their beliefs about the bank’s type based on observations from period 1. To do so, we let the public history at the beginning of period 2 be denoted by \( \theta_1 \), where \( \theta_1 \in \{h, s0, s\bar{v}\} \) where \( \theta_1 = h \) denotes that the bank held its loan in period 1, \( \theta_1 = s0 \) denotes that the bank sold its loan and the loan paid off \( v = 0 \), and \( \theta_1 = s\bar{v} \) denotes that the bank sold its loan and the loan paid off \( v = \bar{v} \).

As in the static game, we focus on the strategic incentives of the HH bank and restrict the strategy sets of the low-cost bank as well as the low-quality, high-cost bank. Specifically, we assume that the low-cost bank must hold its loan and the LH bank must sell its loan. A strategy for the high-quality, high-cost bank is now given by a pair of functions, \( a_1(p_1) \) representing the decision in period 1 and \( a_2(p_2, \theta_1) \) representing the loan decision in period 2 if the bank realizes a high cost in period 2, as a function of offered prices.

Consider next how the buyers in the last period update their beliefs about the bank’s type. This updating depends through Bayes’ rule on the prior belief of the buyers, the loan decision of the bank and the loan return realization if the bank sold, as well as on the first-period strategies chosen by the HH bank and period 1 buyers. From Bayes’ rule, these posterior probabilities are given by

\[
\mu_2(\mu_1, \theta_1 = h, a_1(\cdot), p_1) = \frac{\mu_1 (\alpha + (1 - \alpha)(1 - a_1(p_1)))}{\mu_1 (\alpha + (1 - \alpha)(1 - a_1(p_1))) + (1 - \mu_1)\alpha} \quad (5.6)
\]

\[
\mu_2(\mu_1, \theta_1 = s\bar{v}, a_1(\cdot), p_1) = \frac{\mu_1 a_1(p_1)(1 - \alpha)\bar{\pi}}{\mu_1 a_1(p_1)(1 - \alpha)\bar{\pi} + (1 - \mu_1)(1 - \alpha)\bar{\pi}} \quad (5.7)
\]

\[
\mu_2(\mu_1, \theta_1 = s0, a_1(\cdot), p_1) = \frac{\mu_1 a_1(p_1)(1 - \alpha)(1 - \bar{\pi})}{\mu_1 a_1(p_1)(1 - \alpha)(1 - \bar{\pi}) + (1 - \mu_1)(1 - \alpha)(1 - \bar{\pi})}. \quad (5.8)
\]
For notational convenience, we suppress the dependence on strategies and priors and let $\mu_h$ denote the posterior associated with the bank holding its loan, and $\mu_{s\bar{v}}$ and $\mu_{s0}$ denote the posteriors associated with selling and yielding a high or low return.

Given the updating rules, the period 1 payoffs for the HH bank are given by

$$w_1(a|p) = a[p - q + \beta (\bar{\pi}V_2(\mu_{s\bar{v}}) + (1 - \bar{\pi})V_2(\mu_{s0}))]$$

$$+ (1 - a)[(\bar{\pi}\bar{v} - q(1 + r) - \bar{c}) + \beta V_2(\mu_h)]$$

where $\mu_h, \mu_{s\bar{v}},$ and $\mu_{s0}$ are given by equations (5.6), (5.7), and (5.8). Buyers’ payoffs associated with an accepted price, $p$, in period $t$ are given by

$$u_t(p|\pi, a_t, \mu_t) = \mu_t(1 - \alpha)a_t(p)\bar{\pi} + (1 - \mu_t)(1 - \alpha)\bar{\pi} + (1 - \mu_t)(1 - \alpha)\bar{v} - p.$$ 

A Perfect Bayesian Equilibrium is a first-period price, $p_1$, a first-period loan decision for the high-quality, high-cost bank $a_1(\cdot)$ that maps accepted prices into loan decisions, updating rules $\mu_h, \mu_{s\bar{v}}, \mu_{s0}$ that map observations on loan decisions into posterior beliefs, a second-period price, $p_2$, that maps second-period beliefs into prices, and a second-period loan decision $a_2(\cdot)$ that maps accepted prices and histories into loan decisions such that (i) for all $p$, the HH bank chooses the optimal action in period 1 so that $w_1(a_1(p)|p) \geq \max_{a'} w_1(a|p)$, (ii) for all $p$, the HH bank chooses the optimal action in period 2 so that $w_2(a_1(p)|p) \geq \max_{a'} w_2(a|p)$, (iii) the first-period price, $p_1$, satisfies $p_1 \in \max\{p|u_1(p|a_1) = 0\}$, (iv) the second-period price, $p_2$, satisfies $p_2 \in \max\{p|u_2(p|a_2) = 0\}$, (v) the updating rules, $\mu_h, \mu_{s\bar{v}}, \mu_{s0}$, satisfy Bayes’ rule, namely, (5.6), (5.7), and (5.8).

To show that our model has mutliplicity of equilibria, we begin by showing that the game has two (pure strategy) equilibria when prior beliefs in period 1, $\mu_1$, are equal to the static threshold, $\mu^*_2$. Continuity of payoffs then implies that the game has two equilibria in an interval around the static threshold.

In one equilibrium, labeled the positive reputational equilibrium, the HH bank chooses to sell its loan in period 1. To see that such a choice is part of an equilibrium, note that in this case, the period 1 price is given by

$$\hat{p}(\mu_2) := [\mu_2\bar{\pi} + (1 - \mu_2)\bar{\pi}]\bar{v}. \quad (5.9)$$
Given this price, selling is optimal if the following incentive constraint is satisfied:

\[
(\mu_1 \bar{\pi} + (1 - \mu_1) \bar{\pi}) \bar{v} - q + \beta (\bar{\pi} V_2(\mu_{s\theta}) + (1 - \bar{\pi}) V_2(\mu_{s0})) \geq \bar{\pi} \bar{v} - q (1 + r) - \bar{c} + \beta V_2(\mu_h)
\]

(5.10)

where the posterior beliefs are obtained from (5.6) through (5.8) substituting \(a_1(p_1) = 1\) so that

\[
\mu_h = \mu_1, \quad \mu_{s\theta} = \frac{\mu_1 \bar{\pi}}{\mu_1 \bar{\pi} + (1 - \mu_1) \bar{\pi}}, \quad \text{and} \quad \mu_{s0} = \frac{\mu_1 (1 - \bar{\pi})}{\mu_1 (1 - \bar{\pi}) + (1 - \mu_1)(1 - \bar{\pi})}.
\]

(5.11)

Notice from (5.11) that if a bank holds the loan, the posterior beliefs are unchanged. The reason is that the beliefs of period 2 buyers is that low cost banks of both qualities hold their loans and period 2 buyers receive no information about the return to the loan if it is held.

We show that at \(\mu_2^*\), the incentive constraint (5.10) holds as a strict inequality. Note that using (5.9), (5.10) evaluated at \(\mu_2^*\) can be written as

\[
\beta (\bar{\pi} V_2(\mu_{s\theta}) + (1 - \bar{\pi}) V_2(\mu_{s0})) \geq \beta V_2(\mu_h)
\]

Further, from the updating rules for posterior beliefs, (5.11), \(\mu_{s\theta} > \mu_h = \mu_2^* > \mu_{s0}\). Hence, using the second period payoffs from (5.5), it follows that \(V_2(\mu_{s\theta}) > V_2(\mu_{s0}) = V_2(\mu_h)\) so that (5.10) holds as a strict inequality at \(\mu_2^*\). Thus, the HH bank has a strict incentive to sell. The reason for this strict incentive is that the worst outcome associated with selling is that the loan is unsuccessful and this payoff is the same as that associated with holding the loan. If the loan is successful, the HH bank’s payoff is strictly higher than the payoff to holding the loan. Not surprisingly, this result suggests that for reputation levels in an interval around \(\mu_2^*\), given beliefs that the HH bank sells in period 1, the bank finds it optimal to do so, and hence the model has a positive equilibrium for an interval around \(\mu_2^*\).

In the second type of equilibrium, labeled the negative reputational equilibrium, the HH bank chooses to hold its loan. In this case the equilibrium price is given by \(\bar{\pi} \bar{v}\) using (5.4). A bank holds its loan if and only if

\[
(\mu_1 \bar{\pi} + (1 - \mu_1) \bar{\pi}) \bar{v} - q + \beta (\bar{\pi} V_2(\mu_{s\theta}) + (1 - \bar{\pi}) V_2(\mu_{s0})) \leq \bar{\pi} \bar{v} - q (1 + r) - \bar{c} + \beta V_2(\mu_h),
\]

(5.12)
where
\[ \mu_h = \frac{\mu_1}{\mu_1 + (1 - \mu_1)\alpha}, \] and \( \mu_{s0} = \mu_{s0} = 0. \] (5.13)

Note that in the negative equilibrium, only low quality banks sell, and uninformed agents assign a posterior reputation of zero if the bank sells and rationally disregard the information from the realized value of the loans. Note also that if a bank chooses to hold its loan, buyers perceive that it is more likely to be a high quality bank and the posterior belief rises.

The argument that at \( \mu^{\ast}_2 \), the incentive constraint (5.12) holds as a strict inequality parallels that of the positive equilibrium. Using the updating rules in (5.13), it follows that \( \mu_{s0} = \mu_{s0} < \mu_2^* < \mu_h \). Hence, using the second period payoffs given in (5.5), it follows that \( V_2(\mu_{s0}) = V_2(\mu_{s0}) = V_2(\mu_2^*) < V_2(\mu_h) \) so that the incentive constraint (5.12) holds as a strict inequality at \( \mu_2^* \). This result suggests that for reputation levels in an interval around \( \mu_2^* \), given beliefs that the HH bank holds in period 1, the bank finds it optimal to do so, and hence the model has a negative equilibrium.

Continuity of payoffs implies that (5.10) and (5.12) hold as strict inequalities in some interval of prior beliefs around \( \mu_2^* \) so that our model has multiple equilibria in this interval. In Appendix D.2, we show that our model has unique equilibria outside this interval under the assumption that \( \beta(1 - \alpha) \leq 1 \).

**Proposition 5.2 (Multiplicity of Equilibria)** Suppose \( 0 < \mu_2^* < 1 \). Then, there exist \( \underline{\mu} \) and \( \bar{\mu} \) with \( \underline{\mu} < \mu_2^* < \bar{\mu} \) such that if \( \mu_1 \in [\underline{\mu}, \bar{\mu}] \), the model has two equilibria: in one the HH bank sells its loan, and in the other the HH bank holds its loan in the first period.

In the proposition, we have shown that introducing reputation as a device for mitigating lemons problems results in equilibrium multiplicity, that is, reputation can be both a blessing and a curse. The game has a positive reputational equilibrium in which, encouraged by reputational incentives, banks with a high-quality asset sell their asset. In this equilibrium, reputation helps sustain market activity in a market that would be illiquid without reputational incentives. The game also has a negative reputational equilibrium in which reputational incentives discourage selling and banks with a high-quality asset hold on to their asset. In this equilibrium, reputation helps depress market activity in a market that would be liquid without reputational incentives.
In terms of the relationship to the literature on reputation, our model nests features of the model in [Mailath and Samuelson, 2001] and [Ordoñez, 2008] as well as that of [Ely and Välimäki, 2003]. In [Mailath and Samuelson, 2001] and [Ordoñez, 2008], strategic types are good and want to separate from nonstrategic types, although in [Mailath and Samuelson, 2001] reputation generally fails to deliver this type of equilibria. Nevertheless, in their environments, there is no long-run reputational loss from good behavior. [Ely and Välimäki, 2003] share the property that strategic types are good and want to separate; however, the structure of learning is such that good behavior never implies long-run positive reputational gains, and therefore reputational incentives exacerbate bad behavior in equilibrium.

5.3.3 Sudden Collapses and Increased Inefficiency

In this section, we study the efficiency properties of the positive and negative reputational equilibria. We provide sufficient conditions under which the positive reputational equilibrium Pareto dominates the negative reputational equilibrium in the sense of interim utility (see [Holmström and Myerson, 1983]), and sufficient conditions under which the positive equilibrium dominates the negative equilibrium in the sense of ex ante utility. In this sense, sudden collapses of trade volume in our model due to switches between equilibria are associated with increased inefficiency.

In order to develop these sufficient conditions, suppose that $\mu_1 \in [\mu, \mu^*_2]$. Consider the welfare of the HH bank. Let $\mu^*_h$ denote the posterior beliefs in the negative equilibrium respectively, conditional on future buyers observing a hold decision by a bank in the first period. Suppose $\mu^*_h$ is less than the static cutoff, $\mu^*_2$. (In Appendix D.2, we show that $\mu^*_2 < \left(\frac{\bar{\pi}}{\pi\alpha - \bar{\pi}} + \beta \bar{\pi}\right) / (1 + \beta \bar{\pi}(1 - \alpha))$ and $\mu_1$ close to $\mu$ is a sufficient condition for $\mu^*_h$ to be less than or equal to $\mu^*_2$.) Since $\mu^*_h$ is less than the static cutoff, $\mu^*_2$, using the form of second period payoffs (5.5), it follows that the present value of payoffs in the negative equilibrium is given by the right side of the incentive constraint in the positive equilibrium, (5.10). The left side of (5.10) is the equilibrium payoff in the positive equilibrium. Clearly, the payoff for the HH bank is higher in the positive equilibrium than it is in the negative equilibrium.

Consider next the low quality, high cost, or LH bank. This bank sells in both equilibria in the first period, but receives a higher price in the positive equilibrium than
in the negative equilibrium. In terms of continuation values, note that the reputation level in the negative equilibrium falls to zero and is positive in the positive equilibrium. It follows that this bank is strictly better off in the positive equilibrium than in the negative equilibrium. Since $\mu_h^n \leq \mu_h^p$, the continuation values for low-cost types is the same in the two equilibria, and since they are holding in the first period, their utility levels are the same. Since buyers make zero profits in both equilibria, we have established the following proposition.

**Proposition 5.3** Suppose that $0 < \mu_2^* < \frac{\beta \bar{\pi} - \bar{\pi} \alpha}{1 + \beta \bar{\pi} (1 - \alpha)}$ and that $\mu_2^* < 1$. Then for all $\mu_1$ in some neighborhood of $\mu$, the utility level for each type of bank and the buyers in the positive equilibrium is at least as large as the utility level for the corresponding type of bank and the buyers in the negative equilibrium.

If $\mu_h^n > \mu_h^p$, one can show that the utility level of the low-cost types is lower in the positive reputational equilibrium than in the negative reputational equilibria. Hence, the two equilibria are not comparable in interim utility terms. However, under appropriate sufficient conditions, the positive equilibrium yields a higher ex ante utility than the negative equilibrium. Consider the allocations in the two equilibria in the first period. The only difference in allocations is that in the positive equilibrium the high-quality, high-cost type sells, whereas in the negative equilibrium this type holds. Thus, the difference in ex ante utility (or social surplus) in the first period between the two equilibria is given by $(1 - \alpha)\mu(qr + \bar{c})$. Clearly, first-period utility is higher in the positive equilibrium than in the negative equilibrium. However, in the second period social surplus is higher in the negative equilibrium than in the positive equilibrium because the high-cost types always sell in the negative equilibrium, whereas in the positive equilibrium they hold the asset some fraction of the time – when the signal quality is bad in the first period or after a hold decision in the first period. Therefore, the change in social surplus in the second period is given by $-\mu(1 - \alpha)((1 - \alpha)(1 - \bar{\pi}) + \alpha)(qr + \bar{c})$. Thus, the overall change in the social surplus is given by

$$\mu(1 - \alpha)(1 - \beta(1 - \bar{\pi}(1 - \alpha)))(qr + \bar{c}).$$

Clearly, this overall change is positive if and only if $\beta(1 - \bar{\pi}(1 - \alpha)) < 1$. We have established the following proposition.
Proposition 5.4 Suppose that $\beta(1 - \bar{\pi}(1 - \alpha)) < 1$. Then the ex-ante utility of the bank is higher in the positive reputational equilibrium than in the negative reputational equilibrium and the ex-ante utility of the buyers is the same in the two equilibria.

5.4 Aggregate Shocks and Uniqueness

In this section, we show that with two perturbations our model has a unique equilibrium which is fragile. The perturbations add aggregate shocks to collateral values and assume that past aggregate shocks are imperfectly observed. With these perturbations, we show that the model has a unique equilibrium in which small fluctuations in collateral values in a critical region lead to sudden collapses in the volume of trade.

Adding aggregate shocks with imperfect observability ensures that our model has a unique equilibrium and is, in this sense, a type of refinement. This device is in the spirit of the refinement literature on static coordination games (see, for example, [Carlsson and Van Damme, 1993], [Morris and Shin, 2003]). One reason for using such a refinement is to compare outcomes under various policies. Uniqueness is desirable because such comparison is difficult in models with multiple equilibria. Furthermore, we want to develop a well-defined notion of fragility. In many macroeconomic environments with multiple equilibria, small shocks to the environment can cause sudden changes in behavior. Without a selection device, multiplicity leads to a lack of discipline on how equilibrium behavior changes in response to shocks. Techniques adapted from the literature on coordination games, however, enable us to impose such discipline. We show that small aggregate shocks to collateral values near a critical range induce sudden collapses in trade while similar small shocks far from the critical range do not induce significant changes in the volume of trade.

We assume that aggregate shocks affect collateral values. Specifically, we assume that the collateral value, $v$, is affected by an aggregate shock common for all banks. One example of the situation in which collateral values are subject to aggregate shocks is a mortgage on a residential or a commercial property. The value of real estate is often subject to aggregate shocks.

Consider the following model with aggregate shocks and imperfect observability. In each period $t = 1, 2$, an aggregate shock $v_t \sim F_t(v_t)$ is drawn. These shocks are drawn
independently across periods. Banks and buyers at the beginning of each period observe a noisy signal of \( v_t \) given by \( v_t = v_t + \sigma \varepsilon_t \), where \( \varepsilon_t \sim G(\varepsilon_t) \) with \( E[\varepsilon_t] = 0 \) is i.i.d. across periods. When \( \sigma > 0 \) the aggregate shock is imperfectly observed. We assume that \( F_t \) and \( G \) have full support over \( \mathbb{R} \).

We assume that the distributions \( F_1 \) and \( G \) satisfy a monotone likelihood property. To develop this property note that, when \( \sigma > 0 \), the updating rules for the signal of the aggregate shock are given by

\[
\Pr(v_1 \leq \hat{v}_1 | v_1) = \Pr(v_1 + \sigma \varepsilon_1 \leq \hat{v}_1) = G\left(\frac{\hat{v}_1 - v_1}{\sigma}\right)
\]

\[
\Pr(v_1 \leq \hat{v}_1 | v_1) = \frac{\int_{-\infty}^{\hat{v}_1} f_1(v) g\left(\frac{v_1 - v}{\sigma}\right) dv}{\int_{-\infty}^{\infty} f_1(v) g\left(\frac{v_1 - v}{\sigma}\right) dv} = H(\hat{v}_1 | v_1)
\]

**Assumption 5.5 (Monotone Likelihood Ratio)** The posterior belief function \( H(v_1 | v_1) \) is a decreasing function of \( v_1 \).

This assumption implies that when the signal, \( v_1 \), about the shock is high, the value of the shock, \( v_1 \), is likely to be high. Straightforward algebra can be used to show that this assumption is satisfied if a monotone likelihood ratio property on \( g \) holds, namely, that for any \( v_1 > v'_1 \), \( g(v_1 - v_1)/g(v'_1 - v_1) \) is increasing in \( v_1 \).

The timing of the game is as follows: (i) At the beginning of each period \( t \), agents observe the aggregate shock in the previous period \( v_{t-1} \). Buyers do not observe previous period signals \( v_{t-1} \) or the market price \( p_{t-1} \). (We believe that our uniqueness result goes through if future buyers receive a noisy signal about previous prices.), (ii) The new aggregate state \( v_t \) is drawn, the bank and current period buyers do not observe the current state, \( v_t \), but they do observe the noisy signal, \( v_t \), (iii) Buyers offer prices, \( v_t \), (iv) The bank decides whether to sell or hold.

With aggregate shocks and perfect observability, \( \sigma = 0 \), it is immediate that a version of Proposition 5.2 applies and the two period model has multiple equilibria.

To establish uniqueness in our two period model with imperfect observability we begin from the last period. We will show that in the last period, the unique equilibrium is characterized by a cutoff threshold \( \mu^*_2(v_2) \) such that banks with reputation levels above \( \mu^*_2(v_2) \) sell their loans and banks below this threshold hold their loans and a fall in \( v_2 \) raises \( \mu^*_2(v_2) \). In this sense, a fall in collateral values worsens the adverse selection
problem. To see this result, note that an HH bank sells its loan if and only if
\[ \hat{p}(\mu_2; v_2) - q \geq \bar{\pi} \bar{v} + (1 - \bar{\pi})E[v_2|v_2] - q(1 + r) - \bar{c}, \] (5.14)
where
\[ \hat{p}(\mu_2; v_2) := [\mu_2 \bar{\pi} + (1 - \mu_2) \bar{\pi}] \bar{v} + [\mu_2(1 - \bar{\pi}) + (1 - \mu_2)(1 - \bar{\pi})] E[v_2|v_2]. \] (5.15)
Substituting for \( \hat{p}(\mu_2; v_2) \) from (5.15) into (5.14) and noting that \( E[v_2|v_2] = v_2 \), we obtain that the threshold reputation at which the HH bank is just indifferent between holding and selling is given by
\[ \mu_2^*(v_2) = 1 - \frac{qr + \bar{c}}{(\bar{\pi} - \bar{\pi})(\bar{v} - v_2)}, \] (5.16)
whenever the right hand side of (5.16) is between zero and one and at the appropriate extreme points otherwise. Clearly \( \mu_2^*(v_2) \) is decreasing in \( v_2 \). We summarize this discussion in the following proposition.

**Proposition 5.6** In the second period, given a reputation level \( \mu_2 \) and a default value signal \( v_2 \), there is a unique equilibrium outcome in which the HH bank's decision is to sell if \( \mu_2 \geq \mu_2^*(v_2) \) and to hold otherwise, where
\[ \mu_2^*(v_2) = \max \left\{ \min \left\{ 1 - \frac{qr + \bar{c}}{(\bar{\pi} - \bar{\pi})(\bar{v} - v_2)}, 1 \right\}, 0 \right\}. \]

Given this characterization of the second period equilibrium, we can calculate the payoff to the HH bank before the aggregate shock (as well as the second period signal) or the cost type is realized for every value of reputation at the beginning of the second period. These payoffs are given by
\[ V_2(\mu_2) = \int \int \hat{V}_2(\mu_2, v_2) dG \left( \frac{v_2 - \bar{v}}{\sigma} \right) dF_2(v_2) \] (5.17)
where
\[ \hat{V}_2(\mu_2, v_2) = \alpha [\bar{\pi} \bar{v} - q(1 + r) - \bar{c}] + (1 - \alpha) \max \{ \hat{p}(\mu_2; v_2) - q, \bar{\pi} \bar{v} + (1 - \bar{\pi}) v_2 - q(1 + r) - \bar{c} \}. \]
Next, we use the characterization of the payoffs given in (5.17) to prove that the two period model has a unique equilibrium. Proving that the perturbed game has a unique equilibrium is easiest when \( F_1 \) is an improper uniform distribution, \( U[-\infty, \infty] \). In Section 5.4.2, we prove uniqueness as \( \sigma \to 0 \), when \( F_1 \) is a proper distribution.

We have the following proposition:
Proposition 5.7 For each $\sigma > 0$ and $V_2(\mu_2)$ given by (5.17), the game with uniform improper priors has a unique equilibrium in which in period 1, HH bank’s action is characterized by a cutoff $v_1^*(\sigma) \in \mathbb{R}$ above which the HH bank sells and below which the HH bank holds.

