

Essays on Time Inconsistency

**A THESIS
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY**

Pei-Cheng Yu

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY**

David Rahman, Adviser

July, 2016

© Pei-Cheng Yu 2016
ALL RIGHTS RESERVED

Acknowledgements

Graduate school was exciting and challenging, and I owe this experience to many people. All of them have influenced my intellectual development and none have ever stopped believing in me. First and foremost, I am indebted to David Rahman, who was more than an adviser but a friend. Research is an undulating and difficult road to traverse, David's endless encouragements and unwavering support shone a bright light on the murky path towards discovery. I am also forever grateful for the inquisitive queries of Aldo Rustichini, whose infectious curiosity is something I would try to emulate for the rest of my academic career. I will always remember the conversations I had with Manuel Amador and Erzo Luttmer. Their demand for clarity and high academic standards are watermarks to which I will forever measure my work against. My research was shaped by all participants of the Theory Workshop at the Department of Economics. The openness of the environment helped me find my way in the beginning stages of my research. Last but not least, I could not have asked for a better coauthor than Radek Paluszynski. I won't soon forget our midnight conversations on economics over blueberry pies.

Dedication

To my mom and dad, with boundless love and appreciation.

Abstract

Time inconsistency is of primary concern in dynamic economic environments. People with time-inconsistent preferences have difficulty saving for retirement or avoid accumulating too much debt. Hence, interactions with time-inconsistent agents differ from predictions made by classical economics with time-consistent agents. If agents are time inconsistent, government policies that encourage retirement savings would have to account for this behavior. Contracts in the credit market between creditors and debtors would also be different if debtors are time inconsistent. The first two chapters in the thesis analyze these interactions. Institutions may also display time inconsistency. In the final chapter, my coauthor, Radek Paluszynski, and I document an empirical phenomenon in the pricing of life-insurance. We find that prices are extremely rigid over time, even though the marginal cost is nonstationary and volatile. We show that such pricing can occur when firms have time-inconsistency problem in setting its price.

In Chapter 1, I explore the optimal tax policies with time-inconsistent agents. People with time-inconsistent preferences tend to make intertemporal choices different from their original intentions. In particular, time-inconsistent agents make sub-optimal saving decisions. This chapter studies the optimal savings and insurance policy in an economy with time-inconsistent agents who privately observe their skill. I introduce a mechanism that can elicit private information at zero cost by threatening sophisticated time-inconsistent agents with off-equilibrium path policies that can undo commitments, and fooling naïvely time-inconsistent agents with empty promises. The government can implement the full information efficient optimum, which is better than the constrained optimum obtained with traditional proposals to increase savings, like linear savings subsidies or mandatory savings rules. I show that welfare increases monotonically with the population of time-inconsistent agents. In essence, the presence of time-inconsistent agents improves the government's ability to provide insurance. I characterize the tax policies for decentralization, and also discuss its implications on the design of social security and retirement plans.

In Chapter 2, I analyze price discrimination by firms offering credit contracts in a sequential screening model with sophisticated time-inconsistent consumers. Consumers choose a credit contract after learning about their likely future taste for consumption and proceed to

make consumption decisions after realizing their tastes. I show that a credit company price discriminates by providing immediate gratification combined with large payment penalties to time-inconsistent consumers with high taste shocks, while consumers with lower taste shocks are provided with commitment contracts. This may cause time-inconsistent consumers to experience large welfare losses ex-post without being naïve about their preference reversal. I show that policies, such as the Credit Card Accountability, Responsibility and Disclosure act, aimed at avoiding ex-post welfare loss may induce ex-ante deadweight loss.

In Chapter 3, I explore an interesting price phenomenon exhibited in life insurance contracts. Life insurance premiums have displayed a significant degree of rigidity over the past two decades. On average, prices took over 3 years to adjust and the magnitude of these one-time jumps exceeded 10%. This stands in sharp contrast with the dynamics of the corresponding marginal cost which exhibited considerable volatility since 1990 due to the movements in the interest and mortality rates. We build a model with consumer hold-up problem that captures these empirical findings. Price rigidity arises as an optimal response to the relationship-specific investment the consumers need to make before buying. The optimal contract takes the form of a simple cutoff rule: premiums are rigid for all cost realizations smaller than the threshold, and adjustments must be large and are only possible when cost realizations exceed it.

Contents

Acknowledgements	i
Dedication	ii
Abstract	iii
List of Tables	ix
List of Figures	x
1 Optimal Taxation with Time-Inconsistent Agents	1
1.1 Introduction	1
1.1.1 Related Literature	6
1.2 The Model	9
1.2.1 Setup of the Model	9
1.2.2 The Benchmark	14
1.2.3 Incentive Compatibility	16
1.3 Savings with Non-sophisticated Agents	17
1.3.1 The Fooling Mechanism: Real and Imaginary Allocations	17
1.3.2 Main Result for Non-Sophisticated Agents	19
1.4 Savings with Sophisticated Agents	21
1.4.1 The Threat Mechanism: Real and Threat Allocations	22
1.4.2 Main Result for Sophisticated Agents	23
1.4.3 The Threat Mechanism for Partially Naïve Agents	26
1.5 Diversely Time-Inconsistent Agents	27

1.5.1	Hidden Sophistication	27
1.5.2	Hidden Temptation	29
1.6	Model with Time-Consistent Agents	30
1.6.1	The Threat Mechanism with Time-Consistent Agents	30
1.6.2	The Fooling Mechanism with Time-Consistent Agents	36
1.6.3	Incentives for Self-Awareness	38
1.6.4	Paranoia	39
1.7	Policy Implementation	41
1.7.1	Decentralizing the Fooling Mechanism	41
1.7.2	Decentralizing the Threat Mechanism	43
1.7.3	Decentralization with Time-Consistent Agents	45
1.8	Social Security and Retirement Plans	47
1.8.1	The Timing of Claiming Social Security Benefits	47
1.8.2	Liquidity of Defined Contribution Plans	50
1.9	Discussions	53
1.9.1	Dynamic Stochastic Productivity Shocks	53
1.9.2	Demand for Flexibility	56
1.9.3	Impediments to Implementation	57
1.10	Summary and Conclusion	60
2	Price Discrimination in the Credit Market	61
2.1	Introduction	61
2.2	Model	63
2.2.1	Consumer: Taste and Time-Inconsistency	64
2.2.2	Firm: Menu and Incentive Compatibility	65
2.2.3	Timing	66
2.2.4	Perfect Price Discrimination	67
2.3	Static Information Revelation	67
2.3.1	The Threat Mechanism	67
2.4	Dynamic Information Revelation	69
2.4.1	Public Shocks and Ex-Ante Information Rents	69
2.4.2	Private Shocks and Ex-Post Information Rents	70
2.5	The Credit Market and Policy Intervention	74

2.6	Conclusion	75
3	Pay What Your Dad Paid: Commitment and Price Rigidity in the Market for Life Insurance - <i>with Radek Paluszynski</i>	76
3.1	Introduction	76
3.2	Literature Review	80
3.3	Life Insurance Prices	82
3.3.1	Data Construction	82
3.3.2	Historical Premiums	83
3.3.3	Marginal Cost Estimation	87
3.4	The Model	88
3.4.1	The Setup	89
3.4.2	Characterizing the Demand	93
3.4.3	Incentive Compatibility	97
3.4.4	Equilibrium Definition and the Optimization Problem	98
3.5	Characterization of the Optimal Premium Schedule	99
3.5.1	Qualitative Features of the Optimal Premium Schedule	99
3.5.2	Computing the Optimal Premium Schedule	103
3.6	Alternative Explanations	106
3.6.1	Menu Cost	106
3.6.2	Adverse Selection and Avoiding the ‘Death Spiral’	107
3.7	Conclusion	108
	References	109
4	References	109
	Appendix A. Appendix for Chapter 1	116
A.1	General Model	116
A.1.1	Setup of Model	116
A.1.2	The Effects of Non-sophistication	119
A.1.3	The Effects of Sophistication	124
A.1.4	Model with Diversely Naïve Agents	129
A.2	Proofs for Section 1.6 and Beyond	135

Appendix B. Appendix for Chapter 2	149
B.1 Proofs	149
Appendix C. Appendix for Chapter 3	156
C.1 Estimating the Marginal Cost of a Life Insurance Company	156
C.2 Incentive Compatibility Constraints	158
C.3 Proofs	158

List of Tables

1.1	Summary of Mechanism for Different Sophistication Levels	27
3.1	Structure of an Annual Renewable Term (ART) contract	77
3.2	Price rigidity in the sample	84

List of Figures

1.1	Welfare Comparison with Traditional Policy Instruments	3
1.2	Welfare and Proportion of Time-Inconsistent Agents	5
1.3	Timing of Events	12
1.4	Finding the Imaginary Allocations	21
1.5	Finding the Threat Allocation: Part I	24
1.6	Finding the Threat Allocation: Part II	25
1.7	Limited Fooling	58
1.8	Limited Threats	58
2.1	Timing of Events	67
2.2	Leisure Goods with Time-Inconsistent Consumers	69
2.3	Full Support of Types	71
3.1	Premiums over time for different level-term policies	78
3.2	Distribution of premium durations and adjustment sizes	85
3.3	Distribution of insurance premiums, relative to the cross-sectional average	86
3.4	Stylized structure of a life insurer's expected cash flows	87
3.5	Net premium for an Annual Renewable Term policy over time	88
3.6	Timing of events	92
3.7	Distribution of consumers and their investment decisions	95
3.8	Second period incentive compatible premium profile for a given \hat{c}_1	103
3.9	Premium duration and adjustment size in the life insurance market	107
A.1	Timing of Events	119

Chapter 1

Optimal Taxation with Time-Inconsistent Agents

1.1 Introduction

Empirical evidence shows that time-inconsistent behaviors exhibited in the real world have substantial impact and are pervasive.¹ Such bias is present even in important issues like retirement savings or investment portfolios.² For example, for retirement savings, the National Research Council (2012) finds that up to $\frac{2}{3}$ of the US population is saving inadequately for retirement.³ It is argued that people want to save more, but controlling ones' impulses is difficult.⁴ Furthermore, research implies that models with time-inconsistent

¹DellaVigna and Malmendier (2006) study gym membership data and show that 80% of monthly gym members would have been better off had they chosen to pay per visit. Ausubel (1999) and Shui and Ausubel (2005) find similar biases in the credit card market. Gottlieb and Smetters (2013) also find similar evidence in the life insurance market. DellaVigna (2009) provides an overview of the empirical evidence for behavioral economics.

²Benartzi and Thaler (2001) find that individuals do poorly when it comes to investment diversification. Madrian and Shea (2001) study participation in the 401(k) plan and find evidence of strong default effects on the participation decision of individuals. O'Donoghue and Rabin (2001) show that a model of time-inconsistent individuals with naïveté can explain the influence of the default option on retirement savings.

³Scholz et al. (2006), one of the more conservative studies, estimates at least 20% of the US population are not saving enough for retirement.

⁴Choi et al. (2002) randomly surveyed employees of a large US corporation and found that 67.7% of the respondents felt that their saving rate is too low relative to their ideal saving rate, but only 4% of them took action to increase their saving rate over the next few months.

preference can explain the consumption and savings patterns observed in the data.⁵ As a result, the literature argues for the government to implement policies that could offset the time-inconsistent behavior, for example, savings subsidies and mandatory savings policies have been suggested.⁶ However, the recommended policies for time-inconsistent agents have so far been analyzed in isolation from other government objectives.

In this paper, I introduce new policy tools to increase savings when the government also wishes to insure agents against their productivity realizations. More specifically, this paper considers an environment with present-biased agents who privately observe their productivity. As a result, in addition to the efficiency and equity trade-off in a traditional Mirrlees environment, the government has the additional motive to increase retirement savings. The main contribution of this paper is to describe how the insurance problem changes with the introduction of time-inconsistent preferences, and to characterize the optimal policy. I demonstrate policy implementation with taxation and also highlight its implications on the design of social security benefits and retirement plans. For example, this paper implies that social security benefits for agents who start claiming at 70 to be regressive in income and progressive for when they start at age 62.

This paper also analyzes the problem for a full spectrum of sophistication levels. Agents who are fully aware of their time-inconsistency are *sophisticated*, while those who are completely unaware are *naïve*. Agents who are aware of their time-inconsistency but are wrong about the severity are called *partially naïve*. The distinction among different sophistication levels is important because the recommended policy varies with the sophistication level of the agents.

Traditional policy suggestions, such as linear savings subsidies or mandatory savings, can mitigate the agents' present bias. Such policy proposals offset the bias independent of the surrounding economic environment. Figure 1.1 shows how varying the linear savings subsidy rate can help the government attain the constrained efficient optimum when time-inconsistent agents are present. The government can set the savings subsidy rate to

⁵Angeletos et al. (2001) and Laibson et al. (2007) examine the implications of quasi-hyperbolic discounting on life-cycle intertemporal decisions. They find that introducing quasi-hyperbolic discounting helps explain several patterns that cannot be explained by traditional exponential discounting models, like the co-existence of high credit card debt with a high demand for illiquid assets.

⁶Theoretical models with time-inconsistent preferences have been used to justify the introduction of such policies. See Section 1.1.1 for more details.

exactly offset the agents' present bias and proceed to implement the constrained efficient allocations. Hence, using traditional policy proposals, the government is able to guarantee the constrained efficient welfare level, but it cannot do better.

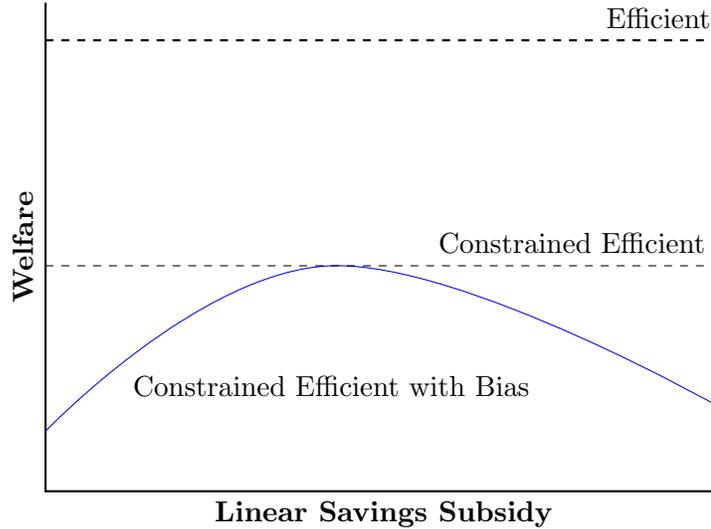


Figure 1.1: Welfare Comparison with Traditional Policy Instruments

I consider a set of more sophisticated policy instruments and find that the optimal set of allocations differs from the constrained efficient allocations. In particular, the government is able to fully insure all agents and avoid any labor distortions. More specifically, the main result shows that, with time-inconsistent agents, the government can implement the full information efficient allocation (Henceforth, referred to as efficient allocation) despite the presence of information asymmetry. Therefore, the government is able to do better than the constrained efficient optimum.

This surprising result is due to the fact that with naïve agents, the government induces efficient labor provision by promising a savings subsidy that increases with income. However, after preference reversal, the agents no longer value savings as much and would instead prefer the proportion of consumption and savings that corresponds to the efficient allocation. Naïve agents do not foresee this change in preference. In other words, the government loads the information rent on savings to induce efficient output, which the agents do not value after preference reversal. The government *fools* the naïve agents into revealing their private information and does not need to deliver on the promise. Thus, the government can

implement the full insurance policy without paying any information rent. Since partially naïve agents hold incorrect beliefs about their bias, deception is effective in implementing the efficient allocation as long as agents are non-sophisticated.

For sophisticated agents, the government exploits the fact that they know about their time-inconsistency and thus have a demand for commitment. The government will provide commitment only if the agents report truthfully. To deter misreports, the government designs a threat allocation that unravels the commitment such that a misreporting agent would be tempted to choose it, while an honest agent would not. This is possible due to differing marginal rates of substitution in labor and consumption for agents of different productivity types. Since credible threats can be constructed and the sophisticated agents can foresee the future consequences of lying, they report truthfully. In other words, efficiency is obtained using credible off-equilibrium path threats. Since partially naïve agents also have a demand for commitment, screening with credible threats also works for partially naïve agents.

In an environment where both sophisticated and naïve agents coexist, the government needs to screen the agents' productivity and their sophistication level. I show that mechanisms designed to fool more naïve agents and threaten more sophisticated ones can be combined without any efficiency loss. Therefore, the multidimensional screening problem does not alter the main result. The same method can be extended to an environment where time-inconsistent agents differ in the degree of present bias.

I also consider an economy where the government is uncertain whether a preference reversal would occur, in essence, some agents could be time-consistent. The environment with time-consistent agents limits the ability of the government to fool or threaten the agents. I find that welfare increases with the population of time-inconsistent agents. Figure 1.2 demonstrates this. This suggests that incentive compatibility causes distortions only in environments with time-consistent agents. This is because total output increases with the proportion of time-inconsistent agents. Hence, the government can provide more information rent per time-consistent agent without using more resources. This relaxes the incentive compatibility constraints and improves welfare. This also suggests that the estimate in Farhi and Werning (2012) for the welfare benefits of adopting sophisticated intertemporal policies is a lower bound. The potential gains for adopting sophisticated intertemporal policies could be much larger depending on the population size of time-inconsistent individuals.

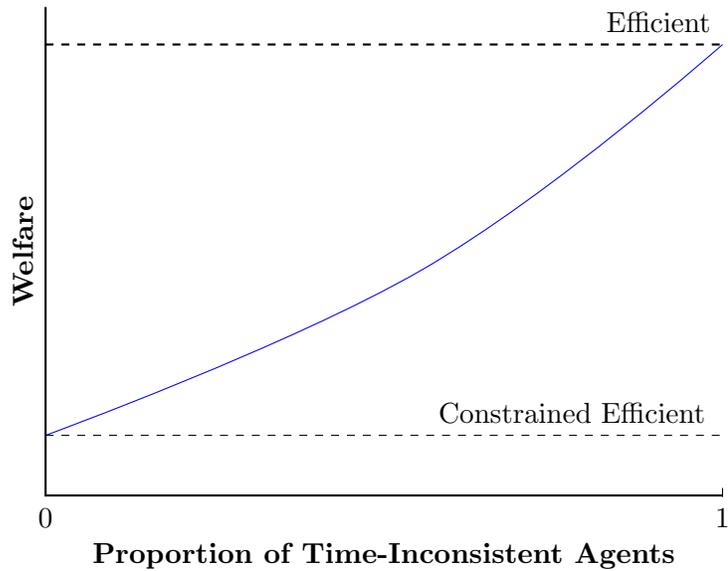


Figure 1.2: Welfare and Proportion of Time-Inconsistent Agents

The model with time-consistent agents also sheds light on whether the government should help the agents become more sophisticated. I compare the welfare of different sophistication levels and find that a higher sophistication level weakly improves welfare. In particular, the government would prefer fully naïve agents to be at least partially aware of their temptation. Meanwhile, the government is indifferent between other levels of sophistication.

With time-inconsistent agents, the government can increase production efficiency through the design of savings policies. This reduces the role income taxes play in the provision of work incentives, so income taxes are used largely for redistributive purposes. For implementation, I demonstrate how the government can use an income regressive non-linear savings subsidy to fool the non-sophisticated agents. For sophisticated agents, the government can implement the efficient allocation using loans with progressive repayment schemes. The loan acts as a commitment device and the repayment schemes are progressive so that a misreporting agent who earns more than initially planned would be punished with a reduction in retirement savings.

I also discuss the design of social security benefits. It has been argued that people in the US are claiming social security benefits too early in life, and should instead retire

later and delay benefits claiming.⁷ Hence, previous literature has focused on policies that encouraged delay in claiming benefits.⁸ Unlike previous work, this paper takes advantage of the behavioral bias and shows how benefits can be designed to improve insurance and smooth consumption. It suggests that social security benefits should be regressive in income when agents claim at the age of 70, and progressive if they claim at the age of 62. An implementation with retirement benefits is also demonstrated. This paper shows that the ability of agents to withdraw early from their defined contribution plans can help implement the efficient allocation. Since this additional flexibility, the ability to withdraw early, can be constructed as a credible off-equilibrium path threat that sustains full efficiency on-equilibrium path. Consequently, my work provides new suggestions on the design of policies to help agents ameliorate their time-inconsistency and simultaneously improve equality.

The results of this paper apply to a very general setting with agents who experience preference changes and may or may not be accurate in their predictions of these changes.⁹ This demonstrates that the results may be applied more generally to different environments with dynamic inconsistencies and private information.¹⁰

1.1.1 Related Literature

This paper tries to combine two strands of literature: the optimal taxation literature and the behavioral contracting literature. Some papers have already attempted to do so.

Farhi and Gabaix (2015) studied optimal taxation with agents who suffer from a wide array of behavioral biases. They analyzed how the optimal tax formulas would be altered under Ramsey, Pigou and Mirrlees taxation. In contrast to their work, this paper attempts to focus on a certain behavioral bias, time-inconsistency, and to analyze the optimal policies from a mechanism design perspective. In particular, this paper adopts a different timing: agents are allowed to commit before making allocation decisions. In Farhi and Gabaix (2015), this ex-ante stage is absent.

Krusell et al. (2010) study the optimal taxation of consumers who suffer from temptation. They find that the government should subsidize savings to correct the agent's

⁷See Coile et al. (2002), Shoven and Slavov (2014) and Knoll et al. (2015).

⁸The National Commission on Fiscal Responsibility and Reform (2010) recommends the use of behavioral economics to nudge people to retire and claim benefits later.

⁹Appendix A provides a general model.

¹⁰Chapter 2 studies the optimal pricing of a monopoly firm facing time-inconsistent consumers.

impatience and tendency to save too little. My work differs from Krusell et al. (2010) in two aspects. Firstly, their environment is a complete information one, while I introduce asymmetric information in productivity. Secondly, I also consider non-sophisticated agents, while their agents are sophisticated.

Amador et al. (2006) examine government policies that could help agents with temptation. They study agents who suffer from temptation and are subject to future taste shocks. Hence, a preference for commitment and flexibility coexists, which causes a trade-off. They find that a minimum savings rule is optimal in this environment. Their work considers a sophisticated agent, and the adverse selection problem is in the agent's taste shock. The main difference is that my work seeks to analyze how government policies aimed at increasing savings of a time-inconsistent agent could affect labor decisions. Therefore, I have a production economy and the government has insurance motives, while Amador et al. (2006) focus on an endowment economy with a government only trying to smooth consumption.

Halac and Yared (2014) apply a repeated model of Amador et al. (2006) with persistent shocks to investigate the trade-off between rules and discretion for a monetary authority. Similar to this paper, Halac and Yared (2014) choose more relaxed policies that tempt the future-self to threaten and discipline present behavior. Such relaxed policies can deter misreporting, but at the expense of exacerbating the present bias of certain types. However, in my setting, I show how it is possible to construct off-equilibrium path policies such that it causes no distortions.

A few papers have studied the optimal taxation problem with hidden productivity and time-inconsistent agents. Lockwood (2015) considers agents with present bias. He considers both naïve and sophisticated agents. Guo and Krause (2015) study an environment where the government does not have full commitment. These papers address the same issue as this one, but they do not exploit the agents' time-inconsistency and are unable to achieve the full information optimum.

Several papers have examined taxation models where individuals are differentiated along

two or more dimensions.¹¹ There are some closely related papers that address optimal savings policies with agents having additional private information in time preference. Diamond and Spinnewijn (2011) discuss a model with heterogeneity in both productivity and time preference (agents are time-consistent with different discount factors). Bassi (2010) considers an environment with time-inconsistent agents where the quasi-hyperbolic discount factor is also non-observable, which creates a two-dimensional screening problem for the government. More recently, Moser and de Souza e Silva (2015) consider a similar two-dimensional screening setup and decentralizes the optimum using social security and retirement accounts. Social security is the main savings instrument for low income individuals, while high income individuals have access to more flexible instruments such as 401(k). The flexibility is part of the information rent for high productivity individuals. This paper is also concerned with the policy on savings, but in addition to heterogeneous productivity, the agents also differ in their awareness of the underlying present bias and in their time-inconsistency. This type of multidimensional screening has yet to be analyzed in public policy or the economic literature.

The paper is also related to several behavioral contracting papers. In particular, Esteban and Miyagawa (2005) examine optimal pricing schemes with time-inconsistent agents and also find that distortions caused by information asymmetry can be averted for a given type of temptation: when agents are tempted to over-consume. Another closely related work is Bond and Sigurdsson (2015). Under certain conditions, they show that it is possible to include off-equilibrium path options in commitment contracts so that a time-inconsistent agent would follow through with an ex-ante plan that accommodates flexible needs. Eliaz and Spiegler (2006) examine a model with diversely naïve agents and found that firms can screen beliefs by bisecting the population into relatively sophisticated and relatively naïve agents. Similar to this paper, they also find that relatively sophisticated agents exert no informational externality on the relatively naïve agents. More recently, Galperti (2015) extends Amador et al. (2006) to a sequential screening model where a mechanism designer first screens an agent's time-consistency and then the realized taste shock. This paper will

¹¹More notably, Cremer et al. (2001) examine a model where both the productivity level and endowments are not observed by the government. Cremer et al. (2003) extend the model to an overlapping generations setting and endogenize individual endowments as inherited wealth. Beaudry et al. (2009) examines an economy where agents could participate in both market and non-market production and have different unobservable productivity levels for both sectors.

consider an economy where the government screens both the agents' productivity and their time-consistency simultaneously.

The paper is organized as follows. Section 1.2 outlines the setup of a three-period model with present bias. Section 1.3 and Section 1.4 work out the results for the three-period model for non-sophisticated and sophisticated agents respectively. Section 1.5 examines the environment with hidden sophistication level and hidden time-inconsistency. Section 1.6 presents an analysis of a model where the government is not certain whether an agent is time-consistent or time-inconsistent. Section 1.7 provides methods for decentralizing the allocations, and Section 1.8 provides ways to implement the optimum with social security and defined contribution plans. Section 1.9 considers extensions with dynamic stochastic productivity shocks and commitment versus flexibility preferences, and discusses some of the impediments to the implementation of the proposed mechanism. Section 1.10 concludes the paper. All proofs of results before Section 1.6 can be found in Appendix A, where I present a general three-period model. Proofs of results contained in Section 1.6 and beyond can be found in Appendix B.

1.2 The Model

This section describes the environment, the welfare criterion, and sets up the mechanism.

1.2.1 Setup of the Model

A continuum of agents of measure one live for three periods. In the first period, the agents realize their productivity and do not consume.¹² They produce and make consumption and savings decisions in the second period, and consume their savings in the third and final period. Agents are differentiated by their production efficiency θ . There are $|M| \geq 2$ types of agents denoted by the set $\Theta = \{\theta_1, \theta_2, \dots, \theta_M\}$, with $\theta_{m+1} > \theta_m$. The types are distributed according to $\Pr(\theta = \theta_m) = \pi_m > 0$, for all $\theta_m \in \Theta$ with $\sum_{m=1}^M \pi_m = 1$.

The production technology is linear and depends only on labor input l and the productivity of the agent: $y = \theta l$. In a competitive equilibrium, the wages are equated to the marginal productivity of labor. There is also a storage technology that transfers one unit of good in the first period to one unit of second period good. (Alternatively, they have access

¹²The main results do not change if the agents require consumption in the first period.

to a bond with interest rate 0.)

As is standard in Mirrlees taxation, both the production efficiency θ of each agent and their labor input l are not observed by the government. The government can only observe output y .

Agent Utility

The agents have the following utility before consumption

$$U(c, k, y; \theta) = u(c) - h\left(\frac{y}{\theta}\right) + w(k),$$

where c is consumption in the second period and k is savings for the final period. I will refer to U as the *ex-ante utility*. I also refer the incarnation of the agent with the ex-ante utility as the *planner*. This is the utility agents use to evaluate their consumption plans. The agents' utility changes when they are consuming to

$$V(c, k, y; \theta) = u(c) - h\left(\frac{y}{\theta}\right) + \beta w(k).$$

I will refer to V as the *ex-post utility*. I also refer the incarnation of the agent with the ex-post utility as the *doer*.¹³ This utility models the tendency of a time-inconsistent agent to deviate from plans when confronted with an allocation decision.