We prove this proposition using a method similar to [Carlsson and Van Damme, 1993]. We begin by restricting attention to switching strategies in which the bank sells for all default values above a threshold and holds for all default values below that threshold. We show that the game has a unique equilibrium in switching strategies. We then prove that the equilibrium switching strategy is the only strategy that survives iterated elimination of strictly dominant strategies so that we have a unique equilibrium.

The intuition for the iterated elimination argument is as follows. Note that we can define equilibrium as a strategy for the bank in period 1, and a belief – about the bank’s action in period 1 – by period 2 buyers used for Bayesian updating. In equilibrium beliefs have to coincide with strategies. Obviously reputational incentives depend on future buyers’ beliefs. When $v_1$ is very large, independent of future buyers’ beliefs, an HH bank sells the asset. Similarly, when $v_1$ is very low, an HH bank holds on to the asset, independent of future beliefs. This argument establishes two bounds $\hat{v}_1 > \tilde{v}_1$, such that any equilibrium strategy must prescribe a sale for $v_1$ higher than $\hat{v}_1$ and holding for $v_1$ lower than $\tilde{v}_1$. This result means that the set of beliefs by future buyers have to satisfy the same property. Limiting the set of beliefs puts tighter upper and lower bounds on reputational incentives, which in turn implies new bounds $\hat{v}_2 > \tilde{v}_2$. We show that iterating in this manner implies that the bounds $\hat{v}_n$ and $\tilde{v}_n$ converge to a common limit.

Here we sketch the key steps of the proof and leave the details to Appendix D.1.

5.4.1 Outline of Proof with Improper Priors

1. Unique Equilibrium in Switching Strategies: We begin by restricting attention to switching strategies of the form:

$$d_k(v_1) = \begin{cases} 1 & v_1 \geq k \\ 0 & v_1 < k, \end{cases}$$
where $k$ represents the switching point. We characterize the best response of the HH bank when future buyers use $d_k$ to form their posteriors over the bank’s type. To do so, we use Bayes rule. Consider an arbitrary belief $\hat{a}_1(\cdot)$ by period 2 buyers about the HH bank’s period 1 action. Based on the observed history and signal $v_1$, Bayes rule implies the following updating formulas:

$$
\mu_{sg}(v_1; \hat{a}_1) = \frac{\mu_1 \pi \int \hat{a}_1(v_1) dG \left( \frac{v_1 - v_1}{\sigma} \right)}{\mu_1 \pi \int \hat{a}_1(v_1) dG \left( \frac{v_1 - v_1}{\sigma} \right) + (1 - \mu_1) \pi}
$$

$$
\mu_{sd}(v_1; \hat{a}_1) = \frac{\mu_1 (1 - \pi) \int \hat{a}_1(v_1) dG \left( \frac{v_1 - v_1}{\sigma} \right) + (1 - \mu_1)(1 - \pi)}{\mu_1 (1 - \pi) \int \hat{a}_1(v_1) dG \left( \frac{v_1 - v_1}{\sigma} \right) + (1 - \mu_1)(1 - \pi)}
$$

$$
\mu_h(v_1; \hat{a}_1) = \frac{\mu_1 \left( (1 - \alpha) \int [1 - \hat{a}_1(v_1)] dG \left( \frac{v_1 - v_1}{\sigma} \right) + \alpha \right)}{\mu_1 \left( (1 - \alpha) \int [1 - \hat{a}_1(v_1)] dG \left( \frac{v_1 - v_1}{\sigma} \right) + \alpha \right) + (1 - \mu_1) \alpha}.
$$

For switching strategies, these formulas simplify to

$$
\mu_{sg}(v_1; d_k) = \frac{\mu_1 \pi \left[ 1 - G \left( \frac{k - v_1}{\sigma} \right) \right]}{\mu_1 \pi \left[ 1 - G \left( \frac{k - v_1}{\sigma} \right) \right] + (1 - \mu_1) \pi}
$$

(5.18)

$$
\mu_{sd}(v_1; d_k) = \frac{\mu_1 (1 - \pi) \left[ 1 - G \left( \frac{k - v_1}{\sigma} \right) \right] + (1 - \mu_1)(1 - \pi)}{\mu_1 (1 - \pi) \left[ 1 - G \left( \frac{k - v_1}{\sigma} \right) \right] + (1 - \mu_1)(1 - \pi)}
$$

$$
\mu_h(v_1; d_k) = \frac{\mu_1 \left[ (1 - \alpha)G \left( \frac{k - v_1}{\sigma} \right) + \alpha \right]}{\mu_1 \left[ (1 - \alpha)G \left( \frac{k - v_1}{\sigma} \right) + \alpha \right] + (1 - \mu_1) \alpha}.
$$

Next, given any belief $\hat{a}_1$ and noting that with improper priors $H(v_1|v_1) = G \left( \frac{v_1 - v_1}{\sigma} \right)$, we define the gain from reputation as

$$
\Delta(v_1; \hat{a}_1) = 
\beta \int \left[ \pi V_2(\mu_{sg}(v_1; \hat{a}_1)) + (1 - \pi) V_2(\mu_{sd}(v_1; \hat{a}_1)) - V_2(\mu_h(v_1; \hat{a}_1)) \right] dG \left( \frac{v_1 - v_1}{\sigma} \right)
$$

In Appendix D.1 we prove the following Lemma, which characterizes the gain from reputation for general strategies and switching strategies.
Lemma 5.8 The gain from reputation $\Delta(v_1; \hat{a}_1)$ is uniformly bounded and strictly increasing in $\hat{a}_1$ according to a point-wise ordering on beliefs. In particular, if $\hat{a}_1$ is a switching strategy, $d_k$, then $\Delta(v_1; d_k)$ is strictly decreasing in $k$. Moreover, when $\hat{a}_1$ is a switching strategy, $\Delta(v_1; \hat{a}_1)$ is strictly increasing in $v_1$.

Facing a switching strategy belief of future buyers, $d_k$, clearly, the HH bank sells if and only if

$$\hat{p}(\mu_1; v_1) - q + \Delta(v_1; d_k) \geq \bar{\pi}v + (1 - \bar{\pi})v_1 - q(1 + r) - \bar{c}. \quad (5.19)$$

Note that the value of selling, given by the left side of (5.19), is increasing in $v_1$ and its partial derivative with respect to $v_1$ is at least the derivative of $\hat{p}(\mu_1; v_1)$, given by $\mu_1(1 - \bar{\pi}) + (1 - \mu_1)\bar{\pi}$. The value of holding, given by the right side of (5.19), is increasing in $v_1$ and its derivative is $1 - \bar{\pi}$. Since the derivative for the value of selling is greater than the value of holding, there exists a unique solution, $b(k)$, that solves the equation

$$\hat{p}(\mu_1; b(k)) - q + \Delta(b(k); d_k) = \bar{\pi}v + (1 - \bar{\pi})b(k) - q(1 + r) - \bar{c}.$$ 

Hence, the best response of the HH bank to a switching strategy belief of future buyers, $d_k$, is a switching strategy, $b(k)$, in which the bank sells for all returns above $b(k)$ and holds for all return values below $b(k)$. An equilibrium in switching strategies must be a fixed point of the above equation, so an equilibrium switching point, $k^*$, satisfies

$$\hat{p}(\mu_1; k^*) - q + \Delta(k^*; d_{k^*}) = \bar{\pi}v + (1 - \bar{\pi})k^* - q(1 + r) - \bar{c}.$$

In Appendix D.1, we prove the following lemma.

Lemma 5.9 The best response function $b(k)$ has a unique fixed point $k^*$ which is globally stable.

Hence, the game with switching strategies has a unique equilibrium.

2. Restriction to Switching Strategies Is without Loss of Generality: We follow [Morris and Shin, 2003] in showing that the restriction to switching strategies is without loss of generality. We do so by showing that regardless of future buyers’ belief functions, the bank has a dominant strategy for extreme values of default values. Consider two numbers $\hat{v} < \bar{v}$. We define an extreme monotone strategy to be a strategy...
that calls for selling when $v_1 \geq \tilde{v}$ and holding for $v_1 \leq \tilde{v}$. We define $A_{\tilde{v}, \bar{v}}$ to be the set of such strategies. Notice that $A_{-\infty, \infty}$ is the set of all strategies. Define the best response set operator on a subset of beliefs, $A$, as

$$BR(A) = \{a_1 | \exists \hat{a}_1 : a_1(v_1) = 1 \Leftrightarrow \hat{p}(\mu_1; v_1) - q + \Delta(v_1; \hat{a}_1) \geq \tilde{\pi}v + (1 - \tilde{\pi})v_1 - q(1 + r) - \bar{c}\}.$$

We show that there exist bounds $\hat{v}^0 < \tilde{v}^0$ such that the HH bank holds for $v_1 \leq \hat{v}^0$ and it sells the asset for $v_1 \geq \tilde{v}^0$, independent of future buyers’ belief function $\hat{a}_1$. That is,

$$\forall \hat{a}_1, v_1 \geq \hat{v}^0; \hat{p}(\mu_1; v_1) - q + \Delta(v_1; \hat{a}_1) \geq \tilde{\pi}v + (1 - \tilde{\pi})v_1 - q(1 + r) - \bar{c} \quad (5.20)$$

$$\forall \hat{a}_1, v_1 \leq \tilde{v}^0; \hat{p}(\mu_1; v_1) - q + \Delta(v_1; \hat{a}_1) \leq \tilde{\pi}v + (1 - \tilde{\pi})v_1 - q(1 + r) - \bar{c}.$$

Using the result from Lemma (5.8) that $\Delta(v_1; \hat{a}_1)$ is uniformly bounded in (5.20), it follows that these bounds exist. We have established that any equilibrium strategy must be an extreme monotone strategy with cutoffs $\hat{v}^0 < \tilde{v}^0$. That is,

$$BR(A_{-\infty, \infty}) \subseteq A_{\hat{v}^0, \tilde{v}^0}.$$

Thus, we can restrict attention to extreme monotone strategies without loss of generality.

Next, we show that the best response set operator is decreasing in the sense that it induces a best response set, which is a strict subset of any arbitrary set of extreme monotone beliefs. Repeatedly applying this operator induces a decreasing sequence of sets, which converges to a unique equilibrium.

To show that the best response set operator is decreasing, we show that for any $\hat{v} < \tilde{v}$, $BR(A_{\hat{v}, \tilde{v}}) \subseteq A_{\hat{b}(\hat{v}), \bar{b}(\tilde{v})} \subset A_{\hat{v}, \tilde{v}}$. Since $\Delta(v_1; \hat{a}_1)$ is increasing in $\hat{a}_1$, for all $\hat{a}_1 \in A_{\hat{v}, \tilde{v}}$ we have

$$\hat{p}(\mu_1; v_1) - q + \Delta(v_1; \hat{d}_\hat{v}) \leq \hat{p}(\mu_1; v_1) - q + \Delta(v_1; \hat{a}_1) \leq \hat{p}(\mu_1; v_1) - q + \Delta(v_1; \hat{d}_\tilde{v})$$

because $\hat{a}_1$ first order stochastically dominates $\hat{d}_\hat{v}$ and is dominated by $\hat{d}_\tilde{v}$. This result implies that

$$\tilde{\pi}v + (1 - \tilde{\pi})v_1 - q(1 + r) - \bar{c} \geq \hat{p}(\mu_1; v_1) - q + \Delta(v_1; \hat{a}_1)$$

if

$$\tilde{\pi}v + (1 - \tilde{\pi})v_1 - q(1 + r) - \bar{c} \geq \hat{p}(\mu_1; v_1) - q + \Delta(v_1; \hat{d}_\tilde{v}).$$
This result implies that if $a_1$ is the best response to $\hat{a}_1$, then

$$\forall v_1 < b(\hat{v}), \quad a_1(v_1) = 0.$$ 

Similarly, we can show that the best response to $\hat{a}_1$ must satisfy $a_1(v_1) = 1$ for all $v_1 \geq b(\tilde{v})$. We have proved that $BR(A_{\hat{v},\tilde{v}}) \subseteq A_{b(\hat{v}),b(\tilde{v})}$. Since $b(k)$ is globally stable, $A_{b(\hat{v}),b(\tilde{v})} \subseteq A_{\hat{v},\tilde{v}}$ so that $BR(A_{\hat{v},\tilde{v}}) \subseteq A_{b(\hat{v}),b(\tilde{v})} \subset A_{\hat{v},\tilde{v}}$. Finally, because $b(k)$ has a unique fixed point, $A^n_{b(\hat{v}),b(\tilde{v})}$ converges to $A_k^* = \{d_k^*\}$ so that $BR^n(A_{-\infty,\infty})$ also converges to $\{d_k^*\}$.

5.4.2 Uniqueness Result with Proper Priors

In this section, we provide a characterization of equilibria in the limiting perturbed game with general proper priors. In particular, we prove that in the perturbed game as $\sigma \to 0$, the set of period 1 equilibrium strategies converges to a unique strategy. We use the method of Laplacian beliefs introduced by [Frankel et al., 2003] and reviewed by [Morris and Shin, 2003] to prove our uniqueness result. In fact, we show that the game described above is equivalent to a game discussed by [Morris and Shin, 2003]. We then use their result to prove the following theorem. The proof is in Appendix D.1.

**Theorem 5.10** Given the value function $V_2(\mu_2)$ given by (5.17), as $\sigma \to 0$ the set of first period equilibrium strategies in the game with proper priors converges to a unique strategy by the HHI bank in which the bank sells if $v_1 \geq v_1^*$ and holds if $v_1 < v_1^*$ where $v_1^*$ satisfies

$$\hat{\rho}(\mu_1; v_1^*) - q + \beta \int_0^1 \left[ \bar{\pi} V_2(\mu_{sg}(l)) + (1 - \bar{\pi}) V_2(\mu_{sd}(l)) - V_2(\mu_{h}(l)) \right] dl = \bar{\pi} \hat{v} + (1 - \bar{\pi}) v_1^* - q(1 + r) - \bar{c}$$

and

$$\hat{\mu}_{sg}(l) = \frac{\mu_1 \bar{\pi} l}{\mu_1 \bar{\pi} l + (1 - \mu_1) \bar{\pi}}$$

$$\hat{\mu}_{sd}(l) = \frac{\mu_1 (1 - \bar{\pi}) l}{\mu_1 (1 - \bar{\pi}) l + (1 - \mu_1) (1 - \bar{\pi})}$$

$$\hat{\mu}_{h}(l) = \frac{\mu_1 [(1 - \alpha) (1 - l) + \alpha]}{\mu_1 [(1 - \alpha) (1 - l) + \alpha] + (1 - \mu_1) \alpha}.$$
5.5 The Multi-Period Model

In this section, we extend the model to many periods. The qualitative properties of the model are very similar to the model with two periods. In particular, we show that the game with noisy signals has a unique equilibrium in the limit as the observation error converges to zero.

The extension of the model to multi periods is as follows: time is discrete and \( t = 1, \ldots, T, T < \infty \). The bank’s quality type is drawn at the beginning of period 1. The bank’s cost type is drawn independently over time and is independent of the quality type. The collateral value \( v_t \) is drawn from a distribution function \( F(v_t) \) and is independent across periods. A new set of buyers arrives each period and lives only for that period. The information structure of the game is as in the two-period model in Section 5.4. In each period before trading occurs, all agents in the economy observe \( v_t = v_t + \sigma_t \varepsilon_t \) where \( \varepsilon_t \) is i.i.d. and distributed according to \( G(\varepsilon) \). They do not, however, observe \( v_t \). Given this information, the agents trade in the market. After the trade, the collateral value \( v_t \) becomes public information. Previous prices are not observed by current buyers. Based on observables, agents update their beliefs at the end of period \( t \).

In Appendix D.1, we recursively construct the payoff of the HH bank and its equilibrium strategy and prove the following proposition.

**Proposition 5.11** Suppose that for some period \( t + 1 \) and for any \( \mu_{t+1} \), the multiperiod model has a unique equilibrium with payoff for the HH bank given by \( V_{t+1}(\mu_{t+1}) \). If \( V_{t+1}(\mu_{t+1}) \) is increasing in \( \mu_{t+1} \), there is a unique equilibrium strategy in period \( t \) as \( \sigma_t \to 0 \) for all \( \mu_t \). The equilibrium strategy for the HH bank in period \( t \) is given by a cutoff strategy in which the HH bank sells if \( v_t \geq v^*_t(\mu_t) \) and holds if \( v_t < v^*_t(\mu_t) \) where \( v^*_t(\mu_t) \) satisfies the following equation

\[
\hat{p}(\mu_t; v^*_t) - q + \beta \int_0^1 [\bar{\pi} V_{t+1}(\hat{\mu}_{sg}(l)) + (1 - \bar{\pi}) V_{t+1}(\hat{\mu}_{sb}(l)) - V_{t+1}(\hat{\mu}_h(l))] \, dl
\]

\[
= \bar{\pi} v + (1 - \bar{\pi}) v^*_t - q(1 + r) - \bar{c}.
\]

Furthermore, the model has a unique equilibrium in the last period.

This proposition shows that the finite horizon version of the model has a unique equilibrium under the assumption that the value function is increasing in the reputation...
of the bank. This assumption can be replaced by assumptions on parameter values. One such assumption is that $\alpha$, the probability that the bank’s cost type is low, is sufficiently small. In the numerical examples described below, we found that the value function is increasing in the reputation of the bank for all of the parameter values we studied.

5.6 Fragility

We think of equilibrium outcomes as fragile in two ways. One notion of fragility is simply that the economy has multiple equilibria so that sunspot-like fluctuations can induce changes in outcomes. A second notion of fragility is that small changes in fundamentals induce large changes in aggregate outcomes.

Equilibrium outcomes in our unperturbed game are clearly fragile under the first notion because that game has multiple equilibria. They are also fragile under the second notion if agents in the model coordinate on different equilibria depending on the realization of the fundamentals and if a large mass of agents have reputation levels in the multiplicity region.

Since our perturbed game has a unique equilibrium, it is not fragile under the first notion. We argue that it is fragile under our second notion. In our multi-period model, the history of past outcomes induces dispersion in the reputation levels of different banks. In order for our equilibrium to display fragility under the second notion, we must have that either banks with a wide variety of reputation levels change their actions in the same way in response to aggregate shocks or that the reputation levels of banks cluster close to each other. We conducted a wide variety of numerical exercises and found that the clustering effect is very strong in our model. This clustering effect clearly depends on the details of the history of exogenous shocks. To abstract from these details, we consider the invariant distribution associated with our model and show that this invariant distribution displays clustering. The invariant distribution is that associated with the infinite horizon limit of our multi-period model. We allow for a small probability of replacement in order to ensure that the invariant distribution is not concentrated at a single point.
Figure 3 displays the cutoff values for each reputation type for the ergodic set associated with the invariant distribution. This ergodic set contains reputation levels between roughly 0.25 and 0.85. For collateral values above the cutoffs shown in Figure 3, banks sell their loans and below the cutoffs banks hold their loans. This figure illustrates that as the collateral value falls, the adverse selection problem worsens in the sense that banks with a wider range of reputations hold their loans. For example, at a collateral value of 5, banks with reputation levels below roughly 0.4 hold their loans and the banks with higher reputation levels sell their loans. At a collateral value of 4, banks with reputation levels below roughly 0.65 hold their loans and banks with higher reputation levels sell their loans. Thus, a fall in collateral values from 5 to 4 induces banks with reputation levels roughly between 0.4 and 0.65 to switch from selling to holding their loans.

Figure 4 displays the invariant distribution of reputation levels for high-quality banks. This figure shows that the invariant distribution displays significant clustering. Roughly 70 percent of high-quality banks have reputation levels between 0.8 and 0.85. Small fluctuations in the default value of loans around the cutoff values for such banks can induce a large mass of banks to alter their behavior.

Figure 5 plots the volume of trade, measured as the fraction of all banks that sell their loans. A decrease in the default value from 1.3 to 1.1 induces a 50 percent decrease in the volume of trade. In this sense, Figure 5 suggests that equilibrium outcomes in our model are fragile under the second notion.

Next we analyze the forces that induce clustering in our model. Bayes’ rule implies that $\frac{1}{\mu_t}$ is a martingale. Since $\frac{1}{\mu_t}$ is a convex function, Jensen’s inequality implies that the reputation of a bank, $\mu_t$, is a submartingale so that $\mu_t$ tends to rise. Conditional on a high-quality, high-cost bank holding, the analysis of our equilibrium implies that the reputation of such a bank also rises. These forces imply that the reputation of a high-quality bank displays an upward trend. This upward trend is dampened by replacement. Since all high-quality banks tend to have an upward trend in their reputations, these reputations tend to cluster toward each other.

---

2 The parameters used in this simulation are the following: $\bar{\pi} = 0.8, \bar{v} = 0.3, \bar{v} = 7, \bar{c} = 0.5, \bar{c} = -3, \alpha = 0.15, q = .1, r = 0.5, \beta(1 - \lambda) = .99, \lambda = .4, \mu_0 = .6$, where $\lambda$ represents the exogenous probability of replacement and $\mu_0$ is the reputation of a newly replaced bank. The distribution of $\bar{\pi}$ is $N(0, 2)$. 
Figure 5.3: Cutoff Thresholds for High-Quality Banks.

Figure 5.4: Invariant Distribution of Reputations of High-Quality Banks.
This reasoning suggests that fragility under the second notion does not depend on the particular equilibrium that we have selected. In both the positive and negative reputational equilibria, the reputations of high-quality banks rise over time and tend to cluster together eventually. This clustering tends to make them react in the same way to fluctuations in the default value of the underlying loans. We conjecture that any continuous selection procedure will produce periods of high volumes of new issuances followed by sudden collapses.

We have analyzed the effect of other aggregate shocks in our model. In particular, we allowed the comparative advantage cost, \( \bar{c} \), to be subject to aggregate shocks. In that version of the model, we found that banks with a wide variety of reputations tend to have cutoffs that are very close to each other. That model displays fragility under our second notion because small fluctuations in holding costs around a critical value induce large changes in actions by banks with a wide variety of reputations. (Details are available upon request.)
5.7 Policy Exercises

In this section, we use our model to evaluate the effects of various policies intended to remedy problems of credit markets—policies that have been proposed since the 2007 collapse of secondary loan markets in the United States. We focus on the effects of policies in which the government would purchase asset-backed securities at prices above existing market value, such as the Public-Private Partnership plan, as well as on policies that decreased the costs of holding loans to maturity, including changes in the Federal Funds target rate, the Term Asset-Backed Securities Loan Facility (TALF), and increased FDIC insurance.

These policies were motivated by perceived inefficiencies in secondary loan markets. For example, the Treasury Department asserts, in its Fact Sheet dated March 23, 2009, releasing details of a proposed Public-Private Investment Program for Legacy Assets,

Secondary markets have become highly illiquid, and are trading at prices below where they would be in normally functioning markets. ([of Treasury, 2009])

Similarly, the Federal Reserve Bank of New York asserts, in a White Paper dated March 3, 2009, making the case for the Term Asset-Backed Securities Loan Facility (TALF),

Nontraditional investors such as hedge funds, which may otherwise be willing to invest in these securities, have been unable to obtain funding from banks and dealers because of a general reluctance to lend. (TALF White Paper 2009)

Note that in our model sudden collapses are associated with increased inefficiency so that our model is consistent with policy makers concerns that the market had become more inefficient. In this sense, our model is an appropriate starting point for analyzing policies intended to remedy inefficiencies.

We first consider policies in which the government attempts to purchase so-called toxic assets at above-market values. Consider the following government policy in the limiting version of the perturbed game as $\sigma \to 0$. The government offers to buy the asset at some price $p$ in the first period.
Suppose first that $p \leq \hat{p}(\mu_1; v_1)$. We claim that the unique equilibrium without government is also the unique equilibrium with this government policy. To see this claim, note that the equilibrium in the second period is the same with and without the government policy so that the reputational gains are the same with and without the government policy. Consider the first period and a realization of first-period return $v_1 < v_1^*$. In the game without the government, the HH bank found it optimal not to sell at a price $\hat{p}(\mu_1; v_1)$. Since the reputational gains are the same with and without the government policy, in the game with the government, it is also optimal for the HH not to sell at this price. A similar argument implies that the equilibrium strategy of the HH bank is unchanged for $v_1 > v_1^*$. Thus, this government policy has no effect on the equilibrium strategy of the HH bank. Of course, under this policy, the government ends up buying the asset from low-quality banks. The only effect of this policy is to make transfers to low-quality banks.

Suppose next that the price set by the government, $p$, is sufficiently larger than $\hat{p}(\mu_1; v_1)$. Then, the HH bank will find it optimal to sell and will enjoy the reputational gain associated with a policy of selling. In this sense, if the government offers a sufficiently high price, it can ensure that reputational incentives work to overcome adverse selection problems. Note, however, that this policy necessarily implies that the government must earn negative profits.

Consider now a policy that reduces interest rates in period 1 and leaves period 2 interest rates unchanged. We begin the analysis with the unperturbed game. Such a policy increases the static payoff in period 1 from holding loans which worsens the static incentives for the HH bank to sell its loan. Specifically, this policy raises both the threshold $\mu$ below which banks find it optimal to hold in the positive reputational equilibrium and the threshold $\bar{\mu}$ below which banks find it optimal to hold their loans in the negative reputational equilibrium. Thus, this policy serves only to aggravate the lemons problem in secondary loans markets.

Consider next a policy under which the government commits to reducing period 2 interest rates but leaves period 1 interest rates unchanged. Obviously, this policy increases incentives for banks to hold their loans in period 2 and thereby increases the threshold below which banks hold their loans, $\mu_2^*$. In this sense, it makes period 2 allocations less efficient. We will show that this policy reduces the region of multiplicity
in period 1 and in this sense can improve period 1 allocations. To show the reduction in the region of multiplicity, consider the reputational gain in the positive reputational equilibrium evaluated at $\mu$:

$$\beta (\bar{\pi} V_2(\mu_{s\theta}) + (1 - \bar{\pi}) V_2(\mu_{s0}) - V_2(\mu_h)).$$

Using (5.5), it is straightforward to see that an arbitrarily small reduction in interest rates of $dr$ in period 2 reduces $V_2(\mu_{s\theta})$ by $\alpha q dr$ since $\mu_{s\theta} > \mu^*$. Moreover, since $\mu_{s0}$ and $\mu_h$ are strictly less than $\mu^*_2$, $V_2(\mu_{s0})$ and $V_2(\mu_h)$ fall by $q dr$. As a result, the reputational gain falls by $\beta \bar{\pi}(1 - \alpha)q dr$. This decline in reputational gain induces an increase in the threshold $\mu$. Similarly, we can show that the policy induces a fall in the threshold $\bar{\mu}$. Thus, the region of multiplicity shrinks and in this sense can improve period 1 allocations. Interestingly, such a policy is time inconsistent because the government has a strong incentive in period 2 not to make period 2 allocations less efficient.

An alternative policy that has not been proposed is to consider forced asset sales in which the government randomly forces banks to sell their loans. Such a policy in our model would mitigate the lemons problem in secondary loan markets by generating a pool of loans in secondary markets consistent with the ex ante mix of loan types. Although this is a standard intervention directed at increasing the price and volume of trade in markets that suffer from adverse selection, in our model such an intervention comes at the cost of misallocating loans to those without comparative advantage. Specifically, some banks with low costs of holding loans will be forced to sell to the marketplace.

It is straightforward to show that a policy under which the government commits to purchase assets in period 2 at prices that are contingent on the realization of the signals can eliminate the multiplicity of equilibria and support the positive reputational equilibrium. Although such a policy would be desirable, the feasibility of such a policy can be analyzed only by developing a model in which private agents cannot commit but the government can.

### 5.8 Conclusion

This paper is an attempt to make three contributions: a theoretical contribution to the literature on reputation, a substantive contribution to the literature on the behavior of
financial markets during crises, and a contribution to analyses of proposed and actual policies during the recent crisis. In terms of the theoretical contribution, we have combined insights from the literature that emphasizes the positive aspects of reputational incentives (see [Mailath and Samuelson, 2001]) with the literature that emphasizes the negative aspects of reputational incentives (see [Ely and Välimäki, 2003]) to show that multiplicity of equilibria naturally arise in reputation models like ours. We have also shown how techniques from the coordination games literature can be adapted to develop a refinement method that produces a unique equilibrium. In terms of the literature on the behavior of financial markets during crises, we have argued that sudden collapses in secondary loan market activity are particularly likely when the collateral value of the underlying loan declines. In terms of policy, we have argued that a wide variety of proposed policy responses would not have averted either the sudden collapse or the associated inefficiency. An important avenue for future work is to analyze policies that might in fact remedy the inefficiencies.

Another important avenue for future work is to introduce loan origination as a choice for banks in the model so that the model can be used to analyze the effects of sudden collapses on investment and other macroeconomic aggregates.
Bibliography


Appendix A

Appendix to Chapter 2

A.1 Proofs

A.1.1 Proof of Proposition 2.1.

Recall the planning problem (P1). Suppose the lagrange multiplier on feasibility at period \( t \) is \( \lambda_t \), the multiplier on the third constraint is given by \( \gamma(\theta) \), the multiplier on (2.11) is given by \( \mu_2(\theta)\lambda_0 f(\theta) \), and the co-state associated with (2.10) is given by \( \mu_1(\theta)f(\theta)\lambda_0 \). Then the Lagrangian for this problem is given by

\[
\mathcal{L} = \int_\theta^\theta \left\{ U(\theta) + \lambda_0 [c_0(\theta) - c_0(\theta) - k_1(\theta)] + \lambda_1 \left[ \theta^\alpha k_1(\theta)^\alpha - \int c_1(\theta)g(y|k_1(\theta),\theta) \right] \\
+ \gamma(\theta) \left[ u(c_0(\theta)) + \beta \int_0^\theta u(c_1(\theta, y))g(y|k_1(\theta),\theta)dy - U(\theta) \right] \\
+ \lambda_0 \mu_1(\theta) \left[ U'(\theta) - \frac{1}{\theta} k_1(\theta)u'(c_0(\theta)) \right] \\
+ \lambda_0 \mu_2(\theta) \left[ \beta \int_0^\theta u(c_1(\theta, y))g_k(y|k_1(\theta),\theta)dy - u'(c_0(\theta)) \right] \right\} f(\theta)d\theta
\]
An application of the integration by parts formula implies that

\[ \mathcal{L} = \int_{\theta}^{\bar{\theta}} \left\{ U(\theta) + \lambda_0 [c_0 - c_0(\theta) - k_1(\theta)] + \lambda_1 \left[ \theta^\alpha k_1(\theta)^\alpha - \int c_1(\theta)g(y|k_1(\theta), \theta) \right] \ight. \\
+ \gamma(\theta) \left[ u(c_0(\theta)) + \beta \int_{\gamma}^{\bar{\theta}} u(c_1(\theta, y))g(y|k_1(\theta), \theta)dy - U(\theta) \right] \\
- \lambda_0 \mu_1(\theta) \left[ \frac{1}{\theta} k_1(\theta)u'(c_0(\theta)) \right] \\
+ \lambda_0 \mu_2(\theta) \left[ \beta \int_{\gamma}^{\bar{\theta}} u(c_1(\theta, y))g_k(y|k_1(\theta), \theta)dy - u'(c_0(\theta)) \right] \right\} f(\theta)d\theta \\
+ \lambda_0 U(\theta)f(\theta)\mu_1(\theta)\int_{\theta}^{\bar{\theta}} \left[ \frac{\beta}{\theta} U(\theta) [\mu_1(\theta)f(\theta) + \mu_1(\theta)f'(\theta)] \right] d\theta \\

By a theorem from [Luenberger, 1969] (section §9.3, Theorem 1), in order for an allocation \( \{c_0(\theta), c_1(\theta, y), U(\theta), k_1(\theta)\}_{(\theta, y) \times \mathbb{R}^2} \) to attain a local optimum in program (P1), the multipliers \( \mu_1, \mu_2 \) and \( \gamma \) must exist such that the Fréchet derivative of the above Lagrangian is zero. More technically, we assume that our underlying space is the space of bounded continuous functions \( \{c_0(\theta), c_1(\theta, y), U(\theta), k_1(\theta)\}_{(\theta, y) \times \mathbb{R}^2} \) and the Fréchet derivative is taken with respect to a member of this Hilbert space. Given the assumption that all allocation are interior, the theorem implies that the following conditions should hold a.e.-\( F \):

\[-\lambda_0 + \gamma(\theta)u'(c_0(\theta)) - \lambda_0 \mu_1(\theta) \frac{1}{\theta} k_1(\theta)u''(c_0(\theta)) - \lambda_0 \mu_2(\theta)u''(c_0(\theta)) = 0 \quad \text{(A.1)}
\]

\[-g(y|k_1(\theta), \theta)\lambda_1 + \beta u'(c_1(\theta, y)) [g(y|k_1(\theta), \theta) + \beta \lambda_0 \mu_2(\theta)g_k(y|k_1(\theta), \theta)] = 0 \quad \text{(A.2)}
\]

\[1 - \gamma(\theta) - \lambda_0 [\mu_1(\theta)f(\theta) + \mu_1(\theta)f'(\theta)] = 0 \quad \text{(A.3)}
\]

\[\lambda_1 \alpha \theta^\alpha k_1(\theta)^{\alpha-1} - \lambda_0 - \lambda_1 \int c_1(\theta, y)g_k(y|k_1(\theta), \theta)dy \quad \text{(A.4)}
\]

\[+ \gamma(\theta)\beta \int_{\gamma}^{\bar{\theta}} u(c_1(\theta, y))g(y|k_1(\theta), \theta)dy
\]

\[-\frac{1}{\theta} u'(c_0(\theta)) \lambda_0 \mu_1(\theta) + \lambda_0 \mu_2(\theta) \beta \int_{\gamma}^{\bar{\theta}} u(c_1(\theta, y))g_k(y|k_1(\theta), \theta)dy = 0 \quad \text{(A.5)}
\]

Moreover, continuity of allocations in \( \theta \) implies that

\[\mu_1(\bar{\theta}) = \mu_1(\theta) = 0\]
Note that by definition $q = \frac{\lambda_1}{\lambda_0}$. From above, we have

$$\gamma(\theta) = \frac{\lambda_0}{u'(c_0(\theta))} + \lambda_0 \frac{u''(c_0(\theta))}{u'(c_0(\theta))} \left[ \frac{1}{\theta} k_1(\theta) \mu_1(\theta) + \mu_2(\theta) \right]$$  \hspace{1cm} (A.6)

$$\gamma(\theta) g(y|k_1(\theta), \theta) = \frac{\lambda_1}{\beta u'(c_1(\theta, y))} g(y|k_1(\theta), \theta) - \lambda_0 \mu_2(\theta) g_k(y|k_1(\theta), \theta)$$

Integrating the last equation over $y$ and using $\int_0^y g_k(y|k_1(\theta), \theta) dy = 0$, we get

$$\gamma(\theta) = \lambda_1 \int_0^\theta \frac{1}{\beta u'(c_1(\theta, y))} g(y|k_1(\theta), \theta) dy$$  \hspace{1cm} (A.7)

Combining (A.6) and (A.7), gives us the desired result.

QED.

A.1.2 Proof of Lemma 2.2.

If we multiply (A.2) by $u(c_1(\theta, y))$ and divide it by $u'(c_1(\theta, y))$, we have – we suppress $\theta$:

$$\lambda_1 u(c_1) \times \frac{1}{u'(c_1)} g = \beta \gamma u(c_1) g + \beta \lambda_0 \mu_2 u(c_1) g_k$$

Integrating over $y$, we get

$$\lambda_1 \int u(c_1) \times \frac{1}{u'(c_1)} g dy = \beta \gamma \int u(c_1) g dy + \beta \lambda_0 \mu_2 \int u(c_1) g_k dy$$

Using (A.7) and (2.11), we get

$$\lambda_1 \int u(c_1) \times \frac{1}{u'(c_1)} g dy = \lambda_1 \int \frac{1}{u'(c_1)} g dy \int u(c_1) g dy + \lambda_0 \mu_2 u'(c_0)$$

and therefore

$$\mu_2 \lambda_0 u'(c_0) = \lambda_1 \left[ \int u(c_1) \times \frac{1}{u'(c_1)} g dy - \int \frac{1}{u'(c_1)} g dy \int u(c_1) g dy \right]$$

$$\Rightarrow \mu_2 = \frac{q}{u'(c_0)} Cov_{\theta}(u(c_1), \frac{1}{u'(c_1)})$$

Replacing the above in (A.5) gives the formula for $\mu_1$.

QED.
A.1.3 Proof of Proposition 2.3.

From above, we know that an optimal allocation must satisfy (A.1)-(A.5). Now suppose to the contrary that $\mu_1(\theta) < 0$ for some $\theta$. Since $\mu_1(\bar{\theta}) = \mu_1(\bar{\theta}) = 0$ and $\mu_1(\theta)$ is continuously differentiable, there must exist $\theta_1 < \theta_2$ such that $\mu_1(\theta_1) = \mu_1(\theta_2) = 0$ and $(\mu_1 f)'(\theta_1) \leq 0 \leq (\mu_1 f)'(\theta_2)$. Hence the equations (A.1)-(A.3) evaluated at $\theta_1$ and $\theta_2$ become

$$-\lambda_0 + \gamma u'(c_0) - u''(c_0)\lambda_0 \mu_2 = 0$$
$$-g\lambda_1 + \beta \gamma u'(c_1)g + \beta \lambda_0 \mu_2 u'(c_1)g_k = 0$$
$$1 - \gamma - \lambda_0 (\mu_1 f)' = 0$$

The last equation implies that $\gamma(\theta_1) \geq 1 \geq \gamma(\theta_2)$. Since $u(c) = -\exp(-\psi c)$, the above can be rewritten as

$$-\lambda_0 - \gamma \psi u(c_0) + \psi u'(c_0)\lambda_0 \mu_2 = 0$$
$$-g\lambda_1 - \beta \gamma \psi u(c_1)g - \beta \mu_2 \lambda_0 \psi u(c_1)g_k = 0$$

Integrating the second equation, we get

$$\lambda_0 \mu_2 = -\frac{\lambda_1}{\psi u'(c_0)} - \frac{\beta}{\psi} \int u(c_1)gdy$$

Replacing the above equation in (A.8)

$$-\lambda_0 - \gamma \psi u(c_0) - \lambda_1 - \psi \gamma \beta \int u(c_1)gdy = 0$$

or

$$u(c_0) + \beta \int u(c_1)gdy = -\frac{\lambda_0 + \lambda_1}{\psi \gamma}$$

Therefore, $U(\theta) = -\frac{\lambda_0 + \lambda_1}{\psi \gamma(\theta)}$ when $\theta = \theta_1, \theta_2$. Hence, we have

$$\gamma(\theta_1) \geq \gamma(\theta_2) \Rightarrow \frac{\lambda_0 + \lambda_1}{\psi \gamma(\theta_1)} \leq \frac{\lambda_0 + \lambda_1}{\psi \gamma(\theta_2)}$$

$$\Rightarrow U(\theta_1) \geq U(\theta_2)$$

(A.9)

Note however that by (2.10), $U(\theta_2) - U(\theta_1) = \int_{\theta_1}^{\theta_2} \frac{1}{\psi} u'(c_0(\hat{\theta})) k_1(\hat{\theta}) d\hat{\theta} > 0$ which is in contradiction with (A.9). This proves that $\mu_1(\theta) \geq 0$. The above analysis also shows
also that $\mu_1(\theta)$ needs to be non-zero for at least a positive measure of $\theta$'s. Otherwise, for almost all $\theta$'s, $\mu_1(\theta) = 0$ and continuous differentiability of $\mu_1(\theta)$ implies that $\mu'_1(\theta) = 0$. In this case, the above analysis implies that $U(\theta)$ is constant which violates the adverse selection incentive constraint (2.10).

QED.

A.1.4 Proof of Theorem 2.9.

The FOC from program (P2) are given by

$$\begin{align*}
-f^t(\theta|\theta_{-1}) + \gamma(\theta)f^t(\theta|\theta_{-1})u'(c(\theta)) + \mu(\theta)f^t(\theta|\theta_{-1})\frac{1}{\theta}k(\theta)u''(c(\theta)) + \mu_2(\theta)u''(c(\theta)) &= 0 \\
f^t(\theta|\theta_{-1})\frac{Q_{t+1}}{Q_t}P_{w}^{t+1}(\theta)g^{t+1}(y|k(\theta), \theta) + \beta\gamma(\theta)f^t(\theta|\theta_{-1})g^{t+1}(y|k(\theta), \theta) \\
-\beta\mu_2(\theta)f^t(\theta|\theta_{-1})g^{t+1}_k(y|k(\theta), \theta) &= 0 \\
\lambda f^t(\theta|\theta_{-1}) + \lambda' f^t_{\theta_{-1}}(\theta|\theta_{-1}) - \gamma(\theta)f^t(\theta|\theta_{-1}) + \mu(\theta)f^t(\theta|\theta_{-1}) + \mu(\theta)f^t_\theta(\theta|\theta_{-1}) &= 0
\end{align*}$$

where $\mu(\theta)$ is the costate associated with (2.27), $\mu_2(\theta)$ is the Lagrange Multiplier on (2.28) and $\gamma(\theta)f^t(\theta|\theta_{-1})$ is the Lagrange Multiplier on the third constraint. Integrating the second equation with respect to $y$ implies that

$$\begin{align*}
\frac{Q_{t+1}}{Q_t} \int_{0}^{\bar{y}} P_{w}^{t+1}(\theta)g^{t+1}(y|k(\theta), \theta)dy + \beta\gamma(\theta) &= 0
\end{align*}$$

Moreover, the first FOC can be written as

$$\gamma(\theta) = \frac{1}{u'(c(\theta))} - \mu(\theta)\frac{1}{\theta}k(\theta)\frac{u''(c(\theta))}{u'(c(\theta))} - \mu_2(\theta)\frac{u''(c(\theta))}{u'(c(\theta))}$$

Hence,

$$\begin{align*}
\frac{1}{u'(c(\theta))} - \frac{\mu_1}{\theta}k_{t+1}u''(c_t) - \frac{\mu_2}{u'(c_t)} + \frac{Q_{t+1}}{Q_t} \int_{0}^{\bar{y}} P_{w}^{t+1}(w_{t+1}, \Delta_{t+1}, \theta_t)g^{t+1}(y|k_{t+1}, \theta_t)dy &= 0
\end{align*}$$

Moreover, from the envelope theorem, we know that $P_{w}^{t} = -\lambda$. Integrating the first
and the last FOC w.r.t $\theta$ implies that

\[- \int_\bar{\theta}^\theta \frac{1}{u'(c(\theta))} f^t(\theta|\theta_{-1}) d\theta + \int_\bar{\theta}^\theta \gamma(\theta) f^t(\theta|\theta_{-1}) d\theta + \int_\bar{\theta}^\theta \mu(\theta) f^t(\theta|\theta_{-1}) \frac{1}{\theta} k(\theta) \frac{u''(c(\theta))}{u'(c(\theta))} d\theta + \int_\bar{\theta}^\theta \mu_2(\theta) \frac{u''(c(\theta))}{u'(c(\theta))} d\theta = 0\]

\[\lambda - \int_\bar{\theta}^\theta \gamma(\theta) f^t(\theta|\theta_{-1}) d\theta = 0\]

Hence, \[
\frac{Q_{t+1}}{P_{t+1}} E_t P_w^{t+1} = P_w^t.
\]

From (2.29), \[
\mu_{2t} = f^t(\theta|\theta_{-1}) \frac{Q_{t+1}}{Q_t} \frac{1}{u'(c_t)} Cov(P_w^{t+1}, w_{t+1}|(\theta^t, y^t)).
\]

Therefore,

\[P_w^t = E \left[ \frac{1}{u'(c_t)} + \mu_t k_{t+1} \frac{u''(c_t)}{\theta_t u'(c_t)} + \frac{Q_{t+1}}{Q_t} \frac{u''(c_t)}{u'(c_t)^2} Cov(P_w^{t+1}, w_{t+1}|(\theta^t, y^t)) \right]
\]

Applying the above formula to period $t + 1$ and using (A.10) gives the MIEE.

QED.

### A.1.5 Proof of Theorem 2.11.

Given the steps provided in the text, we only need to show that the value function has the form

\[P(w) = \frac{A_t}{\psi} \log(-w) + B_t\]

where \[A_t = \frac{1}{Q_t} \sum_{s=t}^T Q_s\] and that inequality (2.31) holds.

To show that the value function has the claimed form, we use induction. Notice that at $t = T$, since there are no shocks, simply, we have

\[P_T(w) = \frac{1}{\psi} \log(-w).
\]

Now suppose that the claim holds for $t + 1$, then we show that it holds for $t$. Given this assumption, planning problem (P2) becomes

\[P_t(w) = \max \int_\Theta \left[ \frac{Q_{t+1}}{Q_t} (\theta k(\theta))^{\alpha} - c(\theta) - k(\theta) + \frac{Q_{t+1}}{Q_t} \frac{A_{t+1}}{\psi} \log(-w'(\theta)) \right] f^t(\theta) d\theta\]

subject to

\[w = \int_\Theta U(\theta) f^t(\theta) d\theta\]

\[U(\theta) = u(c(\theta)) + \beta w'(\theta)\]

\[\frac{\partial}{\partial \theta} U(\theta) = \frac{1}{\theta} k(\theta) u'(c(\theta))\]
subject to

\[ \gamma \]

That is, if \( H(\theta) = -w \cdot \dot{w}(\theta) \), and \( U(\theta) = (-w) \cdot \dot{U}(\theta) \). Under these assumption, \( u(c(w, \theta)) = (-w) \cdot u(c(\theta)) \) and \( u'(c(w, \theta)) = (-w)u'(c(\theta)) \). Hence,

\[
P^t(w) = \max_{\theta} \left[ \int_\Theta \left[ \frac{Q_{t+1}}{Q_t} (\theta k(\theta))^\alpha + \frac{1}{\psi} \log(-w - \dot{\theta}) - k(\theta) \right. \right.
\]

\[
+ \frac{Q_{t+1}}{Q_t} \frac{A_{t+1}}{\psi} \log((-w)(-\dot{w}(\theta))) \left. \right] \right] f^t(\theta) d\theta
\]

subject to

\[
-1 = \int_\Theta \dot{U}(\theta)f^t(\theta)d\theta
\]

\[
\dot{U}(\theta) = u(c(\theta)) + \beta \dot{\theta}
\]

\[
\frac{\partial}{\partial \theta} \dot{U}(\theta) = \frac{1}{\theta} \theta k(\theta)u'(c(\theta))
\]

and the objective becomes

\[
\int_\Theta \left[ \frac{Q_{t+1}}{Q_t} (\theta k(\theta))^\alpha - \dot{c}(\theta) - k(\theta) + \frac{Q_{t+1}}{Q_t} \frac{A_{t+1}}{\psi} \log(-\dot{w}(\theta)) \right] f^t(\theta) d\theta
\]

\[
+ \frac{1}{\psi} \log(-w) + \frac{Q_{t+1}}{Q_t} \frac{A_{t+1}}{\psi} \log(-w)
\]

This means that the above maximization problem is independent of \( w \) and therefore

\[ P^t(w) = \frac{A_t}{\psi} \log(-w) + B_t \text{ where } A_t = 1 + \frac{Q_{t+1}}{Q_t} A_{t+1}. \]

In order to show (2.31), we show that \( \mu_1(\theta) \geq 0 \). In fact, the proof that \( \mu_1(\theta) \geq 0 \) is identical to the proof of Proposition 2.3. That is, if \( \mu_1(\theta) < 0 \), there must exists \( \theta_1 < \theta_2 \) such that \( \mu_1(\theta_1) = \mu_1(\theta_2) = 0 \) and \( \frac{d}{d\theta} (\mu_1(\theta_1) f^t(\theta_1)) \leq 0 \leq \frac{d}{d\theta} (\mu_1(\theta_2) f^t(\theta_2)) \). Note that under this assumption the FOCs evaluated at \( \theta_1 \) and \( \theta_2 \) are given by

\[
-f^t + \gamma u'(c) = 0
\]

\[
f^t \frac{Q_{t+1}}{Q_t} \frac{A_{t+1}}{\psi} \frac{1}{w} + \gamma \beta = 0
\]

\[
\lambda - \gamma - \frac{1}{f^t} \frac{d}{d\theta} (\mu_1 f^t) = 0
\]

Hence \( \gamma(\theta_1) \geq \gamma(\theta_2) \) and therefore \( c(\theta_1) \geq c(\theta_2) \) and \( w(\theta_1) \geq w(\theta_2) \) which contradicts the incentive constraint. Hence, \( \mu_1(\theta) \geq 0 \). Note that from the FOCs we have

\[
-1 + \gamma(\theta)u'(c(\theta)) - \mu_1(\theta) \frac{1}{\theta} k(\theta)u''(c(\theta)) = 0
\]

\[
\frac{Q_{t+1}}{\beta Q_t} P^t_{w+1}(w'(\theta)) + \gamma(\theta) = 0
\]
and therefore
\[-\frac{Q_{t+1}}{\beta Q_t} P_{w_{t+1}}^t(w'(\theta)) = \frac{1}{u'(c(\theta))} + \mu_1(\theta)\frac{1}{\theta} k(\theta) u''(c(\theta)) \]

Since \(\mu_1 \geq 0\) and \(u''(c) < 0\), we have the inequality (2.31).