The period utilities $u : \mathbb{R}_+ \mapsto \mathbb{R}$ and $w : \mathbb{R}_+ \mapsto \mathbb{R}$ are continuously differentiable, unbounded below and satisfy the usual strictly increasing and concavity assumptions: $u', -u'' > 0$ and $w', -w'' > 0$, while the dis-utility from labor $h : \mathbb{R}_+ \mapsto \mathbb{R}$ satisfies $h', h'' > 0$. Furthermore, $u', w' > \epsilon$ for some $\epsilon > 0$. This rules out the case where $u(c)$ or $w(k)$ asymptotically approaches a finite number when c or k approaches infinity, so $u(c)$ and $w(k)$ are unbounded above.

The Spence-Mirrlees single crossing property is automatically satisfied under the assumptions on the utility functions and the production function for both U and V . In other words, the marginal rates of substitution between consumption and output and between savings and output are smaller for more efficient agents: $\frac{\partial}{\partial \theta} \left(-\frac{\partial U}{\partial y} / \frac{\partial U}{\partial c} \right), \frac{\partial}{\partial \theta} \left(-\frac{\partial U}{\partial y} / \frac{\partial U}{\partial k} \right) < 0$ and $\frac{\partial}{\partial \theta} \left(-\frac{\partial V}{\partial y} / \frac{\partial V}{\partial c} \right), \frac{\partial}{\partial \theta} \left(-\frac{\partial V}{\partial y} / \frac{\partial V}{\partial k} \right) < 0$.

¹³The terminologies 'planner' and 'doer' is derived from Thaler and Shefrin (1981), and has subsequently been widely used in the behavioral literature. In this paper, the word 'planner' would refer to the farsighted incarnation of the agent, and should not be confused with the government.

I will focus on the case with present bias, where $\beta < 1$.¹⁴ A smaller β represents a stronger bias for present consumption. I will refer to β as measuring the degree of temptation the agents sufferer from. With $\beta \neq 1$, the marginal rate of substitution between consumption and savings ($MRS_{c,k}$) is different for U and V : $\frac{\partial U}{\partial c} / \frac{\partial U}{\partial k} \neq \frac{\partial V}{\partial c} / \frac{\partial V}{\partial k}$. Since both utilities are separable in consumption and labor, $MRS_{c,k}$ is independent of the agents' labor choice. Along with strictly increasing and concavity assumptions, the difference in $MRS_{c,k}$ between U and V implies a single crossing condition on the indifference curves for the ex-ante and ex-post utilities of c and k . I will refer to preferences that display this difference between ex-ante and ex-post utility as exhibiting *preference reversal*.¹⁵ Preference reversal is key for allowing the government to implement policies that would seem attractive to the planner while remaining undesirable for the doer, or vice versa.

Timing

At date 0, the government designs the tax system. By the law of large numbers, the government knows the measure of each type of agent. At date 1, the agents' types are realized. At date 2, the agents report their types according to reporting strategy $\sigma : \Theta \mapsto \Theta$. They make decisions according to the ex-ante utility from date 0 to date 2. At date 3, just when the agents are making their consumption decisions, their preferences switch, and they make their consumption and labor decisions based on the ex-post utility. The timing of the model is shown in Figure 2.1.

I assume that the government has full commitment. Hence, the revelation principle

¹⁴The main results do not change if $\beta > 1$.

¹⁵In the current setup, the present bias is similar to a temptation shock that triggers immediate gratification and under-weighs the virtues of saving. This formulation has an equivalent quasi-hyperbolic discounting representation. The three period quasi-hyperbolic representation shown in Figure 2.1 is

$$\begin{aligned} U_1(c, k, y; \theta) &= \hat{\beta}\delta \left[u(c) - h\left(\frac{y}{\theta}\right) + \delta w(k) \right], \\ U_2(c, k, y; \theta) &= u(c) - h\left(\frac{y}{\theta}\right) + \hat{\beta}\delta w(k), \\ U_3(k) &= w(k), \end{aligned}$$

with $\delta = 1$. In the first period, the agents enroll in an income-specific savings plan (with several options available in the plan). In the second period, agents work and make consumption and savings decision according to the plan. Finally, in the third period, agents consume their retirement savings. If $\hat{\beta} = \beta$, then the agents are sophisticated and the transformed model is similar to Laibson (1997), and if not, then it is similar to the model with cognitive limitations as presented in O'Donoghue and Rabin (2001).

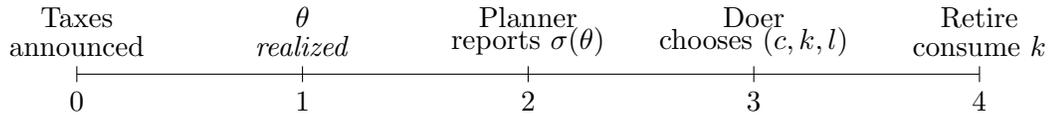


Figure 1.3: Timing of Events

can be applied. For ease of analysis, the current timing focuses on a direct mechanism, where the agents report their productivity types and the government assigns allocations according to the reports. For policy implementation, the timing can be dependent on the sophistication level of the agents.

Modeling Non-Sophistication

Following O’Donoghue and Rabin (2001), partially naïve agents perceive their degree of present bias to be $\hat{\beta} \in (\beta, 1]$. Let $W(c, k, y; \theta)$ denote the non-sophisticated agents’ perceived ex-post utility:

$$W(c, k, y; \theta) = u(c) - h\left(\frac{y}{\theta}\right) + \hat{\beta}w(k).$$

Notice if $\hat{\beta} = 1$, then the planner is fully naïve and unaware of the doer’s present bias. If $\hat{\beta} = \beta$, then the planner is sophisticated and aware of the doer’s bias. Partially naïve agents know they have present bias, $\hat{\beta} < 1$, but their perceived present bias is always strictly greater than the actual present bias, $\hat{\beta} > \beta$. In other words, the planner underestimates the severity of the doer’s temptation problem.¹⁶ I will refer to the perception or belief $\hat{\beta}$ as describing the *sophistication level* of an agent.¹⁷

¹⁶The main results do not change if $\hat{\beta} < \beta$.

¹⁷There are two common ways to model partial naïveté. Loewenstein et al. (2003) and Heidhues and Koszegi (2010) have interpreted partial naïveté as the underestimation of the *magnitude* of temptation. Eliaz and Spiegler (2006) have interpreted partial naïveté as the underestimation of the *likelihood* of temptation occurring. Following Spiegler (2011), I will refer to the former as *magnitude naïveté* and the latter as *frequency naïveté*. The paper focuses on magnitude naïveté and address *frequency naïveté* along with the general model in Appendix A. The main results go though for both types of non-sophistication.

Welfare Criterion

The government tries to help the agents commit to the ex-ante utility. Implicitly, I assume the non-sophisticated agents do not draw any inferences from the policies the government enacts. This is because they do not share the same prior as the government and are dogmatic in their beliefs.

The government evaluates allocations at date 0 according to the following welfare criterion

$$\sum_{m=1}^M \pi_m [\kappa U(c_m, k_m, y_m; \theta_m) + (1 - \kappa) V(c_m, k_m, y_m; \theta_m)], \quad (1.1)$$

where (c_m, k_m, y_m) denotes the allocation type θ_m agent consumes and $\kappa \in (0, 1]$ represents the welfare weight the government places on the planner. Since both U and V are strictly increasing and concave in (c, k) , the government has a desire to insure agents against the realization of θ .

Much of the literature on dynamically inconsistent preferences have evaluated welfare solely with the ex-ante utility, or $\kappa = 1$. I adopt a Pareto criterion and allow the government to place positive welfare weight on the doer as well, as long as $\kappa \in (0, 1]$. Therefore, the government has a desire to smooth the agents' consumption across periods. Regardless of the perspective on the appropriate welfare weights, the main results of the paper are robust to changes in κ .¹⁸

The Availability of Insurance and Commitment

I assume that no private markets exist to insure against productivity shocks so the government has a role in insurance provision.¹⁹ I also assume that there are no markets for illiquid assets or other commitment devices. If such a market exists, then sophisticated agents can use it for commitment. The government would have a limited role in helping

¹⁸The welfare weight on the planner is strictly positive because I adopt the perspective that the ex-ante utility reflects the agents' long-term planning, while the ex-post utility reflects the agents' short-term temptations. In other words, the ex-post utility is not immune to regret and a benevolent government would consider the adverse implications if the agents give in to their urges. As a result, the choice of the welfare criteria is non-arbitrary, since the actions undertaken by the doer can be regarded as a systematic mistake the agents make, as in Bernheim and Rangel (2004).

¹⁹Prescott and Townsend (1984) show that the presence of an efficient market that allows agents to insure against future productivity shocks can make distortionary taxes redundant.

the sophisticated agents smooth consumption. However, for non-sophisticated agents, the presence of such markets does not preclude the need for government intervention.²⁰ In Section 1.9, I discuss how the presence of a market for commitment can affect the results of this paper.

1.2.2 The Benchmark

No Private Information

In the benchmark, no private information case, the government maximizes social welfare (1.1) subject to the feasibility constraint

$$\sum_{m=1}^M \pi_m (\theta_m l_m - c_m - k_m) - G = 0, \quad (1.2)$$

where G is government's external revenue needs. For the remainder of the paper, $G = 0$, so the sole role of the government is to provide insurance and smooth consumption.

With complete information, the government can achieve full insurance regardless of the agents' time-inconsistency or their degree of naïveté. This is because with complete information, the agents work according to their productivity type. The government then chooses an appropriate linear tax to correct the distortion caused by the present bias. The optimal allocation of the government's problem without private information is referred to as the efficient allocation.

Let (c_m^*, k_m^*, y_m^*) denote the efficient allocation for type θ_m . The efficient consumption satisfies $\forall \theta_m \in \Theta$,

$$\frac{\partial [\kappa U + (1 - \kappa)V]}{\partial c_m^*} = \frac{\partial [\kappa U + (1 - \kappa)V]}{\partial k_m^*}.$$

Since $U \neq V$ and $\kappa > 0$, the government can choose linear taxes or subsidies τ^c and τ^k on c and k respectively such that $\frac{1}{1+\tau^c} \frac{\partial V}{\partial c_m^*} = \frac{1}{1+\tau^k} \frac{\partial V}{\partial k_m^*}$, which implements the efficient consumption. Therefore, if the agents save too little, the government can subsidize savings k or tax c . The following proposition characterizes the properties of the efficient allocation.

Proposition 1 *The efficient allocation $\{(c_m^*, k_m^*, l_m^*)\}_{\theta_m \in \Theta}$ satisfies the following:*

1. *Full insurance: For any $\theta_m, \theta_{m'} \in \Theta$, $c_m^* = c_{m'}^*$ and $k_m^* = k_{m'}^*$.*

²⁰Heidhues and Koszegi (2009) demonstrate how commitment devices can do more harm than good for non-sophisticated agents, since they pay the cost of commitment but still suffer from self-control problems.

2. *Consumption smoothing:* For any $\theta_m \in \Theta$, $u'(c_m^*) = [\kappa + (1 - \kappa)\beta] w'(k_m^*)$.
3. *Efficient output:* For any $\theta_m \in \Theta$, $u'(c_m^*) = \frac{1}{\theta_m} h'(l_m^*)$.

Consumption smoothing is achieved since the government puts a strictly positive weight κ on the ex-ante utility. As a special case, if $\kappa = 1$, then I will refer to the efficient allocation as attaining *perfect consumption smoothing*: $u'(c_m^*) = w'(k_m^*)$.

There are three important features of the policy instrument in this environment. First, the taxes or subsidies can be linear. Second, these linear instruments are the same across all productivity types. This is because all productivity types suffer from the same degree of temptation. Finally, the linear taxes or subsidies also work independently of sophistication levels. This is because regardless of the sophistication level, the policy enacted to correct for the taste change does not distort the incentives to work. The government can always ‘force’ the productive agents to work more than the less productive agents.

Private Information without Time-Inconsistency

For the case with private information, the implementable allocations must be incentive compatible. The government maximizes social welfare (1.1) subject to the feasibility constraint (1.2) and the incentive compatibility constraints, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

$$u(c_m) - h\left(\frac{y_m}{\theta_m}\right) + w(k_m) \geq u(c_{\hat{m}}) - h\left(\frac{y_{\hat{m}}}{\theta_m}\right) + w(k_{\hat{m}}). \quad (1.3)$$

The efficient allocations are not incentive compatible, so with private information and time-consistent agents, the government can only implement the constrained efficient allocation. The following proposition characterizes the properties of the constrained efficient allocation.

Proposition 2 *The constrained efficient allocation $\{(c_m^{**}, k_m^{**}, l_m^{**})\}_{\theta_m \in \Theta}$ satisfies the following:*

1. *Partial insurance:* For any $\theta_m, \theta_{m'} \in \Theta$, with $\theta_m > \theta_{m'}$, $c_m^{**} > c_{m'}^{**}$ and $k_m^{**} > k_{m'}^{**}$.
2. *Consumption smoothing:* For any $\theta_m \in \Theta$, $u'(c_m^{**}) = [\kappa + (1 - \kappa)\beta] w'(k_m^{**})$.
3. *Output distortions:* For any $\theta_m < \theta_M$, $u'(c_m^{**}) > \frac{1}{\theta_m} h'(l_m^{**})$.

The constrained efficient allocation distorts the labor decision of all agents except for the most productive agent θ_M . This distortion relaxes the incentive compatibility constraint, which allows the government to provide partial insurance for productivity shocks. Hence, Proposition 2 characterizes the optimal trade-off between efficiency and equity.

With sophisticated time-inconsistent agents, if the government uses a linear savings subsidy to correct the present bias, the constrained efficient optimum can still be obtained, as shown in Figure 1.1. This is because the linear savings subsidy would correct the doer's present bias without changing the set of incentive compatible allocations. Hence, the planner would face the same incentive compatibility constraints and the government can implement the constrained efficient allocations. In the following subsection, I will discuss a more general mechanism which would allow for a richer set of policy instruments.

1.2.3 Incentive Compatibility

The revelation principle applies to this setting, so the analysis will focus on a truth-telling direct mechanism. The government presents a menu of allocations C_m for type θ_m agents defined as

$$C_m = \{(c_m, k_m, y_m), (c'_m, k'_m, y'_m), \dots\}.$$

An agent could be assigned a menu C_m after a report $\sigma(\theta) = \theta_m$ as opposed to an allocation. Previous literature has also highlighted the benefits of using enlarged menus in mechanism design problems with time-inconsistent agents.²¹ Let $(c_m^R, k_m^R, y_m^R) \in C_m$ denote the *real allocation*, which is the optimal allocation the government can implement. Following the timing in Figure 2.1, the government posts $C = \{C_1, \dots, C_M\}$. The agents then proceed to choose a menu from C after learning θ . After temptation sets in, the agents choose an allocation from the menu they initially selected.

Incentive compatibility is characterized by what the planner perceives to be the allocation the doer will choose when he reports truthfully, and when he misreports. Let

$$C_m^{\hat{\beta}} = \left\{ (c_m, k_m, y_m) \in C_m \mid (c_m, k_m, y_m) \in \arg \max_{(c'_m, k'_m, y'_m) \in C_m} W(c'_m, k'_m, y'_m; \theta_m) \right\}.$$

²¹Enlarged menus have been used to exploit time-inconsistent agents in the literature. For instance, Esteban and Miyagawa (2005) showed how enlarging menus could help monopolists achieve perfect price discrimination for a particular type of temptation even when consumer valuations are unobservable. Galperti (2015) used large menus to help separate time-inconsistent agents from the time-consistent agents.

Hence, $C_m^{\hat{\beta}}$ denotes the set of allocations a truthful θ_m agent with sophistication level $\hat{\beta}$ predicts the doer would choose. Let

$$C_{\hat{m}|m}^{\hat{\beta}} = \left\{ (c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}) \in C_{\hat{m}} \mid (c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}) \in \arg \max_{(c'_{\hat{m}}, k'_{\hat{m}}, y'_{\hat{m}}) \in C_{\hat{m}}} W(c'_{\hat{m}}, k'_{\hat{m}}, y'_{\hat{m}}; \theta_m) \right\}.$$

Hence, $C_{\hat{m}|m}^{\hat{\beta}}$ denotes the set of allocations a type θ_m agent with sophistication level $\hat{\beta}$ predicts the doer would choose after he misreports to be a type $\theta_{\hat{m}}$ agent. Incentive compatibility is thus expressed as, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

$$\max_{(c_m, k_m, y_m) \in C_m^{\hat{\beta}}} U(c_m, k_m, y_m; \theta_m) \geq \max_{(c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}) \in C_{\hat{m}|m}^{\hat{\beta}}} U(c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}; \theta_m). \quad (1.4)$$

The incentive compatibility constraints (1.4) make sure the agents choose the menu that is intended for their productivity type given their sophistication level $\hat{\beta}$. Additional constraints are needed to make sure the real allocations are implemented. The executability constraints are, $\forall \theta_m \in \Theta$,

$$(c_m^R, k_m^R, y_m^R) \in \arg \max_{(c_m, k_m, y_m) \in C_m} V(c_m, k_m, y_m; \theta_m). \quad (1.5)$$

The executability constraints (1.5) make sure that the doer would choose the real allocations. In a model with non-common priors, the direct revelation mechanism has to take into account the agents' beliefs through the incentive compatibility constraints and the government's beliefs through the executability constraints.

All other allocations besides the real allocations are considered off-equilibrium path. However, depending on the sophistication level of the agents, the presence of other allocations in the menu can affect the reporting strategy σ .

1.3 Savings with Non-sophisticated Agents

For this section, I will first describe the method the government uses to elicit truth-telling and setup the optimization problem for $\hat{\beta} \in (\beta, 1]$. I will then introduce the first main result which characterizes the optimal allocation for non-sophisticated agents.

1.3.1 The Fooling Mechanism: Real and Imaginary Allocations

Non-sophisticated agents hold erroneous beliefs about their level of temptation, and hence make incorrect predictions about their behavior. This means that their reporting

strategies reflect the expected choices of a fictitious doer, and not the real doer. The government can exploit this misspecified belief by using a *fooling mechanism*.

Definition 1 *A direct fooling mechanism has a menu $C = \{C_m\}_{\theta_m \in \Theta}$ with $C_m = \{(c_m^R, k_m^R, y_m^R), (c_m^I, k_m^I, y_m^I)\}$, satisfying the fooling constraints: $(c_m^I, k_m^I, y_m^I) \in C_m^{\hat{\beta}}$ and $(c_m^I, k_m^I, y_m^I) \in C_{\hat{m}|m}^{\hat{\beta}}, \forall \theta_m, \theta_{\hat{m}} \in \Theta$.*

After the menu is announced, non-sophisticated agents of any type θ_m ‘mentally’ choose a set of allocations (c_m^I, k_m^I, y_m^I) . However, the government intends the agents to ‘actually’ choose allocation (c_m^R, k_m^R, y_m^R) to maximize the ex-post utility. The superscript I represents ‘imaginary,’ since it is never actually selected, but were the perceived choices of the planner before preference reversal. I will set $y_m^R = y_m^I = y_m$, and show that $(c_m^R, k_m^R) \neq (c_m^I, k_m^I)$ is enough to exploit non-sophistication.²²

The planners of non-sophisticated agents underestimate the degree of temptation, so the design of imaginary allocations is important. Planners make their reporting decisions based on $U(c, l)$ while anticipating an ex-post utility of $W(c, l)$. Therefore, to fool the non-sophisticated agents, the government requires the imaginary allocations to be more appealing than the real allocations under $W(c, l)$, and $\sigma(\theta) = \theta$. However, to implement the real allocation, it is required to be more appealing than the imaginary allocation for the actual doer. The analysis in this section will focus on a truth-telling direct revelation mechanism. The following definition defines truthful implementation in a direct fooling mechanism.

Definition 2 *An allocation $\{(c_m^R, k_m^R, y_m^R)\}_{\theta_m \in \Theta}$ is truthfully implementable by a direct fooling mechanism if there exists $\{(c_m^I, k_m^I, y_m^I)\}_{\theta_m \in \Theta}$ such that the*

- i. incentive compatibility constraints, and*
- ii. executability constraints*

are satisfied.

²²The use of imaginary allocations for exploiting non-sophisticates has been explored in the literature before, for instance, Eliaz and Spiegler (2006) and Heidhues and Koszegi (2010). However, it is to the best of this author’s knowledge that this paper is the first to use it to elicit private information without cost.

The government's problem is to choose the menu of allocations, C , to maximize (1.1) subject to the feasibility constraint (1.2) evaluated at the real allocations and the incentive compatibility constraints (1.4), executability constraints (1.5) and the fooling constraints. The imaginary allocations are not required to satisfy the feasibility constraint. This is because the government only cares about the real allocations, and views the imaginary allocations as an empty promise. The government is certain about the degree of the naïveté and present bias of the agents, so it places no weight on a future where it may need to actually honor the delivery of imaginary allocations. Another concern is that the agents may realize that the aggregate imaginary allocation violates the feasibility constraint and doubt the validity of the government's promise. However, each agent is infinitesimally small, and even though an agent believes he would consume the imaginary allocation, he does not consider the belief and behavior of other agents.

1.3.2 Main Result for Non-Sophisticated Agents

It can be shown that the government achieves the efficient allocation by using a fooling mechanism. In other words, surprisingly, private information does not matter in an environment where all agents harbor some naïveté.

Theorem 1 *If $\hat{\beta} \in (\beta, 1]$, then the efficient allocation $\{(c_m^*, k_m^*, y_m^*)\}_{\theta_m \in \Theta}$ is truthfully implementable by a direct fooling mechanism.*

Theorem 1 states that the private information problem can be alleviated if the agents are non-sophisticated. With the imaginary allocations, the government is able to provide the information rents necessary for truth-telling. However, these rents are fictional. After preference reversal, the government implements the efficient allocation without paying information rents. Indeed, it follows that it is optimal for the government to deceive the agents when they are not sophisticated.

Corollary 1 *If $\hat{\beta} \in (\beta, 1]$, it is optimal for the government to implement a fooling mechanism.*

The key to deceiving the agents is to load the rents on savings, $k^I \geq k^R$ and $c^R \geq c^I$, which the planner values relatively more during the reporting stage, but the doer would not value as much. The government can promise a higher return on savings for the imaginary

allocations to elicit truthful reports as long as the agents hold the wrong beliefs on their temptation level. In essence, non-sophisticated agents are willing to trade information rents for empty promises.

It is interesting to note that there is a discontinuity in the optimal welfare with respect to the sophistication level of the agents. The government achieves the efficient welfare level for any sophistication level $\hat{\beta} \in (\beta, 1]$. However, with fully sophisticated agents, the best the fooling mechanism can implement is the constrained efficient allocation which requires information rent for the productive types. This is because the sophisticated agents would perfectly foresee their doers' present bias at the consumption stage and would thus be immune to any deceptions at the reporting stage. As a result, there is a discontinuity in welfare which is similar to the discontinuity in Heidhues and Koszegi (2010).

I will show that the discontinuity is only present if the government restricts attention to a fooling mechanism. For sophisticated agents, the government can also take advantage of their time-inconsistency and implement the efficient optimum. In the next section, I will present such a mechanism.

An Illustration of the Fooling Mechanism

To demonstrate how the fooling mechanism works, consider an economy with two productivity types $\Theta = \{\theta_L, \theta_H\}$, where $\theta_H > \theta_L$, and let $\kappa = 1$, so the government maximizes the sum of ex-ante utility and wants to achieve perfect consumption smoothing. By Theorem 1, $\{(c_m^R, k_m^R, y_m^R)\}_{\theta_m \in \Theta}$ will be the efficient allocation, where $c_H^R = c_L^R = c^*$ and $k_H^R = k_L^R = k^*$ and $y_m^R = y_m^*$ for all $\theta_m \in \Theta$. The efficient allocation satisfies Proposition 1, so $y_H^* > y_L^*$. For simplicity, I will examine the fully naïve case: $\hat{\beta} = 1$.

In Figure 1.4, the flatter solid (blue) curve represents the indifference curve of the ex-ante utility at allocation (c^*, k^*) . The steeper solid (red) curve represents the indifference curve of the ex-post utility at allocation (c^*, k^*) . The imaginary allocations have to be in the area bounded by the solid line indifference curves in the north-west region.²³ Any allocation within this area satisfies the fooling constraint and executability constraint. Furthermore, the incentive compatibility constraints provide an upper and lower bounds to the difference

²³If $\beta > 1$, then the imaginary allocations would be bounded by the solid line indifference curves in the south-east region.

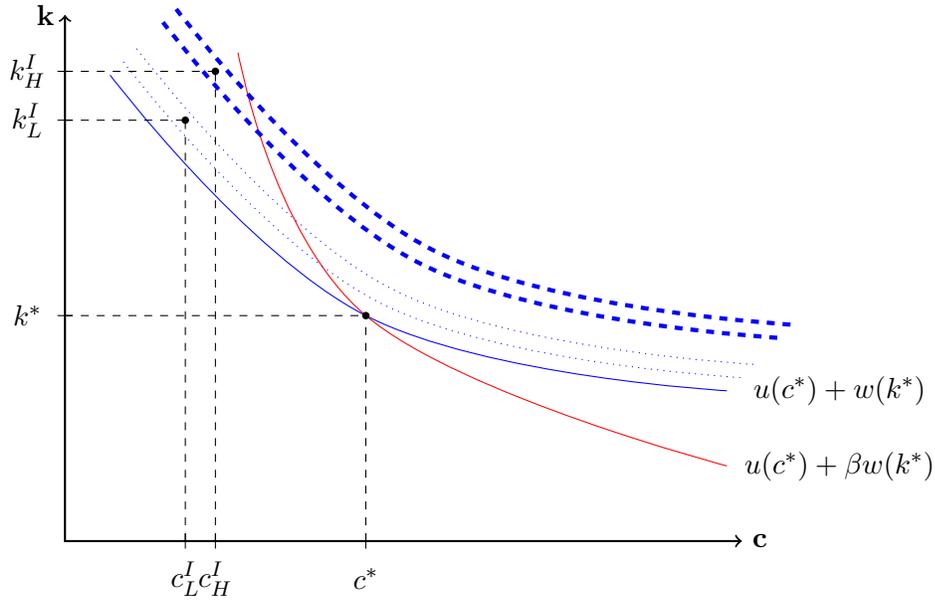


Figure 1.4: Finding the Imaginary Allocations

in ex-ante utility of the two types of agents. In essence,

$$h\left(\frac{y_H^*}{\theta_L}\right) - h\left(\frac{y_L^*}{\theta_L}\right) \geq [u(c_H^I) + w(k_H^I)] - [u(c_L^I) + w(k_L^I)] \geq h\left(\frac{y_H^*}{\theta_H}\right) - h\left(\frac{y_L^*}{\theta_H}\right).$$

Therefore, given y_m^* , the imaginary allocations have to be within the dashed indifference curves, where the high productivity type's imaginary allocation is within the bold dashed area, and the low productivity type's is within the light dotted area.

1.4 Savings with Sophisticated Agents

The fooling mechanism introduced for non-sophisticated agents no longer work for sophisticated agents. For sophisticated agents, the government can instead design an off-equilibrium path threat that will be chosen only if an agent misreports. This threat helps attain the efficient allocation.²⁴

²⁴Adverse selection models with sophisticated agents have been explored in Esteban and Miyagawa (2005), Galperti (2015) and Chung (2015).

1.4.1 The Threat Mechanism: Real and Threat Allocations

When sophisticated agents choose reporting strategy σ , they know that their preferences will change and their doer will be tempted to consume a set of allocations that is inconsistent with the ex-ante utility U . The government can take advantage of the agents' time-inconsistency and their awareness by using a *threat mechanism*.

Definition 3 A direct threat mechanism for sophisticated agents ($\hat{\beta} = \beta$) has $C = \{C_m\}_{\theta_m \in \Theta}$ with $C_m = \{(c_m^R, k_m^R, y_m^R), (c_m^T, k_m^T, y_m^T)\}$ satisfying the threat constraints: $(c_{\hat{m}}^T, k_{\hat{m}}^T, y_{\hat{m}}^T) \in C_{\hat{m}|m}^\beta, \forall \theta_m, \theta_{\hat{m}} \in \Theta$.

I will refer to (c_m^T, k_m^T, y_m^T) as the *threat allocation*. The analysis will focus on a truth-telling direct mechanism. Sophisticated agents have the correct belief about their temptation, so the executability constraints imply that truth-telling is evaluated at the real allocation and the threat constraints imply that misreports are evaluated at the threat allocations. The following definition defines truthful implementation in a direct threat mechanism.

Definition 4 An allocation $\{(c_m^R, k_m^R, y_m^R)\}_{\theta_m \in \Theta}$ is truthfully implementable by a direct threat mechanism if there exists $\{(c_m^T, k_m^T, y_m^T)\}_{\theta_m \in \Theta}$ such that the

- i. incentive compatibility constraints, and
- ii. executability constraints

are satisfied

The threat allocations, (c_m^T, k_m^T, y_m^T) , for type θ_m are designed such that, after preference reversal, a type θ_m planner who reports truthfully would never choose it (by the executability constraint), but a planner who misreports as type θ_m would (by the threat constraint). The superscript T represents 'threat,' because the misreporting planner would consider this allocation to be inferior according to his ex-ante preferences. Adding the threat allocation to the menu deters the agents from misreporting. When agents are sophisticated, a threat allocation that satisfies the threat and executability constraints will be referred to as satisfying the *credible threat constraints*.²⁵

²⁵The government has full commitment, so the credible threat constraints are not to ensure subgame perfection. They merely address the need for the government to find threats that deter misreports and to avoid truthful agents from choosing the threats, so the government is credibly benevolent.

It is important to note that threats are non-credible if $y_m^R = y_m^T$. This is because all agents share the same preference for goods consumption, so a misreporting agent would make the same consumption choice as a truthful agent if $y_m^R = y_m^T$. Hence, agents will not be deterred from misreporting if $y_m^R = y_m^T$. As a result, unlike a fooling mechanism, the real allocations and threat allocations have to be different in both consumption and output.

The government's problem is to choose the menu of allocations, C , to maximize (1.1) subject to the feasibility constraint (1.2) evaluated at the real allocations and the incentive compatibility constraints (1.4) and the credible threat constraints.

1.4.2 Main Result for Sophisticated Agents

The following theorem shows that private information does not matter in an environment with sophisticated agents as well. When the government uses a threat mechanism, the efficient allocation is achievable.

Theorem 2 *If $\hat{\beta} = \beta$, then the efficient allocation $\{(c_m^*, k_m^*, y_m^*)\}_{\theta_m \in \Theta}$ is truthfully implementable by a direct threat mechanism.*

Theorem 2 holds because sophisticated agents are aware that their doers would distort their savings plan, and would desire a commitment device that deters them from doing so. The government provides this commitment device only if the sophisticated agents report truthfully. If not, the threat allocation caters to the temptations of the doers. In essence, the government holds the agents' doers hostage and threatens to distort the savings plan unless agents report truthfully. This threat also helps screening because the threat allocations can be designed such that it separates the productivity of the agents by using the Spence-Mirrlees single crossing property. In essence, sophisticated agents trade information rents for commitment.

An Illustration of the Threat Mechanism

To demonstrate Theorem 2, consider the setting with two productivity types and $\kappa = 1$ investigated previously. In a Mirrlees setting, the productive agent has an incentive to pretend to be a less productive agent to decrease labor supply while enjoying the gains of insurance. In a relaxed problem, the threat is targeted at the productive types to discourage them from pretending to be the less productive. Hence, the government designs the

threat allocation $(c_L^T, k_L^T, y_L^T) \in C_L$. I will focus on the downward incentive compatibility constraint.

By Theorem 2, $\{(c_m^R, k_m^R, y_m^R)\}_{\theta_m \in \Theta}$ will be the efficient allocation, so $c_H^R = c_L^R = c^*$ and $k_H^R = k_L^R = k^*$ and $y_m^R = y_m^*$ for all $\theta_m \in \Theta$. Let $\Phi_{j,k}^i = u(c_j^i) - h(y_j^i/\theta_k)$ and $\Phi_{k,k}^i = \Phi_k^i$, where $i \in \{R, T\}$ and $j, k \in \{L, H\}$. From incentive compatibility and credible threat constraints and from the fact that $\theta_H > \theta_L$ and $\beta < 1$, the efficient and threat allocations have to satisfy

$$\Phi_{L,H}^T > \Phi_{L,H}^R > \Phi_L^R > \Phi_H^R, \text{ and } k^* > k_L^T.$$

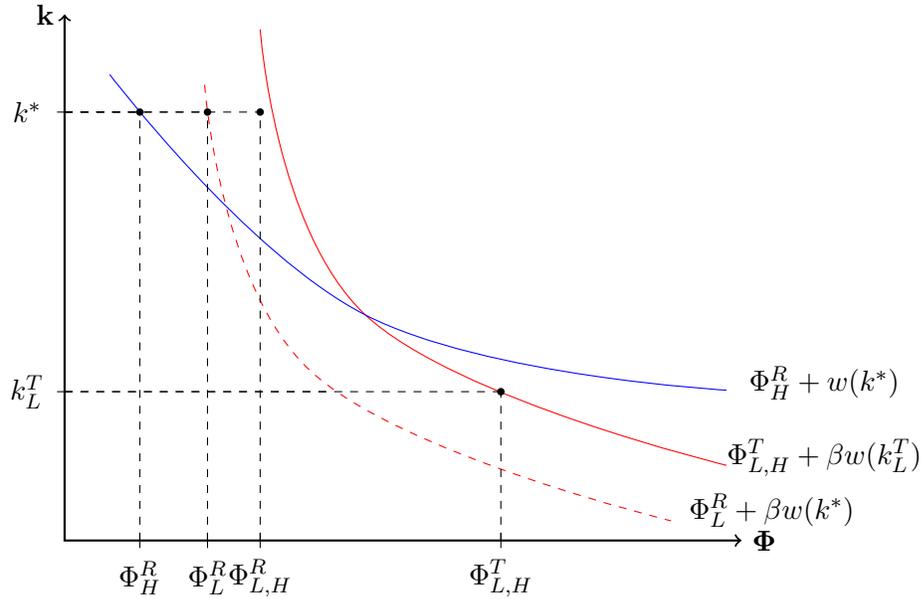


Figure 1.5: Finding the Threat Allocation: Part I

Figure 1.5 shows how the incentive compatibility constraint restricts the set of threat allocations. The steeper solid (red) curve represents the indifference curve of the ex-post utility for the θ_H agent who pretends to be θ_L . The flatter solid (blue) curve represents the indifference curve of the ex-ante utility for the θ_H agent who reports truthfully. The dashed (red) curve represents the indifference curve of the ex-post utility for the truthful θ_L agent. Figure 1.5 shows that the government can choose (c_L^T, k_L^T, y_L^T) such that the incentive compatibility constraint is satisfied by decreasing k_L^T and increasing $\Phi_{L,H}^T$.

By the Spence-Mirrlees single crossing property, the government can increase $\Phi_{L,H}^T$ without violating the credible threat constraints. To see this, fix the choice of k_L^T and $\Phi_{L,H}^T$ at the level shown in Figure 1.5. If the threat satisfies incentive compatibility and credible threat constraints, it implies

$$\Delta w \geq \Phi_{L,H}^T - \Phi_H^R > \Phi_{L,H}^T - \Phi_{L,H}^R \geq \beta \Delta w \geq \Phi_L^T - \Phi_L^R,$$

where $\Delta w \equiv [w(k^*) - w(k_L^T)]$. The problem now is to find c_L^T and y_L^T such that $u(c_L^T) - h\left(\frac{y_L^T}{\theta_H}\right) = \Phi_{L,H}^T$ and satisfies the executability constraint, $\beta \Delta w \geq \Phi_L^T - \Phi_L^R$. Figure 1.6 shows how this could be done.

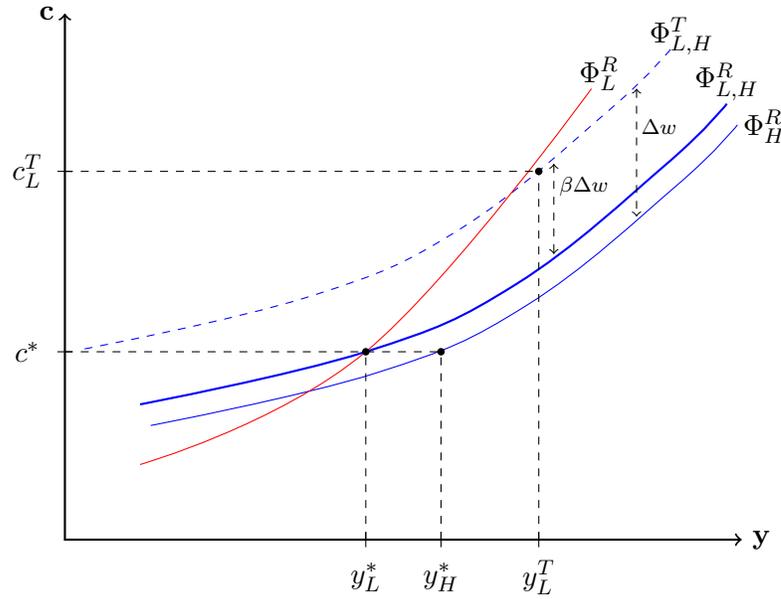


Figure 1.6: Finding the Threat Allocation: Part II

In Figure 1.6, the flatter thick solid (blue) curve represents the indifference curve of Φ for the θ_H agents at allocation (c^*, y_L^*) . The steeper solid (red) curve represents the indifference curve of Φ for the θ_L agents at allocation (c^*, y_L^*) . The dashed (blue) curve represents the indifference curve of Φ for the θ_H agent at allocation (c_L^T, y_L^T) , chosen so that $u(c_L^T) - h\left(\frac{y_L^T}{\theta_H}\right) = \Phi_{L,H}^T$ and the executability constraint holds.

The key to making a threat credible and incentive compatible is to decrease savings and increase present utility from choosing the threat allocation. The doer has a present bias, so a threat allocation would be more preferred than the real allocation. To make sure

that only a misreporting agent is punished, the government exploits the fact that the more productive agents have a lower marginal rate of substitution in consumption and labor than less productive agents. Hence, the threat allocation can always be constructed such that the compensation in consumption for the output is too low for the low productivity agents, but sufficient for the high productivity agents.

1.4.3 The Threat Mechanism for Partially Naïve Agents

Fully naïve agents have to be fooled, since they do not respond to threats. While sophisticated agents have to be threatened, because they can never be deceived. Since partially naïve agents have demand for commitment devices too, they are also susceptible to threats. The following definition defines a direct threat mechanism for agents with sophistication level $\hat{\beta} \in (\beta, 1)$ and truthful implementation in this environment.

Definition 5 *A direct threat mechanism has a menu $C = \{C_m\}_{\theta_m \in \Theta}$ with $C_m = \{(c_m^R, k_m^R, y_m^R), (c_m^T, k_m^T, y_m^T)\}$ satisfying credible threat constraints: $(c_m^R, k_m^R, y_m^R) \in C_m^{\hat{\beta}}$ and $(c_m^T, k_m^T, y_m^T) \in C_{\hat{m}|m}^{\hat{\beta}}, \forall \theta_m, \theta_{\hat{m}} \in \Theta$. The real allocation $\{(c_m^R, k_m^R, y_m^R)\}_{\theta_m \in \Theta}$ is truthfully implementable by a direct threat mechanism if there exists $\{(c_m^T, k_m^T, y_m^T)\}_{\theta_m \in \Theta}$ such that the incentive compatibility constraints (1.4) and the executability constraints (1.5) are satisfied.*

For partial naïveté with sophistication level $\hat{\beta} \in (\beta, 1)$, the threats are evaluated using the erroneous $\hat{\beta}$. Therefore, the credible threat constraints are defined to make sure the planner perceives his doer choosing the real allocation when report is truthful and the threat allocation if he misreported.

The government also needs to make sure that after preference reversal, the truthful agents will indeed choose the real allocation, so the executability constraint is needed. In contrast to Definition 3, the constraint, $(c_m^R, k_m^R, y_m^R) \in C_m^{\hat{\beta}}$, is redundant when agents are sophisticated. The following corollary shows partially naïve agents can be screened using a threat mechanism and Table 1.1 summarizes the applicability of the mechanisms to each sophistication level.

Corollary 2 *It is optimal to threaten the agents when $\hat{\beta} \in [\beta, 1)$.*

	Fully Naïve	Partially Naïve	Sophisticated
Fooling	✓	✓	✗
Threat	✗	✓	✓

Table 1.1: Summary of Mechanism for Different Sophistication Levels

1.5 Diversely Time-Inconsistent Agents

The previous model had agents differing in their productivity while sharing the same degree of temptation and sophistication level. In this section, I consider a setting where the government faces time-inconsistent agents who differ in their degree of temptation and sophistication level, which are both hidden from the government. I will first examine a model with hidden levels of sophistication, and then introduce hidden degrees of temptation.

1.5.1 Hidden Sophistication

In this economy, agents share the same temptation β but have varying sophistication levels. The hidden type of each agent is indexed by $(\theta_m, \hat{\beta})$, so the optimal policy has to solve a multidimensional screening problem.

The sophistication level of agents is distributed within the bounded support of $[\beta, 1]$. Let $\Pi(\theta_m, \hat{\beta})$ denote the joint distribution of productivity and sophistication level. I will refer to mechanisms that attain the efficient allocation as *effective*.

Lemma 1 *A fooling mechanism that is effective for agents of sophistication level $\hat{\beta} \in (\beta, 1)$ is also effective for all agents with $\hat{\beta} \geq \hat{\beta}$. A threat mechanism that is effective for agents of sophistication level $\hat{\beta} \in (\beta, 1)$ is also effective for all agents with $\hat{\beta} \leq \hat{\beta}$.*

Lemma 1 shows that a fooling mechanism designed for sophistication level $\hat{\beta}$ agents can also fool agents who are more naïve. While a threat mechanism for sophistication level $\hat{\beta}$ agents, can also credibly threaten agents who are more sophisticated. This is because providing incentives for the least naïve agents for truth-telling is the most difficult, so any incentives that could separate the productivity of the least naïve agents will also induce truth-telling for more naïve agents. Similarly, a threat mechanism for $\hat{\beta}$ will work for any $\hat{\beta} < \hat{\beta}$, since the less sophisticated agents need a stronger threat to be willing to divulge the truth, so the threat would also work for more sophisticated agents.

With Lemma 1, the government can choose an arbitrary *target sophistication level*, $\tilde{\beta}$, such that all agents who are more sophisticated than $\tilde{\beta}$ are threatened by using the same threat mechanism and those who are more naïve than $\tilde{\beta}$ are fooled by using the same fooling mechanism. I will henceforth refer to the agents with sophistication level $\hat{\beta} \in (\tilde{\beta}, 1]$ as *relatively naïve* and agents with sophistication level $\hat{\beta} \in [\beta, \tilde{\beta})$ as *relatively sophisticated*. Hence, a fooling mechanism designed for agents with sophistication level $\tilde{\beta}$ is applied to relatively naïve agents and a threat mechanism designed for agents with sophistication level $\tilde{\beta}$ is applied to relatively sophisticated agents.

The only concern is whether implementing a fooling mechanism for the relatively naïve would affect the effectiveness of the threat mechanism for the relatively sophisticated. The government can choose a fixed target sophistication level at $\tilde{\beta} \in (\beta, 1)$ and introduce the menu $C = \{C_m\}_{\theta_m \in \Theta}$, with

$$C_m = \{(c_m^R, k_m^R, y_m^R), (c_m^I, k_m^I, y_m^I), (c_m^T, k_m^T, y_m^T)\}.$$

The imaginary and threat allocations are chosen such that agents with sophistication level $\tilde{\beta}$ are fooled and threatened with effective mechanisms. I will refer to this mechanism as a *hybrid* mechanism. The following theorem shows that the efficient allocation is attainable in this environment.

Theorem 3 *For the environment with hidden sophistication, the efficient allocation $\{(c_m^*, k_m^*, y_m^*)\}_{\theta_m \in \Theta}$ is truthfully implementable with a hybrid mechanism.*

Theorem 3 follows immediately from the fact that relatively naïve and relatively sophisticated agents can be costlessly separated using a hybrid mechanism. This is because the relatively naïve agents focus on the imaginary allocations for truth-telling as long as the imaginary benefits from truth-telling are sufficiently appealing, while the relatively sophisticated agents focus on the threat allocations for misreporting. Therefore, the fooling and threat mechanisms do not interact when they are integrated, so the government can consider the relatively naïve agents independently of the relatively sophisticated agents. Along with Lemma 1, the government can choose any arbitrary target sophistication level $\tilde{\beta}$, so the population that is being fooled and threatened can be arbitrary.

Another feature of Lemma 1 and Theorem 3 is that it relies on very little information about the economic environment. The government does not need to know the joint distribution $\Pi(\theta_m, \hat{\beta})$, which is an integral information in usual multi-dimensional screening

mechanisms.²⁶ The only information the government needs is the distribution of productivity to determine the efficient allocation.

1.5.2 Hidden Temptation

Here, I will consider an environment where all agents are time-inconsistent, but vary in the degree of their temptation β and sophistication level $\hat{\beta}$. Let $\mathcal{B} = [\underline{\beta}, \bar{\beta}] \subset [0, 1]$ be the set of possible levels of temptation, with $\bar{\beta} > \underline{\beta}$ and $\bar{\beta} < 1$. In this economy, an agent's type is represented by $(\theta_m, \beta, \hat{\beta})$, where $\hat{\beta} \in [\beta, 1]$. Let $\Pi(\theta_m, \beta, \hat{\beta})$ denote the joint distribution of productivity, temptation level and sophistication level.

Notice the government is unable to ascertain whether an agent with reported belief $\hat{\beta}$ is sophisticated or non-sophisticated. It is thus impossible to simultaneously screen for the agents' degree of temptation and sophistication levels. With an additional hidden dimension, a mechanism with an arbitrary target level of sophistication, like above, may no longer work. For example, if the government chose a cutoff $\tilde{\beta} \in (\underline{\beta}, \bar{\beta})$ and attempted to threaten all agents more sophisticated and fool all agents more naïve, then agents with temptation $\beta > \tilde{\beta}$ and sophistication level $\hat{\beta} \geq \beta$ will not be fooled.

However, when $\kappa = 1$, the government can choose $\tilde{\beta} \in [\bar{\beta}, 1)$ and design both a threat mechanism and fooling mechanism for sophistication level $\tilde{\beta}$. All agents with $\hat{\beta} \leq \tilde{\beta}$ would be threatened, and all agents with $\hat{\beta} \geq \tilde{\beta}$ would be fooled. This fact immediately follows from Lemma 1. Lemma 1 can be applied with this cutoff, because the temptation level of the agents can be ignored, since all agents have $\beta \leq \tilde{\beta}$. The subtlety lies in how the threats and empty promises are constructed to satisfy the executability constraints. For the imaginary allocations, the government needs to ensure the agents with the least temptation $\beta = \bar{\beta}$ would prefer the real allocations, then all agents with stronger temptation $\beta < \bar{\beta}$ would strictly prefer the real allocation. For the threat allocations, the government needs to ensure the agents with the most temptation $\beta = \underline{\beta}$ would prefer the real allocations, then all agents with less temptation $\beta > \underline{\beta}$ would strictly prefer the real allocation. This gives us the following theorem.

Theorem 4 *For the environment with hidden temptation and sophistication, if $\kappa = 1$, the efficient allocation $\{(c_m^*, k_m^*, y_m^*)\}_{\theta_m \in \Theta}$ is truthfully implementable with a hybrid mechanism.*

²⁶For an introduction to the multi-dimensional screening model, see Armstrong and Rochet (1999)

Theorem 4 is also robust to changes in the joint distribution $\Pi(\theta_m, \beta, \hat{\beta})$. In addition to the primitives introduced in Section 1.2, the government does not need to know more than the support of \mathcal{B} . Theorem 4 also demonstrates how the efficient allocation is implementable as long as all agents are time-inconsistent and the welfare criterion only takes the planner's utility into consideration, regardless of the sophistication or temptation. If $\kappa \in (0, 1)$, then the government cares about the doer's utility. In this case, Theorem 4 no longer holds since the government would need to screen the agents' level of temptation and thus provide information rents to the doer.

1.6 Model with Time-Consistent Agents

In the previous sections, all agents in the economy were time-inconsistent (henceforth, TI). In this section, I will explore an economy where some agents are time-consistent (henceforth, TC). The presence of TC agents causes distortions, because without preference reversal, TC agents can follow through with their consumption plan and choose the imaginary allocation and avoid the threat allocation. This restricts the ability of the government to provide insurance.

By Theorem 4, it is without loss of generality to focus on TI agents with the same temptation level. The government is uncertain whether the agents are TC ($\beta = 1$) or TI ($\beta < 1$), with probability $\Pr(TI) = \phi$. The TC agents know their time-consistency (sophisticated), while TI agents could be non-sophisticated.²⁷ I assume the distribution of productivity is independent of the agents' consistency level. For this section, I will focus on the case with $\kappa = 1$, so the government puts all of the welfare weights on the planner. Thus, the welfare of both TC and TI agents are measured using the same utility.

1.6.1 The Threat Mechanism with Time-Consistent Agents

Suppose TI agents have a fixed sophistication level of $\hat{\beta} \in [\beta, 1)$, the government can implement a threat mechanism. The government introduces the following menu for TC agents of productivity θ_m :

$$C_m^{TC} = \{(c_m^P, k_m^P, y_m^P); (c_m^D, k_m^D, y_m^D)\}$$

²⁷I will relax this assumption and discuss the case with paranoid TC agents, TC agents believing they may be TI agents, in the last subsection of this section.

and the following menu for TI agents of productivity θ_m :

$$C_m^{TI} = \{(c_m^R, k_m^R, y_m^R); (c_m^T, k_m^T, y_m^T)\}.$$

The allocation (c_m^P, k_m^P, y_m^P) will be referred to as the *persistent allocation*, and it is the allocation the government implements for TC agents with θ_m productivity. The allocation (c_m^D, k_m^D, y_m^D) is referred to as the *deterrent allocation*, and it is meant to deter the TI agents from misreporting as TC agents of θ_m productivity.

The deterrent allocation has $k_m^D < k_m^P$ and $c_m^D > c_m^P$, which appeals to the doer's present bias. This deters the TI agent's planner from misreporting the consistency level. The deterrent allocation can also be constructed so that the TC agents would never prefer it over the persistent allocation. The construction of a deterrent allocation will be shown.

The government offers the following menu at date $t = 0$, $C = \{C_m^{TC}, C_m^{TI}\}_{\theta_m \in \Theta}$. The planner first chooses a menu C_m^j from C , and then an allocation from C_m^j is chosen by the doer. The mechanism is meant to separate agents along two dimensions: productivity and level of consistency. Let

$$C_{\hat{m}|m}^{1|\hat{\beta}} = \left\{ (c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}) \in C_{\hat{m}}^{TC} \mid (c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}) \in \arg \max_{(c'_{\hat{m}}, k'_{\hat{m}}, y'_{\hat{m}}) \in C_{\hat{m}}^{TC}} W(c'_{\hat{m}}, k'_{\hat{m}}, y'_{\hat{m}}; \theta_m) \right\}.$$

Hence, $C_{\hat{m}|m}^{1|\hat{\beta}}$ denotes the set of allocations a TI agent with productivity θ_m and sophistication level $\hat{\beta}$ predicts his doer will choose after he misreports to be a TC agent with productivity $\theta_{\hat{m}}$. Let

$$C_m^1 = \left\{ (c_m, k_m, y_m) \in C_m^{TC} \mid (c_m, k_m, y_m) \in \arg \max_{(c'_m, k'_m, y'_m) \in C_m^{TC}} U(c'_m, k'_m, y'_m; \theta_m) \right\}.$$

Hence, C_m^1 denotes the set of allocations a truth-telling TC agent would choose. Let $C_m^{\hat{\beta}}$ and $C_{\hat{m}|m}^{\hat{\beta}}$ be defined as before. The following definition defines a direct threat mechanism with TC agents.

Definition 6 A direct threat mechanism with TC agents has $C = \{C_m^{TC}, C_m^{TI}\}_{\theta_m \in \Theta}$, with $C_m^{TC} = \{(c_m^P, k_m^P, y_m^P); (c_m^D, k_m^D, y_m^D)\}$ and $C_m^{TI} = \{(c_m^R, k_m^R, y_m^R); (c_m^T, k_m^T, y_m^T)\}$ satisfying:

- i. credible threat constraints: $(c_m^R, k_m^R, y_m^R) \in C_m^{\hat{\beta}}$ and $(c_{\hat{m}}^T, k_{\hat{m}}^T, y_{\hat{m}}^T) \in C_{\hat{m}|m}^{\hat{\beta}}, \forall \theta_m, \theta_{\hat{m}} \in \Theta$,
- ii. deterrent constraints: $(c_m^P, k_m^P, y_m^P) \in C_m^1$, and $(c_{\hat{m}}^D, k_{\hat{m}}^D, y_{\hat{m}}^D) \in C_{\hat{m}|m}^{1|\hat{\beta}}, \forall \theta_m, \theta_{\hat{m}} \in \Theta$.

To define the set of truthfully implementable allocations that can be implemented by a direct threat mechanism with TC agents, note that the incentive compatibility constraints for the TI agents are, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

$$\max_{(c_m, k_m, y_m) \in C_m^{\hat{\beta}}} U(c_m, k_m, y_m; \theta_m) \geq \max \left\{ \begin{array}{l} \max_{(c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}) \in C_{\hat{m}|m}^{\hat{\beta}}} U(c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}; \theta_m), \\ \max_{(c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}) \in C_{\hat{m}|m}^{1|\hat{\beta}}} U(c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}; \theta_m) \end{array} \right\}. \quad (1.6)$$

By the incentive compatibility constraints (1.6), it is optimal for the TI agents to report truthfully about their productivity and consistency level. Let

$$C_{\hat{m}|m}^{\hat{\beta}|1} = \left\{ (c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}) \in C_{\hat{m}}^{TI} \mid (c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}) \in \arg \max_{(c'_{\hat{m}}, k'_{\hat{m}}, y'_{\hat{m}}) \in C_{\hat{m}}^{TI}} U(c'_{\hat{m}}, k'_{\hat{m}}, y'_{\hat{m}}; \theta_m) \right\},$$

$$C_{\hat{m}|m}^1 = \left\{ (c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}) \in C_{\hat{m}}^{TC} \mid (c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}) \in \arg \max_{(c'_{\hat{m}}, k'_{\hat{m}}, y'_{\hat{m}}) \in C_{\hat{m}}^{TC}} U(c'_{\hat{m}}, k'_{\hat{m}}, y'_{\hat{m}}; \theta_m) \right\}.$$

Hence, $C_{\hat{m}|m}^{\hat{\beta}|1}$ denotes the set of allocations a TC agent with productivity θ_m would select from a menu meant for TI agents with productivity $\theta_{\hat{m}}$, while $C_{\hat{m}|m}^1$ denotes the set of allocations a TC agent would pick if he misreports his productivity as $\theta_{\hat{m}}$ and is truthful about his consistency. The incentive compatibility constraints for the TC agents are, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

$$\max_{(c_m, k_m, y_m) \in C_m^1} U(c_m, k_m, y_m; \theta_m) \geq \max \left\{ \begin{array}{l} \max_{(c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}) \in C_{\hat{m}|m}^1} U(c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}; \theta_m), \\ \max_{(c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}) \in C_{\hat{m}|m}^{\hat{\beta}|1}} U(c_{\hat{m}}, k_{\hat{m}}, y_{\hat{m}}; \theta_m) \end{array} \right\}. \quad (1.7)$$

Incentive compatibility constraints (1.7) discourage the TC agents from misreporting about their productivity and level of consistency. The executability constraints for the real allocation are defined by (1.5).