QED.

\section*{A.1.6 Proof of Proposition 2.12.}

The recursive problem with safe returns is given by
\[
P^t(w) = \max \int_\Theta \left[ \frac{Q_{t+1}}{Q_t} (\theta k(\theta))^\alpha - c(\theta) - k(\theta) + \frac{Q_{t+1}}{Q_t} P_{w_{t+1}}^t(w'(\theta)) \right] f^t(\theta) d\theta
\]
subject to
\[
w = \int_\Theta U(\theta) f^t(\theta) d\theta
\]
\[
U(\theta) = u(c(\theta)) + \beta w'(\theta)
\]
\[
\frac{\partial}{\partial \theta} U(\theta) = \frac{1}{\theta} k(\theta) u'(c(\theta))
\]  \hfill (A.11)

The Envelope condition (A.11) can be written as
\[
u'(c(\theta)) c'(\theta) + \beta \frac{d}{d\theta} w'(\theta) = \frac{1}{\theta} k(\theta) u'(c(\theta))
\]
By assumption, \(\frac{d}{d\theta} w'(\theta) > 0\). Hence, we must have \(\frac{1}{\theta} k(\theta) > c'(\theta)\). The term \(\frac{1}{\theta} k(\theta) - c'(\theta)\) can be thought of as incremental increase in consumption in the current period when the agent lies locally. Hence, when \(w'(\theta)\) is increasing, lying downward increases consumption.

The FOC of the above planning problem are given by
\[-f^t(\theta) + \gamma(\theta) f^t(\theta) u'(c(\theta)) - \mu_1(\theta) f^t(\theta) \frac{1}{\theta} k(\theta) u''(c(\theta)) = 0
\]
\[\lambda f^t(\theta) - \gamma(\theta) f^t(\theta) - \frac{d}{d\theta} (\mu_1(\theta) f^t(\theta)) = 0
\]
Combining these equations implies that
\[-f^t(\theta|\theta_{-1}) + \left[ \lambda f^t(\theta) - \frac{d}{d\theta} (\mu_1(\theta) f^t(\theta)) \right] u'(c(\theta)) - \mu_1(\theta) f^t(\theta) \frac{1}{\theta} k(\theta) u''(c(\theta)) = 0
\]
Moreover, by assumption $\mu_1(\theta) > 0$. Hence, we must have

$$\mu_1(\theta) \frac{1}{\theta} k(\theta) u''(c(\theta)) < \mu_1(\theta) c'(\theta) u''(c(\theta))$$

Therefore,

$$-f^t(\theta) + \left[ \lambda f^t(\theta) - \frac{d}{d \theta} \left( \mu_1(\theta) f^t(\theta) \right) \right] u'(c(\theta)) - \mu_1(\theta) f^t(\theta) c'(\theta) u''(c(\theta)) < 0$$

or

$$-f^t(\theta) + \lambda f^t(\theta) u'(c(\theta)) - \frac{d}{d \theta} \mu_1(\theta) f^t(\theta) u'(c(\theta)) < 0.$$

Integrating the above inequality over $\theta$,

$$-1 + \lambda \int u'(c(\theta)) f^t(\theta) d\theta - \mu_1(\theta) f^t(\theta) u'(c(\theta)) |^\theta_{\bar{\theta}} < 0$$

and since $\mu_1(\bar{\theta}) = \mu_1(\theta) = 0$, hence

$$\int u'(c(\theta)) f^t(\theta) d\theta < \frac{1}{\lambda}$$

Note that by Envelope theorem, $\lambda = -P_{w}^{t}(w)$, hence

$$\int u'(c(\theta)) f^t(\theta) d\theta < -\frac{1}{P_{w}^{t}(w)}$$

The rest is identical to the prove given in the text.

QED.

A.1.7 Proof of Lemma 2.16.

Note that since the problem is stationary, i.e., $Q_{t+1}/Q_t = \dot{q}$, the policy functions are time independent and therefore

$$c_t(w_t) = -\frac{1}{\psi} \log(-w_t) + c^*$$

(A.12)

$$c_{t+1}(w_t, y_{t+1}) = -\frac{1}{\psi} \log(-w_t) - \frac{1}{\psi} \log(-w^*(y_{t+1})) + c^*$$

Hence, the intertemporal wedge is given by

$$1 - \tau_s = \frac{q u'(c_t)}{\beta E_t u'(c_{t+1})} = \frac{q(-w_t) u'(c^*)}{\beta(-w_t) u'(c^*) \int_{\bar{y}} \left( -w^*(y_{t+1}) \right) g(y_{t+1}|k^*) dy_{t+1}}$$

$$= \frac{\beta \int_{\bar{y}} \left( -w^*(y_{t+1}) \right) g(y_{t+1}|k^*) dy_{t+1}}{\beta \int_{\bar{y}} \left( -w^*(y_{t+1}) \right) g(y_{t+1}|k^*) dy_{t+1}}$$
Let \( 1 + \hat{R} = \frac{1 - \tau}{1 - \frac{1}{\beta} \int_0^y (-w^*(y_{t+1}))g(y_{t+1}|k^*)dy_{t+1}} \). Note by definition that

\[
\beta \int_0^y (-w^*(y_{t+1}))g(y_{t+1}|k^*)dy_{t+1} = 1 + u(c^*) < 1.
\]

Hence, \( \hat{R} > 0 \). Next, for any convex and smooth function \( T(y) \), consider the following recursive formulation of (P4):

\[
V(a) = \max u(c) + \beta \int_0^y V(y - T(y) - (1 + \hat{R})B')g(y|k')dy
\]

subject to

\[
c + k' - B' = a
\]

Using a guess and verify method, we show that \( V(a) = e^{-\frac{\hat{R}}{1 + \hat{R}} \psi_0 + \varphi} \) for a constant number \( \varphi \) that depends on the choice of \( T(y) \). Moreover, the policy functions implied by the above maximization problem is

\[
c(a) = \frac{\hat{R}}{1 + \hat{R}} a + \zeta_1
\]

\[
B'(a) = \frac{\hat{R}}{1 + \hat{R}} a + \zeta_2
\]

for constants \( \zeta_1 \) and \( \zeta_2 \). Furthermore, \( k' \) solves the following equation:

\[
\psi \beta \hat{R} \int e^{-\frac{\hat{R}}{1 + \hat{R}} \psi[y_0 - T(y_0)]} g(y|k')dy = - \int e^{-\frac{\hat{R}}{1 + \hat{R}} \psi[y_0 - T(y_0)]} g_k(y|k')dy
\]

and hence independent of \( a \). Since \( a = y - T(y) - (1 + \hat{R})B \), we must have

\[
c_t = \frac{\hat{R}}{1 + \hat{R}} [y_t - T(y_t) - (1 + \hat{R})B_t] + \zeta_1.
\]

Note that from (A.12), \( c_t = -\frac{1}{\psi} \log(-w_{t-1}) - \frac{1}{\psi} \log(-w^*(y_t)) + c^* \) Hence, if the tax function \( T(\cdot) \) is to implement the optimal allocation, we must have

\[
T(y) = y + \frac{1 + \hat{R}}{\hat{R}} \frac{1}{\psi} \log(-w^*(y)) + \kappa
\]

for some constant \( \kappa \). Therefore, to complete the proof we need to find \( B_0 \) and \( \kappa \) given the value of income realization at \( t = 0 \) and \( w_0 \). Note that for the implementation to work, we must have

\[
V(y_0 - T(y_0) - (1 + \hat{R})B_0) = e^{-\frac{\hat{R}}{1 + \hat{R}} \psi[y_0 - T(y_0) - (1 + \hat{R})B_0] + \varphi} = w_0
\]
Further analysis of the recursive problem implies that we must have
\[ e^{-\psi c} = \beta \hat{R} \int e^{-\frac{\beta}{1+\hat{R}} \psi[y_{0}-T(y_{0})-(1+\hat{R})(c+k')]} + \varphi g(y|k') \, dy \]
Taking log from both sides
\[ -\psi c = \log(\beta \hat{R}) + \psi \hat{R} c - \hat{R} \psi k' + \varphi + \log \int e^{-\frac{\beta}{1+\hat{R}} \psi[y_{0}-T(y_{0})]} g(y|k') \, dy \]
Hence,
\[ -\psi c = \frac{1}{1 + \hat{R}} \log(\beta \hat{R}) \int e^{-\frac{\beta}{1+\hat{R}} \psi[y_{0}-T(y_{0})]} + \varphi g(y|k') \, dy - \psi \frac{\hat{R}}{1 + \hat{R}} a \]
Replacing in the value function, we get
\[ e^\varphi = (1 + \frac{1}{\hat{R}}) \left( \beta \hat{R} \right)^{\frac{1}{1+\hat{R}}} \int e^{-\frac{\beta}{1+\hat{R}} \psi[y_{0}-T(y_{0})]} + \frac{\varphi}{1+\hat{R}} g(y|k') \, dy \]
Therefore,
\[ e^{\varphi} = (1 + \frac{1}{\hat{R}}) \left( \beta \hat{R} \right)^{\frac{1}{1+\hat{R}}} \int (-w^*(y))^{\frac{1}{1+\hat{R}}} e^{\frac{\beta}{1+\hat{R}} \psi \kappa} g(y|k') \, dy \]
which gives \( \varphi \) as a function of \( \kappa \). Hence, for any value of \( \kappa \), \( B_0 \) must be chosen so that
\[ (-w^*(y))^{\frac{\beta}{1+\hat{R}} \psi \kappa - \hat{R} B_0 + \varphi} = w_0 \]
This completes the proof.
QED.

A.2 Sufficient Conditions for FOA

In this section, we provide sufficient conditions for validity of the FOA. To gain insight, we start from a two shock example and extend the derived sufficient conditions to the general case.

A Two Shock Example.

To address the adverse selection problem, we consider a simple example in which the FOA is evidently valid regarding the moral hazard problem. Suppose that the output from the project can only take two value \( \{0, \bar{y}\} \) where \( \Pr(y = \bar{y}|\theta, k) = \theta^\alpha k^\alpha \). In this case, the local incentive compatibility constraints become
\[
\begin{align*}
    u'(c_0(\theta)) &= \beta \alpha \theta^\alpha k_1(\theta)^{\alpha-1} [u(c_1(\theta, \bar{y})) - u(c_1(\theta, 0))] \\
    u'(c_0(\theta)) [c_0'(\theta) + k_1'(\theta)] &= \beta \theta^\alpha k_1(\theta)^\alpha u'(c_1(\theta, \bar{y})) c_{1\theta}(\theta, \bar{y}) \\
    &+ \beta (1 - \theta^\alpha k_1(\theta)^\alpha) u'(c_1(\theta, 0)) c_{1\theta}(\theta, 0) = 0
\end{align*}
\]
In this case, if an agent of type $\theta$ pretends to be $\hat{\theta}$, his optimal investment is given by $\tilde{k}(\theta, \hat{\theta})$ where

$$u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}(\theta, \hat{\theta})) = \beta\theta^\alpha \tilde{k}(\theta, \hat{\theta})^{\alpha-1} \left[ u(c_1(\hat{\theta}, \bar{y})) - u(c_1(\hat{\theta}, 0)) \right] \quad (A.15)$$

We claim that in this case, if $c_0(\theta) + k_1(\theta)$ and $u(c_1(\hat{\theta}, \bar{y})) - u(c_1(\hat{\theta}, 0))$ are both increasing functions of $\theta$, then the FOA is valid. To show the validity of FOA under this assumption, we must show that

$$u(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}(\theta, \hat{\theta})) + \beta\theta^\alpha \tilde{k}(\theta, \hat{\theta})^{\alpha} u(c_1(\hat{\theta}, \bar{y})) + \beta(1 - \theta^\alpha \tilde{k}(\theta, \hat{\theta})^{\alpha}) u(c_1(\hat{\theta}, 0)) \leq U(\theta)$$

We illustrate this for the case where $\hat{\theta} > \theta$ by showing that the LHS is decreasing in $\theta$. The case with $\hat{\theta} < \theta$ can be shown in the same way. To do this, we first show that $\tilde{k}(\theta, \hat{\theta}) \leq k_1(\hat{\theta})$. Suppose not. Then,

$$u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}(\theta, \hat{\theta})) \geq u'(c_0(\hat{\theta}))$$

Note that from (A.15) and (A.13), we must have

$$\theta^\alpha \tilde{k}(\theta, \hat{\theta})^{\alpha-1} \geq \hat{\theta}^\alpha k_1(\hat{\theta})^{\alpha-1}$$

or

$$\frac{k_1(\hat{\theta})}{\tilde{k}(\theta, \hat{\theta})} \geq \left( \frac{\hat{\theta}}{\theta} \right)^\frac{\alpha}{1-\alpha} > 1$$

which is a contradiction. Hence $\tilde{k}(\theta, \hat{\theta}) < k_1(\hat{\theta})$. Note that given this, $\theta^\alpha \tilde{k}(\theta, \hat{\theta})^{\alpha-1} < \hat{\theta}^\alpha k_1(\hat{\theta})^{\alpha-1}$ and hence, $\theta^\alpha \tilde{k}(\theta, \hat{\theta})^{\alpha} < \hat{\theta}^\alpha k_1(\hat{\theta})^{\alpha}$. This result is intuitive. In fact, if an agent with a lower productivity pretends to have a higher productivity, since his marginal return is lower, he will invest less than the agent with high productivity and will enjoy more consumption today. Therefore, since $c_0'(\hat{\theta}) + k_1'(\hat{\theta}) \geq 0$ by assumption,

$$u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}(\theta, \hat{\theta})) \left[ c_0'(\hat{\theta}) + k_1'(\hat{\theta}) \right] \leq u'(c_0(\hat{\theta})) \left[ c_0'(\hat{\theta}) + k_1'(\hat{\theta}) \right]$$

Moreover, since $u(c_1(\theta, \bar{y}) - u(c_1(\theta, 0)))$ is increasing, $u'(c_1(\theta, \bar{y})) c_1(\theta, \bar{y}) > u'(c_1(\theta, 0)) c_1(\theta, 0)$. Hence,

$$\beta\theta^\alpha \tilde{k}(\theta, \hat{\theta})^{\alpha} u'(c_1(\hat{\theta}, \bar{y})) c_1(\theta, \bar{y}) + \beta(1 - \theta^\alpha \tilde{k}(\theta, \hat{\theta})^{\alpha}) u'(c_1(\hat{\theta}, 0)) c_1(\theta, 0)$$

$$\leq \beta \hat{\theta}^\alpha k_1(\hat{\theta})^{\alpha} u'(c_1(\hat{\theta}, \bar{y})) c_1(\theta, \bar{y}) + \beta(1 - \hat{\theta}^\alpha k_1(\hat{\theta})^{\alpha}) u'(c_1(\hat{\theta}, 0)) c_1(\theta, 0)$$
The above inequalities together with (A.14) implies that

\[ u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}(\hat{\theta}, \theta)) \left[ c'_0(\hat{\theta}) + k'_1(\hat{\theta}) \right] + \beta \theta \alpha \tilde{k}(\hat{\theta}, \theta)^\alpha u'(c_1(\hat{\theta}, \bar{y})) c_{1\theta}(\hat{\theta}, \bar{y}) \]

\[ + \beta (1 - \theta \alpha \tilde{k}(\hat{\theta}, \theta)^\alpha) u'(c_1(\hat{\theta}, 0)) c_{1\theta}(\hat{\theta}, 0) \leq 0 \]

The above expression coincides with \( \frac{\partial}{\partial \hat{\theta}} U(\theta, \hat{\theta}) \). Hence, for all \( \hat{\theta} > \theta \), \( \frac{\partial}{\partial \hat{\theta}} U(\theta, \hat{\theta}) \leq 0 \) and therefore, \( U(\theta, \hat{\theta}) \leq U(\theta) \). A similar argument works for \( \hat{\theta} < \theta \). So the required assumptions on endogenous variables that can be checked are

1. Total transfers in the first period must be increasing with type, i.e., \( c'_0(\hat{\theta}) + k'_1(\hat{\theta}) \geq 0 \),
2. \( u(c_1(\theta, \bar{y})) - u(c_1(\theta, 0)) \) must be increasing in \( \theta \).

QED.

The above example is useful since it identifies the main forces leading to validity of the FOA regarding adverse selection. In fact, there are two steps in proving the validity of the FOA. First, we need to show that the optimal choice of \( \tilde{k}(\theta, \hat{\theta}) \) when agent of type \( \theta \) pretends to be \( \hat{\theta} \) is monotone decreasing in \( \hat{\theta} \). This imposes certain restrictions on the schedule \( c_1(\theta, y) \). Second, we should show that given the monotonicity of \( \tilde{k}(\theta, \hat{\theta}) \), \( \int_0^{\bar{y}} u'(c_1(\hat{\theta}, y)) c_{1\theta}(\hat{\theta}, y) g(y|k(\hat{\theta}, \theta), \theta) dy \) is monotone decreasing in \( \hat{\theta} \). In this case, a sufficient assumption is for \( u'(c_1(\theta, y)) c_{1\theta}(\theta, y) \) to be increasing in \( y \). We can summarize this discussion in the following lemma:

**Lemma A.1** Suppose that an allocation \( \{c_0(\theta), c_1(\theta, y), k_1(\theta)\}_{(\theta, y) \in [\theta, \bar{\theta}] \times [0, \bar{y}]} \) satisfies the following:

1. The function \( \frac{\partial}{\partial k} \int_0^{\bar{y}} u'(c_1(\hat{\theta}, y)) g(y|k, \theta) dy \) is increasing in \( \theta \) and decreasing in \( k \), for all \( \hat{\theta}, \theta \), and \( k \),
2. Transfers in first period, \( c_0(\theta) + k_1(\theta) \), is increasing in \( \theta \),
3. The function \( u'(c_1(\theta, y)) c_{1\theta}(\theta, y) \) is increasing in \( y \) for all \( \theta, y \),
4. The allocation is locally incentive compatible,

Then, the allocation is incentive compatible.
Moreover, hence, from assumption 3, \( \tilde{k}(\theta, \hat{\theta}) \) is unique and is given by

\[
u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}(\theta, \hat{\theta})) = \beta \int_0^{\tilde{y}} u(c_1(\hat{\theta}, y))g_k(y|\tilde{k}(\theta, \hat{\theta}), \theta)dy \]

We first prove the claim for the case with \( \hat{\theta} > \theta \). As in the example in the text, we start by showing \( \tilde{k}(\theta, \hat{\theta}) \leq k_1(\theta) \). Suppose not. That is, \( \tilde{k}(\theta, \hat{\theta}) > k_1(\theta) \). This implies that

\[u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}(\theta, \hat{\theta})) > u'(c_0(\hat{\theta}))\]

Therefore,

\[
\int_0^{\tilde{y}} u(c_1(\hat{\theta}, y))g_k(y|\tilde{k}(\theta, \hat{\theta}), \theta)dy > \int_0^{\tilde{y}} u(c_1(\hat{\theta}, y))g_k(y|k_1(\hat{\theta}), \theta)dy 
\]  

(A.16)

By Assumption 1, the function \( \Psi(k, \theta; \hat{\theta}) = \int_0^{\hat{\theta}} u(c_1(\hat{\theta}, y))g_k(y|k, \theta)dy \) is increasing in \( \theta \) and decreasing in \( k \). Since \( \hat{\theta} > \theta \) and \( \tilde{k}(\theta, \hat{\theta}) > k_1(\theta) \), \( \Psi(k_1(\theta), \hat{\theta}) \geq \Psi(\tilde{k}(\theta, \hat{\theta}), \theta) \) which is a contradiction to (A.16). Therefore, we must have \( \tilde{k}(\theta, \hat{\theta}) \leq k_1(\theta) \). Therefore, from assumption 2 we must have

\[u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}(\theta, \hat{\theta})) \left[ c_0'(\hat{\theta}) + k_1'(\hat{\theta}) \right] \leq u'(c_0(\hat{\theta})) \left[ c_0'(\hat{\theta}) + k_1'(\hat{\theta}) \right]\]

Moreover, \( \tilde{k}(\theta, \hat{\theta}) \leq k_1(\hat{\theta}) \) and \( \hat{\theta} > \theta \) implies that \( G(y|k_1(\hat{\theta}), \hat{\theta}) \geq_{FOSD} G(y|\tilde{k}(\theta, \hat{\theta}), \theta) \).