Definition 7 *The allocation $\{(c_m^P, k_m^P, y_m^P), (c_m^R, k_m^R, y_m^R)\}_{\theta_m \in \Theta}$ is truthfully implementable by a direct threat mechanism if there exists $\{(c_m^D, k_m^D, y_m^D), (c_m^T, k_m^T, y_m^T)\}_{\theta_m \in \Theta}$ such that the*

i. incentive compatibility constraints, and

ii. executability constraints

are satisfied.

The government maximizes welfare

$$\sum_{\theta_m \in \Theta} \pi_m [\phi U(c_m^R, k_m^R, y_m^R; \theta_m) + (1 - \phi) U(c_m^P, k_m^P, y_m^P; \theta_m)], \quad (1.8)$$

subject to the incentive compatibility constraints ((1.6) and (1.7)), executability constraints (1.5), credible threat constraints, deterrent constraints and the feasibility constraint

$$\sum_{\theta_m \in \Theta} \{ \phi \pi_m [y_m^R - c_m^R - k_m^R] + (1 - \phi) \pi_m [y_m^P - c_m^P - k_m^P] \} = 0. \quad (1.9)$$

A suitably chosen set of threat and deterrent allocations can help relax (1.6). As a result, the government only needs to deter misreporting from the TC agents. The following theorem shows how the government takes advantage of the TI agents.

Theorem 5 *There exists $\bar{\theta} \in \Theta \setminus \theta_1$ such that*

- i. for productivity types $\theta_m \geq \bar{\theta}$, $(c_m^P, k_m^P) \gg (c_m^R, k_m^R)$ with $y_m^P < y_m^R$,*
- ii. for productivity types $\theta_1 < \theta_m < \bar{\theta}$, $(c_m^P, k_m^P, y_m^P) \ll (c_m^R, k_m^R, y_m^R)$,*
- iii. for θ_1 , $(c_1^P, k_1^P, y_1^P) = (c_1^R, k_1^R, y_1^R)$.*

By Theorem 5, full insurance is no longer incentive compatible when TC agents are in the economy. The high productivity ($\theta_m \geq \bar{\theta}$) TC agents require information rents, so they have lower marginal utilities from consumption and lower disutility from effective labor y . Since TI agents are manipulable, the government exploits the higher productivity TI agents by requiring them to work more and consume less, which increases the resources available for redistribution. As a result, TC agents with $\theta_m \geq \bar{\theta}$ would never misreport to be TI agents of the same productivity. For agents with lower productivity ($\theta_m < \bar{\theta}$), the government would exploit the TI agents by requiring them to work more, but also compensate them with more consumption. The consumption is limited by the incentive compatibility constraint for the higher productivity agents. This increase in production more than offsets the increase in consumption, so it also increases the resources available for redistribution. The government refrains from exploiting the TI agents with the lowest productivity by bunching them with the TC agents, because they are already receiving the least utility and any exploitation

would only lead to less insurance. The only binding incentive compatibility constraints are the downward adjacent incentive compatibility constraints for the TC agents. For $\theta_m > \theta_1$, TI agents have strictly lower lifetime utility than TC agents of the same productivity.

To construct the deterrent allocation, let (c^D, k^D, y_m^D) with $y_m^D = y_m^P$ be the deterrent allocation satisfying:

$$\min_{\theta_m, \theta_{\hat{m}} \in \Theta} \left[u(c^D) - h\left(\frac{y_m^D}{\theta_m}\right) + \hat{\beta}w(k^D) \right] \geq \max_{\theta_m, \theta_{\hat{m}} \in \Theta} \left[u(c_{\hat{m}}^P) - h\left(\frac{y_{\hat{m}}^P}{\theta_m}\right) + \hat{\beta}w(k_{\hat{m}}^P) \right], \quad (1.10)$$

and

$$\min_{\theta_m \in \Theta} \left[u(c_m^R) - h\left(\frac{y_m^R}{\theta_m}\right) + w(k_m^R) \right] \geq \max_{\theta_m, \theta_{\hat{m}} \in \Theta} \left[u(c^D) - h\left(\frac{y_m^D}{\theta_m}\right) + w(k^D) \right]. \quad (1.11)$$

By inequality (1.10), any TI agent who misreports his consistency level would select the deterrent allocation over the persistent allocation. Inequality (1.11) guarantees the TI agents would prefer to report their time consistency truthfully. If (1.11) is satisfied, the TC agents would never choose the deterrent allocations over the persistent allocations. If (1.10) and (1.11) hold, then TI agents of all productivity types would never misreport to be TC agents, and it is possible to construct the deterrent allocations such that (1.10) and (1.11) are satisfied.²⁸

Let the intertemporal wedge be defined as $\tau_K = 1 - \frac{U_c(c, k, \frac{y}{\theta})}{U_k(c, k, \frac{y}{\theta})}$, and the labor wedge as $\tau_L = 1 + \frac{U_y(c, k, \frac{y}{\theta})}{U_c(c, k, \frac{y}{\theta})}$. The following theorem helps characterize the optimal allocation and wedges in an environment with TC agents.

Theorem 6 *For a threat mechanism, if $\phi \in (0, 1)$ and $\hat{\beta} \in (\beta, 1)$, the optimal allocation has the following properties*

- i. $\tau_K = 0$ for all agents.*
- ii. $\tau_L = 0$ for all TI agents with $\theta_m \geq \bar{\theta}$ and TC agents of productivity θ_M .*
- iii. $\tau_L \geq 0$ for all TI agents with $\theta_m < \bar{\theta}$ and $\tau_L > 0$ for all TC agents with $\theta_m \in \Theta \setminus \theta_M$.*

²⁸To see how, first choose (c^D, k^D) so (1.11) holds. Next, increase c^D and decrease k^D such that $u(c^D) + w(k^D)$ remains unchanged, and since $1 > \hat{\beta}$, then it is possible to find (c^D, k^D) such that (1.10) holds.

The usual trade-off between insurance and output efficiency is present in this economy. Theorem 6 demonstrates how the the output of the less productive agents is distorted downwards. This result is standard in Mirrlees taxation. The government is able to provide consumption smoothing for all agents.

The next corollary shows that as long as TI agents not fully naïve, then the threat mechanism can implement the same optimal allocation for all levels of sophistication. This follows from the fact that the optimal allocation in a threat mechanism does not depend on the sophistication level of the TI agents.

Corollary 3 *In a threat mechanism, the optimal allocation is the same for any sophistication level $\hat{\beta} \in [\beta, 1)$.*

The usual Mirrlees taxation with TC agents occurs when $\phi = 0$, and the constrained efficient optimum is achieved. This paper has shown that the full information efficient optimum is attainable when the economy is populated by TI agents ($\phi = 1$). Let $W^T(\phi, \hat{\beta})$ denote the welfare under a threat mechanism with measure ϕ of TI agents with sophistication level $\hat{\beta} \in [\beta, 1)$. It is natural to presume that social welfare increases as the proportion of TI agents increases, as shown in Figure 1.2. Theorem 7 confirms this intuition.

Theorem 7 *In a threat mechanism, $W^T(\phi, \hat{\beta})$ increases continuously with ϕ from the constrained efficient optimum to the full information efficient optimum.*

As the mass of TI agents increases, the first order effect is an increase in the available resources for redistribution, which comes from the allocation patterns in Theorem 5. The second order effect is that with less TC agents, the government can provide each TC agent more information rent using fewer resources. This relaxes the incentive compatibility constraints (1.7). Hence, the government is able to provide better insurance when more agents are time-inconsistent.

Also of note, Theorem 7 shows that continuity in welfare can be achieved with respect to the proportion of TC agents. However, in an economy with homogeneous temptation β , there is a discontinuity in welfare as $\beta \rightarrow 1$. This discontinuity is due to the assumption on utility $u(\cdot)$ and $w(\cdot)$. Since $u(\cdot)$ and $w(\cdot)$ are unbounded below and above, the government is always able to construct possibly extreme threats (very small k and very large c) that deter TI agents. In Section 1.9, I will discuss the effects of relaxing this assumption on utility.

1.6.2 The Fooling Mechanism with Time-Consistent Agents

The government can also implement a fooling mechanism when $\hat{\beta} \in (\beta, 1)$. The government introduces the following menu for agents of productivity $\theta_m : C_m = \{C_m^{TC}, C_m^{TI}\}$, where C_m^{TC} consists of the persistent and deterrent allocations and C_m^{TI} consists of the real and imaginary allocations.

Definition 8 *A direct fooling mechanism with TC agents has $C = \{C_m^{TC}, C_m^{TI}\}_{\theta_m \in \Theta}$, with $C_m^{TC} = \{(c_m^P, k_m^P, y_m^P); (c_m^D, k_m^D, y_m^D)\}$ and $C_m^{TI} = \{(c_m^R, k_m^R, y_m^R); (c_m^I, k_m^I, y_m^I)\}$ satisfying:*

- i. fooling constraints: $(c_m^I, k_m^I, y_m^I) \in C_m^{\hat{\beta}}$ and $(c_{\hat{m}}^I, k_{\hat{m}}^I, y_{\hat{m}}^I) \in C_{\hat{m}|m}^{\hat{\beta}}, \forall \theta_m, \theta_{\hat{m}} \in \Theta$,*
- ii. deterrent constraints: $(c_m^P, k_m^P, y_m^P) \in C_m^1$, and $(c_m^D, k_m^D, y_m^D) \in C_{\hat{m}|m}^{1|\hat{\beta}}, \forall \theta_m, \theta_{\hat{m}} \in \Theta$.*

The allocations that are truthfully implementable by a direct fooling mechanism with TC agents is bounded by the incentive compatibility constraints (1.6) and (1.7) and the executability constraints (1.5). Thus, the definition of truthfully implementable allocations in this mechanism is similar to Definition 7.

Definition 9 *The allocation $\{(c_m^P, k_m^P, y_m^P), (c_m^R, k_m^R, y_m^R)\}_{\theta_m \in \Theta}$ is truthfully implementable by a direct fooling mechanism if there exists $\{(c_m^D, k_m^D, y_m^D), (c_m^I, k_m^I, y_m^I)\}_{\theta_m \in \Theta}$ such that the*

- i. incentive compatibility constraints, and*
- ii. executability constraints*

are satisfied.

The following theorem shows that, for a given $(\phi, \hat{\beta})$, the government is indifferent between implementing a fooling or a threat mechanism when agents are partially naïve.

Theorem 8 *For $\hat{\beta} \in (\beta, 1)$, the optimal allocation in a direct fooling mechanism is equivalent to the optimal allocation in a direct threat mechanism.*

Theorem 8 implies that the properties derived for the optimal allocation in a threat mechanism with TC agents also applies to the fooling mechanism. Combined with Corollary 3, for a given ϕ , the optimal allocation and welfare are the same for any sophistication level $\hat{\beta} \in (\beta, 1)$ regardless of whether a threat or a fooling mechanism is implemented. The caveat

here is that this result is only true when TI agents are partially naïve, in essence, $\hat{\beta} \in (\beta, 1)$. However, when agents are completely naïve, the optimal allocation and maximal welfare is different.

Fully Naïve Time-Inconsistent Agents

When $\hat{\beta} \in (\beta, 1)$, the government was able to use deterrent allocations to prevent the TI agents from pretending to be TC agents. However, separation in consistency is no longer possible when $\hat{\beta} = 1$.²⁹ When $\hat{\beta} = 1$, the government introduces the following menu at date $t = 0$, $C = \{C_m\}_{\theta_m \in \Theta}$, with

$$C_m = \{(c_m^R, k_m^R, y_m^R), (c_m^P, k_m^P, y_m^P)\}.$$

The government is unable to separate the TC agents from the TI agents at date $t = 0$, so it expects agents of the same productivity to pick the same menu. After preference reversal, the government expects the TI agents to choose the real allocations and the TC agents to choose the persistent allocations. Here, the persistent allocations serve a similar purpose as the imaginary allocations to the TI agents. The TI agents evaluate their reporting strategy using the persistent allocations, but consume the real allocations after preference reversal.

The definition of a direct fooling mechanism for full naiveté with TC agents and the corresponding definition of truthful implementation is presented below.

Definition 10 *A direct fooling mechanism with TC agents and $\hat{\beta} = 1$ has $C = \{C_m\}_{\theta_m \in \Theta}$, with $C_m = \{(c_m^R, k_m^R, y_m^R), (c_m^P, k_m^P, y_m^P)\}$ satisfying the fooling constraints: $(c_m^P, k_m^P, y_m^P) \in C_m^{\hat{\beta}}$ and $(c_{\hat{m}}^P, k_{\hat{m}}^P, y_{\hat{m}}^P) \in C_{\hat{m}|m}^{\hat{\beta}}, \forall \theta_m, \theta_{\hat{m}} \in \Theta$. The allocation $\{(c_m^R, k_m^R, y_m^R), (c_m^P, k_m^P, y_m^P)\}_{\theta_m \in \Theta}$ is truthfully implementable by a direct fooling mechanism if the allocation satisfies the incentive compatibility constraints (1.4) and the executability constraints (1.5).*

Note that since $\hat{\beta} = 1$, the incentive compatibility constraints (1.4) are the same for the TC and TI agents. The government maximizes the utilitarian welfare (1.8) subject to the fooling constraints, the incentive compatibility constraints (1.4), executability constraints (1.5), and feasibility (1.9). The following theorem characterizes the intertemporal wedge in a fully naïve environment and the allocations for the top and bottom of the productivity distribution.

²⁹This point was also made in Galperti (2015), but its implications were not fully explored.

Theorem 9 *For a fooling mechanism, if $\phi \in (0, 1)$ and $\hat{\beta} = 1$, the optimal allocation has the following properties:*

- i. $\tau_K < 0$ for all TC agents with $\theta_m > \theta_1$ and $\tau_K > 0$ for all TI agents with $\theta_m > \theta_1$.*
- ii. For θ_M , the allocation satisfies $c_M^P < c_M^R$, $k_M^P > k_M^R$ and $y_M^P > y_M^R$.*
- iii. For θ_1 , $(c_1^P, k_1^P, y_1^P) = (c_1^R, k_1^R, y_1^R)$ with $\tau_K = 0$.*

Theorem 9 shows how the government provides consumption smoothing for agents with $\theta_m > \theta_1$ in expectation. In essence, the best the government can do is to require the TC agents to over-save and the TI agents to under-save. This is true even for the most productive agents, so there will be distortion at the top.

The reason for this distortion is due to the binding executability constraints (1.5). If the government implements the allocation described in Theorem 5, then the executability constraints would be violated since the TI agents would pick the persistent allocation. This stems from the government's inability to separate the TI agents from the TC agents, so the persistent allocation acts as a proxy for the imaginary allocation when TI agents are fully naïve. The next section discusses the welfare implications of increasing the TI agents' awareness of their foibles.

1.6.3 Incentives for Self-Awareness

Comparing the optimal welfare for each sophistication level helps answer the question of whether the government wants to raise the TI agents' awareness of their present bias. It is clear from the setting where all agents were TI, the government is indifferent among the various sophistication levels of the agents. However, when TC agents are present, this is no longer the case. Theorem 8 and Corollary 3 show that when agents are aware of their present bias, the government is indifferent between implementing a threat mechanism or a fooling mechanism, which would both achieve the same optimal allocation and welfare regardless of the sophistication level of the TI agents. Thus, the question is whether it is in the best interest of the government to alert the fully naïve agents of their present bias.

For a given ϕ , let $W(\hat{\beta} < 1)$ denote the social welfare when agents are sophisticated or partially naïve, and $W(\hat{\beta} = 1)$ be the welfare when agents are fully naïve. The following theorem shows that the government would want the TI agents to be at least partially naïve.

Theorem 10 *If $\phi \in (0, 1)$, then $W(\hat{\beta} < 1) > W(\hat{\beta} = 1)$.*

The difference in welfare in Theorem 10 is due to the inability of the government to separate the fully naïve TI agents from the TC agents, while separation is possible if TI agents are at least partially naïve. The inability to separate along consistency levels causes two problems. First, the government is unable to design a deterrent allocation to dissuade TI agents from mimicking TC agents. Second, the persistent allocation serves as a proxy for the imaginary allocation, so the allocation used for fooling the TI agents are no longer empty promises and are not off-equilibrium path when TC agents are present. As a result, when $\hat{\beta} = 1$, the optimal allocation distorts the intertemporal wedge, which is an additional distortion not present when TI agents are at least partially naïve.

Theorem 10 has new implications on the education policy for TI agents. Educating non-sophisticated TI agents is only weakly welfare increasing. This result contrasts with the policy recommendations from the literature.³⁰

1.6.4 Paranoia

Previous analysis assumed the TC agents were sophisticated ($\hat{\beta} = \beta = 1$). Sophisticated TC agents do not respond to threats or empty promises, and require strictly positive information rents for truth-telling. Here, I will briefly discuss how paranoia affects the behavior of TC agents, and consequently how government policy can be adjusted to take advantage of this.

A paranoid agent is a non-sophisticated TC agent who believes he is time-inconsistent. Paranoid agents respond to threats, and can also be fooled. It is trivial to see that a threat mechanism can extract all information rents from paranoid agents, since the government can construct the threat allocation using similar methods employed for the TI agents. The logic for the fooling mechanism is more subtle.

To see how a fooling mechanism achieves the full information optimum in an economy with only paranoid agents, $\hat{\beta} < \beta = 1$, consider the example with two productivity types $\Theta = \{\theta_L, \theta_H\}$. Let $C_m = \{(c^*, k^*, y_m^*), (c_m^I, k_m^I, y_m^*)\}$ and the allocations satisfy the fooling

³⁰Previous literature has focused on the pitfalls of being non-sophisticated. For example, behavioral contract theory has focused on how limited awareness of self control problems makes consumers vulnerable to exploitation by firms, see Spiegler (2011) or Koszegi (2014) for examples.

constraints

$$u(c_m^I) + \hat{\beta}w(k_m^I) \geq u(c^*) + \hat{\beta}w(k^*),$$

the executability constraints

$$u(c^*) + w(k^*) \geq u(c_m^I) + w(k_m^I),$$

and the incentive compatibility constraints, which implies the following

$$h\left(\frac{y_H^*}{\theta_L}\right) - h\left(\frac{y_L^*}{\theta_L}\right) \geq [u(c_H^I) + w(k_H^I)] - [u(c_L^I) + w(k_L^I)] \geq h\left(\frac{y_H^*}{\theta_H}\right) - h\left(\frac{y_L^*}{\theta_H}\right).$$

The fooling and executability constraints imply $c_m^I > c^*$ and $k_m^I < k^*$. Combined with the incentive compatibility constraints, it must be that $c_L^I > c_H^I$ with $k_L^I < k_H^I$. In essence, the government fools the paranoid agents by choosing the imaginary allocations to exacerbate their fears. A paranoid agent would predict choosing the imaginary allocations even though he is strictly worse off by choosing it, because he does not think the real allocation is attainable. The government takes advantage of this by making the imaginary allocation for the θ_L agent even worse. Hence, the paranoid θ_H agent produces efficiently because he is afraid that by misreporting, he would have even less savings.

The way the fooling mechanism works for paranoid agents is in stark contrast to the logic presented in the previous sections. Non-sophisticated TI agents were fooled by empty promises, but paranoid agents were fooled by empty threats.³¹

If the economy has both paranoid and TI agents, then using a fooling mechanism could be problematic for the government. For example, imagine an economy with paranoid TC agents with incorrect belief $\hat{\beta} < 1$, which corresponds to the belief of the non-sophisticated TI agents with temptation $\beta < \hat{\beta}$. If the government tries to fool the agents, then it is not possible for the government to separate agents along consistency level. In this particular case, depending on who the government chooses to fool, either the paranoid TC agent would end up selecting the imaginary allocation used to fool the TI agents or the TI agent would choose the imaginary allocation used to fool the TC agents. The resulting welfare would be lower compared to when TC agents are sophisticated. This is analogous to the case with sophisticated TC agents and fully naïve TI agents. However, such a problem does not arise

³¹In the previous sections, it was sufficient to set $(c_L^I, k_L^I, y_L^*) = (c^*, k^*, y_m^*)$. For paranoid agents, however, it must be the case that $(c_L^I, k_L^I, y_L^*) \neq (c^*, k^*, y_L^*)$, but the government can set $(c_H^I, k_H^I, y_H^*) = (c^*, k^*, y_H^*)$.

when the government uses a threat mechanism. This is because the same threats for the TI agents can also deter the paranoid agents from misreporting.³²

1.7 Policy Implementation

In the previous sections, a direct revelation mechanism was used to characterize the optimal allocations. In this section, I explore a decentralized setting and the policy tools that could implement the optimum.

The government introduces a policy \mathcal{P} , and the agents earn before-tax income y , choose consumption c and purchase risk-free bond b with gross rate $R = 1$ in response to \mathcal{P} . Let $\hat{A} = \left\{ \left(\hat{c}_m, \hat{k}_m, \hat{y}_m \right) \right\}_{\theta_m \in \Theta}$ denote the resulting equilibrium allocation given \mathcal{P} . The following definition specifies what implementation means in a decentralized environment.

Definition 11 *A government policy \mathcal{P} implements an optimum \tilde{A} if the equilibrium allocation given \mathcal{P} is $\hat{A} = \tilde{A}$.*

The section will proceed by analyzing different policies that could implement the full information efficient optimum. For illustrative purposes, the government evaluates welfare only by the ex-ante utility, in essence, $\kappa = 1$. The outline of the policy implementation does not change with κ .

1.7.1 Decentralizing the Fooling Mechanism

For non-sophisticated agents ($\hat{\beta} > \beta$), the government can introduce a savings plan with a default savings subsidy set at τ^* and an income contingent savings subsidy $\tau(y)$ to incentivize work. Though an agent who earns income y is qualified for the savings subsidy $\tau(y)$, it is costly to switch from τ^* . Let the income dependent tax for switching be $T^s(y) > 0$. An agent with income y is only eligible for $\tau(y)$ and the default τ^* . The savings subsidy is combined with a standard income tax $T^e(y)$ independent of the subsidy. The gross income tax will be $T^e(y) + T^s(y)$ if the agent switches from the default subsidy, and $T^e(y)$ if no switching occurs. The government policy is $P^n = (\tau^*, \tau(y), T^e(y), T^s(y))$.

³²This is true because the credible threat constraints and the incentive compatibility constraints are the same for both the TI and TC agents who share the same beliefs. Finally, the executability constraint is more relaxed for the TC agents than the TI agents.

For the timing, the agents are required to work before making consumption and savings decisions. They choose work effort by deducing future consumption and savings given policy P^n and belief $\hat{\beta}$. However, due to the incorrect beliefs, the government expects the agents to stick with the default option of τ^* , which does not vary with income. In essence, the savings subsidy would be designed such that non-sophisticated agents would be fooled into working efficiently by $\tau(y)$, but their doer would instead prefer τ^* . The agents face the following budget constraints:

$$c + (1 - \tau)b \leq y - T^e(y) - \mathbf{1}_{\tau \neq \tau^*} T^s(y), \text{ and } k \leq b,$$

where $\mathbf{1}_{\tau \neq \tau^*}$ is an indicator function that equals to one if $\tau \neq \tau^*$. At the optimum, $k = b$, so the sequential budget constraints can be rewritten as

$$c + (1 - \tau)k \leq y - T^e(y) - \mathbf{1}_{\tau \neq \tau^*} T^s(y). \quad (1.12)$$

An agent solves the following problem:

$$\max_y U(c(y), k(y), y; \theta)$$

subject to

$$(c(y), k(y)) \in \arg \max_{(c,k)} W(c, k, y; \theta) \text{ s.t. (1.12)}.$$

Proposition 3 *If $\hat{\beta} \in (\beta, 1]$, then the efficient allocation $\{(c_m^*, k_m^*, y_m^*)\}_{\theta_m \in \Theta}$ can be implemented by $P^n = (\tau^*, \tau(y), T^e(y), T^s(y))$, where savings subsidy $\tau(y)$ is an increasing step function and the gross income tax is strictly increasing ($T^e(y)$ is strictly increasing and $T^s(y)$ is an increasing step function).*

The savings subsidy is regressive in income, while the gross income tax increases with income. This is because the government increases the savings subsidy for higher incomes to encourage efficient output. With a regressive savings subsidy, it is necessary for the gross income tax to increase with income, so the income effect can offset the increase in utility caused by the substitution effect from the higher subsidies $\tau(y) > \tau^*$. Consequently, given an increasing $\tau(y)$, the increasing $T^s(y)$ guarantees the implementation of the efficient allocation after preference reversal. In other words, the doer would prefer the default savings subsidy τ^* instead.

O’Donoghue and Rabin (2001) have shown that naiveté and present bias can cause status quo bias. If agents suffer from status quo bias and with τ^* as the default savings subsidy, their doers may find switching to be inherently costly. For reasons outside of the model, the government may be able to utilize the status quo bias to soften the need for an increasing $T^s(y)$.³³

There are other policies that could implement the efficient outcome for non-sophisticated agents, especially since the timing of when the agents work is flexible in the implementation for the relatively naïve. Other policies that require agents to commit to a menu of subsidy plans before working can also be conceived. In the next section, I explore a different implementation of the fooling mechanism with social security benefits.

1.7.2 Decentralizing the Threat Mechanism

I propose a policy implementation for partially naïve and sophisticated agents ($\hat{\beta} < 1$) with loans and flexible repayment plans. The government provides agents with *commitment loans*, where its associated repayment plans help smooth consumption. Let $(L(y^e), R(b, y^e, y))$ be the commitment loans and repayment pairs in the government plan P^s , where each loan is a function of projected income y^e and the repayment depends on savings b , realized income y and the original projection.

Agents take out commitment loans $L(y^e)$ after productivity is realized and before allocation decisions are made, with associated repayment plan $R(b, y^e, y)$. Notice the repayment plan depends on the savings and income of the agent, so the planner needs to predict what the doer would choose before committing to a plan. The government policy is comprised of the commitment loans and repayment plan with a standard income tax: $P^s = (L(y^e), R(b, y^e, y), T(y))$. It is assumed that there will be no interest earned from savings b .

The agents face the following sequential budget constraints after choosing $L(y^e)$

$$c + b \leq y + L(y^e) - T(y), \tag{1.13}$$

$$k \leq b - R(b, y^e, y). \tag{1.14}$$

³³Other policy proposals have suggested utilizing the status quo bias to help increase retirement savings. For instance, Thaler and Benartzi (2004) utilizes present bias and status quo bias to increase 401(k) enrollment and raise contribution rates.

Since selecting a commitment loan is equivalent to choosing a projected income y^e , a planner solves the following problem:

$$\max_{y^e} U(c(y^e), k(y^e), y(y^e); \theta)$$

subject to

$$(c(y^e), k(y^e), y(y^e)) \in \arg \max_{(c,k,y)} W(c, k, y; \theta) \text{ s.t. (1.13) and (1.14).}$$

The loan is referred to as a commitment loan, because the agents do not use the loan to invest or consume. In contrast to usual loans, the purpose of a commitment loan is to facilitate commitment to increasing savings. Consequently, the commitment loan is purely a commitment device for the agents.

The agents obtain commitment from the structure of the repayment plan, which is constructed to incentivize savings. Another essential component of the repayment plan is its dependence on realized income and projected income. The repayment plan is designed so that if the realized income does not match the original projection, then the government is able to punish the agent by exacerbating his present bias. It also awards the agent with a consumption smoothing plan if the projection matches with the actual realization.

Proposition 4 *If $\hat{\beta} \in [\beta, 1)$, then the efficient allocation $\{(c_m^*, k_m^*, y_m^*)\}_{\theta_m \in \Theta}$ can be implemented through commitment loans $L(y^e)$ with associated repayment $R_m(b, y^e, y)$ contingent on savings, expected income and realized income.*

Since the repayment plan depends on realized income, the government is always able to verify whether an agent chose the appropriate loans plan by observing an agent's income. If an agent projects an income of y^e and has an actual income not significantly higher than y^e , then the repayment plan provides a subsidy on bond savings that would offset the present bias. However, a more productive agent who selects a commitment loan meant for less productive agents would be tempted to work significantly more than y^e in exchange for a subsidy on consumption and tax on savings. This would lower the ex-ante utility since the off-equilibrium path policy aggravates the present bias, so each agent chooses the commitment loan designed for their productivity.

The commitment loans and the associated repayment plan proposed here is not unusual. The agents have to take out the loan before earning income but after realizing their future

possible productivity, much like student loans for medical or law school. Current government funded student loans programs also have repayment plans that depend on income. For instance, it is possible to select an income-driven repayment plan for federal student loans. Therefore, only minor adjustments, such as adding savings incentives into the repayment plan, need to be made for the implementation suggested above.

1.7.3 Decentralization with Time-Consistent Agents

In this subsection, I will present a policy for the TC agents which can be combined with both fooling and threat mechanisms in an environment with partially naïve agents.

The persistent allocation can be implemented with a standard income tax. Let $Y^P = \{y_1^P, \dots, y_M^P\}$ be the set of constrained efficient output for the TC agents. Since the constrained efficient consumption $\{(c_m^P, k_m^P)\}_{\theta_m \in \Theta}$ depends on the productivity θ_m only through y_m^P , it is possible to define the following consumption mapping $c^P : [y_1^P, \infty) \mapsto \mathbb{R}$ such that $c_m^P = c^P(y_m^P)$ and $c^P(y) = c^P(y_m^P)$ for any $y \in (y_m^P, y_{m+1}^P)$. The mapping for savings $k^P(y)$ is defined in a similar fashion. Since the constrained efficient consumption is monotonically increasing, the functions $c^P(y)$ and $k^P(y)$ are increasing step functions. The government can define an income tax schedule $T^P(y)$ for the TC agents as a function of y :

$$T^P(y) = \begin{cases} y - (c^P(y) + k^P(y)) & \text{if } y \geq y_1^P \\ y & \text{if } y < y_1^P. \end{cases}$$

Given the income tax schedule $T^P(y)$, TC agents of productivity θ_m would solve the following problem

$$\max_{c, k, b} U(c, k, y; \theta_m) \text{ s.t. } c + b \leq y - T^P(y) \text{ and } k \leq b.$$

Note the TC agents are allowed to save freely (uninhibited use of the storage technology). It is easy to show that the income tax can implement the constrained efficient allocation for the TC agents, provided that they do not pretend to be TI agents. This is because given the income tax $T^P(y)$, agents would always at least produce y_1^P , and combined with the consumption and savings functions, $c^P(y)$ and $k^P(y)$, the agents would never produce $y \notin Y^P$. Also, since the allocations satisfy incentive compatibility, a TC agent with productivity θ_m would choose to produce output y_m^P . Finally, since the agents are

time-consistent, they would consume and save the constrained efficient amount given the after-tax income.

The deterrent allocation is a penalty for TI agents pretending to be TC agents. It can be implemented with a high interest rate loan combined with bankruptcy insurance χ . Agents are qualified for this policy if they did not take out a commitment loan presented in Section 1.7.2, nor applied for the savings subsidy introduced in Section 1.7.1. Let (c^D, k^D) satisfy (1.10) and (1.11). Agents take out a loan $L(y) = c^D - [c^P(y) + k^P(y)]$ and are required to repay $R(y) = c^P(y) + k^P(y) + L$. The bankruptcy insurance is set at $\chi = k^D$. An agent taking out this loan would solve

$$\max_{c, k, y, b} U(c, k, y; \theta_m) \text{ s.t. } c + b \leq y - T^P(y) + L(y) \text{ and } k \leq \max\{b - R(y), \chi\}.$$

Agents taking out this loan would find it optimal to work y_1^P , save $b = 0$, and proceed to file for bankruptcy to receive χ . Since (c^D, k^D) satisfies (1.10) and (1.11), the TC agents would never choose this loans plan, and only the TI agents would choose this sub-optimal plan. Hence, depending on their sophistication level, TI agents would either choose a commitment loan (relatively sophisticated) or a savings subsidy (relatively naïve).

If an agent takes out a commitment loan, then he chooses $L(y^e)$ with the associated repayment plan $R(b, y^e, y)$ to maximize the ex-ante utility. The income tax facing the agents who take out a commitment loan is defined as

$$T^R(y) = \begin{cases} y - (c^R(y) + \beta k^R(y)) & \text{if } y \geq y_1^R \\ y & \text{if } y < y_1^R. \end{cases}$$

where the consumption functions depends on y_m^R . The government defines the following consumption mapping $c^R : [y_1^R, \infty) \mapsto \mathbb{R}$ such that $c_m^R = c^R(y_m^R)$ and $c^R(y) = c^R(y_m^R)$ for any $y \in (y_m^R, y_{m+1}^R)$. The mapping for savings $k^R(y)$ is defined in a similar way. Following the blueprint for constructing commitment loans in Section 1.7.2, the policy can be constructed so that TI agents who choose the commitment loan meant for their productivity will work, consume and save the constrained efficient amount, while the agents who do not choose the appropriate commitment loan would be worse off. Similarly, by using the construction outlined in Section 1.7.1, it is also possible to use regressive savings subsidy to implement the constrained optimum.

1.8 Social Security and Retirement Plans

This section will provide recommendations for policy implementation that deviates from the previous section in two respects. Firstly, the previous section focused on implementation with tax policies. In this section, I will discuss the implications of the paper on the design of the social security system and retirement plans, such as defined contribution plans. Secondly, the previous section focused on savings, while social security and certain retirement plans, such as defined benefit plans, are not considered savings.³⁴ This section provides new perspectives on the implementation of social security benefits and retirement plans as prescribed by the lessons of this paper. I will highlight the current debate on the design of the social security system and retirement plans, and proceed to implement the fooling mechanism using social security and the threat mechanism with retirement plans.

1.8.1 The Timing of Claiming Social Security Benefits

Social security reforms have generated a lot of discussion in the United States. This is because a majority of the US population relies on social security benefits as their primary source of income post-retirement.³⁵ Also, while the US population is living longer, the average retirement age has remained steady for the past decade.³⁶ This increases the duration of claiming social security benefits. Consequently, discussions on social security reforms to improve post-retirement welfare is currently an especially relevant issue for a savings constrained and aging US population.

One of the most important decisions for a retiree in the US is the timing of claiming social security benefits. The earliest age possible for receiving benefits is 62. However, a person can delay claiming and receive higher monthly benefits. For example, according to the Social Security Administration, the average monthly social security benefit for a beneficiary who started claiming at the age of 62 in 2014 is \$1,098. However, if the same beneficiary waited till the age of 70 to start claiming benefits (this is the oldest enrollment

³⁴Skinner (2007) defines retirement savings as assets that agents manage, such as 401(k) plans, IRAs, business equities, stock investments and so forth. This does not include social security and pension plans.

³⁵According to the Social Security Administration, nine out of ten individuals aged 65 or older receive social security benefits. Also, among the elderly beneficiaries, over half of the households receive over 50% or more of their income from social security.

³⁶Munnell (2015) showed that, in 2015, the average retirement age is 64 for men and 62 for women, which is roughly the same as a decade ago.

age possible), then the monthly benefits would increase to \$1932. Knoll and Olsen (2014) find that the age of 62 is the most frequent enrollment age, and the age of 70 to be the least frequent. Only 2% of the population choose to delay benefits till 70. Several papers have shown it is optimal for most people to delay benefits claiming, and that average Americans are receiving benefits too early.³⁷

More recently, the National Commission on Fiscal Responsibility and Reform (2010) suggested the Social Security Administration to adopt behavioral economic approaches to encourage the US population to retire later and save more. Knoll et al. (2015) show that people expect to retire and claim benefits later, but many end up retiring and claiming benefits earlier than they have initially planned. This suggests that time-inconsistency with present bias could explain the tendency to claim early. It also suggests that people are non-sophisticated. They proceed to devise effective choice architectures to delay claiming.³⁸ I propose a new approach on this issue.

This paper takes the perspective that, whether the government knows it or not, it is already fooling the agents. Given the social security benefits structure, the labor decisions of the agents are made according to the benefits received for later enrollment. However, agents claim their benefits earlier than planned due to present bias. This paper suggests that the government can utilize this time-inconsistent behavior to also achieve better insurance. Consequently, social security reforms should take the time-inconsistent behavior as given and try to design the monthly benefits to increase equality. Choice architecture alone cannot accomplish this.

To implement full efficiency with social security benefits, the timing would be different from the implementation with taxes. Agents simultaneously retire and start claiming benefits earlier than planned. In other words, the mechanism has to fool the agents into planning to work more and claim higher retirement benefits through delay.

To illustrate how the fooling mechanism works, consider the following three period model, where the agents work in the first period, and decide to continue working or retire in the second period (at age 62), and then retire (stay in retirement if already retired

³⁷Coile et al. (2002) show that delays in claiming benefits can be optimal. Shoven and Slavov (2014) argue that it is optimal, in certain cases, to wait till the age of 70 to start claiming benefits.

³⁸For example, they showed that process intervention (by asking people to consider the benefits of delaying before considering the benefits of claiming early) can postpone enrollment by 9.4 months.

in the second period) in the third period (at age 70).³⁹ Productivity in the first and second periods are the same. Social security as a policy with income tax T_t is defined as $P^{ss} = (k^l(y_1), k^s(y_1, y_2), T_1(y_1), T_2(y_2))$. The social security benefits for agents who claim early is k^l , and k^s for those who claim late. Both depend on period income y_t . The lifetime utility of agents with productivity θ_m who retire late is

$$\left[u_1(c_1(y_{m,1})) - h_1\left(\frac{y_{m,1}}{\theta_m}\right) \right] + \left[u_2(c_2(y_{m,2})) - h_2\left(\frac{y_{m,2}}{\theta_m}\right) \right] + w(k^s(y_{m,1}, y_{m,2})),$$

and the lifetime utility for those who retire early is

$$\left[u_1(c_1(y_{m,1})) - h_1\left(\frac{y_{m,1}}{\theta_m}\right) \right] + 2w(k^l(y_{m,1})).$$

I will assume it is efficient for early retirement and denote the full information efficient allocation as $\{(c_m^*, k_m^*, y_m^*)\}_{\theta_m \in \Theta}$, where $c_m^* = c_{m,1}(y_{m,1}^*)$ and $y_m^* = y_{m,1}^*$.

Let $\tilde{y}_{m,2}$ be the optimal second period income from a type θ_m partially naïve planner's perspective, and $\hat{y}_{m,2}$ denotes the optimal second period income the type θ_m doer would choose. To fool the agents and implement the efficient allocation, the following inequalities have to hold for each $\theta_m \in \Theta$,

$$\begin{aligned} \left[u_2(c_2(\tilde{y}_{m,2})) - h_2\left(\frac{\tilde{y}_{m,2}}{\theta_m}\right) \right] + \hat{\beta}w(k^s(y_{m,1}^*, \tilde{y}_{m,2})) &\geq (1 + \hat{\beta})w(k^l(y_{m,1}^*)), \\ (1 + \beta)w(k^l(y_{m,1}^*)) &\geq \left[u_2(c_2(\hat{y}_{m,2})) - h_2\left(\frac{\hat{y}_{m,2}}{\theta_m}\right) \right] + \beta w(k^s(y_{m,1}^*, \hat{y}_{m,2})). \end{aligned}$$

Suppose the government chooses k^s and T_2 such that $\tilde{y}_{m,2} = \hat{y}_{m,2}$, then this implies the following relationship between utility from working and retiring in the second period, and also between k^l and k^s

$$\hat{\beta} \left[w(k_m^s) - w(k_m^l) \right] \geq w(k_m^l) - \left[u_2(c_{m,2}) - h_2\left(\frac{\tilde{y}_{m,2}}{\theta_m}\right) \right] \geq \beta \left[w(k_m^s) - w(k_m^l) \right],$$

where $c_{m,2} = c_2(\tilde{y}_{m,2})$, $k_m^s = k^s(y_{m,1}^*, \tilde{y}_{m,2})$ and $k_m^l = k^l(y_{m,1}^*)$. Benefits from delaying have to be sufficiently large, in essence $k_m^l \leq k_m^s$, for any income level, which increases the incentives for the agents to plan on retiring late. However, since $\hat{\beta} > \beta$ and the short-term utility from retirement $w(k_m^l)$ is higher than the utility from working $u_2(c_{m,2}) - h_2\left(\frac{\tilde{y}_{m,2}}{\theta_m}\right)$, the agents would be tempted to retire earlier than planned.

³⁹People can start claiming social security benefits at any age between 62 and 70. The model simplifies the decision by modeling it as a choice between early or late enrollment.

The incentive compatibility constraints are $\forall \theta_m \in \Theta$,

$$\begin{aligned} & \left[u_1 (c_1 (y_{m,1}^*)) - h_1 \left(\frac{y_{m,1}^*}{\theta_m} \right) \right] + \left[u_2 (c_{m,2}) - h_2 \left(\frac{y_{m,2}}{\theta_m} \right) \right] + w (k_m^s) \\ & \geq \left[u_1 (c_1 (y_{\hat{m},1}^*)) - h_1 \left(\frac{y_{\hat{m},1}^*}{\theta_m} \right) \right] + \left[u_2 (c_{\hat{m},2}) - h_2 \left(\frac{y_{\hat{m},2}}{\theta_m} \right) \right] + w (k_{\hat{m}}^s). \end{aligned}$$

In essence, the agents choose to work according to their retirement plan. The following proposition demonstrates how social security can implement the full information optimum in a decentralized environment.

Proposition 5 *If $\hat{\beta} \in (\beta, 1]$, then the efficient allocation $\{(c_m^*, k_m^*, y_m^*)\}_{\theta_m \in \Theta}$ can be implemented by $P^{ss} = (k^l (y_1), k^s (y_1, y_2), T_1 (y_1), T_2 (y_2))$, where*

- i. $k^s (y_1, y_2)$ is increasing and regressive in income y_1 and y_2 , and*
- ii. $k^l (y_1) = k^*$ for any income y_1 .*

The proof of Proposition 5 follows from the proof for Proposition 3. Proposition 5 shows how social security can implement the efficient allocation. The social security benefits for later enrollment k^s is regressive in income to incentivize productive agents to produce efficiently. The benefits for early enrollment k^l is a lump-sum transfer that provides full insurance and consumption smoothing for early retirees. The agents are time-inconsistent, so they would retire early and claim k^l . Moreover, since the agents are non-sophisticated, they imagine claiming k^s and would thus work efficiently in the first period. The current social security benefits are progressive in income for both early and late claimants. A reform along the lines proposed in Proposition 5 would make the benefits even more progressive for early claimants but regressive for late claimants. Given the fact that the US population is already claiming earlier than planned, such a reform can help increase output efficiency and improve social insurance.

1.8.2 Liquidity of Defined Contribution Plans

The design of defined contribution plans is of growing interest. Over the years, the percentage of workers covered by defined contribution (DC) plans has been increasing, while coverage by defined benefit (DB) plans has been in steady decline.⁴⁰ The literature

⁴⁰From 1980 to 2008, the proportion of workers participating in DB plans fell from 38% to 20%, while participation in DC plans rose from 8% to 31%. (See Butrica et al. (2009).)

has focused on how to influence DC plan enrollment behavior.⁴¹ Other aspects of the design of DC plans has also gained attention. In particular, Beshears et al. (2015b) showed the DC plan in the US to be very liquid compared to countries like the UK, Germany, Canada, Australia and Singapore. After separating from their employer, workers in the US can move their DC account balance to an IRA or Roth IRA and withdraw for any reason from the account before the eligibility age of 59.5 subject to a tax penalty of 10%. Such liquidation before eligibility is forbidden in many countries, such as Germany, Singapore and the UK, except under certain situations, such as a debilitating injury. Consequently, DC plans in the US are very flexible and can meet the transitory needs of a worker, and this has indeed been the case. Argento et al. (2015) find 45% of contributions to retirement accounts among participants under the age of 55 in 2010 were offset by early withdrawals, which is higher than years prior to the Great Recession. However, such flexibility may not be a desirable feature if workers withdraw early due to present bias.⁴²

While making the DC plan more illiquid for time-inconsistent agents can be an effective commitment device, this paper provides an alternative view. The liquidity of DC plans can be redesigned to depend on income and act as a threat to sophisticated time-inconsistent agents, which could incentivize agents to produce efficiently and allow the government to provide better insurance.

To implement full efficiency with liquid retirement accounts, consider the following timing: agents work and deposit s into their retirement accounts in the first and second periods, and are allowed to withdraw d early from their retirement plans in the second period, and agents retire in the third period. The first and second period productivity are perfectly correlated. Let $\tau(y_1, y_2)$ be the pre-eligible withdrawal penalty, which is an off-equilibrium path threat and depends on the first and second period incomes. Retirement plan as a policy with income tax T_t is defined as $P^r = (\tau(y_1, y_2), \xi(y_1), \rho^*, T_1(y_1), T_2(y_2))$, where $\xi(y_1) > 0$ is a lump-sum transfer if an agent withdraws early and ρ^* is a savings subsidy.

In the first period, an agent of productivity θ_m and sophistication level $\hat{\beta}$ predicts the

⁴¹For example, Madrian and Shea (2001) and Thaler and Benartzi (2004) have used the default bias to increase initial enrollment. Carroll et al. (2009) propose the adoption of *active decision* and show that it is better than automatic enrollment in certain situations.

⁴²See Ashraf et al. (2006) and Beshears et al. (2015a).

doer would solve the following problem

$$\max_{c_2, s_2, y_2, d} u_2(c_2) - h_2\left(\frac{y_2}{\theta_m}\right) + \hat{\beta}w(k)$$

subject to

$$c_2 + \frac{s_2}{1 + \rho^*} \leq \mathbf{1}_{early}(d + \xi(y_1)) + y_2 - T_2(y_2),$$

$$k \leq \mathbf{1}_{early}(1 - \tau(y_1, y_2))(s_1 + s_2 - d) + (1 - \mathbf{1}_{early})(s_1 + s_2). \quad (1.15)$$

Let $\rho^* = \frac{1}{\beta} - 1$ to incentivize the present-biased agent to save, and $\mathbf{1}_{early}$ be the indicator of whether the agent withdrew early. Inequality 1.15 shows that the retirement savings k decreases by the withdrawal amount d and the withdrawal penalty $\tau(y_1, y_2)$, where

$$\tau(y_1, y_2) = \begin{cases} \hat{\rho} & \text{if } y_2 \geq \bar{y}(y_1) \\ 1 & \text{if } y_2 < \bar{y}(y_1) \end{cases}.$$

The penalty is structured so that if y_2 is commensurate with y_1 , then the agent would not be tempted to withdraw from the retirement account. Therefore, the retirement account is illiquid and the agent is committed to having sufficient savings for retirement. However, if y_2 is much greater than y_1 (at least $\bar{y}(y_1)$), then the present-biased agent will withdraw early and be penalized by having lower retirement savings. Only agents that produce earn an inefficiently low output in the first period with respect to their productivity would be tempted to withdraw early in the second period. Consequently, by backward induction, the agents in the first period would produce efficiently to avoid the temptation of withdrawing early in the second period. The following proposition states that more flexible retirement plans can help implement the full information optimum.

Proposition 6 *If $[\beta, 1)$, then the efficient allocation $\{(c_m^*, k_m^*, y_m^*)\}_{\theta_m \in \Theta}$ can be implemented by $P^r = (\tau(y_1, y_2), \xi(y_1), \rho^*, T_1(y_1), T_2(y_2))$, where*

- i. $\tau(y_1, y_2)$ is applied to the residual amount left in the savings account, and*
- ii. those who qualify for early withdrawal have $y_1 < y_2$.*

The proof of Proposition 6 follows from the proof for Proposition 4. Currently, retirement plans in the US are more liquid the lower the income, which is also what Proposition

6 proposes. However, there are two main differences between current retirement plans and the plan proposed in Proposition 6. First, liquidity in the proposed retirement plan is not indiscriminately available for all low income agents. To be able to withdraw early, an agent must earn a sufficiently larger income than the previous period to qualify. This is in sharp contrast to the current system, where early withdrawal helps provide agents with short-term liquidity needs. Proposition 6 shows that the current system could be detrimental for time-inconsistent agents and induce inefficiently low output.

Another difference between current retirement plans and the proposed retirement plan is that in Proposition 6, the early withdrawal penalty tax τ is applied to the residual amount left in the savings account. This penalty tax threatens the agents to produce efficiently, or else their future-selves would be tempted to withdraw early and decrease the savings available in the retirement account. The current system has the early withdrawal penalty tax on the withdrawal amount. Though the current system discourages early withdrawals and helps smooth consumption, it does not encourage efficient output.

1.9 Discussions

In this section, I address some immediate extensions, which include environments with dynamic stochastic productivity shocks and commitment versus flexibility preferences. I also discuss some limitations of the main results.

1.9.1 Dynamic Stochastic Productivity Shocks

The model abstracts from concerns with capital taxation (work does not occur in the retirement stage). However, even with dynamic productivity, the efficient optimum is still implementable.

Consider a three period model, similar to Section 3 of Kocherlakota (2005), where work occurs in the first and second period, while the agent retires in the third period. To illustrate, let the set of possible second period productivity be $\Theta_2 = \{\theta_L, \theta_H\}$, with distribution $\Pr(\theta = \theta_m) = \pi_m \in (0, 1)$. The set of possible first period productivity is a singleton $\Theta_1 = \{\theta_1\}$. In each period, the agents have access to a one-period bond with

interest rate zero. The expected lifetime utility of the agents at $t = 0$ is

$$u_1(c_1) - h\left(\frac{y_1}{\theta_1}\right) + \delta \sum_{\theta_m \in \Theta_2} \pi_m \left[u_2(c_m) - h\left(\frac{y_m}{\theta_m}\right) + \delta u_3(k_m) \right].$$

This is also the ex-ante utility for the agents' planners at $t = 0$. The agents suffer from time-inconsistency, and they make decisions in each period according to a quasi-hyperbolic discounting model. The agents' doers have the following ex-post utility at each period

$$U_1 = u_1(c_1) - h\left(\frac{y_1}{\theta_1}\right) + \beta \delta \sum_{\theta_m \in \Theta_2} \pi_m \left[u_2(c_m) - h\left(\frac{y_m}{\theta_m}\right) + \delta u_3(k_m) \right],$$

$$U_{2,m} = u_2(c_m) - h\left(\frac{y_m}{\theta_m}\right) + \beta \delta u_3(k_m),$$

$$U_{3,m} = u_3(k_m).$$

I will assume that $\delta = 1$ and $\beta < 1$. Similar to previous sections, in each period, the doer is triggered when consumption and savings decisions are made.

If all agents were time-consistent, then by the first order condition, the constrained efficient optimal allocation has to satisfy the inverse Euler equation.⁴³ This implies the agents are savings constrained at the optimum. In essence, if agents are allowed to save freely, they would want to save more for the second period than what the optimum prescribes.

If all agents are time-inconsistent and not fully naïve, in the second period, the government can either adopt a fooling mechanism or a threat mechanism. In a threat mechanism, the planner in the first period knows that regardless of the productivity realization in the second period, the planner in the second period would reveal the true productivity. Since all agents have the same productivity in the first period, for this specific model, a linear savings subsidy set at $1 - \beta$ for first period savings along with a threat mechanism in the second period can implement the full information efficient optimum. In essence, the agents can save freely at the savings subsidy augmented interest rate.

The problem is more subtle for a fooling mechanism. In the first period, the agents believe they will consume the at imaginary consumption levels in the second and third periods. As a result, if the agents are allowed to save freely, they would save according to an incorrect projection of their future consumption. In this case, the linear savings subsidy in the first period would also have to take this mis-specification into account.

⁴³From the first order conditions, it is possible to show that $1/u'(c_1^{**}) = \pi_L/u'(c_L^{**}) + \pi_H/u'(c_H^{**})$. See Kocherlakota (2005) for details.

In a model with multiple periods and stochastic productivity shocks, it is still possible to implement both the fooling and threat mechanisms. However, the implementation of a fooling mechanism has certain subtleties. More specifically, at the beginning of each period, the non-sophisticated agents learn their productivity and make decisions based on the imaginary allocation and an imaginary promised utility. When the agents make consumption and savings decisions, another imaginary promised utility is used by the agents to evaluate the merits of choosing the real allocations. In essence, within each period, the planners use a different imaginary promised utility than the doers to make intertemporal decisions: the promised utility for the planners is higher than the one for the doers. As a result, the fooling mechanism is dynamic, since the planners for the next period would use the imaginary allocations contained in the promised utility for the previous period's doers.

The threat mechanism does not have the complications described above. With a threat mechanism, the dynamic taxation problem only needs to keep track of the history of reports. In fact, if the productivity shocks are independent across periods, the government can implement the same threat mechanism every period. In this case, the dynamic taxation problem is essentially a static problem for each period.

In a multi-period model, other concerns may arise, chief among them is the agents' ability to learn. In a fooling mechanism, the doers consume differently from what the planners expected to consume. This could trigger learning and the agents could become more sophisticated.⁴⁴ Therefore, in a dynamic fooling mechanism, the government may also need to account for the evolving sophistication level of the agents. The threat mechanism does not have such concerns. This is because a threat mechanism confirms the planners' initial beliefs about their doers' bias, regardless of whether the beliefs were correct or not. In other words, partially naïve agents never have an opportunity to learn about their true degree of temptation in a threat mechanism, and the government does not need to worry about the effects of learning in a threat mechanism.

⁴⁴For instance, Ali (2011) has the naïve planner learn about the doer's present bias through Bayesian updating by observing the outcome of an intertemporal task with noise.

1.9.2 Demand for Flexibility

In the previous sections, the agents only have a demand for commitment. Recent developments in the literature have highlighted the trade-off between commitment and flexibility.⁴⁵ In these models, agents suffer from temptation but would also like to accommodate their intertemporal taste shock. The demand for commitment comes from the temptation problem. The demand for flexibility comes from the taste shock, which is an unresolved uncertainty.

I examine a two period model as in Section 1.2. I include a taste shock and augment the timing so that the agents receive additional private information over time. To be explicit, the taste shock γ is distributed according to probability distribution $f(\cdot|\theta)$ (CDF $F(\cdot|\theta)$) within a bounded support of $[\underline{\gamma}, \bar{\gamma}]$, with $0 < \underline{\gamma} < \bar{\gamma} < \infty$. The taste shock is realized when the agent makes consumption and savings decision. The agent does not know the extent of the taste shock when productivity is learned, but I allow the distribution of the taste shock to vary with productivity. Hence, both the government and the agents do not know the true taste shock, but the agent is better informed than the government.

The ex-ante utility of the agent is

$$U(c, k, y; \theta) = E_{\gamma} \left[\gamma u(c) - h\left(\frac{y}{\theta}\right) + w(k) \right].$$

The ex-post utility is

$$V(c, k, y; \theta) = \gamma u(c) - h\left(\frac{y}{\theta}\right) + \beta w(k).$$

The tension between commitment and flexibility arises from the fact that the agents would like to save according to the realization of γ , but the act of saving triggers the present bias.

This is a sequential screening problem. The government asks agents to report productivity first, and then the taste shock. As a benchmark, if the taste shock is public and all agents are time-inconsistent to the same degree, then by using the appropriate fooling or threat mechanism, the government can learn about the productivity of the agents and prescribe the optimal savings plan for each agent according to their observed taste shock. The government can implement the efficient allocation if the taste shock was public.

As was shown in the previous sections, by implementing a fooling mechanism, it is possible to use imaginary allocations to fool the non-sophisticated agents and elicit their

⁴⁵Amador et al. (2006), Ambrus and Egorov (2013), Bond and Sigurdsson (2015) and Galperti (2015) are some of the papers that have explored the trade-off in the context of time-inconsistent preferences.

actual productivity. For sophisticated agents, a threat mechanism can costlessly reveal the productivity type. If all agents are time-inconsistent, then after learning the productivity, the government can implement a savings dependent transfer $T(k)$. By Proposition 1 in Galperti (2015), an optimally chosen $T(k)$, with $T(k)$ strictly increasing and $T(k) > 0$ above some threshold \bar{k} and $T(k) \leq 0$ below it, can help the government achieve the efficient allocation. This is because the transfer aligns the incentives of the planner and the doer. More importantly, since a fooling or threat mechanism elicits all of the private information from the agent before consumption and savings in a costless manner, the government can use the savings dependent transfer to ensure the doer saves according to the realized γ .

However, if the consistency of the agents is unobservable by the government (for example, if time-consistent agents exist in the economy), then the optimal allocations may not only be distorted in output, as in Section 1.6, but also intertemporally, as in Galperti (2015). It should be interesting future work to explore the optimal sequential mechanism for screening both consistency and productivity in the first period and then the taste shock in the second.