Hence, from assumption 3,

\[
\int_0^{\tilde{y}} u'(c_1(\hat{\theta}, y))c_{1\theta}(\hat{\theta}, y)g(y|\tilde{k}(\theta, \hat{\theta}), \theta)dy \leq \int_0^{\tilde{y}} u'(c_1(\hat{\theta}, y))c_{1\theta}(\hat{\theta}, y)g(y|k_1(\hat{\theta}), \theta)dy
\]

The above inequalities together with the local incentive constraint (2.6),

\[
u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}(\theta, \hat{\theta})) \left[ c_0'(\hat{\theta}) + k_1'(\hat{\theta}) \right] + \int_0^{\tilde{y}} u'(c_1(\hat{\theta}, y))c_{1\theta}(\hat{\theta}, y)g(y|\tilde{k}(\theta, \hat{\theta}), \theta)dy \leq 0
\]

(A.17)

Note that if we define \( U(\theta, \hat{\theta}) \) as follows

\[
U(\theta, \hat{\theta}) = \max_k u(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}) + \int_0^{\tilde{y}} u'(c_1(\hat{\theta}, y))g(y|\tilde{k}, \theta)dy
\]

\[
= u(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}(\theta, \hat{\theta})) + \int_0^{\tilde{y}} u'(c_1(\hat{\theta}, y))g(y|\tilde{k}(\theta, \hat{\theta}), \theta)dy
\]
then $\frac{\partial}{\partial \hat{\theta}} U(\theta, \hat{\theta})$ is given by the LHS of the above inequality (A.17). Hence, for all $\hat{\theta} > \theta$, $\frac{\partial}{\partial \hat{\theta}} U(\theta, \hat{\theta}) \leq 0$ and therefore

$$U(\theta, \hat{\theta}) \leq U(\theta, \theta) = U(\theta)$$

When $\hat{\theta} < \theta$, a similar argument as above shows that $\tilde{k}(\theta, \hat{\theta}) > k_1(\hat{\theta})$. Therefore, $G(y|k_1(\hat{\theta}), \hat{\theta}) \preceq_{FOSD} G(y|\tilde{k}(\theta, \hat{\theta}), \theta)$ and hence

$$\int_{0}^{\tilde{y}} u'(c_1(\hat{\theta}, y)) c_{1\theta}(\hat{\theta}, y) g(y|\tilde{k}(\theta, \hat{\theta}), \theta) dy \geq \int_{0}^{\tilde{y}} u'(c_1(\hat{\theta}, y)) c_{1\theta}(\hat{\theta}, y) g(y|k_1(\hat{\theta}), \hat{\theta}) dy$$

Moreover,

$$u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}(\theta, \hat{\theta})) \left[ c_0'(\hat{\theta}) + k_1'(\hat{\theta}) \right] \geq u'(c_0(\hat{\theta})) \left[ c_0(\hat{\theta}) + k_1'(\hat{\theta}) \right]$$

Hence

$$u'(c_0(\hat{\theta}) + k_1(\hat{\theta}) - \tilde{k}(\theta, \hat{\theta})) \left[ c_0'(\hat{\theta}) + k_1'(\hat{\theta}) \right] + \int_{0}^{\tilde{y}} u'(c_1(\hat{\theta}, y)) c_{1\theta}(\hat{\theta}, y) g(y|\tilde{k}(\theta, \hat{\theta}), \theta) dy \geq 0$$

That is, for $\hat{\theta} < \theta$, $\frac{\partial}{\partial \hat{\theta}} U(\theta, \hat{\theta}) \geq 0$. Hence, $U(\theta, \hat{\theta}) \leq U(\theta, \theta) = U(\theta)$ for all $\hat{\theta} < \theta$.

QED.

Note that condition 1 implies that given a report $\hat{\theta}$ (possibly $\theta$), there is a unique level of investment that maximizes utility. Hence, this assumption resolves the moral hazard issue as well. In the two shock example given above, it is satisfied since for any $c_1(\hat{\theta}, y) > c_1(\hat{\theta}, 0)$, since the function $\frac{\partial}{\partial k} \{\theta^\alpha k^\alpha u(c_1(\hat{\theta}, y)) + (1 - \theta^\alpha k^\alpha) u(c_1(\hat{\theta}, 0))\}$ is increasing in $\theta$ and decreasing in $k$ due to decreasing returns to scale. See [Jewitt, 1988] for an extensive discussion of assumptions on fundamentals that lead to assumption 1.

Unfortunately, condition 1 is a complicated condition that cannot be easily checked. Below, we provide further restriction on the distribution function $g(y|k, \theta)$ that makes checking condition 1 easier.

**Lemma A.2** Suppose that for all $\hat{\theta}$, $c_1(\hat{\theta}, y)$ is increasing in $y$ and $g(\cdot|k, \theta)$ satisfies the following:

1. The function $\frac{\partial}{\partial y} g(y|k, \theta)$ is increasing in $y$,
2. The function $\frac{\partial}{\partial k} g(y|k, \theta)$ is decreasing in $y$. 

Then, condition 1 in lemma A.1 is satisfied.

**Proof.** First, note that
\[ \int_0^y g(y|k, \theta) \, dy = 1 \]
Therefore
\[ \int_0^y g_k(y|k, \theta) \, dy = \int_0^y g_{k\theta}(y|k, \theta) \, dy = \int_0^y g_{kk}(y|k, \theta) \, dy = 0 \]
Now define \( \Psi(k, \theta; \hat{\theta}) = \int_0^y u(c_1(\hat{\theta}, y)) g_k(y|k, \theta) \, dy \). Then,
\[
\Psi_k(k, \theta; \hat{\theta}) = \int_0^y u(c_1(\hat{\theta}, y)) g_{kk}(y|k, \theta) \, dy = \text{Cov}(u(c_1(\theta, y)), \frac{g_{kk}(y|\theta, k)}{g(y|\theta, k)}) \\
\Psi_\theta(k, \theta; \hat{\theta}) = \int_0^y u(c_1(\hat{\theta}, y)) g_{k\theta}(y|k, \theta) \, dy = \text{Cov}(u(c_1(\theta, y)), \frac{g_{k\theta}(y|\theta, k)}{g(y|\theta, k)})
\]
Therefore, the above assumptions imply that \( \Psi_k < 0 \) and \( \Psi_\theta > 0 \).

Q.E.D.
Appendix B

Appendix to Chapter 3

B.1 Proofs

B.1.1 Proof of Proposition 3.6

We first prove the following lemma:

**Lemma B.1** Suppose Assumptions 3.3 and 3.4 hold, then the value function and the policy functions satisfy the following properties:

\[
\begin{align*}
\lim_{w \to -\infty} v(w) &= - \sum_i \pi_i \theta_i \\
\lim_{w \to -\infty} c(w, \theta_i) &= 0 \\
\lim_{w \to -\infty} n(w, \theta_i) &= 0 \\
\lim_{w \to -\infty} l(w, \theta_i) &= 1
\end{align*}
\]

**Proof.**

Consider the following set of function:

\[
S = \left\{ \hat{v} ; \hat{v} \in C(\mathbb{R}_-) , : \hat{v} \text{ weakly increasing } : \lim_{w \to -\infty} \hat{v}(w) = - \sum_i \pi_i \theta_i \right\}
\]

Moreover define the following mapping on \( S \) as

\[
T\hat{v}(w) = \min_j \sum \pi_j \left[ c_j - \theta_j l_j + \frac{1}{R} n_j \hat{v}(w'_j) \right]
\]

192
\[
\sum_j \pi_j \left[ u(c_j) + h(1 - l_j - b n_j) + \beta n_j^n w'_j \right] \geq w
\]

\[
u(c_j) + h(1 - l_j - b n_j) + \beta n_j^n w'_j \geq u(c_i) + h(1 - \frac{\theta_i l_i}{\theta_j} - b n_i) + \beta n_i^n w'_i, \quad \forall j > i
\]

\[
l_j + b n_j \leq 1
\]

\[
e_j, l_j, n_j \geq 0
\]

We first show that the solution to the above program has the claimed property for the policy function and that \( \hat{T} \) satisfies the claimed property. Then, since \( S \) is closed and \( T \) preserves \( S \), by Contraction Mapping Theorem we have that the fixed point of \( T \) belongs to \( S \).

Now, suppose the claim about policy function for fertility, does not hold. Then there exists a sequence \( w_n \to -\infty \) such that for some \( i \), \( n(w_n, \theta_i) \to \bar{n}_i > 0 \). For each \( j \neq i \), define \( \bar{n}_j = \liminf_{n \to \infty} n(w_n, \theta_j) \), then we must have

\[
\liminf_{n \to \infty} T \hat{\nu}(w_n) \geq \sum_j \pi_j \left[ -\theta_j (1 - b \bar{n}_j) + \frac{1}{R} \bar{n}_j \left[ -\sum_k \pi_k \theta_k \right] \right]
\]

Note that by Assumption 3.4, we have

\[
b \theta_j > \frac{1}{R} \sum_k \pi_k \theta_k, : \forall j
\]

and therefore, if \( \bar{n}_j \geq 0 \), we must have

\[
-\theta_j + b \bar{n}_j \theta_j - \frac{1}{R} \bar{n}_j \sum_k \pi_k \theta_k \geq -\theta_j
\]

with equality only if \( \bar{n}_j = 0 \). This implies that

\[
\liminf_{n \to \infty} T \hat{\nu}(w_n) > -\sum_j \pi_j \theta_j
\]

since \( \bar{n}_i > 0 \). Now, we construct a sequence of allocation and show that the above cannot be an optimal one. Consider a sequence of numbers \( \epsilon_m \) that converges to zero. Define
\[c_m(\theta_i) = u^{-1}(-\epsilon_m^n)\]
\[n_m(\theta_i) = (n - i)\frac{1}{n}\epsilon_m\]
\[w'_m(\theta_i) = \tilde{w} < 0\]

If \(h\) is bounded above and below, define
\[l_m(\theta_i) = 1 - \epsilon_m - bn_m(\theta_i)\]

By construction,
\[c_m(\theta_i) \to 0\]
\[n_m(\theta_i) \to 0\]
\[l_m(\theta_i) \to 1\]

Moreover,
\[u(c_m(\theta_j)) + \beta n_m(\theta_j)^n w'_m(\theta_j) - u(c_m(\theta_i)) - \beta n_m(\theta_i)^n w'_m(\theta_i) = \beta \tilde{w}\epsilon_m^n(i - j), \forall j > i\]

This expression converges to \(\infty\) and therefore, since \(h\) is bounded above and below for \(m\) large enough, the allocations are incentive compatible.

When, \(h\) is unbounded below, since the utility of deviation is bounded away from \(-\infty\), it is possible to construct a sequence for \(l_m\) that converges to 1. Find \(l_m(\theta_i)\) such that
\[h(1 - l_m(\theta_i) - bn_m(\theta_i)) = \frac{1}{2} \tilde{w}\epsilon_m^n\]

Hence, we have that
\[u(c_m(\theta_j)) + h(1 - l_m(\theta_j) - bn_m(\theta_j)) + \beta n_m(\theta_j)^n w'_m(\theta_j) - u(c_m(\theta_i)) - \beta n_m(\theta_i)^n w'_m(\theta_i) = \tilde{w}\epsilon_m^n(i - j + \frac{1}{2})\]

converges to \(\infty\). Moreover, by definition \(l_m(\theta_i)\) converges to 1 and \(n_m(\theta_i)\) converges to zero and therefore the deviation value for leisure, \(h(1 - \frac{\theta_i l_m(\theta_i)}{\theta_j} - bn_m(\theta_i))\), converges to \(h(1 - \frac{\theta_i}{\theta_j})\). This implies that for \(m\) large enough
\[u(c_m(\theta_j)) + h(1 - l_m(\theta_j) - bn_m(\theta_j)) + \beta n_m(\theta_j)^n w'_m(\theta_j) - u(c_m(\theta_i)) - \beta n_m(\theta_i)^n w'_m(\theta_i) \geq h(1 - \frac{\theta_i l_m(\theta_i)}{\theta_j} - bn_m(\theta_i))\]
and for $m$ large enough the allocation is incentive compatible.

Therefore, the utility from the constructed allocation is the following:

$$\hat{w}_m = c_n^n \left[-1 + \beta \sum_{j} A_j^n \hat{w} + \sum_{j} \pi_j [h(1 - l_m(\theta_j) - b m_m(\theta_j))]\right]$$

It is clear that $\hat{w}_m$’s converge to $-\infty$ and the allocation’s cost converges to $-\sum_j \pi_j \theta_j$.

Now since $\hat{w}_m$ and $w_n$ converge to $-\infty$, there exists subsequences $\hat{w}_{mk}$ and $w_{nk}$ such that $\hat{w}_{mk} \geq w_{nk}$ and therefore by optimality:

$$\sum_j \pi_j \left[c_{mk}(\theta_j) - \theta_j l_{mk}(\theta_j) + \frac{1}{R} n_{mk}(\theta_j) \hat{v}(\hat{w})\right] \geq T \hat{v}(w_{nk})$$

and therefore,

$$-\sum_k \pi_k \theta_k \geq \lim inf_{n \to \infty} T \hat{v}(w_n) > -\sum_k \pi_k \theta_k$$

and we have a contradiction. This completes the proof.

Q.E.D.

Since $h$ is unbounded below, given the above lemma for $w \in \mathbb{R}_-$ low enough, allocations should be interior and since $v$ is differentiable, positive lagrange multipliers $\lambda, \mu(i, j)_{i > j}$ must exist such that

$$u'(c(w, \theta_i)) \left[\pi_i \lambda(w) + \sum_{j < i} \mu(i, j; w) - \sum_{j > i} \mu(j, i; w)\right] = \pi_i$$

$$\beta n(w, \theta_i)^{n-1} \left[\pi_i \lambda(w) + \sum_{j < i} \mu(i, j; w) - \sum_{j > i} \mu(j, i; w)\right] = \pi_i \frac{1}{R} v'(w'(w, \theta_i))$$

$$h'(1 - l(w, \theta_i)) - bn(w, \theta_i) \left[\pi_i \lambda(w) + \sum_{j < i} \mu(i, j; w)\right]$$

$$- \sum_{j > i} \mu(j, i; w) \frac{\theta_i}{\theta_j} h'(1 - \frac{\theta_i l(w, \theta_i)}{\theta_j} - bn(w, \theta_i)) = \pi_i \theta_i$$
\[
\{ -bh'(1-l(w, \theta_i) - bn(w, \theta_i)) + \beta \eta n(w, \theta_i)^{n-1} w'(w, \theta_i) \} \left[ \pi_i \lambda(w) + \sum_{j<i} \mu(i, j; w) \right]
\]

\[- \sum_{j>i} \mu(j, i; w) \left\{ -bh'(1- \frac{\theta_i l(w, \theta_i)}{\theta_j} - bn(w, \theta_i)) + \beta \eta n(w, \theta_i)^{n-1} w'(w, \theta_i) \right\} \]

\[= \pi_i \frac{1}{R} v(w'(w, \theta_i)) \]

By Lemma B.1, we must have:

\[\lim_{w \to -\infty} c(w, \theta_j) = 0\]
\[\lim_{w \to -\infty} n(w, \theta_j) = 0\]
\[\lim_{w \to -\infty} l(w, \theta_j) = 1\]

Then for every \(\epsilon > 0\), there exists \(W\) such that for all \(w < W\), we have \(u'(c(w, \theta_j)) > \frac{N}{\epsilon}\).

This implies that

\[\lambda(w) = \sum_j \frac{\pi_j}{u'(c(w, \theta_j))} < \frac{\epsilon}{N}\]
\[\pi_n \lambda(w) + \sum_{j<n} \mu(n, j; w) = \frac{\pi_n}{u'(c(w, \theta_n))} < \frac{\epsilon}{N}\]
\[\Rightarrow \mu(n, j; w) < \frac{\epsilon}{N}\]

In addition,

\[\pi_{n-1} \lambda(w) + \sum_{j<n-1} \mu(n - 1, j; w) - \mu(n, n - 1; w) = \frac{\pi_{n-1}}{u'(c(w, \theta_{n-1}) < \frac{\epsilon}{N}}\]
\[\Rightarrow \mu(n - 1, j; w) < \frac{2\epsilon}{N}\]

By an inductive argument, we have

\[\mu(i, j; w) < \frac{a_i \epsilon}{N}\]

where \(a_{n-1} = 1, a_{n-2} = 2, a_{n-i} = a_{n-1} + \cdots + a_{n-i+1} + 1\). If we pick \(N\) so that \(a_1 < N\), we have that

\[\mu(i, j; w) < \epsilon, \forall w < W.\]

Next, we define the type specific resetting values, \(w_i\), as the values of \(w\) that solve the following equations:

\[\eta v'(w) - v(w) = bR \theta_i.\]
Under our convexity assumptions, the left hand side \(- \eta v'(w)w - v(w)\) is strictly increasing in \(w\), so that if a solution exists, it is unique.

From Proposition 3.5 (resetting at top) we know that there is a \(w_0\) such that:

\[ \eta v'(w)w - v(w) = bR\theta_1. \]

Moreover, from the first order conditions, we know that

\[ \eta v'(w'(w, \theta_i))w'(w, \theta_i) - v(w'(w, \theta_i)) \leq bR\theta_1. \]

Therefore, by the intermediate value theorem, there exists a unique \(w_i > -\infty\) which satisfies

\[ \eta v'(w_i)w_i - v(w_i) = bR\theta_i. \]

Moreover, by substituting first order conditions, we get

\[
\begin{align*}
\pi_i &\geq \pi_i \frac{1}{R} \eta v'(w'(w, \theta_i))w'(w, \theta_i) - \pi_i \frac{1}{R} v(w'(w, \theta_i)) \\
&= \pi_i \theta_1 - b \sum_{j > i} \left(1 - \frac{\theta_i}{\theta_j}\right) \mu(i, j) h'(1 - \frac{\theta_i l(w, \theta_i)}{\theta_j} - bn(w, \theta_i)) \\
&> \pi_i \theta_1 - b \sum_{j > i} \left(1 - \frac{\theta_i}{\theta_j}\right) h'(1 - \frac{\theta_i l(w, \theta_i)}{\theta_j} - bn(w, \theta_i))
\end{align*}
\]

Since hours converges to 1, the term multiplied by \(\epsilon\) in the above expression is bounded away from \(\infty\) as \(w \to -\infty\). From this it follows that

\[ \lim_{w \to -\infty} \eta v'(w'(w, \theta_i))w'(w, \theta_i) - v(w'(w, \theta_i)) = bR\theta_i. \]

Continuity of \(v'\) implies that

\[ \lim_{w \to -\infty} w'(w, \theta_i) = w_i. \]

\[ \blacksquare \]

**B.2 Implementation**

**B.2.1 Distortions**

The constrained efficient allocation \(c^*_1(\theta), l^*(\theta), n^*(\theta), c^*_2(\theta)\) solves the following problem

\[
\sum_{i=H,L} \pi_i \left[ u(c_i) + h(1 - l_i - bn_i) + \beta n_i^\eta u(c_{2i}) \right]
\]
s.t. \[
\sum_{i=H,L} \pi_i \left[ c_{1i} + \frac{1}{R} n_i c_{2i} \right] \leq \sum_{i=H,L} \pi_i \theta_i l_i + RK_0
\]
\[
u(c_{1H}) + h(1 - l_H - bn_H) + \beta n_H^\eta u(c_{2H}) \geq \nu(c_{1L}) + h(1 - \frac{\theta_l l_L}{\theta_H} - bn_L) + \beta n_L^\eta u(c_{2L}).
\]

Assuming the solution is interior, it satisfies the following first order conditions:
\[
u'(c_{1H}^*) (1 + \frac{\mu}{\pi_H}) = \lambda
\]
\[
u'(c_{1L}^*) (1 - \frac{\mu}{\pi_L}) = \lambda
\]
\[
\theta_H h'(1 - l_H^* - bn_H^*) (1 + \frac{\mu}{\pi_H}) = \lambda \theta_H
\]
\[
\theta_L h'(1 - l_L^* - bn_L^*) (1 - \frac{\mu}{\pi_L} h'(1 - \frac{\theta_l l_L^*}{\theta_H} - bn_L^*)) = \lambda \theta_L
\]
\[
\left[ -bh'(1 - l_H^* - bn_H^*) + \beta \eta n_H^* \eta^{-1} u(c_{2H}^*) \right] (1 + \frac{\mu}{\pi_H}) = \lambda \frac{1}{R} c_{2H}^*
\]
\[
\left[ -bh'(1 - l_L^* - bn_L^*) + \beta \eta n_L^* \eta^{-1} u(c_{2L}^*) \right] \frac{\mu}{\pi_L} = \lambda \frac{1}{R} c_{2L}^*
\]

Now suppose that we want to implement the above allocation with a tax in first period of the form \(T(y, n)\). Then consumer’s problem is the following:
\[
\max u(c_1) + h(1 - l - bn) + \beta n^\eta u(c_2)
\]

s.t.
\[
c_1 + k_1 \leq Rk_0 + \theta l - T(\theta l, n, c_2)
\]
\[
t c_2 \leq Rk_1
\]

As a first step, we assume that \(T\) is differentiable and that \(y\) is interior for both types.

Then the FOCs are the following:
\[ u'(c_1) = \lambda_1 \]
\[ h'(1 - l - bn) = \lambda_1 \theta(1 - T_y(\theta l, n, c_2)) \]
\[ R\lambda_1 = \lambda_2 \]
\[ -bh'(1 - l - bn) + \beta \eta n^{\eta-1} u(c_2) = \lambda_2 c_2 + \lambda_1 T_n(\theta l, n, c_2) \]
\[ \beta n^\eta u'(c_2) = n\lambda_2 + \lambda_1 T_{c_2}(\theta l, n, c_2) \]

Comparing the FOC’s for the planner with these, we see immediately that \( T_n(\theta_H l_H, n_H, c_{2H}) = T_y(\theta_H l_H, n_H, c_{2H}) = T_{c_2}(\theta_H l_H, n_H, c_{2H}) = 0 \) – there are no (marginal) distortions on the decisions of the agent with the high shock. Moreover, from the FOC’s of the planner’s problem we get:

\[
\left[ -bh'(1 - l_L^* - bn_L^*)\pi_L + bh'(1 - \frac{\theta L^* L}{\theta H} - bn_L^*)\mu \right] \frac{1}{\pi_L - \mu} + \beta \eta n^{\eta-1} u(c_{2L}^*)
\]
\[
= \frac{1}{R} u'(c_{1L}^*) c_{2L}^*.
\]

We know that

\[
1 - \frac{\theta L^* L}{\theta H} - bn_L^* > 1 - l_L^* - bn_L^*
\]
\[
\Rightarrow h'(1 - \frac{\theta L^* L}{\theta H} - bn_L^*)\mu < h'(1 - l_L^* - bn_L^*)\mu
\]
\[
bh'(1 - \frac{\theta L^* L}{\theta H} - bn_L^*)\mu - bh'(1 - l_L^* - bn_L^*)\pi_L < bh'(1 - l_L^* - bn_L^*)\mu - bh'(1 - l_L^* - bn_L^*)\pi_L
\]
\[
\left[ -bh'(1 - l_L^* - bn_L^*)\pi_L + bh'(1 - \frac{\theta L^* L}{\theta H} - bn_L^*)\mu \right] \frac{1}{\pi_L - \mu} < -bh'(1 - l_L^* - bn_L^*).
\]

Hence,

\[
\frac{1}{R} u'(c_{1L}^*) c_{2L}^* - \beta \eta n^{\eta-1} u(c_{2L}^*) < -bh'(1 - l_L^* - bn_L^*).\]

From the FOC of the consumer’s problem, we have

\[
0 < T_n(\theta L^* L, n_L^*, c_{2L}^*) = -bh'(1 - l_L^* - bn_L^*) - \frac{1}{R} u'(c_{1L}^*) c_{2L}^* + \beta n^{\eta-1} u(c_{2L}^*).\]
Next, we turn to $T_y(\theta_L l^*_L, n^*_L, c^*_2L)$. From above, we have:

\[
\begin{align*}
    h'(1 - l^*_L - b n^*_L) \pi_L - \mu \frac{\theta_L}{\theta_H} h'(1 - \frac{\theta_L l^*_L}{\theta_H} - b n^*_L) & = \lambda \theta_L \pi_L \\
    h'(1 - l^*_L - b n^*_L) \left[ \pi_L - \mu \frac{\theta_L}{\theta_H} \right] & < \lambda \theta_L \pi_L \\
    h'(1 - l^*_L - b n^*_L) \left[ \pi_L - \mu \frac{\theta_L}{\theta_H} \right] & < \theta_L u'(c^*_1L)(\pi_L - \mu) \\
    h'(1 - l^*_L - b n^*_L) & < \theta_L u'(c^*_1L) \frac{(\pi_L - \mu)}{\pi_L - \mu \frac{\theta_L}{\theta_H}} \\
    h'(1 - l^*_L - b n^*_L) & < \theta_L u'(c^*_1L)
\end{align*}
\]

Thus, from the FOC’s of the agent’s problem, we see that $T_y(\theta_L l^*_L, n^*_L, c^*_2L) > 0$.

Finally, since there are only shocks in the first period, it is never optimal to distort the savings decision for either type. Because of this, it follows that $T_c(\theta_L l^*_L, n^*_L, c^*_2L) = 0$.