1.9.3 Impediments to Implementation

This paper has focused on a setting where the government can screen time-inconsistent agents without impunity. In Section 1.6, I showed how the presence of time-consistent agents would impede the screening process and make distortions necessary. Here, I discuss other situations where full efficiency may not be achievable.

Limited Promises and Punishments

With non-sophisticated agents, if the government is limited in the amount of empty promise it can make, then full efficiency might not be achievable. With sophisticated agents, if the government is limited in the amount of punishment it can dole out for misreporting agents, then the efficient allocation might not be implementable. This can arise from the preferences of the agents.

In the previous sections, in addition to the usual assumptions, I assumed the period utility functions were unbounded below and above. This assumption provides a non-empty set of consumption bundles that could deceive non-sophisticated agents for any amount of information rent. Consider the case where the utility function is bounded below. Figure 1.7 illustrates how fooling can be limited for fully naïve agents in the two types example.

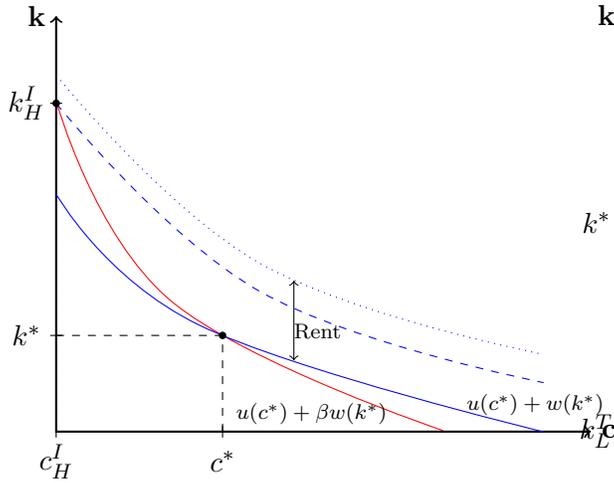


Figure 1.7: Limited Fooling

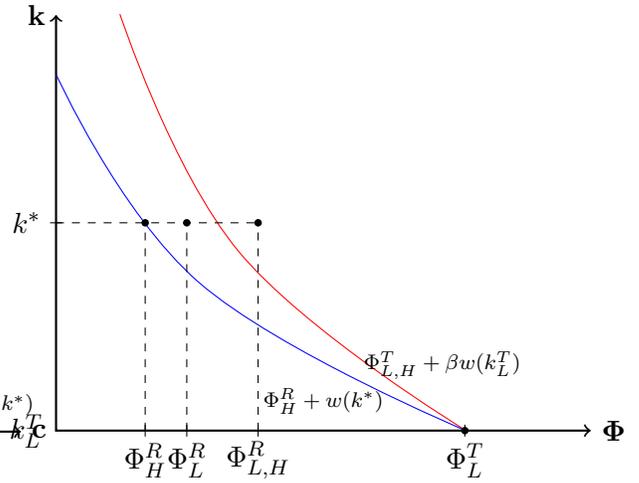


Figure 1.8: Limited Threats

The flatter solid (blue) curve represents the indifference curve of the ex-ante utility and the steeper solid (red) curve represents the indifference curve of the ex-post utility, both evaluated at allocation (c^*, k^*) . The dotted (blue) curve indicates the minimum information rent necessary for the productive agents to be truthful. However, the best the government can do is to set the imaginary allocation at the boundary as indicated in Figure 1.7, which does not provide enough rents to the productive agents to elicit their true type without distorting the allocations for the low productivity type agents. Hence, there will be distortions caused by asymmetric information in the optimal allocation to provide these rents.

Similarly, for sophisticated agents, if the period utility functions are bounded below, then the punishment for misreporting will be limited. Figure 1.8 illustrates how credible threats are limited in a fully sophisticated case with two productivity types. The flatter solid (blue) curve represents the indifference curve of the ex-ante utility for the productive truthful agents who reports truthfully at efficient allocation. The productive agents have an incentive to misreport and receive $(\Phi_{L,H}^R, k^*)$ instead. However, the harshest threat the government can issue is to choose the threat allocation indicated in Figure 1.8. The steeper solid (red) curve represents the indifference curve of the ex-post utility for the efficient agents at the threat allocation. This is clearly not enough to deter the agents from misreporting. Hence, the efficient allocation is not achievable and the optimal allocation

will involve distortions.⁴⁶

However, this does not detract from the main message of the paper, which is threat or fooling mechanisms can improve welfare. Even with limitations on the extent of threats or fooling, the government is still able to improve welfare above the constrained efficient optimum by using the mechanisms introduced in this paper.

Outside Commitment Devices

In reality, people with self-control problems can choose from a wide array of commitment devices available in the market, for example, illiquid assets.⁴⁷ In the case of sophisticated agents, if commitment devices are available and its usage is unobservable, then threats become much less potent. This is because an agent can purchase illiquid assets and bind himself to an intertemporal allocation. This reduces the effectiveness of a threat, because the ex-post utility is maximized over a smaller consumption set. Therefore, private information could matter when commitment devices for sophisticated or partially naïve agents are available.

However, fooling mechanisms for non-sophisticated agents could override the demand for commitment devices. The government can always choose imaginary allocations that make buying an outside commitment device undesirable.

Political Economy

In an economy with political constraints, the incentives to be re-elected may affect the set of implementable policies. Even for benevolent political candidates, if the primary goal is to win the election, political incentives would distort the choice of policies.

The intertemporal rate of consumption could be distorted. This is especially true when elections are held after preference reversal. Suppose an election occurs after the agents' preferences change, then the political candidates have an incentive to undo policies that encourage savings. The competition for votes could force the candidates to pander to the voters' desire for present consumption and undermine the implementation of optimal savings

⁴⁶Similar arguments can be made for fooling and threat mechanisms if the period utility functions were bounded above.

⁴⁷There is also a growing market for commitment devices. For example, StickK, Pact and Beeminder are some recent websites that offer contracts contingent on the completion of stated goals.

policies. The timing of elections has shown to be of crucial importance in models with time-inconsistent voters. Bisin et al. (2015) showed how political candidates would exploit the voters' present bias and undo the incentives for private commitment when elections are held in tandem with the intertemporal decisions of the agents. This fact is true regardless of the agents' sophistication level.

1.10 Summary and Conclusion

In this paper, I demonstrate how government policies can harness the present bias of agents and improve welfare in a Mirrlees taxation model. Contrary to traditional policy proposals, where the primary goal was to mitigate the present bias, I provide methods on utilizing the agents' time-inconsistency to the government's advantage. The methods I developed are also appealing because they can be implemented with familiar policy instruments. For sophisticated time-inconsistent agents, the government can use loans and progressive income repayment plans to increase welfare from the constrained efficient optimum. For naïve agents, the government can use regressive savings subsidy.

The results presented in this paper could be applied to other settings (for example, industrial organization) and to other biases (for example, overconfidence). The concept of fooling and issuing threats could potentially be used in a wider array of mechanism design problems with dynamically inconsistent agents or other biases, like the design of dynamic health or life insurance policies.

It may seem that this paper is against the enactment of commitment devices that could help agents ameliorate their time-inconsistency, since giving agents access to commitment devices would undermine the government's ability to exploit the bias. However, this is not meant to be the message of the paper. This paper highlights the policy concerns and welfare opportunities of the government when the optimal savings policy for time-inconsistent agents is derived in a richer setting. The focus on savings has obscured other costs associated with being time-inconsistent, such as inadequate human capital development. There are substantial costs of having time-inconsistent preferences not included in this model.

Chapter 2

Price Discrimination in the Credit Market

2.1 Introduction

Many goods and services are consumed over time. As an example, credit cards are often mentioned. Two chief concerns arise in the design of credit contracts in dynamic environments. First, in the absence of a commitment device, a consumer who suffers from time-inconsistency may find it difficult to exert self-control and end up with more debt than originally intended. Also, the consumer's preference for credit may also change over time. Due to these two concerns, government policies are enacted to help shield consumers from exploitation, while maintaining sufficient flexibility to meet demand. For example, one of the more well-known policies was the Credit Card Accountability, Responsibility and Disclosure act of 2009 (known as the Credit CARD act of 2009). Despite the popularity of such policies, very little theoretical analysis is done on the effects on welfare.^{1 2} This paper analyzes the optimal design of credit contracts with sequential screening and time-inconsistent consumers. It shows that the optimal contracts within this environment contain features that policies are trying to regulate. However, these features facilitate price discrimination and policies aimed to eliminate them can lead to new distortions and decrease welfare.

¹The Credit CARD act of 2009 was passed with bipartisan support in both the House of Representatives and the Senate.

²For discussions about the welfare effects of policies that restrict credit contracts, see Heidhues and Koszegi (2010) and Tam (2011).

With time-inconsistent consumers and static information revelation, the paper shows that it is optimal for the firm to adopt a *threat mechanism*. A threat mechanism exploits the time-inconsistency of consumers with credible threats off the equilibrium path to exacerbate the self-control problem unless the consumers divulge their private information. Hence, a credit contract will include features that threaten to exacerbate self-control problems to screen consumers. While on the equilibrium path, the time-inconsistent consumers are provided with a commitment device. A threat mechanism is able to achieve perfect price discrimination by increasing the menu of contracts regardless of the degree of time-inconsistency. Therefore, there is a discontinuity in profits with respect to time-inconsistency.³

In an environment where consumers learn new information sequentially, the firm can still use a threat mechanism, but threats are executed on the equilibrium path. As a result, I show that, under certain conditions, the best the firm can do is to virtually implement perfect price discrimination. In essence, the firm perfectly price discriminates almost all of the consumers. Consumers may suffer from large ex-post welfare losses when the threats are executed. These threats provide the consumers with immediate gratification combined with large fees at a later date. The literature has attributed such features to the exploitation of non-sophisticated consumers, while this paper relies on fully sophisticated consumers.

Finally, this paper discusses some policy related issues. Since the threats cause large ex-post welfare losses but arbitrarily small ex-ante welfare losses, policies aimed at restricting the firm's contract space to raise ex-post welfare could lower ex-ante welfare. I demonstrate this by analyzing the welfare effects of the limitation on overdraft charges imposed by the Credit CARD Act.

The recent research on contracting with time-inconsistent preferences has yielded several new insights. There has also been a growing interest in the mechanism design literature on dynamic private information. The present paper is intimately related to the two literatures. This paper contributes to a growing body of research that constitutes the emerging field of behavioral contract theory. Several other papers have focused on the effects of preference reversal on optimal contract design. DellaVigna and Malmendier (2004) focused on a contract design problem for a firm facing partially naïve or sophisticated consumers with time-inconsistent preferences and one-sided commitment. Heidhues and Koszegi (2010)

³This result is similar to Esteban and Miyagawa (2005) with temptation preferences developed in Gul and Pesendorfer (2001).

look at large late fees in credit markets. For a comprehensive introduction to this area of work, Spiegler (2011) and Koszegi (2014) provide overviews of the recent developments in incorporating bounded rationality in industrial organization and contract theory. The paper most related to the present paper is Esteban and Miyagawa (2005). Their paper examines an optimal non-linear pricing model with self-control preferences (Gul-Pesendorfer preferences). They find that despite the consumers having unobserved taste, a firm is able to perfectly price discriminate when consumers suffer from preference reversal in a certain direction, while it is not possible with the opposite direction. More recently, Galperti (2015) contributes to the commitment versus flexibility literature by extending Amador et al. (2006) to a sequential screening model where the principal first screens the agent's time-inconsistency and then proceeds with the agent's taste shock. This paper also deals with sequential screening, but the agent's type is revealed over time.

The paper is also related to the burgeoning literature on sequential screening with the focus on dynamic private information.⁴ More relevant is the recent focus on the role of ex-post and ex-ante information on information rents. Eso and Szentes (2007b) show that ex-post information can be extracted without information rents when the ex-ante information is smooth and continuous. Krahmer and Strausz (2015) show that when the ex-ante information is discrete, the principal is no longer able to extract ex-post information without cost. This paper shows that it is possible to extract both the ex-ante and ex-post information costlessly when agents are time-inconsistent.

The paper is organized as follows. Section 2.2 introduces the model. Section 2.3 analyzes the environment with static information revelation and introduces the threat mechanism. Section 2.4 is the main section and analyzes the sequential screening environment. Section 2.5 applies the results to explain characteristics often observed in credit contracts, and discusses some policy related insights. Section 2.6 summarizes and concludes the paper.

2.2 Model

The model consists of two players, a consumer and a firm. The firm provides a credit contract for the consumer to consume in the present and pay in the future. The relationship lasts for three periods, $t = 0, 1, 2$.

⁴See Courty and Li (2000), Eso and Szentes (2007a), Eso and Szentes (2007b) and Krahmer and Strausz (2015).

2.2.1 Consumer: Taste and Time-Inconsistency

The consumer has private information on his taste which is represented by $\theta \in [\underline{\theta}, \bar{\theta}] \equiv \Theta \subset \mathbb{R}_+$. The consumer receives a private signal $\sigma \in \Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_N\}$ at $t = 0$ with $\sigma_n < \sigma_{n+1}$ and probability $\Pr(\sigma_n) = \gamma_n$. After receiving the signal, the consumer proceeds to sign a contract with the firm before the realization of θ . The consumer receives the outside option $\underline{u} = 0$ if contract is not signed. The signal σ is informative of the distribution of θ . Tastes are drawn from the conditional distribution with full support $\pi(\theta|\sigma)$ at $t = 1$, with cumulative distribution function $\Pi(\theta|\sigma)$. Also, for any $\theta \in \Theta$, $\Pi(\theta|\sigma)$ strictly dominates $\Pi(\theta|\sigma')$ in terms of likelihood ratio if $\sigma > \sigma'$, in essence, $\frac{\pi(\theta|\sigma)}{\pi(\theta|\sigma')}$ is strictly increasing in θ .⁵ Furthermore, for illustrative purposes, I will assume an additional assumption on the distribution of types: For any $\sigma > \sigma'$,

$$\lim_{\theta \rightarrow \bar{\theta}} \frac{\pi(\theta|\sigma)}{\pi(\theta|\sigma')} = \infty.$$

This assumption makes detection of downward deviations in σ detectable almost surely as θ increases.⁶

At $t = 0$, after receiving σ , a consumer has the following *ex-ante* utility

$$U(q, p; \sigma) = \mathbb{E}[\theta q - p|\sigma]$$

from consuming q units of the good bought at price p . The consumer consumes amount q at $t = 1$ and receives utility θq but pays p at $t = 2$.⁷ The ex-ante utility has the consumer discounting with a discount factor of 1. However, a consumer lacking self-control may deviate from intended plans and succumb to temptation. To model this tendency, the consumer is time-inconsistent and makes consumption decisions at $t = 1$ by discounting with a discount factor of $\beta < 1$. Hence, the *ex-post* utility at $t = 1$ after realizing θ is

$$V(q, p; \theta) = \theta q - \beta p.$$

When $\beta \neq 1$, the consumer is dynamically inconsistent and does not consume the amount originally intended. In particular, with $\beta < 1$, the consumer tends to over-consume, because

⁵Note that likelihood ratio dominance implies first order stochastic dominance, so this is stronger than the usual assumption used in the sequential screening literature.

⁶Mirrlees (1999) uses the same assumption on likelihood ratio to show how extreme punishments for large losses in insurance contracts can improve efficiency in a principal-agent setting.

⁷In DellaVigna and Malmendier (2004), goods or services that have such timing are referred to as leisure goods.

the benefit of immediate gratification is magnified while the disutility from delayed payment is discounted.

2.2.2 Firm: Menu and Incentive Compatibility

The firm produces q units of the good at cost $c(q)$. The cost function $c : \mathbb{R}_+ \mapsto \mathbb{R}$ is strictly increasing and strictly convex. Also, for any $q \geq 0$, there exists $0 < M < \infty$ such that $M \geq c'(q)$. This guarantees the firm has unlimited production capacity, and as production increases, the firm operates as if it is producing with constant marginal cost. The firm has full commitment.

By the revelation principle, it is without loss of generality to focus on a truth-telling direct revelation mechanism. The firm offers a contract (a collection of (q, p) pairs) for each σ . A contract for σ is

$$C_\sigma = \{(q_\sigma(\theta), p_\sigma(\theta)), (q'_\sigma(\theta), p'_\sigma(\theta)), \dots\}_{\theta \in \Theta}.$$

A contract is composed of options (q_σ, p_σ) , with $q_\sigma : \Theta \mapsto \mathbb{R}_+$ and $p_\sigma : \Theta \mapsto \mathbb{R}$ for every $\sigma \in \Sigma$. Let $(q_\sigma^R(\theta), p_\sigma^R(\theta))$ denote the option the firm intends a type θ consumer with signal σ to choose. I will refer to this as the *real allocation*. All other options are considered off-equilibrium path.

The game proceeds as follows. The firm posts a menu $C = \{C_\sigma\}_{\sigma \in \Sigma}$. A consumer chooses the contract $C_\sigma \in C$ after receiving the signal σ . After realizing θ , temptation sets in and the consumer chooses an option $(q_\sigma, p_\sigma) \in C_\sigma$ and finalize on the quantity and payment $(q_\sigma(\theta), p_\sigma(\theta))$.

The firm chooses C to maximize expected profit $E[p_\sigma^R(\theta) - c(q_\sigma^R(\theta))]$ subject to the participation constraints, the incentive compatibility constraints and the implementability constraints. Since information is revealed gradually to the consumer, the firm needs to ensure that the contract is dynamically incentive compatible. Let $C_\sigma^*(\theta)$ denote the set of options a truthful consumer at $t = 0$ with signal σ would choose at $t = 1$ when he learns his type is θ :

$$C_\sigma^*(\theta) = \arg \max_{(q'_\sigma(\hat{\theta}), p'_\sigma(\hat{\theta})) \in C_\sigma} V(q'_\sigma, p'_\sigma; \theta).$$

If a consumer with signal σ misreports to be σ' , the consumer would choose an option from

the set $C_{\sigma'|\sigma}^*(\theta)$ when he learns his type is θ in $t = 1$, where

$$C_{\sigma'|\sigma}^*(\theta) = \arg \max_{(q'_{\sigma'}, p'_{\sigma'}(\theta)) \in C_{\sigma'}} V(q'_{\sigma'}, p'_{\sigma'}; \theta).$$

The *ex-post incentive compatibility constraint* ensures that the consumer reports his type truthfully at $t = 1$ provided that he reported truthfully at $t = 0$: $\forall \theta \in \Theta$,

$$(q_{\sigma}^R(\theta), p_{\sigma}^R(\theta)) \in \arg \max_{(q_{\sigma}, p_{\sigma}) \in C_{\sigma}} V(q_{\sigma}, p_{\sigma}; \theta). \quad (2.1)$$

The *ex-ante incentive compatibility constraint* ensures that the consumer reports the signal truthfully at $t = 0$: $\forall \sigma, \sigma' \in \Sigma$

$$\int_{\Theta} \left[\max_{(q_{\sigma}, p_{\sigma}) \in C_{\sigma}^*(\theta)} \theta q_{\sigma} - p_{\sigma} \right] \pi(\theta|\sigma) d\theta \geq \int_{\Theta} \left[\max_{(q_{\sigma'}, p_{\sigma'}) \in C_{\sigma'|\sigma}^*(\theta)} \theta q_{\sigma'} - p_{\sigma'} \right] \pi(\theta|\sigma) d\theta. \quad (2.2)$$

The participation constraints are standard. It requires that each consumer would at least be indifferent between choosing $C_{\sigma} \in C$ from the firm and not buying from the firm: $\forall \sigma \in \Sigma$,

$$\int_{\Theta} \left[\max_{(q_{\sigma}, p_{\sigma}) \in C_{\sigma}^*(\theta)} \theta q_{\sigma} - p_{\sigma} \right] \pi(\theta|\sigma) d\theta \geq 0. \quad (2.3)$$

Since the consumers are sophisticated, they are able to anticipate the preference reversal. Therefore, by (2.1), $(q_{\sigma}^R(\theta), p_{\sigma}^R(\theta)) \in C_{\sigma}^*(\theta)$. However, the dynamic revelation principle does not specify the reporting behavior at $t = 1$ if the consumer misreports at $t = 0$.⁸ It only requires truth-telling on the equilibrium path. This is why (2.2) needs to take into account the possibility of the consumer misreporting at $t = 0$ and then coordinating a future lie at $t = 1$. Hence, a typical element in $C_{\sigma'|\sigma}^*(\theta)$ would be $(q_{\sigma'}(\theta_{\sigma'|\sigma}(\theta)), p_{\sigma'}(\theta_{\sigma'|\sigma}(\theta))) \in C_{\sigma'}$, where $\theta_{\sigma'|\sigma} : \Theta \mapsto \Theta$. In essence, the consumer misreports the signal and then optimally distorts his type according to $\theta_{\sigma'|\sigma}(\theta)$ to maximize his utility.

2.2.3 Timing

To formally capture the time-inconsistency of the consumer and the dynamic information revelation, the timing of the model is shown in Figure 2.1.

At $t = 0$, the firm announces the menu of contracts. The consumer proceeds to choose a contract from the menu according to U after observing σ . At $t = 1$, the consumer realizes θ and chooses a (q, p) pair from the contract according to V . The consumer receives immediate benefit of θq from consumption at $t = 1$. At $t = 2$, the consumer makes the payment p .

⁸See Borgers (2015), Chapter 11.

Chapter 1, the firm can add additional options in the menu that are designed to punish the time-inconsistent consumers only when they misreport in $t = 0$.

The menu is presented so that $(q_\theta^R, p_\theta^R) \in C_\theta^*$, but $(q_{\theta'|\theta}^T, p_{\theta'|\theta}^T) \in C_{\theta'|\theta}^*$. In other words, a truthful type θ consumer would be able to consume (q_θ^R, p_θ^R) , because his future-self would choose it. If he misreports his type to be θ' instead, his future-self would choose the *threat* option $(q_{\theta'|\theta}^T, p_{\theta'|\theta}^T)$. It is referred to as the threat option, because the firm can choose it in such a way that the consumer would avoid misreporting. A threat mechanism has the following contract: $\forall \theta, \in \Theta$,

$$C_\theta = \left\{ (q_\theta^R, p_\theta^R), (q_{\theta'|\theta}^T, p_{\theta'|\theta}^T)_{\theta' \in \Theta} \right\}.$$

The following theorem shows that a threat mechanism can help the firm achieve perfect price discrimination.

Theorem 11 *The firm achieves perfect price discrimination through a threat mechanism, in essence, $(q_\theta^R, p_\theta^R) = (q_\theta^*, p_\theta^*)$, for all $\theta \in \Theta$.*

When $\beta < 1$, the consumer is conscientious of his future-self over-consuming at $t = 1$, so he would like to prevent this. The firm can take advantage of the preference reversal by introducing an allocation that indulges the consumer's temptation. The threat option is designed so that if the consumer deviates to a lower type, he would choose the threat option, which caters to his temptation, over the first best option, which the consumer wished he would be able to consume. The logic is illustrated in Figure 2.2.

Figure 2.2 depicts the two types case with linear utilities. The firm only needs to deter the H -type consumer from pretending to be a L -type consumer. As a result, the threat option needs to be part of the L -type menu: $(q_L^T, p_L^T) \in C_L$. The threat option can keep the H -type consumer honest. If the H -type consumer reports to be L -type, then after the preference change, he would choose (q_L^T, p_L^T) over (q_L^R, p_L^R) , which would give him a lower utility than truth-telling. More importantly, the threat option is also credible, since the L -type consumer would never choose (q_L^T, p_L^T) over (q_L^R, p_L^R) .

Also, there is a discontinuity as β approaches 1. In essence, it is possible to implement perfect price discrimination as long as the consumers are time-inconsistent. However, when $\beta = 1$, the off-equilibrium threats are no longer effective. This result is quite typical in adverse selection with preference reversal, which has been discussed in Esteban and Miyagawa (2005) and Chapter 1.

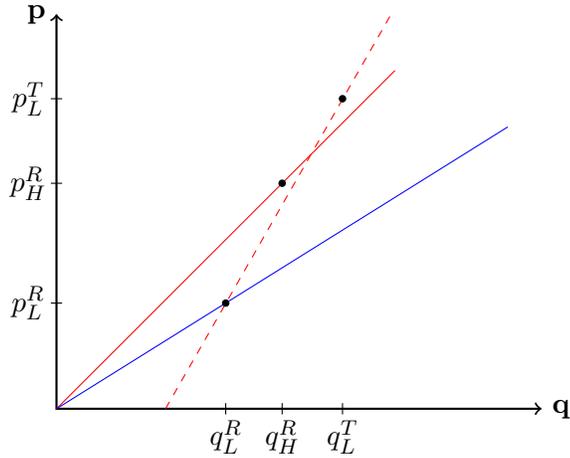


Figure 2.2: Leisure Goods with Time-Inconsistent Consumers

2.4 Dynamic Information Revelation

In this section, the consumer learns an informative signal σ at $t = 0$ and his type θ at $t = 1$. The consumer's information rents can be decomposed into two independent components: the ex-ante private information and ex-post private information.

The consumer's type θ contains additional information not contained in σ . I will refer to σ as the *ex-ante information* and the additional information as *ex-post information*. Following Eso and Szentes (2007b), define $\mu = \Pi(\theta|\sigma)$. Notice μ is distributed uniformly on the unit interval, $\mu \sim U[0, 1]$, and is orthogonal of σ . Let μ be the ex-post information, which can be interpreted as a shock independent of σ occurring at $t = 1$. The consumer would know θ from learning σ and μ , so $\theta = \theta(\sigma, \mu) = \Pi^{-1}(\mu|\sigma)$. From the assumptions on the distribution, $\partial\theta/\partial\sigma, \partial\theta/\partial\mu > 0$. The firm's contract can thus be expressed as $q : \Sigma \times [0, 1] \mapsto \mathbb{R}_+$ and $p : \Sigma \times [0, 1] \mapsto \mathbb{R}$.

2.4.1 Public Shocks and Ex-Ante Information Rents

To define ex-ante information rent, first consider an environment where the ex-post information μ is public and only the ex-ante information σ is private. Following Krahmer and Strausz (2015), this environment will be referred to as the *public shock environment*. The minimum information rent required for separation at $t = 0$ in a public shock environment

will be referred to as the *ex-ante information rent*.

The firm can also use a threat mechanism in a public shock environment to achieve perfect price discrimination. The threat mechanism is defined as $\forall \sigma, \sigma' \in \Sigma$

$$C_\sigma = \left\{ (q_\sigma^R(\mu), p_\sigma^R(\mu)), (q_{\sigma|\sigma'}^T(\mu), p_{\sigma|\sigma'}^T(\mu)) \right\}_{\sigma' \in \Sigma}_{\mu \in [0,1]}.$$

Corollary 4 *In a public shock environment, it is possible for the firm to achieve perfect price discrimination using a threat mechanism.*

By Corollary 4, the firm extracts the whole consumer surplus and does not need to relinquish any information rents to the consumer. Hence, ex-ante information rent is zero. This is in sharp contrast to the usual result in sequential screening with time-consistent preferences. In Eso and Szentes (2007b) and Krahmer and Strausz (2015), the ex-ante information rent is strictly positive.

This result is due to the fact that when μ is public information, the firm can condition the contract on the realization of the shock. This prevents the consumer from attempting to double deviate, so the screening problem is a static one. Therefore, the reasoning from Theorem 11 applies.

It is important to note that in both the static information revelation and the public shock environment, the firm is able to expropriate all of the consumer surplus. However, the consumer does not suffer from any welfare losses.⁹ As a result, in such an environment, policy intervention is unnecessary.

2.4.2 Private Shocks and Ex-Post Information Rents

In a *private shock environment*, the consumer observes μ privately. Hence, the consumer can potentially coordinate a misreport in $t = 0$ with another in $t = 1$ to maximize their utility. In an environment with time-consistent consumers, the firm may need to provide additional information rents in addition to the ex-ante information rents for screening at $t = 0$. This additional rent is referred to as the *ex-post information rent*.

In the private shock environment, the firm needs to design a different set of threat allocations to deter misreports in $t = 0$. This is because for any signals $\sigma, \sigma' \in \Sigma$ with

⁹This is similar to the conclusions of DellaVigna and Malmendier (2004) and Heidhues and Koszegi (2010) for sophisticated consumers, though they do not consider an environment with private information.

$\sigma' > \sigma$ and $\mu \in [0, 1]$ there is a unique $\mu^* \in [0, 1]$ such that $\theta(\sigma', \mu) = \theta(\sigma, \mu^*)$. In essence, if the firm designs a threat allocation to deter a consumer with signal σ' from mimicking a consumer with signal σ , it is unable to deter some truthful consumers at $t = 0$ from choosing the threat allocation. This is because even though a consumer with signal σ' may be deterred by a threat allocation for type $\theta(\sigma', \mu)$, a truthful consumer with type $\theta(\sigma, \mu^*)$ would choose the threat allocation over the real allocation. Figure 2.3 shows how the full support of types makes ex-ante deviations undetectable.

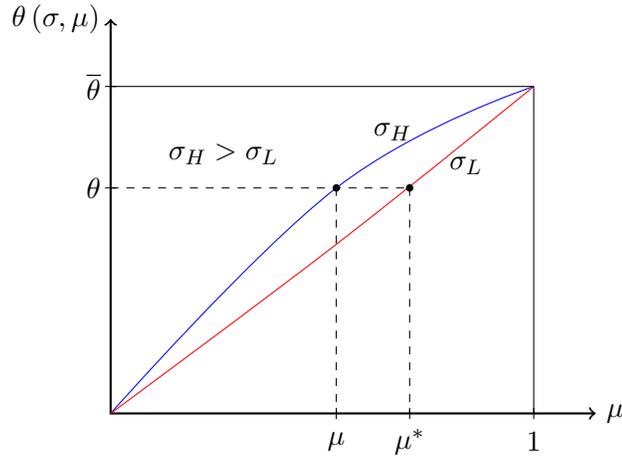


Figure 2.3: Full Support of Types

As a result, if threats are implemented, it is not possible for it to be off the equilibrium path. However, the firm can still exploit both the preference reversal and the difference in ex-ante information by using threats. The firm can introduce the following threat mechanism: $\forall \sigma, \sigma' \in \Sigma$

$$C_\sigma = \left\{ (q_\sigma^R(\mu), p_\sigma^R(\mu))_{\mu \in [0, \bar{\mu}_\sigma]}, (q_{\sigma|\sigma'}^T(\mu), p_{\sigma|\sigma'}^T(\mu))_{\sigma' \in \Sigma, \mu \in (\bar{\mu}_\sigma, 1]} \right\},$$

where $\bar{\mu}_\sigma \in (0, 1]$, for any $\sigma \in \Sigma$. As before, the threat allocations exploit the difference in present and future preferences. The choice of $\bar{\mu}_\sigma$ exploits the difference in distribution of types between higher signals σ' and a lower signal σ . In essence, when a consumer with signal σ' misreports as $\sigma < \sigma'$, he knows that it would be more likely for him to choose the threat allocation than a truthful consumer, because by first order stochastic dominance: $1 - \Pi(\theta(\sigma', \bar{\mu}_\sigma) | \sigma') > 1 - \Pi(\theta(\sigma, \bar{\mu}_\sigma) | \sigma)$. Therefore, this type of threat would punish ex-ante downward misreports more than ex-ante truthful reports.

The optimal contract would require ex-post incentive compatibility. The following lemma characterizes the set of allocations that satisfy ex-post incentive compatibility (2.1) for the real allocations, whenever $\mu \in [0, \bar{\mu}_\sigma]$.

Lemma 2 *The real allocations $\left\{ (q_\sigma^R(\mu), p_\sigma^R(\mu))_{\mu \in [0, \bar{\mu}_\sigma]} \right\}_{\sigma \in \Sigma}$ are ex-post incentive compatible if and only if $V(q_\sigma^R(\mu), p_\sigma^R(\mu); \mu)$ is absolutely continuous in μ and for all $\sigma \in \Sigma$:*

- i. $q_\sigma^R(\mu)$ is non-decreasing in μ , (Monotonicity)*
- ii. $\theta(\sigma, \mu) \frac{\partial q_\sigma^R(\mu)}{\partial \mu} = \beta \frac{\partial p_\sigma^R(\mu)}{\partial \mu}$. (Local incentive compatibility)*

The monotonicity and local incentive compatibility constraints are part of the standard characterization of ex-post incentive compatibility, and it characterizes the real allocations. Since the threat allocations are now on the equilibrium path, it has to satisfy the ex-post incentive compatibility constraint (2.1) as well. The following lemma characterizes the ex-post incentive compatible threat allocations.

Lemma 3 *If the real allocations are ex-post incentive compatible, then the threat allocations $\left\{ (q_\sigma^T(\mu), p_\sigma^T(\mu))_{\mu \in [\bar{\mu}_\sigma, 1]} \right\}_{\sigma \in \Sigma}$ are ex-post incentive compatible if and only if for all $\sigma \in \Sigma$:*

- i. $q_\sigma^T(\mu) = q_\sigma^T(\mu') = q_\sigma^T$ and $p_\sigma^T(\mu) = p_\sigma^T(\mu') = p_\sigma^T$ for all $\mu, \mu' \in [\bar{\mu}_\sigma, 1]$, and*
- ii. $\theta(\sigma, \bar{\mu}_\sigma) q_\sigma^T - \beta p_\sigma^T = \theta(\sigma, \bar{\mu}_\sigma) q_\sigma^R(\bar{\mu}_\sigma) - \beta p_\sigma^R(\bar{\mu}_\sigma)$,*

with $q_\sigma^T \geq q_\sigma^R(\mu)$ and $p_\sigma^T \geq p_\sigma^R(\mu)$ for any $\mu \in [0, \bar{\mu}_\sigma]$.

By Lemma 2 and Lemma 3, it is possible to prove the main result of this paper and that is the threat mechanism can *virtually* implement perfect price discrimination. Let W^* denote the profit of the firm for perfect price discrimination and $W(\bar{\mu})$ be the profit of the firm with private shock where $\bar{\mu} = (\bar{\mu}_1, \bar{\mu}_2, \dots, \bar{\mu}_N)$.

Theorem 12 *In a private shock environment, the firm can implement a threat mechanism such that for any $\epsilon > 0$, there exists a $\bar{\mu}$ so that*

$$W^* - W(\bar{\mu}) < \epsilon.$$

In essence, the firm can virtually implement perfect price discrimination.

Even though it is not possible for the firm to achieve perfect price discrimination in a private shock environment with time-inconsistent consumers, according to Theorem 12, it is able to come arbitrarily close to it. The firm gets arbitrarily close to perfect price discrimination by decreasing the probability of consumers choosing the threat allocation and simultaneously increasing the harshness of the punishment from consuming the threat allocation. The former effect raises welfare, while the latter effect lowers it. However, due to the assumption that as types increase, the firm is able to attribute it to higher signals almost surely, the former effect dominates the latter effect. This is because the firm doesn't need to issue threats that are too severe to deter downward deviations when the likelihood of distinguishing downward misreports from truthful reports is easy.

Note that likelihood ratio dominance, and the assumption on the limit of likelihood ratio is much stronger than previous assumptions in the sequential screening literature. Krahmer and Strausz (2015) assumed first order stochastic dominance. However, even though it may not be possible for the firm to virtually implement perfect price discrimination without the stronger assumptions, the firm can still screen its consumers at $t = 0$ without any information rents with the assumption of first order stochastic dominance. Hence, it is possible for both ex-ante and ex-post information rents to be zero, and this is independent of the stronger assumption on the distribution and is entirely due to the preference reversal of the consumers.

The threshold $\bar{\mu}$ is chosen to exploit the fact that consumers with higher signals find it more likely to choose the threat allocation than consumers with lower signals. Even though the difference in distribution is independent of the consumer's preference reversal, exploiting this difference to elicit ex-post information is not possible for time-consistent consumers. The reason lies in Lemma 3 and the ex-ante incentive compatibility constraint. With time-consistent consumers, the firm is unable to choose threats such that Lemma 3 holds while maintaining a dissonance between the real and threat allocations for the consumer at $t = 0$. Therefore, a time-consistent consumer can benefit from coordinating misreports in both ex-ante and ex-post information and the difference in distributions has less impact on screening than for time-inconsistent consumers. As a result, the firm cannot fully exploit the difference in distributions without the difference in preferences among oneself.

Theorem 12 also utilizes the fact that consumers are risk-neutral. The participation constraint for all consumers bind despite the presence of the threat by increasing their

utility from consuming the real allocations. With risk-averse consumers, the firm would need to provide consumers with a risk premium for the spread between real and threat allocations.

2.5 The Credit Market and Policy Intervention

This section explores the credit market and how the lessons in the previous section can be applied to policy analysis, in particular, the credit CARD act.

By analyzing the contract selection and spending habits in the credit card market, Ausubel (1999) and Shui and Ausubel (2005) find evidence of time-inconsistency. Furthermore, credit contracts often contain features that seem to suggest credit firms preying on the naiveté of time-inconsistent consumers. Heidhues and Köszegi show how credit cards can use teaser rates combined with large fees to take advantage of naïve consumers. Therefore, policymakers have accused the firms of exploitation. Warren (2011) made the following remark a year after the passing of the credit CARD act of 2009:

Some issuers advertised an understated price up front, counting on interest rate re-pricing, fees, penalties, and other often-unexpected charges to let them impose a much higher total cost for the card than implied by the price advertised. The result was a total cost of credit far more expensive than many consumers had anticipated.

– Elizabeth Warren, *The CARD Act: One Year Later*

This has prompted the government to enact policies that protect consumers from potentially welfare reducing credit contracts, with the most well-known being the Credit CARD act of 2009. Even though it has been shown that these policies increase welfare if consumers were non-sophisticated (Heidhues and Köszegi (2010)), it is not known how they would affect the welfare of sophisticates.

Theorem 12 suggests that creditors may construct terms that encourage consumers to over-borrow and then impose a large fee as threats that help screen. Hence, credit contracts can have such features with sophisticates as well. An implementation of the threat allocation would be a credit limit and a corresponding overdraft charge. The overdraft protection of the Credit CARD act limits the assessment of over-limit fees. The law prohibits credit card

companies from charging over-limit fees unless users elected to allow the creditor to approve over-limit transactions. This could raise ex-post welfare by eliminating over-consumption and decrease debt accumulation. However, by limiting the assessment of such fees, the government imposes constraints on the design of the threats. This increases the cost for creditors to screen debtors.

As an illustration, suppose the regulation rules out the use of threats, then it is no longer possible for the creditor to virtually implement the first best welfare. For screening, the creditor would have to provide the consumers with information rents, which introduces additional distortions and decreases ex-ante welfare.

2.6 Conclusion

Intuitively, a firm's optimal pricing strategy must be sensitive to the consumer's preferences. This paper explores the optimal pricing strategy of a firm facing consumers who lack self-control and have dynamic private information. It has shown how a firm facing time-inconsistent consumers can achieve perfect price discrimination in a static screening environment, and virtually implement perfect price discrimination in a sequential screening environment. The optimal contract contains features that can explain seemingly exploitative clauses in a credit contract. This suggests that policies that restrict credit contracts may exacerbate adverse selection, introduce new distortions and decrease welfare.

Chapter 3

Pay What Your Dad Paid: Commitment and Price Rigidity in the Market for Life Insurance - *with Radek Paluszynski*

3.1 Introduction

Traditional economic theories predict that prices in a competitive economic environment evolve accordingly to changes in the underlying marginal cost of production. However, empirical literature provides evidence that this process is often sluggish. In response, a vast sticky price literature has emerged attempting to provide understanding for the observed movements in prices. These models are often based on simple mechanical frictions that lack theoretical underpinnings. In this paper, motivated by empirical findings from the life insurance market, we propose a novel theory of price rigidity in which the company optimally commits to a constant pricing schedule for a certain range of marginal cost variations.

Life insurance data is particularly suitable to test pricing theories. The contracts are simple, the data on historical premiums is readily available and the marginal cost for that industry can be estimated easily. Our attention focuses on the renewable level-term form of insurance. These contracts require a down payment of premium at the moment of signing and stay in force for a pre-defined period, typically between one and twenty years. After the

term expires, customers face a premium schedule that increases with age and are allowed to renew the policy without undergoing a medical reclassification. Thus, as it is pointed out by Hendel and Lizzeri (2003), level-term contracts are characterized by one-sided commitment of the company, but not the consumer. Table 3.1 presents the structure of a typical level-term insurance, commonly referred to as the Annual Renewable Term (ART), for the first 10 policy years. It is important to notice that the insurer only commits to an upper bound on the future premiums (“Guaranteed Maximum” column), which vastly exceed the amounts that can be expected in a market equilibrium. At the same time, however, the contract stipulates a projected path of premiums based on the rates currently offered to older individuals in the same category (“Non-Guaranteed Current” column). This schedule is not binding though, and the company may change it at any point in the future. The only question is: will it?

Table 3.1: Structure of an Annual Renewable Term (ART) contract

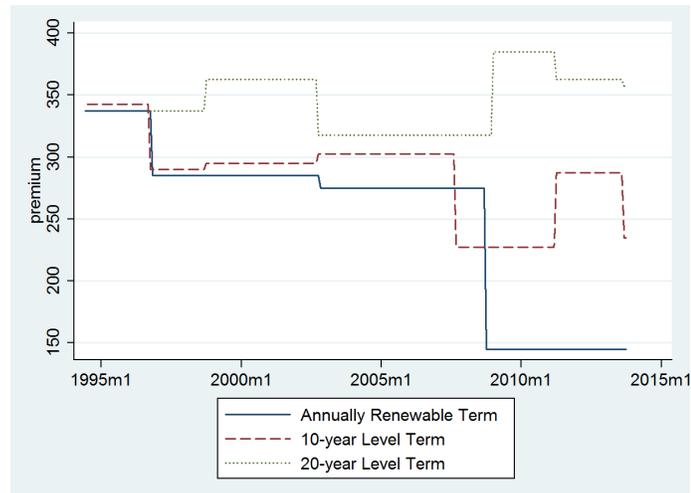
Age	Face Value	Guaranteed Maximum Contract Premium	Non-Guaranteed Current Contract Premium
30	250,000	265.00	265.00*
31	250,000	517.50	267.50*
32	250,000	517.50	267.50*
33	250,000	530.00	270.00*
34	250,000	557.50	272.50*
35	250,000	587.50	280.00*
36	250,000	627.50	292.50*
37	250,000	672.50	307.50*
38	250,000	722.50	325.00*
39	250,000	780.00	350.00*

Sample contract offered by United Heritage Life Insurance Company.

Source: Compulife Software, October 2013.

In this paper we document that life insurance companies in the US have displayed a

striking commitment to their own non-binding promises since 1990. In a data set we construct from an insurance quotation software, the premiums are extremely rigid over time, with the overall probability of a monthly premium change amounting to just 2.56%. This implies an average premium duration of roughly 39 months, placing life insurance on the far-right tail of the price change frequency distribution documented by Bils and Klenow (2004). Figure 3.1 presents an example of premium evolution for different types of policies. These prices are characterized by long periods of rigidity and infrequent, but sharp, adjustments. Remarkably, in the data set we observe a few policies that have kept a constant premium for at least 20 years, rendering it possible, as the title of this paper goes, that a son could pay the same amount for his life insurance as the father used to! We furthermore use the data on historical interest and mortality rates to estimate the underlying marginal cost for the life insurance industry. We find that this cost is non-stationary and highly volatile over time. These empirical patterns from the life insurance market therefore present a puzzle in the light of elementary economic reasoning and create a need for a new theory.



Sample premiums offered by the National Life Insurance Company of Vermont.

Source: Compulife Software, 1990-2013.

Figure 3.1: Premiums over time for different level-term policies

We propose a new theoretical framework that could explain the phenomenon exhibited in

the data. The model addresses the issue of premium rigidity by introducing a transaction cost invested by new consumers before signing the contract. At the same time, existing policyholders have already made an investment in their relationship with the insurance company from previous premium payments. Once the investment is sunk, the consumers are locked in and the insurance company has incentive to increase premiums from their originally promised value by exactly the amount invested. This increase would not decrease the demand for life insurance once the investments have been made. This is because the transaction cost creates a kink in the demand for insurance, introducing a time inconsistency problem for the insurance company. The insurance company also faces stochastic cost shocks that are not observable to the consumers. Therefore, consumers are unsure whether premium changes are due to being held-up or the cost shock. As a result, the insurance companies design premium profiles or schedules that leave no doubt to the consumers that deviations in the premium from its promised value are due to changes in the underlying cost. In other words, the premium schedule has to be incentive compatible for the insurance company.

The main predictions of the model are that optimal premiums display rigidity, and that premium hikes need to be sufficiently large to be incentive compatible. To see why premiums are rigid, notice that if we neglect incentive compatibility, the optimal change in premiums under a mild cost shock is small. However, small changes in premiums do not affect the demand for life insurance once the relationship-specific investments are sunk, so the only incentive compatible premium profile involves rigidity over mild cost shocks. The insurance company is able to regain flexibility in premiums once the cost shocks are sufficiently large, since the corresponding optimal change in premiums is large and would also adversely affect the demand. Therefore, the premium hike needs to be of significant magnitude so that the negative effect on demand is severe and it can only be attributed to changes in cost and not because of any ulterior motives. We show that the optimal premium profile has a simple cutoff rule, where premiums are rigid for cost realizations below a threshold, but flexible above it. Our model explains why the level-term insurance sold by insurance companies have a non-guaranteed premium schedule that afford them the room to be flexible, but at the same time the finalized premiums rarely deviate from it.

The remainder of the paper is structured as follows. Section 3.2 presents a comprehensive literature review. Section 3.3 discusses the construction of our data set and summarizes the

main findings about price rigidity in the life insurance market. Section 3.4 introduces the theoretical model to study the phenomena documented in Section 3.3. Section 3.5 presents the solution to the model and discusses its main predictions. In Section 3.6, we address the alternative hypotheses that could also explain our empirical phenomenon and present some evidence to rule them out. Section 3.7 concludes our findings and discusses the broader implications of the theory developed in the paper.

3.2 Literature Review

Our paper builds upon several strands of economic literature, which we discuss briefly in the following section. Recently there has been much interest in the life insurance market, starting with Hendel and Lizzeri (2003). They use the data on life insurance to test predictions of the Harris-Holmstrom type of model with symmetric learning about the evolution of the insured's health and lack of commitment for the buyer. Because of learning on the company's side, short-term policies induce a risk of reclassification - should the consumer's health deteriorate, the new contract will involve much higher premiums. On the other hand, long-term policies are infeasible due to one-sided commitment on the policyholder's side. The solution, as predicted by the model and confirmed by the data, is front-loading of premiums. Hendel and Lizzeri show that virtually all life insurance policies available on the market exhibit some degree of front-loading which consequently affects lapsation (i.e. rate of voluntary termination of coverage) in a negative way. Daily and Lizzeri (2008) and Fang and Kung (2010) have pursued a similar line of research by considering the effects of a life settlement market on the optimal life insurance contract. Alternatively, Gottlieb and Smetters (2013) develop a model with naive policyholders who underestimate the extent of their income shocks to explain the front-loading of premiums. When policyholders are hit with an unexpected income shock early in their life-cycle and respond by lapsing, the insurance company can make a profit if the premium structure is front-loaded.

The aforementioned papers developed theoretical models for the premium structure faced by a *fixed individual* over time. In contrast, our work tries to explain the premium evolution of a life insurance contract for a *fixed age group* over time. To the best of our knowledge, very few papers have explored this dimension of life insurance contracts, and none have documented the phenomenons highlighted in this paper.

More recently, Kojien and Yogo (2015) investigate the unusual pricing behavior of insurance companies following the financial crisis of 2008. They find that life insurers sold long-term policies at greatly reduced premiums relative to their actuarially fair value, resulting in negative average markups across different companies. The authors propose a theory according to which this phenomenon can be explained as a consequence of financial frictions and statutory reserve regulation. Essentially, life insurers were able to improve their required capital holdings by selling discounted long-term policies. In a subsequent contribution, Kojien and Yogo (2016) develop a similar framework to assess the effect of the so-called shadow insurance on the market outcome observed in the US. Insurance companies have recently been ceding large amounts of liabilities to their affiliated reinsurers, situated in both on- and offshore locations characterized by less restrictive capital regulation. By doing so, the operating insurance companies are able to avoid the cost of keeping the required level of risk-based capital. As a result, the supply of policies in the market for life insurance is significantly expanded, at the cost of much higher impairment probabilities than the ones considered by rating agencies. The findings from these two papers help us understand some of the price drops observed in our sample in the late 2000s, as documented in Section 3.

The theory presented in this paper relates to Klemperer (1987), where consumer switching costs was a factor differentiating the ex ante identical products. The existence of a transaction cost (which we also assume throughout this paper) leads to consumer lock-in and allows the company to extract future monopoly rents. Because of that, firms may often be interested in pricing their products for new customers below marginal cost in order to attract as many of them as possible. In our model, the transaction cost incurred before the contract is signed can also act as a switching cost in future periods. However, price rigidity is not consistent with the standard switching cost model.

Our theoretical model is also related to several papers in the mechanism design literature. Amador et al. (2006) consider the optimal consumption and savings decision for a two-period problem with time inconsistency and taste shocks. Their analysis differs from ours in two dimensions. In their framework, it is not optimal for the agent to choose consumption and savings flexibly because of a systematic indulgence in present consumption which they regret ex-post. In our model, the insurance company suffers from time inconsistency because of a hold-up problem which is inherent in the market for life insurance.

Secondly, their agents do not incur a cost for indulgence, so it is optimal to set an upper

bound for present consumption. However, in our model, we show that for firms it is possible to act optimally when the shock to cost is high, since the loss of consumers from a higher associated price is costly and would help establish credibility. In other words, for high cost shocks, the optimal level of profit can be attained since the prices can be credibly adjusted in a way that the insurance company fully internalizes the consequences of a price hike. This logic is similar to Koszegi (2014). Also related is Athey et al. (2005), who investigate the optimal discretion that should be allowed for a monetary authority. Similar to Amador et al. (2006), and in contrast to our analysis, it is optimal to set an upper bound on the inflation rate for the monetary authority because they lack credibility.

3.3 Life Insurance Prices

3.3.1 Data Construction

We construct a sample of life insurance premiums from Compulife Software, a commercial quotation system used by insurance agents. The updated programs are released monthly, spanning the period from May 1990 until October 2013. For each of the 282 months collected, we recover the premiums for 1-, 5-, 10- and 20-year renewable term policies offered by different companies¹. Even though Compulife is not a complete data set, it covers most of the major life insurers with an A.M. Best rating of at least A-. As the default customer profile we use a 30-year-old male, in the “regular” health category, purchasing a policy at a face value of \$250,000 in California. The choice of this particular state is by Compulife’s recommendation, due to a relatively large population and wide representation of insurance companies. The obtained sample consists of 55,829 observations² on the premium levels for 578 different policies offered by 234 insurance companies. Naturally, over the course of 23 years these firms tend to disappear or merge, as well as discontinue their old products and launch new ones. For this reason, even though we keep track of such transformations whenever possible, each product is observed on average for just over 96

¹Insurance firms often offer several policies of the same type in parallel. In such case, we keep the lowest price assuming it would be the consumer’s optimal choice.

²Because of frequent incompleteness of Compulife databases (especially in the 1990s), we impute the prices whenever a discontinuity appears for up to most 12 months. As a result, we have a total of 562 imputations which constitute roughly 1% of the final sample size. We also drop all the products that are observed for less than 12 continuous months.

months (with a median of 84).

3.3.2 Historical Premiums

Table 3.2 provides a statistical description of price rigidity in our data set. Among 578 distinct insurance products that appear for at least 12 continuous months in the sample, only 369 change their premium amount at all. In total there are just 1432 price changes, consisting of 580 hikes and 852 drops. It is also important to notice that these premium changes, whenever they occur, tend to be of large magnitude, on average amounting to over 10%. The probability of a price change in any month is 2.56%, resulting in an average premium duration of roughly 39 months. This figure includes a vast number of companies that do not adjust prices even once. This may be a deliberate business strategy, but it may also result from other, not market-oriented factors³. Hence, we also calculate the statistics for the subsample of insurance policies that undergo at least one price change. Among those products the probability of a monthly price adjustment increases slightly, but still remains low at 3.5%, resulting in average duration of almost 28 months. It should furthermore be noticed though that products whose price changes in the data set, also tend to stay around for a longer time (114 months as opposed to an unconditional average of 96 months). This observation suggests that certain insurance policies tend to be discontinued (and supposedly replaced by new ones) rather than deviate from the previously promised premium schedule.

³For instance, an insurance company that has no interest in selling certain types of policies may still offer them as a reference for tax authority.

Table 3.2: Price rigidity in the sample

Total number of observations	55,829
Total number of insurance products observed	578
Total number of products that change price	369
Total number of price changes	1432
Total number of price hikes	580
Total number of price drops	852
Average magnitude of price change (in %)	10.74
Average magnitude of price hike (in %)	10.58
Average magnitude of price drop (in %)	10.85
Whole sample:	
probability of a monthly price change (in %)	2.56
median probability of a price change (in %)	1.73
average number of observations per product	96.59
Excluding the companies that never adjust:	
probability of a monthly price change (in %)	3.41
median probability of a price change (in %)	3.23
average number of observations per product	113.82

In order to visualize these findings, Figures 3.2(a) and 3.2(b) take a closer look at the distribution of premium durations and adjustment magnitudes. The first chart depicts a standard view of a distribution of durations with significant positive skewness and a long right tail reaching up to 20 years! Each bin on the histogram represents 6 months, which means that a majority of premiums last between 6 and 30 months, but only a few adjust sooner than that. The second chart presents the distribution of relative sizes of price adjustments. As it is clear from the summary statistics in Table 3.2, premium drops occur more often and are of slightly larger magnitude. The adjustments reach as much as 50% in both directions.

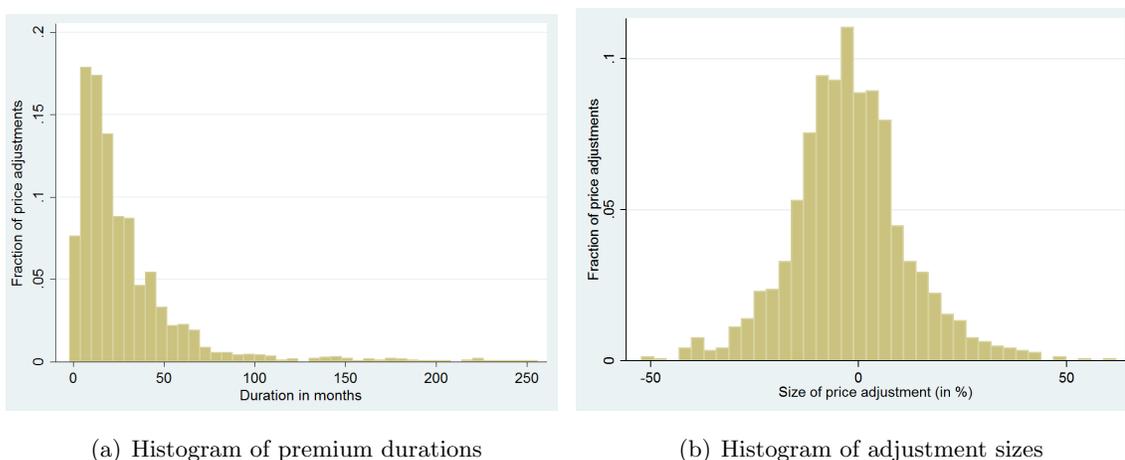


Figure 3.2: Distribution of premium durations and adjustment sizes

In a final piece of data analysis, we explore the distribution in insurance premiums in our sample by examining the relative price dispersion. Figure 3.3 sketches a histogram of all prices, where the average premium in every month and of every term length is normalized to 100. The most striking feature of the graph is the long right tail which implies that some life insurance policies are offered at a price 2.5 times as high as the average in that category, at the given point in time. More generally, even though life insurance may seem to be a rather homogeneous financial product, we observe a significant dispersion across different policies. This may be attributed to varying terms and conditions of different policies (we aggregate

all products under the category “renewable level-term”), as well as imperfectly competitive economic environment in which life insurance companies operate. These imperfections may include search frictions (Hortacsu and Syverson (2004)), information frictions or product differentiation (e.g. with respect to company reputation or brand loyalty).