### B.2.2 Proof of Remark 3.12

First we show, using incentive compatibility, that $T_L(y^*_L, n^*_L, c^*_2L) = T_H(y^*_L, n^*_L, c^*_2L)$. We know that at the constrained efficient allocation, type $\theta_H$ is indifferent between the allocations $(c^*_1H, y^*_H, n^*_H, c^*_2H)$ and $(c^*_1L, y^*_L, n^*_L, c^*_2L)$. Hence we have the following equality:

\[
\bar{u}_H = u(c^*_1H) + h(1 - \frac{y^*_H}{\theta_H} - b n^*_H) + \beta n^*H u(c^*_2H) = u(c^*_1L) + h(1 - \frac{y^*_L}{\theta_H} - b n^*_L) + \beta n^*L u(c^*_2L)
\]

Replace for $c^*_1H$ and $c^*_1L$ from budget constraints to get

\[
\bar{u}_H = u(y^*_H - T_H(y^*_H, n^*_H, c^*_2H) - \frac{1}{R} n^*_H c^*_2H) + h(1 - \frac{y^*_H}{\theta_H} - b n^*_H) + \beta n^*H u(c^*_2H)
\]

\[
= u(y^*_L - T_L(y^*_L, n^*_L, c^*_2L) - \frac{1}{R} n^*_L c^*_2L) + h(1 - \frac{y^*_L}{\theta_H} - b n^*_L) + \beta n^*L u(c^*_2L)
\]

Moreover, from the definition of $T_H$ we know that

\[
\bar{u}_H = u(y^*_L - T_H(y^*_L, n^*_L) - \frac{1}{R} n^*_L c^*_2L) + h(1 - \frac{y^*_L}{\theta_H} - b n^*_L) + \beta n^*L u(c^*_2L)
\]

Hence, the last two equalities imply that $T_L(y^*_L, n^*_L, c^*_2L) = T_H(y^*_L, n^*_L, c^*_2L)$. 

We can also show that \( T_H(y_H^*, n_H^*, c_{2H}^*) > T_L(y_H^*, n_H^*, c_{2H}^*) \). We show that this holds as long as the upward incentive constraint is slack \( \theta_L \) strictly prefers the allocation \((c_{1L}^*, y_{1L}^*, n_{1L}^*, c_{2L}^*)\) to \((c_{1H}^*, y_{1H}^*, n_{1H}^*, c_{2H}^*)\), i.e.:

\[
\bar{u}_L = u(c_{1L}^*) + h(1 - \frac{y_{1L}^*}{\theta_L}) - bn_{1L}^* + \beta n_{1L}^n u(c_{2L}^*) > u(c_{1H}^*) + h(1 - \frac{y_{1H}^*}{\theta_L} - bn_{1H}^*) + \beta n_{1H}^n u(c_{2H}^*).
\]

Using the budget constraints, we get

\[
\bar{u}_L = u(y_L^* - T_L(y_L^*, n_L^*, c_{2L}^*) - \frac{1}{R} n_L^* c_{2L}^*) + h(1 - \frac{y_{1L}^*}{\theta_L}) - bn_{1L}^* + \beta n_{1L}^n u(c_{2L}^*) > u(y_H^* - T_L(y_H^*, n_H^*, c_{2H}^*) - \frac{1}{R} n_H^* c_{2H}^*) + h(1 - \frac{y_{1H}^*}{\theta_L} - bn_{1H}^*) + \beta n_{1H}^n u(c_{2H}^*).
\]

By the definition of \( T_L \) we have

\[
\bar{u}_L = u(y_H^* - T_L(y_H^*, n_H^*, c_{2H}^*) - \frac{1}{R} n_H^* c_{2H}^*) + h(1 - \frac{y_{1H}^*}{\theta_L} - bn_{1H}^*) + \beta n_{1H}^n u(c_{2H}^*) \]

Hence, we have that \( T_H(y_H^*, n_H^*, c_{2H}^*) > T_L(y_H^*, n_H^*, c_{2H}^*) \).

Given the tax function, the consumer’s problem is the following:

\[
\max_{c_1, y, n, c_2} u(c_1) + h(1 - \frac{y}{\theta} - bn) + \beta n^u(c_2)
\]

s.t. \( c_1 + \frac{1}{R} nc_2 \leq y - T(y, n) \)

From above, we know that \( T(y_H^*, n_H^*, c_{2H}^*) = T_H(y_H^*, n_H^*, c_{2H}^*) \). Hence, type \( \theta_H \) can afford \((c_{1H}^*, y_{1H}^*, n_{1H}^*, c_{2H}^*)\) and \( u(c_{1H}^*, y_{1H}^*, n_{1H}^*, c_{2H}^*; \theta_H) = \bar{u}_H \). Let \((c_1, y, n, c_2)\) be any allocation that satisfies \( c_1 + \frac{1}{R} nc_2 = y - T(y, n, c_2) \). Then,

\[
c_1 + \frac{1}{R} nc_2 = y - \max\{T_L(y, n, c_2), T_H(y, n, c_2)\} \leq y - T_H(y, n, c_2)
\]

But by definition of \( T_H \), \( u(c_1, y, n, c_2; \theta_H) \) can be at most \( \bar{u}_H \).

Using a similar argument we can show that type \( \theta_L \) can afford \((c_{1L}^*, y_{1L}^*, n_{1L}^*, c_{2L}^*)\) and \( u(c_{1L}^*, y_{1L}^*, n_{1L}^*, c_{2L}^*; \theta_L) = \bar{u}_L \). Moreover, any allocation that satisfies the budget constraint has utility at most \( \bar{u}_L \).

The differentiability of \( T \) and properties of marginal taxes follow from the discussion above.

Q.E.D.
Appendix C

Appendix to Chapter 4

C.1 Proofs

C.1.1 Proof of Lemma 4.2

The first order condition associated with the planning problem (4.7) is given by (we are suppressing $\theta$)

\begin{align*}
g - \alpha - \mu' &= 0 \quad \text{(C.1)} \\
\alpha - \lambda f &= 0 \quad \text{(C.2)} \\
-\psi' \frac{y(t)\gamma - 1}{\varphi(t)\gamma} \alpha + \lambda f - \mu \psi \gamma \frac{y(t)\gamma - 1}{\varphi(t)\gamma} \varphi'(t) &= 0, \forall t \leq R \quad \text{(C.3)} \\
- \left[ \psi \frac{y(R)\gamma}{\gamma \varphi(R)\gamma} + \eta \right] \alpha + y(R) \lambda f - \left[ \psi \frac{\varphi'(R) \gamma y(R)\gamma}{\varphi(R) \varphi'(R)\gamma} - \eta' \right] \mu &= 0 \quad \text{(C.4)} \\
\mu(\theta) = \bar{\mu}(\theta) &= 0
\end{align*}

where $\alpha(\theta)$ is the multiplier on the first constraint, $\lambda$ is the multiplier on feasibility, and $\mu(\theta)$ is the multiplier on local incentive constraint. Integrating over equation (C.1) and using the boundary conditions, we have

$$
\int_{\theta}^{\bar{\theta}} g(\theta) \, d\theta - \lambda \int_{\theta}^{\bar{\theta}} f(\theta) \, d\theta = 0
$$

202
and hence $\lambda = 1$, since $G(\bar{\theta}) = F(\bar{\theta}) = 1$. Moreover, we also have
\[
\mu(\theta) = \int_{\bar{\theta}}^{\theta} \left[ g(\theta') - f(\theta') \right] d\theta' = G(\theta) - F(\theta)
\]
and we can rewrite the equation (C.3) as
\[
\psi y(t)^{\gamma-1} \left[ 1 + \frac{G(\theta) - F(\theta)}{f(\theta)} \frac{\varphi_\theta(t)}{\varphi} \right] = 1 \quad (C.5)
\]
which implies the first result in Lemma 4.2. Moreover, since marginal utility of consumption is 1, labor wedge here is defined by $\tau_l = 1 - \psi \frac{y(t)^{\gamma-1}}{\varphi(t)}$ and hence the second result in the lemma follows.

Q.E.D.

C.1.2 Proof of Lemma 4.3

If we replace equation (C.5) evaluated at $t = R(\theta)$ in (C.4), we have
\[
\frac{y(R)}{\gamma} - y(R) = \eta(\theta) - \frac{G(\theta) - F(\theta)}{f(\theta)} \eta'(\theta)
\]
which proves the lemma.

Q.E.D.

C.1.3 Proof of Proposition 4.5

It is sufficient to derive equation (4.12), the rest of the argument is described in the text. Note that wedges are defined as
\[
\tau_r(\theta) y(R(\theta), \theta) = y(R(\theta), \theta) - \left[ \psi \frac{y(R(\theta), \theta)^\gamma}{\gamma \varphi(R(\theta), \theta)^\gamma} + \eta(\theta) \right]
\]
\[
\tau_l(R(\theta), \theta) y(R(\theta), \theta) = y(R(\theta), \theta) - \frac{\psi y(R(\theta), \theta)^\gamma}{\varphi(R(\theta), \theta)^\gamma}
\]
and hence
\[
\tau_r(\theta) y(R(\theta), \theta) - \frac{1}{\gamma} \tau_l(R(\theta), \theta) y(R(\theta), \theta)
\]
\[
= \frac{\gamma - 1}{\gamma} y(R(\theta), \theta) - \eta(\theta)
\]
\[
= - \frac{G(\theta) - F(\theta)}{f(\theta)} \eta'(\theta)
\]
where the last equality follows form Lemma 4.3.

Q.E.D.

C.1.4 Proof of implementation

In the body of the paper, we show that \( \hat{\theta} = \theta \) satisfies the first order conditions. Here, we show that it also satisfies the second order condition and hence it is a local maximizer. Then, we show it is a global maximizer as well.

**Lemma C.1** Suppose that \( R'(\theta) \geq 0 \) and \( \eta'(\theta) \leq 0 \). Then the choice of \( \hat{\theta} = \theta \) is a local maximizer in (4.15).

**Proof.** Note that given the proof of Proposition 4.7, the first order condition evaluated at \( \hat{\theta} = \theta \), is given by

\[
c'(\theta) - \left[ \frac{\psi}{\gamma} y(R(\hat{\theta}), \theta)^{\gamma} + \eta(\theta) \right] R'(\theta) - \int_0^{R(\theta)} \frac{\partial}{\partial \eta} y(t, \theta) \left[ 1 - \frac{\partial}{\partial y} T(t, y(t, \theta)) \right] dt = 0
\]

Moreover, the second derivative of (4.15) at \( \hat{\theta} = \theta \) is given by

\[
\begin{aligned}
&\left[ y(R(\hat{\theta}), \theta) - T(R(\hat{\theta}), y(R(\hat{\theta}), \theta)) - \frac{\psi}{\gamma} y(R(\hat{\theta}), \theta)^{\gamma} - \eta(\theta) \right] R''(\hat{\theta}) \\
&+ \left[ \frac{\partial}{\partial t} y(R(\hat{\theta}), \theta) - \left[ \frac{\partial}{\partial t} + \frac{\partial y(R(\hat{\theta}), \theta)}{\partial t} \times \frac{\partial}{\partial y} \right] T(R(\hat{\theta}), y(R(\hat{\theta}), \theta)) \\
&- \frac{\psi}{\gamma} \frac{y(R(\hat{\theta}), \theta)^{\gamma-1}}{\varphi(R(\hat{\theta}), \theta)} \left\{ \frac{\partial}{\partial t} y(R(\hat{\theta}), \theta) - \frac{y(R(\hat{\theta}), \theta)}{\varphi(R(\hat{\theta}), \theta)} \frac{\partial}{\partial t} \varphi(R(\hat{\theta}), \theta) \right\} \right] R'(\hat{\theta})^2 \\
+c''(\hat{\theta}) - \left[ y(R(\hat{\theta}), \theta) - T(R(\hat{\theta}), y(R(\hat{\theta}), \theta)) \right] R''(\hat{\theta}) \\
- \left[ \frac{\partial}{\partial t} y(R(\hat{\theta}), \hat{\theta}) - \left[ \frac{\partial}{\partial t} + \frac{\partial y(R(\hat{\theta}), \hat{\theta})}{\partial t} \times \frac{\partial}{\partial y} \right] T(R(\hat{\theta}), y(R(\hat{\theta}), \hat{\theta})) \right] R'(\hat{\theta})^2 \\
- \left[ \frac{\partial}{\partial \theta} y(R(\hat{\theta}), \hat{\theta}) - \frac{\partial}{\partial \theta} y(R(\hat{\theta}), \hat{\theta}) \frac{\partial}{\partial y} T(R(\hat{\theta}), y(R(\hat{\theta}), \hat{\theta})) \right] R'(\hat{\theta}) \\
- \frac{\partial}{\partial \theta} y(R(\hat{\theta}), \hat{\theta}) \left[ 1 - \frac{\partial}{\partial y} T(R(\hat{\theta}), y(R(\hat{\theta}), \hat{\theta})) \right] R'(\hat{\theta}) \\
- \int_0^{R(\theta)} \left\{ \frac{\partial^2}{\partial \theta^2} y(t, \hat{\theta}) \left[ 1 - \frac{\partial}{\partial y} T(t, y(t, \hat{\theta})) \right] - \left( \frac{\partial}{\partial \theta} y(t, \hat{\theta}) \right)^2 \frac{\partial^2}{\partial y^2} T(t, y(t, \hat{\theta})) \right\} dt
\end{aligned}
\]
Evaluating the above at \( \hat{\theta} = \theta \) gives us the following expression

\[
\left[ -\frac{\psi}{\gamma} \frac{y(R(\theta), \theta)^{\gamma}}{\varphi(R(\theta), \theta)^{\gamma}} - \eta(\theta) \right] R''(\theta) \\
+ \left[ -\frac{\psi}{\gamma} \frac{y(R(\theta), \theta)^{\gamma-1}}{\varphi(R(\theta), \theta)^{\gamma}} \left\{ \frac{\partial}{\partial t} y(R(\theta), \theta) - \frac{y(R(\theta), \theta)}{\varphi(R(\theta), \theta)} \frac{\partial}{\partial t} \varphi(R(\theta), \theta) \right\} \right] R'(\theta)^2 \\
+ c''(\theta) \\
- 2 \frac{\partial}{\partial \theta} y(R(\theta), \theta) \left[ 1 - \frac{\partial}{\partial y} T(R(\theta), y(R(\theta), \theta)) \right] R'(\theta) \\
- \int_{0}^{R(\theta)} \left\{ \frac{\partial^2}{\partial t^2} y(t, \theta) \left[ 1 - \frac{\partial}{\partial y} T(t, y(t, \theta)) \right] - \left( \frac{\partial}{\partial \theta} y(t, \theta) \right)^2 \frac{\partial^2}{\partial y^2} T(t, y(t, \theta)) \right\} dt = 0
\]

If we take derivative of the incentive constraint in its local form with respect to \( \theta \), we have

\[
c''(\theta) - \left[ \frac{\psi}{\gamma} \frac{y(R(\theta), \theta)^{\gamma}}{\varphi(R(\theta), \theta)^{\gamma}} + \eta(\theta) \right] R''(\theta) \\
+ \left[ -\frac{\psi}{\gamma} \frac{y(R(\theta), \theta)^{\gamma-1}}{\varphi(R(\theta), \theta)^{\gamma}} \left\{ \frac{\partial}{\partial t} y(R(\theta), \theta) - \frac{y(R(\theta), \theta)}{\varphi(R(\theta), \theta)} \frac{\partial}{\partial t} \varphi(R(\theta), \theta) \right\} \right] R'(\theta)^2 \\
+ \left[ -\frac{\psi}{\gamma} \frac{y(R(\theta), \theta)^{\gamma-1}}{\varphi(R(\theta), \theta)^{\gamma}} \left\{ \frac{\partial}{\partial \theta} y(R(\theta), \theta) - \frac{y(R(\theta), \theta)}{\varphi(R(\theta), \theta)} \frac{\partial}{\partial \theta} \varphi(R(\theta), \theta) \right\} - \eta'(\theta) \right] R'(\theta) \\
- \frac{\partial}{\partial \theta} y(R(\theta), \theta) \left[ 1 - \frac{\partial}{\partial y} T(R(\theta), y(R(\theta), \theta)) \right] R'(\theta) \\
- \int_{0}^{R(\theta)} \left\{ \frac{\partial^2}{\partial t^2} y(t, \theta) \left[ 1 - \frac{\partial}{\partial y} T(t, y(t, \theta)) \right] - \left( \frac{\partial}{\partial \theta} y(t, \theta) \right)^2 \frac{\partial^2}{\partial y^2} T(t, y(t, \theta)) \right\} dt = 0
\]
We can regroup the terms in the second order condition and use the above expression to further simplify

\[
\begin{align*}
& c''(\theta) + \left[ -\frac{\psi}{\varphi} \frac{y(R(\theta),\theta)}{\varphi(R(\theta),\theta)} - \eta(\theta) \right] R''(\theta) \\
& + \left[ -\frac{\psi}{\varphi} \frac{y(R(\theta),\theta)}{\varphi(R(\theta),\theta)} \left\{ \frac{\partial}{\partial t} y(R(\theta),\theta) - \frac{y(R(\theta),\theta)}{\varphi(R(\theta),\theta)} \frac{\partial}{\partial t} \varphi(R(\theta),\theta) \right\} \right] R'(\theta)^2 \\
& -2 \frac{\partial}{\partial \theta} y(R(\theta),\theta) \left[ 1 - \frac{\partial}{\partial y} T(R(\theta),y(R(\theta),\theta)) \right] R'(\theta) \\
& - \int_0^{R(\theta)} \left\{ \frac{\partial^2}{\partial \theta^2} y(t,\theta) \left[ 1 - \frac{\partial}{\partial y} T(t,y(t,\theta)) \right] - \left( \frac{\partial}{\partial y} y(t,\theta) \right)^2 \frac{\partial^2}{\partial y^2} T(t,y(t,\theta)) \right\} dt \\
& = -\frac{\partial}{\partial \theta} y(R(\theta),\theta) \left[ 1 - \frac{\partial}{\partial y} T(R(\theta),y(R(\theta),\theta)) \right] R'(\theta) \\
& - \left[ -\frac{\psi}{\varphi} \frac{y(R(\theta),\theta)}{\varphi(R(\theta),\theta)} \left\{ \frac{\partial}{\partial \theta} y(R(\theta),\theta) - \frac{y(R(\theta),\theta)}{\varphi(R(\theta),\theta)} \frac{\partial}{\partial \theta} \varphi(R(\theta),\theta) \right\} - \eta'(\theta) \right] R'(\theta) \\
& = \left[ -\frac{\psi}{\varphi} \frac{y(R(\theta),\theta)}{\varphi(R(\theta),\theta)} \frac{\partial}{\partial \theta} \varphi(R(\theta),\theta) + \eta'(\theta) \right] R'(\theta)
\end{align*}
\]

Since \( R'(\theta) \geq 0, \eta'(\theta) \leq 0, \) and that \( \frac{\partial}{\partial \theta} \varphi(R(\theta),\theta) \geq 0, \) the above expression must be negative and therefore \( \hat{\theta} = \theta \) is the local maximizer of (4.15).

Q.E.D.

**Proposition C.2** Suppose that \( R'(\theta) \geq 0 \) and \( \eta'(\theta) \leq 0. \) Then \( \theta = \hat{\theta} \) is a global maximizer of (4.15).

**Proof.** Let \( U(\theta, \hat{\theta}) \) be given by (4.15). From lemma C.1, we know that \( \frac{\partial}{\partial \theta} U(\theta, \theta) = 0. \) Then, we want to show that \( U(\theta, \theta) \geq U(\theta, \hat{\theta}) \). One can rewrite this condition as

\[
\int_{\hat{\theta}}^{\theta} \frac{\partial}{\partial \theta} U(\theta, \hat{\theta}) \, d\hat{\theta} \geq 0
\]

Here, we focus on the case where \( \theta > \hat{\theta} \). In this case, it is sufficient to show that

\[
\frac{\partial}{\partial \theta} U(\theta, \hat{\theta}) \geq \frac{\partial}{\partial \hat{\theta}} U(\hat{\theta}, \hat{\theta}) = 0
\]
Using the formula in (4.15), we have

\[
\frac{\partial}{\partial \hat{\theta}} U (\theta, \hat{\theta}) = R' (\hat{\theta}) \left[ y \left( R (\hat{\theta}), \theta \right) - T \left( R (\hat{\theta}), y \left( R (\hat{\theta}), \theta \right) \right) \right] \\
- \left[ \psi y \left( \frac{R (\hat{\theta})}{\gamma}, \theta \right) + \eta (\theta) \right] \\
+ c' (\hat{\theta}) - R' (\hat{\theta}) \left[ y \left( R (\hat{\theta}), \hat{\theta} \right) - T \left( R (\hat{\theta}), y \left( R (\hat{\theta}), \hat{\theta} \right) \right) \right] \\
- \int_{0}^{R(\hat{\theta})} \frac{\partial}{\partial \theta} y \left( t, \hat{\theta} \right) \left[ 1 - T_{y} \left( t, y \left( t, \hat{\theta} \right) \right) \right] dt
\]

As it can be seen, in order to show that \( \frac{\partial}{\partial \hat{\theta}} U (\theta, \hat{\theta}) \geq \frac{\partial}{\partial \hat{\theta}} U \left( \hat{\theta}, \hat{\theta} \right) \), we need to show that

\[
y \left( R (\hat{\theta}), \theta \right) - T \left( R (\hat{\theta}), y \left( R (\hat{\theta}), \theta \right) \right) - \frac{\psi y \left( \frac{R (\hat{\theta})}{\gamma}, \theta \right)}{\gamma} \geq \eta (\theta)
\]

\[
y \left( \hat{\theta} \right) - T \left( \hat{\theta}, y \left( \hat{\theta}, \hat{\theta} \right) \right) - \frac{\psi y \left( \frac{R (\hat{\theta})}{\gamma}, \hat{\theta} \right)}{\gamma} \geq \eta (\hat{\theta})
\]

since \( R' (\hat{\theta}) \geq 0 \). Note that by assumption \( \eta (\theta) \leq \eta (\hat{\theta}) \) and hence we need to show that

\[
y \left( R (\hat{\theta}), \theta \right) - T \left( R (\hat{\theta}), y \left( R (\hat{\theta}), \theta \right) \right) - \frac{\psi y \left( \frac{R (\hat{\theta})}{\gamma}, \theta \right)}{\gamma} \geq 0
\]

\[
y \left( \hat{\theta} \right) - T \left( \hat{\theta}, y \left( \hat{\theta}, \hat{\theta} \right) \right) - \frac{\psi y \left( \frac{R (\hat{\theta})}{\gamma}, \hat{\theta} \right)}{\gamma} \geq 0
\]
The difference between the two sides of this inequality can be written as

$$\int_{\theta}^{\theta'} \frac{\partial}{\partial \theta} y(R(\hat{\theta}), \theta') \left[ 1 - \frac{\partial}{\partial y} T(R(\hat{\theta}), y(R(\hat{\theta}), \theta')) \right] d\theta'$$

$$- \int_{\theta}^{\theta'} \psi \frac{y(R(\hat{\theta}), \theta')^{\gamma-1}}{\varphi(R(\hat{\theta}), \theta')} \left[ \frac{\partial}{\partial \theta} y(R(\hat{\theta}), \theta') - \frac{y(R(\hat{\theta}), \theta')}{\varphi(R(\hat{\theta}), \theta')} \frac{\partial}{\partial \theta} \varphi(R(\hat{\theta}, \theta')) \right] d\theta'$$

$$= \int_{\theta}^{\theta'} \left\{ \frac{\partial}{\partial \theta} y(R(\hat{\theta}), \theta') \left[ 1 - \frac{\partial}{\partial y} T(R(\hat{\theta}), y(R(\hat{\theta}), \theta')) \right] \right. \right.$$

$$- \left. \psi \frac{y(R(\hat{\theta}), \theta')^{\gamma-1}}{\varphi(R(\hat{\theta}), \theta')} \frac{\partial}{\partial \theta} y(R(\hat{\theta}), \theta') \right\} d\theta'$$

$$+ \int_{\theta}^{\theta'} \psi \frac{y(R(\hat{\theta}), \theta')^{\gamma}}{\varphi(R(\hat{\theta}), \theta')^{\gamma+1}} \frac{\partial}{\partial \theta} \varphi(R(\hat{\theta}, \theta')) d\theta'$$

By the first order associated with (4.13), the first integral is zero and since $\varphi(t, \theta)$ is increasing in $\theta$, the second term is positive and hence the whole difference is a positive number.

Q.E.D.

C.2  Existence of retirement age

Proposition C.3 In the solution to (4.7), a type $\theta$ prefers to work if and only if $t \leq R(\theta)$.

Proof.

To show this, it is sufficient to show that in the solution to (4.7), $y(t, \theta)$ is decreasing in $t$ when $t > t^*(\theta)$ where $t^*(\theta)$ is the point at which $\varphi(t, \theta)$ is maximized. To see this, consider equation (C.5)

$$y(t, \theta) = \psi^{\frac{1}{\gamma+1}} \frac{\varphi(t, \theta)^{\gamma}}{[1 + \gamma \frac{G(\theta) - F(\theta)}{f(\theta)} \frac{\varphi(t, \theta)}{\varphi(t, \theta')}]^{\gamma+1}}$$
Note that when \( t > t^* (\theta) \), \( \varphi (t, \theta) \) is a decreasing function of \( t \). Moreover, \( \frac{\varphi (t, \theta)}{\varphi (t, \theta_i)} \) is an increasing function of \( t \) since

\[
\frac{d}{dt} \frac{\varphi (t, \theta)}{\varphi (t, \theta_i)} = \frac{\varphi (t, \theta)}{\varphi (t, \theta_i)} - \frac{\varphi (t, \theta) \varphi_i (t, \theta)}{\varphi (t, \theta_i)^2}
\]

By assumption 4.1, \( \varphi_i \theta \geq 0 \). Moreover, \( \varphi_i < 0 \) and \( \varphi \theta \geq 0 \). Hence, the above expression is positive. This means that \( y (t, \theta) \) is given by a decreasing function divided by an increasing function and therefore, when \( t > t^* (\theta) \), \( y (t, \theta) \) is decreasing in \( t \) and hence agents retire optimally.

Q.E.D.

C.3 Sufficiency of first-order approach

**Proposition C.4** Suppose that \( \eta (\theta) \) is a decreasing function of \( \theta \), \( \varphi (t, \theta) \) is an increasing function of \( \theta \) and that

\[
\left[ 1 + \gamma \frac{G (\theta) - F (\theta) \varphi (t, \theta)}{f (\theta) \varphi (t, \theta)} \right] \gamma \varphi (t, \theta) \frac{\gamma}{1 - \gamma}
\]

is increasing in \( \theta \). Then the solution to the relaxed problem (4.7) is also a solution to the more restricted problem (4.5).

We first show the following lemma:

**Lemma C.5** Suppose an allocation satisfies (4.6) and is such that \( y (t, \theta) \) and \( R (\theta) \) is increasing in \( \theta \). Then this allocation is incentive compatible.

**Proof.** We want to show that an allocation that satisfies the above conditions satisfies the following

\[
U (\theta, \theta) = c (\theta) - \int_0^{R (\theta)} \left[ \psi \frac{y (t, \theta)^\gamma}{\gamma \varphi (t, \theta)^\gamma} + \eta (\theta) \right] dt \geq c (\hat{\theta}) - \int_0^{R (\hat{\theta})} \left[ \psi \frac{y (t, \hat{\theta})^\gamma}{\gamma \varphi (t, \theta)^\gamma} + \eta (\theta) \right] dt = U (\theta, \hat{\theta}) ; \forall \theta, \theta
\]
To show this, we show that $U_2(\theta, \hat{\theta}) \geq 0$ whenever $\theta \geq \hat{\theta}$ and that $U_2(\theta, \hat{\theta}) \leq 0$ when $\theta \leq \hat{\theta}$. This would imply that

$$U_2(\theta, \hat{\theta}) = \int_\theta^\hat{\theta} U_2(\theta, \hat{\theta}) d\hat{\theta} \geq 0, \text{ if } \theta \geq \hat{\theta}$$

$$U_2(\theta, \hat{\theta}) \leq 0, \text{ if } \theta \leq \hat{\theta}$$

To show that $U_2(\theta, \hat{\theta}) \geq 0$ whenever $\theta \geq \hat{\theta}$, we have the following

$$U_2(\theta, \hat{\theta}) = c'(\hat{\theta}) - \int_0^{R(\hat{\theta})} \psi y_0(t, \hat{\theta}) \frac{y(t, \hat{\theta})^{\gamma-1}}{\varphi(t, \hat{\theta})^\gamma} \, dt$$

$$- \left[ \psi \varphi \left( R(\hat{\theta}), \hat{\theta} \right)^\gamma + \eta(\hat{\theta}) \right] R'(\hat{\theta})$$

Since the allocation satisfies (4.6),

$$c'(\hat{\theta}) = \int_0^{R(\hat{\theta})} \psi y_0(t, \hat{\theta}) \frac{y(t, \hat{\theta})^{\gamma-1}}{\varphi(t, \hat{\theta})^\gamma} \, dt + \left[ \frac{y \left( R(\hat{\theta}), \hat{\theta} \right)^\gamma}{\gamma \varphi \left( R(\hat{\theta}), \hat{\theta} \right)^\gamma} + \eta(\hat{\theta}) \right] R'(\hat{\theta})$$

Hence $U_2(\theta, \hat{\theta})$ can be written as

$$U_2(\theta, \hat{\theta}) = \int_0^{R(\hat{\theta})} \psi y_0(t, \hat{\theta}) \frac{y(t, \hat{\theta})^{\gamma-1}}{\varphi(t, \hat{\theta})^\gamma} t \gamma \varphi \left( R(\hat{\theta}), \hat{\theta} \right)^\gamma + \eta(\hat{\theta}) \right] R'(\hat{\theta})$$

$$+ \left[ \frac{y \left( R(\hat{\theta}), \hat{\theta} \right)^\gamma}{\gamma \varphi \left( R(\hat{\theta}), \hat{\theta} \right)^\gamma} + \eta(\hat{\theta}) \right] R'(\hat{\theta})$$

Given that $R'(\theta) \geq 0$, $\eta(\theta)$ is decreasing, $y(t, \theta)$ is increasing in $\theta$ and $\varphi(t, \theta)$ is increasing in $\theta$, the above expression is positive when $\theta \geq \hat{\theta}$ and negative when $\theta \leq \hat{\theta}$. That completes the proof of the lemma.

Q.E.D.

Given the above lemma, and the formulas provided in the paper, under the provided assumptions above $y(t, \theta)$ is increasing in $\theta$ as well as $R(\theta)$. Hence the sufficient condition for the lemma are satisfied.

Q.E.D.
D.1 Proofs

D.1.1 Proof of Lemma 5.8

First, consider the set $A = \{ v_1; \hat{a}_1(v_1) > \hat{a}'_1(v_1) \}$. Then

$$\int \hat{a}_1(v_1) dG \left( \frac{v_1 - v_1}{\sigma} \right) - \int \hat{a}'_1(v_1) dG \left( \frac{v_1 - v_1}{\sigma} \right) = \int_A dG \left( \frac{v_1 - v_1}{\sigma} \right) \geq 0$$

with equality only if $A$ is measure zero. Given the Bayesian updating formulas, this inequality implies that for any $v_1$,

$$\mu_{sg}(v_1; \hat{a}_1) \geq \mu_{sg}(v_1; \hat{a}'_1), \mu_{sd}(v_1; \hat{a}_1) \geq \mu_{sd}(v_1; \hat{a}'_1), \mu_h(v_1; \hat{a}_1) \leq \mu_h(v_1; \hat{a}'_1)$$

with strict inequalities only if $A$ is zero measure. Therefore, for each $v_1$, the integrand in (5.19) is higher for $\hat{a}_1$ and therefore $\Delta(v_1; \hat{a}_1) \geq \Delta(v_1; \hat{a}'_1)$ with equality only if $A$ is measure zero.

Second, if $\hat{a}_1$ is a switching strategy with switching point $k$, from (5.18) it is straightforward to see that $\mu_{sg}(v_1; \hat{a}_1), \mu_{sd}(v_1; \hat{a}_1)$ are strictly increasing and $\mu_h(v_1; \hat{a}_1)$ is strictly decreasing in $v_1$. Thus the integrand in (5.19) is increasing in $v_1$. Since we have assumed that $H(\hat{v}_1|v_1)$ is decreasing in $v_1$, from first-order stochastic dominance, it follows that $\Delta(v_1; \hat{a}_1)$ is strictly increasing.

Finally, to show boundedness, we first show that for all $\mu_2$, $V_2(\mu_2)$ is well defined and continuous. Since $\mu_2$ lies in a compact set, it follows that $V_2(\mu_2)$ is bounded. To
show continuity, note that when $v_2 \geq (\mu^*)^{-1}(\mu_2)$, $V_2(\mu_2, v_2) = \hat{p}(\mu_2; v_2) - q$ and if $v_2 < (\mu^*)^{-1}(\mu_2)$, $V_2(\mu_2, v_2) = \bar{\pi}\bar{v} + (1 - \bar{\pi})v_2 - q(1 + r) - \bar{c}$. Therefore,

$$V_2(\mu_2) = \int_{\infty}^{\infty} \left[ \int_{-\infty}^{\mu_2^{-1}(\mu_2)} \{\hat{p}(\mu_2; v_2) - q\} dG\left(\frac{v_2 - v_2}{\sigma}\right) \right.\left. + \int_{-\infty}^{\mu_*^{-1}(\mu_2)} \{\bar{\pi}\bar{v} + (1 - \bar{\pi})v_2 - q(1 + r) - \bar{c}\} dG\left(\frac{v_2 - v_2}{\sigma}\right) \right] dF(v_2)$$

$$= \{\hat{p}(\mu_2; v_2) - q\} + \int_{-\infty}^{\infty} \int_{-\infty}^{\mu_*^{-1}(\mu_2)} \{(1 - \mu_2)(\bar{\pi} - \bar{\pi})(\bar{v} - v_2) - qr - \bar{c}\} dG\left(\frac{v_2 - v_2}{\sigma}\right) dF(v_2).$$

Using our assumption that the random variable $v_2$ has a finite mean with respect to $G$ in (D.1), it follows that $V_2(\mu_2)$ is bounded. Continuity follows by inspection of (D.1) noting that so that $G$ and $F$ are continuous functions. Thus, there exist bounds $\Delta \leq \bar{\Delta}$ such that for any $v_1, \hat{a}_1$

$$\Delta \leq \bar{\pi}V_2(\mu_{s\hat{v}}(v_1; \hat{a}_1)) + (1 - \bar{\pi})V_2(\mu_{s0}(v_1; \hat{a}_1)) - V_2(\mu_h(v_1; \hat{a}_1)) \leq \bar{\Delta}.$$

$Q.E.D.$

### D.1.2 Proof of Lemma 5.9.

We start by showing that $b(k)$ is continuous and strictly increasing. Note that $b(k)$ satisfies the following:

$$\hat{p}(\mu_1; b(k)) - q + \Delta(b(k); d_k) = \bar{\pi}\bar{v} + (1 - \bar{\pi})b(k) - q(1 + r) - \bar{c} \quad \text{(D.2)}$$

Since $\Delta(b; d_k)$ is continuous in $b$ and $k$, it is obvious that $b(k)$ is continuous. An increase in $k$ causes the function $\Delta(b; d_k)$ to decrease by Lemma 5.8. Since $\hat{p}(\mu_1; b) - (1 - \bar{\pi})b$ is increasing in $b$, from (D.2), $b(k)$ must be an increasing function of $k$.

Next, we show that the fixed point of $b(k)$ is unique. To see this, note that any fixed point of $b(k)$, $v_1^*$ must satisfy

$$\hat{p}(\mu_1; v_1^*) - q + \Delta(v_1^*; d_{v_1^*}) = \bar{\pi}\bar{v} + (1 - \pi)v_1^* - q(1 + r) - \bar{c}.$$
Now, notice that under \(d_v\), from the Bayesian updating rules, the updating rules are functions of only \(1 - G\left(\frac{v^*_1 - v_1}{\sigma}\right)\). Therefore, we can rewrite \(\Delta(v^*_1; d_v)\) as the following:

\[
\Delta(v^*_1; d_v) = \beta \int_{-\infty}^{\infty} \left\{ \bar{\pi} V_2 \left( \mu_s g \left( 1 - G \left( \frac{v^*_1 - v_1}{\sigma} \right) \right) \right) 
+ (1 - \bar{\pi}) V_2 \left( \mu_{sd} \left( 1 - G \left( \frac{v^*_1 - v_1}{\sigma} \right) \right) \right) 
- V_2 \left( \mu_h \left( 1 - G \left( \frac{v^*_1 - v_1}{\sigma} \right) \right) \right) \right\} dG \left( \frac{v^*_1 - v_1}{\sigma} \right)
\]

Let \(l = 1 - G\left(\frac{v^*_1 - v_1}{\sigma}\right)\). Then the above integral becomes

\[
\Delta(v^*_1; d_v) = \beta \int_0^1 [\bar{\pi} V_2 (\mu_s g (l)) + (1 - \bar{\pi}) V_2 (\mu_{sd} (l)) - V_2 (\mu_h (l))] dl
\]

and \(v^*_1\) must satisfy

\[
-q + \beta \int_0^1 [\bar{\pi} V_2 (\mu_s g (l)) + (1 - \bar{\pi}) V_2 (\mu_{sd} (l)) - V_2 (\mu_h (l))] dl = \bar{\pi} \bar{v} + (1 - \bar{\pi}) v^*_1 - \hat{p}(\mu_1; v^*_1) - q(1 + r) - \bar{c}.
\]

The left side of the above equation does not depend on \(v^*_1\) and the right side is strictly decreasing in \(v^*_1\). Since the right side ranges from plus infinity to minus infinity, there exist a unique \(v^*_1\) that satisfies the above equation. Now, notice that under \(d_v\), from the Bayesian updating rules, the updating rules are functions of only \(1 - G\left(\frac{v^*_1 - v_1}{\sigma}\right)\). Therefore, we can rewrite \(\Delta(v^*_1; d_v)\) as the following:

\[
\Delta(v^*_1; d_v) = \beta \int_{-\infty}^{\infty} \left\{ \bar{\pi} V_2 \left( \mu_s g \left( 1 - G \left( \frac{v^*_1 - v_1}{\sigma} \right) \right) \right) 
+ (1 - \bar{\pi}) V_2 \left( \mu_{sd} \left( 1 - G \left( \frac{v^*_1 - v_1}{\sigma} \right) \right) \right) 
- V_2 \left( \mu_h \left( 1 - G \left( \frac{v^*_1 - v_1}{\sigma} \right) \right) \right) \right\} dG \left( \frac{v^*_1 - v_1}{\sigma} \right)
\]

and \(v^*_1\) must satisfy

\[
-q + \beta \int_0^1 [\bar{\pi} V_2 (\mu_s g (l)) + (1 - \bar{\pi}) V_2 (\mu_{sd} (l)) - V_2 (\mu_h (l))] dl = \bar{\pi} \bar{v} + (1 - \bar{\pi}) v^*_1 - \hat{p}(\mu_1; v^*_1) - q(1 + r) - \bar{c}.
\]
The left side of the above equation does not depend on \( v^*_1 \) and the right side is strictly decreasing in \( v^*_1 \). Since the right side ranges from plus infinity to minus infinity, there exists a unique \( v^*_1 \) that satisfies the above equation.

Finally, we conclude by showing that when \( k > v^*_1 \), \( b(k) < k \) and when \( k < v^*_1 \), \( b(k) > k \). Suppose \( k < v^*_1 \) and \( b(k) \leq k \). Since \( \lim_{k \to -\infty} b(k) = \hat{v}^0 > -\infty \). Then by continuity of \( b(\cdot) \), there must exist \( k \in (-\infty, k] \) such that \( b(\hat{k}) = \hat{k} \), contradicting part 2. Similarly, we can show that for all \( k > v^*_1 \), \( b(k) < k \). \( Q.E.D. \)

D.1.3 Proof of Theorem 5.10.

We show that our environment can be mapped into that described in [Morris and Shin, 2003] and show that their requirements for existence of a unique equilibrium in the limit are satisfied.

Given a value function \( V_2(\mu_2) \), consider an equilibrium strategy profile in the first period \((a_1(\cdot), \hat{a}_1(\cdot), p_1(\cdot))\). In a game with full information about shocks to returns, when agents in period 2 believe that the HH bank sells with probability \( l \) in the first period 1, the HH bank’s differential gain from selling is given by

\[
\hat{\pi}(v_1, l) = \hat{p}(\mu_1; v_1) + q r + c - \tilde{v} - (1 - \pi) v_1 + \beta [\tilde{\pi} V_2(\hat{\mu}_{sg}(l)) + (1 - \pi) V_2(\hat{\mu}_{sd}(l)) - V_2(\hat{\mu}_{h}(l))].
\]

Then, in the game with private information, \( l = \int \hat{a}_1(v_1) dH(v_1|v_1) \) is a random variable. We then show that \( \hat{\pi} \) satisfies the conditions A1–A3, A4*, A5, and A6 in [Morris and Shin, 2003]. We then can apply Theorem 2.2 in [Morris and Shin, 2003], and that completes the proof of our Proposition. It is easy to see that \( \hat{\mu}_{sg}(l) \) and \( \hat{\mu}_{sd}(l) \) are increasing in \( l \) and \( \hat{\mu}_{h}(l) \) is decreasing in \( l \). Since \( V_2(\mu_2) \) is nondecreasing in \( \mu_2 \), \( \hat{\pi}(v_1, l) \) is nondecreasing in \( l \) – condition A1. Obviously \( \hat{\pi}(v_1, l) \) is increasing in \( v_1 \) – condition A2. Since \( \hat{\pi}(v_1, l) \) is separable in \( v_1 \) and \( l \), and \( \hat{\pi}(v_1, l) \) is linearly increasing in \( v_1 \), there must exist a unique \( v^*_1 \) such that \( \int \hat{\pi}(v^*_1, l) dl = 0 \) – condition A3. Since \( V_2(\mu_2) \) is a continuous function over a compact set \([0, 1] \), \( \beta [\tilde{\pi} V_2(\hat{\mu}_{sg}(l)) + (1 - \pi) V_2(\hat{\mu}_{sd}(l)) - V_2(\hat{\mu}_{h}(l))] \) is bounded above and below by \( \Delta \) and
\[ \bar{\Delta}, \text{respectively. Now let } \tilde{v}_1 \text{ and } \hat{v}_1 \text{ be defined by} \]
\[ 0 = -\hat{p}(\mu_1; \tilde{v}_1) - qr + \pi \tilde{v} + (1 - \pi)\tilde{v}_1 - \bar{c} - \Delta - \varepsilon, \]
\[ 0 = -\hat{p}(\mu_1; \hat{v}_1) - qr + \pi \tilde{v} + (1 - \pi)\hat{v}_1 - \bar{c} - \Delta + \varepsilon. \]

Then, if \( v_1 \leq \tilde{v}_1, \hat{\pi}(c_1,l) \leq -\varepsilon \) for all \( l \in [0,1] \). Moreover, if \( v_1 \geq \tilde{v}_1, \hat{\pi}(v_1,l) \geq -\varepsilon \) for all \( l \in [0,1] \) – condition A4*. Continuity of \( V_2 \) implies that \( \hat{\pi}(v_1,l) \) is a continuous function of \( v_1 \) and \( l \). Therefore, \( \int_0^1 g(l)\hat{\pi}(v_1,l)dl \) is a continuous function of \( g(\cdot) \) and \( v_1 \) – condition A5. Moreover, by definition of \( F(\cdot) \) and \( G(\cdot) \), noisy signal \( v_1 \) has a finite expectation, \( E[v_1] \in R \) – condition A6. Therefore, we can rewrite Proposition 2.2 in [Morris and Shin, 2003] for our environment as follows:

Proposition Let \( v^*_1 \) satisfy \( \int \hat{\pi}(v^*_1,l)dl = 0 \). For any \( \delta > 0 \), there exists a \( \bar{\sigma} > 0 \) such that for all \( \sigma \leq \bar{\sigma} \), if strategy \( a_1 \) survives iterated elimination of dominated strategies, then \( a_1(v_1) = 1 \) for all \( v_1 \geq v^*_1 + \delta \) and \( a_1(v_1) = 0 \) for all \( v_1 \leq v^*_1 - \delta \).

Q.E.D.

D.1.4 Proof of Proposition 5.11.

We proceed by induction. As described in Proposition 5.1, the game has a unique equilibrium in period \( T \). The equilibrium strategy in the last period is a cutoff strategy with cutoff \( v^*_T(\mu_T) \) given by

\[ v^*_T(\mu_T) = \bar{v} - \frac{qr + \bar{c}}{(1 - \mu_T)(\bar{\pi} - \pi)}. \]

Using the equilibrium strategy, we define the last period’s ex-ante value function, \( V_T(\mu_T) \) according to

\[ V_T(\mu_T) = (1 - \alpha) \int_{-\infty}^{v^*_T(\mu_T)} \{ \pi \bar{v} + (1 - \pi)v_t - q(1 + r) - \bar{c} \} dF(v_t) \]
\[ + (1 - \alpha) \int_{v^*_T(\mu_T)}^{\infty} \{ \hat{p}(\mu_T; v_t) - q \} dF(v_t). \]
From Theorem 5.10, as $\sigma_{T-1}$ converges to zero, the set of equilibrium strategies in period $T - 1$ converges to a cutoff strategy with cutoff $v_{T-1}^*(\mu_{T-1})$ given by

$$v_{T-1}^*(\mu_{T-1}) = \bar{v} - \frac{qr + c + \beta \int_0^1 [\pi V_T(\hat{\mu}_{sg}(l; \mu_{T-1})) + (1 - \bar{\pi}) V_T(\hat{\mu}_{sg}(l; \mu_{T-1})) - V_T(\hat{\mu}_h(l; \mu_{T-1}))]}{(1 - \mu_{T-1})(\bar{\pi} - \pi)} \]$$

Notice that for $\sigma_{T-1}$ small and given the above cutoff strategy, the value function at period $T - 1$, $V_{T-1}(\mu_{T-1}; \sigma_{T-1})$ is given by

$$V_{T-1}(\mu_{T-1}; \sigma_{T-1}) = (1 - \alpha) \int_{\underline{\psi}_l} \int_{-\infty}^{v_{T-1}^* (\mu_{T-1})} \{\bar{\pi} \bar{v} + (1 - \bar{\pi}) \psi_l - q(1 + r) - c\} + \beta V_T \left(\hat{\mu}_h \left(1 - G \left(\frac{v_{T-1}^* (\mu_{T-1}) - \psi_l}{\sigma_{T-1}}\right)\right)\right) dG(\varepsilon_{T-1}) dF(\psi_l)$$

$$+ (1 - \alpha) \int_{\underline{\psi}_l} \int_{-\infty}^{v_{T-1}^* (\mu_{T-1})} \{\hat{\mu}_{sg} (1 - G \left(\frac{v_{T-1}^* (\mu_{T-1}) - \psi_l}{\sigma_{T-1}}\right))\} + \beta(1 - \bar{\pi}) V_T \left(\hat{\mu}_{sh} \left(1 - G \left(\frac{v_{T-1}^* (\mu_{T-1}) - \psi_l}{\sigma_{T-1}}\right)\right)\right) dG(\varepsilon_{T-1}) dF(\psi_l)$$

and hence, the above formula becomes the following as $\sigma_{T-1} \to 0$:

$$V_{T-1}(\mu_{T-1}) = (D.3)$$

$$(1 - \alpha) \int_{-\infty}^{v_{T-1}^* (\mu_{T-1})} \{\bar{\pi} \bar{v} + (1 - \bar{\pi}) \psi_l - q(1 + r) - c + \beta V_T(\hat{\mu}_h(0))\} dF(\psi_l)$$

$$+ (1 - \alpha) \int_{\underline{\psi}_l} \int_{-\infty}^{v_{T-1}^* (\mu_{T-1})} \{\hat{\mu}_{sg}(1) + \beta(1 - \bar{\pi}) V_T(\hat{\mu}_{sd}(1))\} dF(\psi_l)$$

$$+ \alpha \int_{-\infty}^{v_{T-1}^* (\mu_{T-1})} \{\bar{\pi} \bar{v} + (1 - \bar{\pi}) \psi_l - q(1 + r) - c + \beta V_T(\hat{\mu}_h(0))\} dF(\psi_l)$$

$$+ \alpha \int_{v_{T-1}^* (\mu_{T-1})}^{\infty} \{\bar{\pi} \bar{v} + (1 - \bar{\pi}) \psi_l - q(1 + r) - c + \beta V_T(\hat{\mu}_h(1))\} dF(\psi_l)$$
Similarly, suppose for some period \( t + 1 \) and any \( \mu_{t+1} \), the multi period model has a unique equilibrium with payoff for the HH bank given by \( V_{t+1}(\mu_{t+1}) \). If \( V_{t+1}(\mu_{t+1}) \) is increasing in \( \mu_{t+1} \), then the proof of Theorem 5.10 can be applied. As a result, as \( \sigma_t \to 0 \), the set of equilibrium strategies in period \( t \) converges to a cutoff strategy with cutoff \( v_t^*(\mu_t) \) satisfying the properties defined in Proposition 5.11. In addition, this cutoff strategy can be used to construct the value function in period \( t \), \( V_t(\mu_t) \) in fashion similar to (D.3).

**Q.E.D.**

**Proof of Proposition 5.3.** We shall prove that when
\[
\mu^*_2 < \frac{\beta \bar{\pi} - \frac{\pi}{\pi a - \bar{\pi}}}{1 + \beta \bar{\pi}(1 - \alpha)}
\]
then if \( \mu_1 = \mu \), we must have \( \mu^n_h < \mu^*_2 \). Note that from (5.13),
\[
\mu^n_h = \frac{\beta \bar{\pi} - \frac{\pi}{\pi a - \bar{\pi}}}{1 + \beta \bar{\pi}(1 - \alpha)} < \mu^*_2
\]
\[
\iff \mu < \mu^*_2 [\mu + (1 - \mu)\alpha]
\]
\[
\iff \mu (1 - \mu^*_2 (1 - \alpha)) < \mu^*_2 \alpha
\]
\[
\iff \mu < \frac{\mu^*_2 \alpha}{1 - \mu^*_2 + \mu^*_2 \alpha}
\]
(D.4)

Hence, we must show that the above inequality holds. Notice that from (5.10), \( \underline{\mu} \) is defined by
\[
\dot{p}(\mu) + \beta [\bar{\pi} V(\mu_{s\bar{b}}) + (1 - \bar{\pi}) V(\mu_{s\bar{b}})] = \bar{\pi} \bar{v} - \bar{c} - qr + \beta V(\mu_h).
\]

Since \( \dot{p}(\mu^*_2) = \bar{\pi} \bar{v} - \bar{c} - qr \) and \( V(\mu_h) = V(\mu_{s\bar{b}}) = V(\mu^*_2) \), the above equality can be written as
\[
\dot{p}(\mu) + \beta \bar{\pi} [V(\mu_{s\bar{b}}) - V(\mu^*_2)] = \dot{p}(\mu^*_2).
\]

Moreover, since low cost types always hold their assets, we must have
\[
V(\mu_{s\bar{b}}) - V(\mu^*_2) = (1 - \alpha) [\dot{p}(\mu_{s\bar{b}}) - \dot{p}(\mu^*_2)].
\]

Therefore, (5.10) becomes
\[
\dot{p}(\mu) + \beta \bar{\pi}(1 - \alpha) [\dot{p}(\mu_{s\bar{b}}) - \dot{p}(\mu^*_2)] = \dot{p}(\mu^*_2),
\]
Using the fact that, $\hat{p}(\cdot)$ is a linear function and definition of $\mu_{eq}$ from (5.11),

$$\mu + \beta \bar{\pi}(1 - \alpha) \left[ \frac{\mu}{\mu + (1 - \mu) \frac{\pi}{\bar{\pi}}} - \mu_2^* \right] = \mu_2^*$$

Given that the right hand side of the above equation is increasing in $\mu$, (D.4) is equivalent to the following inequality

$$\frac{\mu_2^* \alpha}{1 - \mu_2^* + \mu_2^* \alpha} + \beta \bar{\pi}(1 - \alpha) \left[ \frac{\mu_2^* \alpha}{1 - \mu_2^* + \mu_2^* \alpha} + \frac{(1 - \mu_2^*) \frac{\pi}{\bar{\pi}}}{\mu_2^* \alpha + (1 - \mu_2^*) \frac{\pi}{\bar{\pi}}} - \mu_2^* \right] > \mu_2^*$$

The above inequality can be further simplified in the following steps:

$$\frac{\mu_2^* \alpha}{1 - \mu_2^* + \mu_2^* \alpha} + \beta \bar{\pi}(1 - \alpha) \left[ \frac{\mu_2^* \alpha}{\mu_2^* \alpha + (1 - \mu_2^*) \frac{\pi}{\bar{\pi}}} - \mu_2^* \right] > \mu_2^*$$

$$\Leftrightarrow \beta \bar{\pi}(1 - \alpha) \mu_2^* \frac{\alpha(1 - \mu_2^*) - (1 - \mu_2^*) \frac{\pi}{\bar{\pi}}}{\mu_2^* \alpha + (1 - \mu_2^*) \frac{\pi}{\bar{\pi}}} > \mu_2^* \frac{1 - \mu_2^* - (1 - \mu_2^*) \frac{\pi}{\bar{\pi}}}{1 - \mu_2^* + \mu_2^* \alpha}$$

$$\Leftrightarrow \beta \bar{\pi}(1 - \alpha) \mu_2^* \frac{\alpha(1 - \mu_2^*) - (1 - \mu_2^*) \frac{\pi}{\bar{\pi}}}{\mu_2^* \alpha + (1 - \mu_2^*) \frac{\pi}{\bar{\pi}}} > \mu_2^* \frac{1 - \mu_2^* - (1 - \mu_2^*) \frac{\pi}{\bar{\pi}}}{1 - \mu_2^* + \mu_2^* \alpha}$$

Since $0 < \mu_2^* < 1$, we can divide both sides of the above inequality by $\mu_2^*(1 - \mu_2^*)$ and we have

$$\beta \bar{\pi}(1 - \alpha) \frac{\alpha - \frac{\pi}{\bar{\pi}}}{\mu_2^* \alpha + (1 - \mu_2^*) \frac{\pi}{\bar{\pi}}} > \frac{1 - \alpha}{1 - \mu_2^* + \mu_2^* \alpha}$$

$$\Leftrightarrow \beta \bar{\pi} \frac{\alpha - \frac{\pi}{\bar{\pi}}}{\mu_2^* \alpha + (1 - \mu_2^*) \frac{\pi}{\bar{\pi}}} > \frac{1}{1 - \mu_2^* + \mu_2^* \alpha}$$

$$\Leftrightarrow \beta \bar{\pi} (\alpha - \frac{\pi}{\bar{\pi}}) (1 - \mu_2^*(1 - \alpha)) > \mu_2^* \left( \alpha - \frac{\pi}{\bar{\pi}} \right) + \frac{\pi}{\bar{\pi}}$$

$$\Leftrightarrow \beta \bar{\pi} (\alpha - \frac{\pi}{\bar{\pi}}) - \frac{\pi}{\bar{\pi}} > \mu_2^* \left( \alpha - \frac{\pi}{\bar{\pi}} \right) [1 + \beta \bar{\pi}(1 - \alpha)]$$

The above inequality is equivalent to

$$\frac{\beta \bar{\pi} - \frac{\pi}{\bar{\pi}}}{1 + \beta \bar{\pi}(1 - \alpha)} > \mu_2^*$$

and this completes the proof. Q.E.D.

**D.2 Full Characterization of Equilibria in Two Period Game**

**Proposition D.1** Suppose $\beta (1 - \alpha) \leq 1$ and $0 < \mu_2^* < 1$. Then, there exist $\underline{\mu}$ and $\bar{\mu}$ with $\underline{\mu} < \mu_2^* < \bar{\mu}$ such that
1. if $\mu_1 \in [\underline{\mu}, \bar{\mu})$, the model has two equilibria: in one the HH bank sells its loan, and in the other the HH bank holds its loan,

2. if $\mu_1 < \underline{\mu}$, the model has a unique equilibrium in which the HH bank holds its loan in period 1,

3. if $\mu_1 \geq \bar{\mu}$, the model has a unique equilibrium in which the HH bank sells its loan in period 1.

**Proof.** We show that our economy has a positive reputational equilibrium. As an implication of Bayes Rule, if the HH bank sells its loan in the first period, the reciprocal of the posterior beliefs is a martingale. Formally, we have

$$\frac{\bar{\pi}}{\mu_{s\bar{v}}} + \frac{1 - \bar{\pi}}{\mu_{s0}} = \frac{1}{\mu_1} = \frac{1}{\mu_h}$$

Since $1/\mu$ is a convex function, it follows that

$$\bar{\pi}\mu_{s\bar{v}} + (1 - \bar{\pi})\mu_{s0} \geq \mu_1 = \mu_h. \quad (D.5)$$

Let the reputational gain be defined as

$$\Delta^g(\mu_1) = \beta (\bar{\pi}V_2(\mu_{s\bar{v}}) + (1 - \bar{\pi})V_2(\mu_{s0}) - V_2(\mu_h))$$

Recall from (5.5) that $V_2$ is a convex and increasing function, so that

$$\bar{\pi}V_2(\mu_{s\bar{v}}) + (1 - \bar{\pi})V_2(\mu_{s0}) \geq V_2(\bar{\pi}\mu_{s\bar{v}} + (1 - \bar{\pi})\mu_{s0}).$$

This convexity together with (D.5) implies that $\Delta^g(\mu_1) \geq 0$.

Next we show that there is some critical value of $\mu_1$ denoted $\mu_g < \mu_1^*$ such that for all $\mu_1$ in the interval $\mu_g < \mu_1 \leq \mu_1^*$, $\Delta^g(\mu_1)$ is strictly positive and increasing in $\mu_1$ and $\Delta^g(\mu_1) = 0$ for $\mu_1 \leq \mu_g$. To obtain these results, define $\mu_g$ implicitly by

$$\mu_g^2 = \frac{\mu_g\bar{\pi}}{\mu_g\bar{\pi} + (1 - \mu_g)\bar{\pi}}.$$ 

That is $\mu_g$ denotes that initial reputation level such that if the HH bank sells and receives a good signal, its reputation level would rise to $\mu_g^2$. Since $\bar{\pi} > \bar{\pi}$, $\mu_g < \mu_1^*$. To see that for all $\mu_g < \mu_1 \leq \mu_1^*$, $\Delta^g(\mu_1)$ is strictly positive and increasing in $\mu_1$, rewrite the reputational gain as

$$\Delta^g(\mu_1) = \beta (\bar{\pi}(V_2(\mu_{s\bar{v}}) - V_2(\mu_h)) + (1 - \bar{\pi})(V_2(\mu_{s0}) - V_2(\mu_h))).$$
Since $\mu_h = \mu_1$ and $\mu_{s0} < \mu_1$, from Proposition 1 it follows that for all $\mu_g < \mu_1 \leq \mu_h^*, V_2(\mu_{s0}) = V_2(\mu_h)$, and it follows that $\Delta^b(\mu_1)$ is strictly positive and since $\mu_{s0}$ is strictly increasing in $\mu_1$, it follows that $\Delta^b(\mu_1)$ is strictly increasing in $\mu_1$. Since $\Delta^b(\mu_1)$ is strictly increasing in $\mu_1$ it follows that $\Delta^b(\mu_1)$ is strictly increasing in $\mu_1$. Thus, there is a unique value of $\mu_1$ for which (D.6) is strictly increasing in $\mu_1$. Since $\Delta^b(\mu_1) = 0$ for $\mu_1 \leq \mu_g$, note that $\mu_{s0} \leq \mu_2^*$ so that $V_2(\mu_{s0}) = V_2(\mu_h)$.

Next, rewrite (5.10) as

\[
(\mu_1 \bar{\pi} + (1 - \mu_1)\bar{\pi}) \bar{v} - q + \Delta^g(\mu_1) \geq \bar{\pi} \bar{v} - q(1 + r) - \bar{c}
\]  

(D.6)

Consider $\mu_1 \leq \mu_2^*$. Since $\Delta^g(\mu_1)$ is a nondecreasing function of $\mu_1$ in this range and $(\mu_1 \bar{\pi} + (1 - \mu_1)\bar{\pi}) \bar{v}$ is a strictly increasing function of $\mu_1$, it follows that the left side of (D.6) is strictly increasing in this range. Since $\Delta^g(\mu_1^*)$ is strictly positive, using (5.3) the left side of (D.6) is strictly greater than the right side of this inequality at $\mu_1^*$. Since $\Delta^g(\mu_1^*) = 0$ and $\mu_g \leq \mu_2^*$, the left side is strictly less than the right side at $\mu_g$. Thus, there is a unique value of $\mu_1$ at which (D.6) holds as an equality. For $\mu_1 > \mu_2^*$, $(\mu_1 \bar{\pi} + (1 - \mu_1)\bar{\pi}) \bar{v} - q > \bar{\pi} \bar{v} - q(1 + r) - \bar{c}$ and $\Delta^g(\mu_1) \geq 0$ so that (D.6) is satisfied.

We have established that our model has an equilibrium in which all HH banks with reputation levels above $\mu_1 \geq \mu_2$ sell.

To obtain the negative reputational equilibrium, define $\mu_b$ implicitly by

\[
\mu_2^* = \frac{\mu_b}{\mu_b + (1 - \mu_b)\alpha}.
\]

That is $\mu_b$ denotes that initial reputation level such that if the HH bank holds, its reputation level would rise to $\mu_2^*$. Clearly $\mu_b < \mu_2^*$.

Since $\mu_h = \mu_1/(\mu_1 + (1 - \mu_1)\alpha)$ is greater than $\mu_1$, it follows that $\Delta^b(\mu_1)$ is negative for $\mu_1 > \mu_b$. If $\mu_1 \in [\mu_b, \mu_2^*]$, selling has a static cost, i.e. $\hat{p}(\mu_2) - q \leq \bar{\pi} \bar{v} - q(1 + r) - \bar{c}$ as well as a loss from reputation, i.e. $\Delta^b(\mu_1) < 0$ so that the HH bank prefers to hold the asset.

If $\mu_1 \in (\mu_2^*, 1]$, there are benefits from selling the asset, i.e. $\hat{p}(\mu_2) - q \geq \bar{\pi} \bar{v} - q(1 + r) - \bar{c}$, while there is a loss from reputation $\Delta^b(\mu_1) < 0$. Our assumption that $\beta(1 - \alpha) \leq 1$ ensures that when $\mu_1 = 1$, the static benefit outweighs the loss from reputation, i.e. (5.12) is reversed at $\mu_1 = 1$. Moreover, since $\mu_h = \mu_1/(\mu_1 + (1 - \mu_1)\alpha)$, it is easy to show that $(\mu_2 \bar{\pi} + (1 - \mu_2)\bar{\pi}) \bar{v} - q + \Delta^b(\mu_1)$ is a strictly convex function of $\mu_1$ for $\mu_1 \in [\mu_2^*, 1]$.

Since the value of this function is strictly less than $\bar{\pi} \bar{v} - q(1 + r) - \bar{c}$ at $\mu_1 = \mu_2^*$ and weakly higher when $\mu_1 = 1$, there exists a unique $\bar{\mu} \in (\mu_2^*, 1)$, at which (5.12) holds with equality. For $\mu_1 \leq \bar{\mu}$, (5.12) holds and for $\mu_1 > \bar{\mu}$ (5.12) is violated.

Q.E.D.
D.3 Strategic Types

**Proposition D.2** Suppose $\beta(1 - \alpha) \leq 1$ and

$$ (\bar{\pi} - \pi) \bar{v} + qr + \max_{\mu_1 \in [0,1]} \Delta^g(\mu_1) < -\zeta. \tag{D.7} $$

Then the unique equilibrium of the static game described in Proposition 1 and the multiple equilibria of the dynamic game described in Proposition 2 are also equilibria of the associated games when all bank types behave strategically.

**Proof.** Consider the static game. It is sufficient to show that given the constructed equilibrium and specified strategies for all agents, there is no profitable deviation by any agent. Note that in the proof of Proposition 2 we show that $\Delta^g(\mu_1) \geq 0$ for all $\mu_1 \in [0,1]$. Hence, (D.7) implies that

$$ \mu_1 (\bar{\pi} - \pi) \bar{v} + qr < -\zeta $$

or

$$ [\mu_1 \bar{\pi} + (1 - \mu_1) \bar{\pi}] \bar{v} - q < \bar{\pi} v - q(1 + r) - \zeta \tag{D.8} $$

Inequality (D.8) implies that facing break even prices the low cost type bank would like to hold. Moreover a deviation by a buyer must attract these types of bank and (D.8) implies that buyers must offer a price higher than the actuarially fair price. Hence, there is no deviation by any buyer or a low cost bank type. Moreover, an LH bank wants to sell even at the lowest possible price, $\bar{\pi} v$, since $\bar{c} > 0$. Thus there are no profitable deviation from the specified strategies in the static game.

Consider the positive equilibrium of the dynamic game. Given future beliefs, the value of selling to a low quality bank adjusted by the future reputational gain from holding is given by

$$ [\mu_1 \bar{\pi} + (1 - \mu_1) \bar{\pi}] \bar{v} - q + \beta [\bar{\pi} V_2(\mu^g_{\bar{\pi} v}) + (1 - \bar{\pi}) V_2(\mu^g_{\bar{\pi} 0}) - V_2(\mu)] $$

where $\mu^g_{\bar{\pi} v} = \bar{\pi} \mu_1 / (\mu_1 \bar{\pi} + (1 - \mu_1) \bar{\pi})$ and $\mu^g_{\bar{\pi} 0} = (1 - \bar{\pi}) \mu_1 / ((1 - \bar{\pi}) \mu_1 + (1 - \bar{\pi})(1 - \mu_1))$. The value of selling to a high quality bank is given by

$$ [\mu_1 \bar{\pi} + (1 - \mu_1) \bar{\pi}] \bar{v} - q + \Delta^g(\mu_1) $$
From (D.7) and \( \beta \left[ V_{\mu 2}(\mu_{s2}) + (1 - \bar{\pi})V_{\mu 2}(\mu_{s0}) - V_{2}(\mu) \right] = \Delta^g(\mu_1) \), we have

\[
[\mu_1 \bar{\pi} + (1 - \mu_1)\bar{\pi}] \bar{v} - q + \beta \left[ V_{\mu 2}(\mu_{s2}) + (1 - \bar{\pi})V_{\mu 2}(\mu_{s0}) - V_{2}(\mu) \right] \leq \bar{v} \pi - q(1 + r) - \zeta
\]

\[
[\mu_1 \bar{\pi} + (1 - \mu_1)\bar{\pi}] \bar{v} - q + \Delta^g(\mu_1) \leq \bar{v} \pi - q(1 + r) - \zeta
\]

Hence, there is no profitable deviation by the low cost types. As for the LH type bank, note that in the positive equilibrium

\[
[\mu_1 \bar{\pi} + (1 - \mu_1)\bar{\pi}] \bar{v} - q + \beta \left[ V_{\mu 2}(\mu_{s2}) + (1 - \bar{\pi})V_{\mu 2}(\mu_{s0}) - V_{2}(\mu) \right] \geq \bar{v} \pi - q(1 + r) - \bar{c} \quad \text{(D.9)}
\]

We use the above inequality to show that the LH type bank does not have a profitable deviation. There are two possible cases: Case 1. \( \bar{c} + qr \geq (\bar{\pi} - \pi)\bar{v} \). In this case, \( \mu_2^* = 0 \) and \( V_{2}(\mu) \) is a constant function. Therefore, \( \Delta^g(\mu_1) = 0 \) for all \( \mu_1 \) and \( \beta \left[ \bar{\pi}V_{2}(\mu_{s2}) + (1 - \bar{\pi})V_{2}(\mu_{s0}) - V_{2}(\mu) \right] = 0 \). In this case, we are back to the static game and as we have shown before, the LH bank finds it optimal to sell always. Case 2. \( \bar{c} + qr < (\bar{\pi} - \pi)\bar{v} \). In this case, we have

\[
\beta \left[ V_{2}(\mu_{s2}) - V_{2}(\mu_{s0}) \right] \leq \beta(1 - \alpha) \left\{ [\mu_{s0} \bar{\pi} + (1 - \mu_{s0})\bar{\pi}] \bar{v} - q - \bar{v} \pi + q(1 + r) + \bar{c} \right\}
\]

\[
= \beta(1 - \alpha) \left\{ -(1 - \mu_{s0})(\bar{\pi} - \pi)\bar{v} + qr + \bar{c} \right\}
\]

The last expression is increasing in \( \mu_1 \) and therefore maximized at \( \mu_1 = 1 \). Hence, we must have

\[
\beta \left[ V_{2}(\mu_{s2}) - V_{2}(\mu_{s0}) \right] \leq \beta(1 - \alpha)(qr + \bar{c}) < \bar{v}
\]

Therefore,

\[
-\beta(\bar{\pi} - \pi) \left[ V_{2}(\mu_{s2}) - V_{2}(\mu_{s0}) \right] > -\bar{v}(\bar{\pi} - \pi)
\]

Adding this inequality to (D.9), we get

\[
[\mu_1 \bar{\pi} + (1 - \mu_1)\bar{\pi}] \bar{v} - q + \beta \left[ \bar{\pi}V_{2}(\mu_{s2}) + (1 - \bar{\pi})V_{2}(\mu_{s0}) - V_{2}(\mu) \right] \geq \bar{v} \pi - q(1 + r) - \bar{c}
\]

which implies that the LH type bank does not have a profitable deviation in the constructed equilibrium.

As for the negative equilibrium, it is clear that a bank with low cost does not want to sell its loan, since selling only punishes the bank. Therefore, it is sufficient to show that the LH bank wants to sell its loan. That is, we need to show that for all \( \mu_1 \in [0, \bar{\mu}] \), we have

\[
\bar{\pi} \pi - q + \beta \left[ V_{2}(0) - V_{2}(\mu_{h}^b) \right] \geq \bar{v} \pi - q(1 + r) - \bar{c} \quad \text{(D.10)}
\]
where \( \mu_h^b = \mu_1/(\mu_1 + (1 - \mu_1)\alpha) \). To do so, we first show that this inequality is satisfied at \( \mu_1 = \bar{\mu} \). Now, since \( \Delta^b(\mu_1) = \beta[V_2(0) - V_2(\mu_h^b)] \) is decreasing, this implies that (D.10) holds for all \( \mu_1 \in [0, \bar{\mu}] \). By definition, \( \bar{\mu} \) satisfies

\[
\bar{\pi} - q + \beta[V_2(0) - V_2(\mu_h^b)] = \bar{\pi} - q(1 + r) - \bar{c}
\]

Obviously, this equality leads to the above inequality. Therefore, we have shown that LH bank still finds it optimal to sell in the negative equilibrium.

Q.E.D.