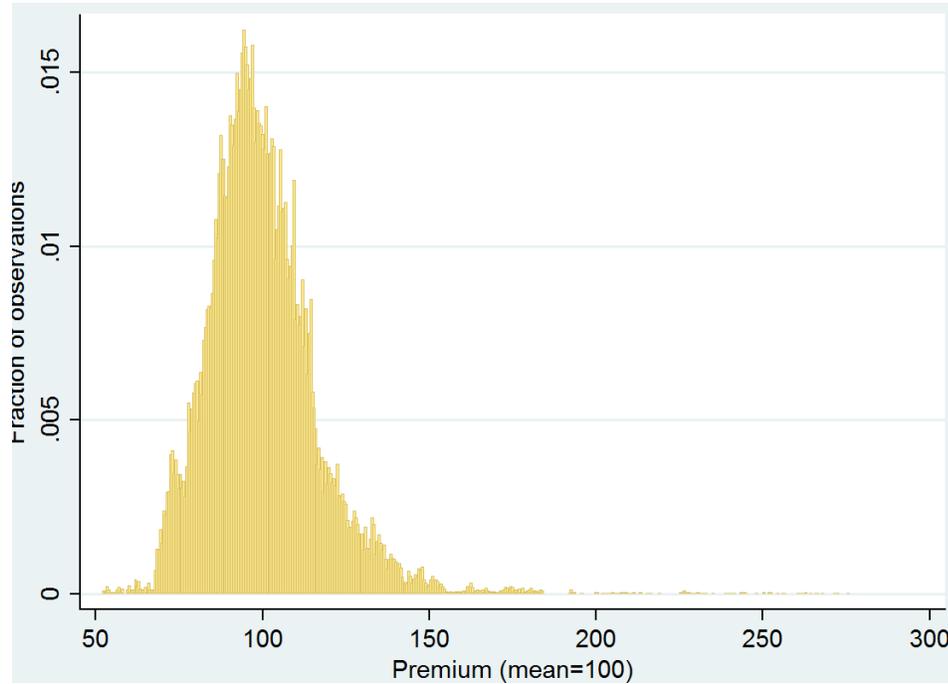


Figure 3.3: Distribution of insurance premiums, relative to the cross-sectional average

Thus far we have illustrated the rigidity of life insurance premiums for a fixed profile of consumers. This means that every month a new 30-year-old non-smoking male, in “regular” health category, can purchase an ART policy for the same price. It is important to understand however, that the premium level this customer expects to pay at renewal is equal to the amount currently offered to the corresponding 31-year-olds (the “Non-Guaranteed Current Contract Premium” column of Table 3.1). Because life insurers only adjust the entire premium schedule, rather than individual rates for selected ages, in most cases these non-guaranteed promises are being kept by the companies. More generally, price rigidity in the market for life insurance can be thought of in two dimensions - one, where a new

customer of the same age is offered the same price every period; and two, where an existing policyholder renews the policy for the same amount as, at the moment of buying, his older counterparts did.

3.3.3 Marginal Cost Estimation

In this section we contrast the historical premiums discussed so far with the marginal cost faced by life insurance companies over the time. We will approximate the life insurance company's marginal cost by estimating the net premium of a level-term policy (also known as the actuarially fair value). A precise description of the method based is provided in the Appendix. Intuitively, a net premium can be thought of as outcome of a zero-profit condition faced by the insurance company. Figure 3.4 presents a stylized illustration of an insurer's cash flow structure. A level-term insurance policy is effective from the moment the first premium is paid, period t , and stays in force for as long as the customer keeps renewing it. Beyond the predefined term, premiums are increasing with age and the benefit is paid out by the company at the moment of death of the insured, denoted $t+n$. In order to break-even, the company must acquire a portfolio of risk-free assets to replicate the present expected value of its future balances. P_t is therefore such that the expected present value of cash flows between the company and the policyholder are equalized, i.e.

$$\sum_{s=0}^N \frac{E_t[P_{t+s}]}{R_t^s} = \sum_{s=0}^N \frac{E_t[B_{t+s}]}{R_t^s}.$$

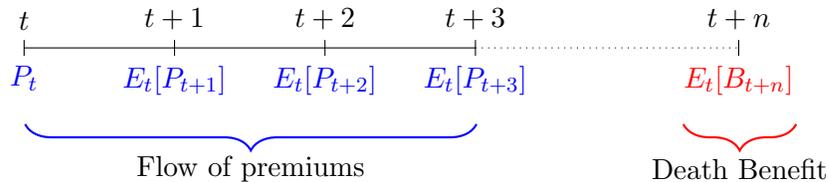


Figure 3.4: Stylized structure of a life insurer's expected cash flows

Figure 3.5 plots the evolution of net premium for an ART policy, from May 1990 until October 2013, together with the HP trend. It ranges from as low as \$118 (in February 2011) up to \$173 (in December 2008), with a standard deviation of 10.5. The net premium exhibits considerable fluctuations over time that depend, by and large, on movements in the interest and mortality rates. In particular, a downward trend can be noticed in the early 2000s (due to a relatively large drop in mortality rates), as well as sharp hikes after November 2008 caused by the interest rate shocks. The recent financial crisis episode contributes to this volatility significantly. While the standard deviation of the HP-filtered net premium is 5.33 for the entire sample, it drops to 3.19 if we disregard all observations after October 2008.

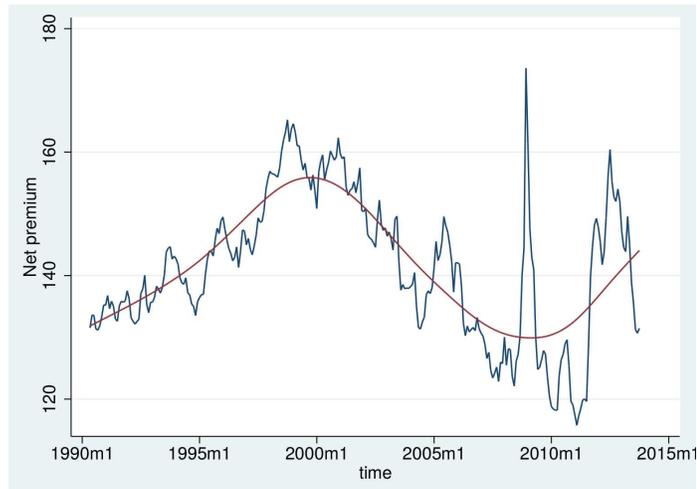


Figure 3.5: Net premium for an Annual Renewable Term policy over time

3.4 The Model

In this section, we present a dynamic life insurance pricing model. We start by briefly presenting the main idea of the model along with its basic intuition. The two key ingredients of the model are the presence of a one-sided hold-up problem of the consumers, and the private stochastic cost faced by the insurance company.

The hold-up problem comes from the consumers' need to make an investment (time forgone while searching, attending medical exams and filling out questionnaires, as well as

the risk of having claims denied during the contestability period) before formally purchasing life insurance. We will call this investment the transaction cost. This creates an incentive for the life insurance company to increase its premiums after the transaction cost is sunk. Another source of the hold-up problem arises from a risk of health reclassification. An existing policyholder may face the risk of being reclassified in the future by other insurance companies after switching. With adverse selection, the insurance company would want to increase the premiums since the existing pool of policyholders would likely face higher premiums if they choose to switch due to health deterioration. As a result, the life insurance company is confronted with a time inconsistency problem, and it can respond by setting a legally binding guaranteed premium schedule to demonstrate commitment.

However, the life insurance company would also like to retain a certain degree of flexibility to respond to the private stochastic cost it faces. Due to this problem, it may not be optimal for the life insurance company to commit to a guaranteed premium schedule. Instead, it may wish to retain some degree of flexibility in the premium schedule in response to potentially large movements in the cost.

These two features of the model highlight the main trade-off. The life insurance company would need to announce an incentive compatible premium schedule such that consumers are convinced that upward premium movements are due to cost shocks and not opportunistic behavior. The solution for the life insurance company is to commit to a rigid premium for small cost shocks, and to change its premiums only when the cost shocks are large enough. The life insurance company can afford to change its premiums only when the increase in premiums decrease the demand for insurance to an extent that it couldn't possibly be profitable to do so for any other reason.

3.4.1 The Setup

Health, Demand and Transaction Cost

Consider a two period model where a continuum of consumers choose whether to purchase insurance in the first period and whether to renew in the second period. In each period, the consumers face a mortality risk of $m_t \in (0, 1)$, where $t \in \{1, 2\}$, which is common knowledge.

Consumers are heterogeneous in their private valuations of the life insurance contract

$r = (r_1, r_2)$. Variable r_t can be thought of as linear utility from owning a policy in period t . The private valuations are distributed according to a continuous and differentiable joint distribution function $h(\cdot, \cdot)$, with bounded support $[\rho, R]^2$ and $\rho > 0$. We assume that R is sufficiently large such that there still is positive demand even when a firm facing the largest possible cost realization charges the optimal monopoly premium. We will denote the cumulative marginal distribution function as $H_t(\cdot)$ for period t private valuations. We will also assume that the hazard rate is increasing for H_t , for $t \in \{1, 2\}$, which is a common assumption to ensure a downward sloping demand curve. We will assume that only the distribution of consumers' valuations is common knowledge, so that life insurance companies are unable to write individual specific contracts. We normalize the value of not owning life insurance to zero.

We focus on consumers who seek to obtain life insurance coverage for multiple periods. This is obtained by assuming that the expected value of renewing a policy in the second period is greater than the value of dropping out, normalized to zero, or switching. The reason behind excluding the consumers who only demand life insurance for one period is that they are not held-up by the company, and they would most likely purchase a term policy without the option to renew. These so-called “non-renewable” contracts are usually cheaper than renewable level-term insurance and thus are more likely to attract such customers.

Before becoming a policyholder, a consumer needs to invest a transaction cost of $\mu > 0$. As mentioned before, there are various costs captured within the transaction cost parameter μ . For simplicity, we assume that μ does not vary with age or time. The policyholder also needs to invest the transaction cost if he decides to switch companies in the second period, but would not need to if he renews. We will elaborate on the outside options when we discuss the commitment problem of the consumers.

The model implicitly assumes the existence of other firms providing life insurance coverage. However, we do not explicitly model the strategic interaction among firms. This modeling choice enables us to focus on the relationship between the insurance company and its customers and on the contracts they sign.

Cost Shocks and Premium Schedules

For each period, the life insurance company faces a stochastic unit cost c_t , which is

randomly drawn from a continuous and differentiable cumulative distribution function $G_t(\cdot)$ with bounded support $[\underline{c}, \bar{c}]$, for $t \in \{1, 2\}$. Notice that the unit cost does not vary with the size of the insured pool. We will assume that \bar{c} is sufficiently large, so that large negative cost shocks have a strictly positive probability of occurring. We also assume that the cost realizations are independent.

The insurance company chooses premium schedules $P_1(c_1)$ and $P_2(c_2; c_1)$ as a function of the possible cost realizations. The second period premium $P_2(c_2; c_1)$ depends on c_1 because the first period premium affects the pool of existing policyholders before the second period. In essence, the premium schedule is a mapping $P_1 : [\underline{c}, \bar{c}] \rightarrow \mathbb{R}$ and $P_2 : [\underline{c}, \bar{c}]^2 \rightarrow \mathbb{R}$. A life insurance contract is defined as the premium schedules and the face value of the contract $\{P_1(c_1), P_2(c_2; c_1), F\}$. For simplicity, we take the face value as given, so the equilibrium contract that we solve for is $\{P_1(c_1), P_2(c_2; c_1)\}$, with a predefined F . For illustrative purposes, we also assume that c_1 is known, which simplifies the contract to $\{P_1, P_2(c_2), F\}$.

One-sided Commitment and Selection

A policyholder can choose to switch to another company in the second period, if there are other companies that offer better expected premiums. We will denote the expected premium of the outside option to be \underline{P}_2 . We assume that the policyholder cannot revert back to its original contract if he is reclassified to a different risk group after switching.

An existing policyholder may experience health deterioration and hence face higher premiums in the second period. We assume that the expected premium of the outside option is correlated with the second period valuation r_2 . More specifically, the premium of the outside option is denoted as $\underline{P}_2(r_2)$, a stochastic variable for any given second period valuation r_2 . Furthermore, we assume that the biased selection only applies to consumers who were not covered in the first period.

We also assume that a consumer cannot find a better insurance contract than the one offered unless he experiences some changes in health status after the first period. This can only happen if the consumer is a policyholder in the first period. In essence, a consumer who chooses to forego purchasing insurance in the first period, or a newcomer who chooses to forego purchasing in the second period, will not find a better insurance deal in the market, and hence receive no coverage for that period.

Timing

Figure 3.6 provides a timeline with a detailed account of the sequence of events in the model. The insurance company's decisions are shown at the top of the timeline, while the consumer's decisions are shown at the bottom. Shocks are marked in italics. The dashed line represents the underlying events that take place implicitly before and following the story described in our model. The solid part of the timeline depicts the sequence of events that we focus on in this framework.

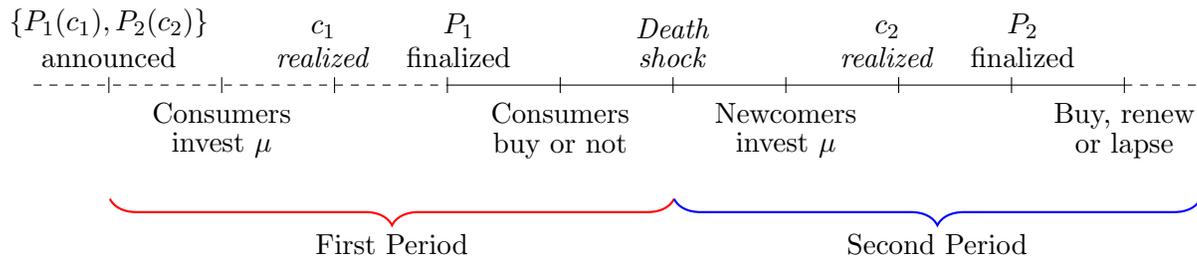


Figure 3.6: Timing of events

In the first period, the insurance company announces the schedule for period t premiums that depends on the realization of c_t . Consumers proceed to make their investment decision. After the cost shock c_1 , the existing consumers decide whether to buy the policy as insurance against the impending death risk at the end of the first period. The model will only describe the purchasing decision of the consumer and take the first period cost shock c_1 as given.

At the beginning of the second period, consumers who did not purchase in the first period can choose to invest μ in order to buy in the second period. After the investment μ is sunk, the cost shock is realized and the insurance company finalizes their premium P_2 . Having observed P_2 , the newcomers can choose to actually sign the contract or not, while the existing policyholders decide whether to renew, switch or lapse.

3.4.2 Characterizing the Demand

Let $P_1(c_1)$ and $P_2(c_2; c_1)$ be the premium schedules announced by the life insurance company, where c_1 is given. For simplicity, we will refer to the premium schedules as $\{P_1, P_2(c_2)\}$. We use backward induction to characterize the demand for life insurance. Consumers, irrespective of whether they are newcomers or existing policyholders, would purchase life insurance in the second period if the realized second period premium is less than their second period valuation: $r_2 \geq P_2(c_2)$. Therefore, newcomers would choose to invest if the following inequality holds

$$V_2^N(r_2) \equiv \Pr(P_2(c_2) \leq r_2) E[r_2 - P_2(c_2) \mid P_2(c_2) \leq r_2] - \mu \geq 0. \quad (3.1)$$

In other words, the newcomers choose to invest in the second period if the expected payoff from owning life insurance less the transaction cost, denoted $V_2^N(r_2)$, is greater than the outside value of 0.

We focus on the premiums that are efficient. In essence, the expected payoff for the consumers monotonically increases with r_2 . Afterwards, we will show that the optimal premium schedules are indeed efficient in Proposition 7. We call a second period valuation for which inequality (3.1) binds the second period threshold valuation of the newcomers, and denote it as \bar{r}_2^N which is unique. This implies that consumers with $r_2 \geq \bar{r}_2^N$ would invest μ in the second period, if they didn't purchase in the first period.

A consumer would also evaluate the merits of renewing against the benefit of switching prior to purchasing in the first period. The expected value of switching is

$$V_2^S(r_2) \equiv \Pr(\underline{P}_2(r_2) \leq r_2) E[r_2 - \underline{P}_2(r_2) \mid \underline{P}_2(r_2) \leq r_2] - \mu.$$

We assume that for any second period valuation r_2 , $V_2^S(r_2)$ is strictly positive. Since our model focuses on consumers with coverage needs that extend to multiple periods, a consumer who purchases life insurance in the first period would have to satisfy the following inequality

$$V_2^R(r_2) \equiv \Pr(P_2(c_2) \leq r_2) E[r_2 - P_2(c_2) \mid P_2(c_2) \leq r_2] \geq V_2^S(r_2). \quad (3.2)$$

In other words, for an existing policyholder with valuation r_2 the second period expected payoff from renewing is at least as high as the expected payoff from switching. Notice that an existing policyholder who renews does not need to invest μ . As a result, combining (3.1)

and (3.2), we have the following relationship between the payoffs for a newcomer and for an existing policyholder

$$V_2^R(r_2) = V_2^N(r_2) + \mu.$$

We call a second period valuation for which inequality (3.2) binds the second period threshold valuation for policyholders, and denote it as \bar{r}_2^E . This implies that consumers who bought coverage in the first period have at least a second period valuation of \bar{r}_2^E . If not, then they wouldn't have invested in the first period. Assuming that $V_2^R(r_2)$ and $V_2^S(r_2)$ only cross each other once, \bar{r}_2^E is unique.

In the first period, following the investment of μ , a consumer with valuation (r_1, r_2) such that $r_2 \geq \bar{r}_2^E$ will actually purchase life insurance if the following inequality is satisfied

$$(r_1 - P_1) + (1 - m_1) V_2^R(r_2) \geq (1 - m_1) \max \{V_2^N(r_2), 0\}. \quad (3.3)$$

First note that $V_2^R(r_2)$ is always greater or equal to zero. The inequality states that consumers would purchase life insurance in the first period if the value of purchasing coverage in the first period with the expected benefit of renewing in the second period is greater than the expected payoff from delaying the purchasing to next period. Notice the consumers who purchase in the first period do not need to invest the transaction cost μ again in the second period, which factors into the consumer's decision to purchase. Also, since we assumed that a consumer without coverage in the first period would not experience a change in health status, the expected payoff from delaying the purchase to the second period does not include $V_2^S(r_2)$.

Next, we examine the purchasing decision of consumers with different second period valuations. For a consumer with $r_2 \geq \bar{r}_2^N$, inequality (3.3) can be expressed as $r_1 \geq P_1 - (1 - m_1) \mu$. On the other hand, for consumers with $r_2 < \bar{r}_2^N$, inequality (3.3) can be written as $r_1 \geq P_1 - (1 - m_1) V_2^R(r_2)$, which is dependent on the second period valuation r_2 . For $r_2 \geq \bar{r}_2^N$, let $\delta = P_1 - (1 - m_1) \mu$ which is independent of r_2 . For $r_2 < \bar{r}_2^N$, let $\chi(r_2) = P_1 - (1 - m_1) V_2^R(r_2)$ which is monotonically decreasing in r_2 under our efficiency assumption. Since if $r_2 < \bar{r}_2^N$ we have $V_2^N(r_2) < \mu$, then a consumer with lower second period valuation requires a higher first period valuation for him to purchase first period coverage.

We assume $\mu > V_2^S(\bar{r}_2^E)$. Under this assumption, it immediately follows that $\bar{r}_2^N > \bar{r}_2^E$ and there exists a benefit for early coverage, since the cost for postponing is high. As a

result, there exists a positive mass of policyholders who will not buy insurance in the second period unless they can renew the policy purchased in the first period. In other words, some consumers with $r_2 < \bar{r}_2^N$ and sufficiently high r_1 buy in the first period to avoid the higher cost incurred by postponing.⁴

Figure 3.7 presents a stylized graphical illustration of the consumers' investment decisions depending on their private valuation (r_1, r_2) . In the graph, the red area represents the mass of consumers who pay the transaction cost in period one. The blue area represents the newcomers who make the investment of μ in the second period. It should be noticed that the consumers' eventual purchase decision depends on the announced prices in both periods. Also, $\chi(\cdot)$ is represented on the graph as a straight line for simplicity, in fact it may be a nonlinear function.

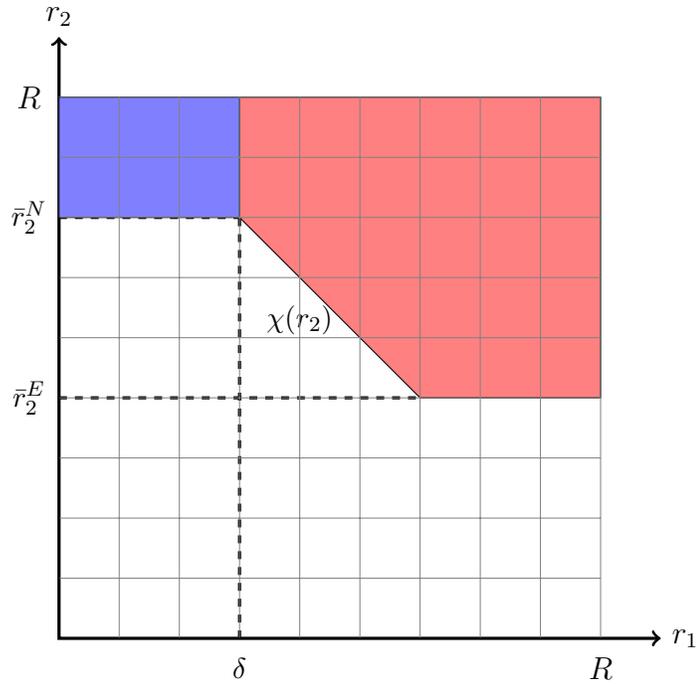


Figure 3.7: Distribution of consumers and their investment decisions

⁴Another case is $\mu \leq V_2^S(\bar{r}_2^E)$. For this case, $\bar{r}_2^N \leq \bar{r}_2^E$ and many consumers with low first period valuation would prefer to defer purchasing coverage till the second period. There is no benefit in purchasing in the first period unless the consumer's first period and second period valuations are both sufficiently high. It should be noted that the main results on the optimal premium schedule would still hold for this case.

We derive the first period demand function by integrating the probability distribution function of the consumers

$$D_1(P_1, P_2(c_2)) = \int_{\bar{r}_2^N}^R \int_{\delta}^R h(r_1, r_2) dr_1 dr_2 + \int_{\bar{r}_2^E}^{\bar{r}_2^N} \int_{\chi(r_2)}^R h(r_1, r_2) dr_1 dr_2. \quad (3.4)$$

The first part of the demand represents the consumers with $r_2 \geq \bar{r}_2^N$, while the second part stands for the consumers with $r_2 < \bar{r}_2^N$. The demand is weakly decreasing in P_1 .

To characterize the demand in the second period, we first focus on the demand of the newcomers, which can be expressed as

$$D_2^{buy}(P_1, P_2(c_2)) = (1 - m_1) \int_{\max\{\bar{r}_2^N, P_2(c_2)\}}^R \int_0^{\delta} h(r_1, r_2) dr_1 dr_2. \quad (3.5)$$

Similarly, the demand function is weakly decreasing in the second period premium $P_2(c_2)$. In particular, the newcomers' demand stays constant when the premiums are below \bar{r}_2^N . In other words, the company would not attract new consumers by lowering the second period premiums below \bar{r}_2^N . However, they may still choose to do so in order to attract more policyholders to renew their contract.

The demand of the existing policyholders for renewing is

$$\begin{aligned} D_2^{renew}(P_1, P_2(c_2)) = & (1 - m_1) \int_{\max\{\bar{r}_2^N, P_2(c_2)\}}^R \int_{\delta}^R h(r_1, r_2) dr_1 dr_2 \\ & + (1 - m_1) \int_{\max\{\bar{r}_2^E, \min\{P_2(c_2), \bar{r}_2^N\}\}}^{\bar{r}_2^N} \int_{\chi(r_2)}^R h(r_1, r_2) dr_1 dr_2. \end{aligned} \quad (3.6)$$

The first part of the demand represents the policyholders with $r_2 \geq \bar{r}_2^N$, while the second part of the demand stands for the policyholders with $r_2 < \bar{r}_2^N$. The *min* function in the lower integral limit makes sure the demand is non-negative. The demand is weakly decreasing in $P_2(c_2)$. In particular, the policyholders' demand is constant when the premiums are below \bar{r}_2^E .

We define the second period demand as the aggregate demand of the newcomers and the existing policyholders

$$D_2(P_1, P_2(c_2)) = D_2^{buy}(P_1, P_2(c_2)) + D_2^{renew}(P_1, P_2(c_2)).$$

As intimated above, the company cannot increase demand by lowering the premiums below \bar{r}_2^E . As a result, the demand becomes perfectly inelastic for any second period premium smaller than the threshold.

3.4.3 Incentive Compatibility

The insurance company faces a different demand once the consumers' investment μ is sunk. Before the investment of μ , the demand in the second period is downward sloping for all prices above the investment cost μ and the insurance company would like to charge a monopoly price. We refer to this pre-investment demand as the *ex-ante* demand and the demand following the investment of μ as the *ex-post* demand. (The kink in the *ex-ante* demand comes from the fact that even if the insurance company sets a premium of zero, only the consumers with valuation of at least μ would consider buying.) The *ex-ante* and *ex-post* demand are different which creates a time inconsistency problem for the insurance company. In particular, after the investment of μ , the insurance company loses the incentive to announce low premiums since it cannot attract any new consumers. Furthermore, the existing policyholders partially revealed their second period valuation by investing and purchasing in the first period. By taking advantage of this opportunity, the insurance company has incentive to increase second period premiums up to \bar{r}_2^E , because it knows the demand is inelastic for any prices below it.

The disparity between the *ex-ante* demand and the *ex-post* demand is the reason why the insurance company needs to commit to keeping its promises. However, private cost shocks create an incentive to tailor the premium according to the shocks. This tension generates a trade-off between commitment and flexibility. To resolve this tension, in our model the life insurance company disciplines its pricing behavior by setting incentive compatible premium schedules. In essence, the company's choice of the finalized premium amount is restricted to the promised schedule that corresponds to the true cost realization.

To begin, we divide the possible cost realizations into three regions. We can define the following cost regions for the second period cost shocks

$$\begin{aligned}\mathcal{C}_2^h &= \{c_2 \mid P_2(c_2) \geq \bar{r}_2^N\}, \\ \mathcal{C}_2^m &= \{c_2 \mid \bar{r}_2^E < P_2(c_2) < \bar{r}_2^N\}, \\ \mathcal{C}_2^l &= \{c_2 \mid P_2(c_2) \leq \bar{r}_2^E\},\end{aligned}$$

and let $\underline{\mathcal{C}}_2 = \mathcal{C}_2^l$ and $\bar{\mathcal{C}}_2 = \mathcal{C}_2^m \cup \mathcal{C}_2^h$.

We will proceed by formulating the incentive compatibility constraints for the second period. Using the newly defined cost regions we can express the incentive compatibility

constraints as

$$[P_2(c_2) - c_2] D_2(P_1, P_2(c_2)) \geq [P_2(\tilde{c}_2) - c_2] D_2(P_1, P_2(\tilde{c}_2)), \quad (3.7)$$

$\forall c_2 \in \mathcal{C}_2^i, \tilde{c}_2 \in \mathcal{C}_2^j$ and $i, j \in \{h, m, l\}$. There are a total of nine second period incentive compatibility constraints. Three of those are to deter deviations within the same cost regions (when $i = j$). Three are to deter downward deviations, and the rest are to deter upward deviations. The incentive compatibility constraints for the second period are written out in full in the appendix.

We will now briefly describe how the incentive compatibility works in the context of our model. The constraints serve to reduce the possible set of premiums the life insurance company can choose from in each period. This set is announced at the beginning of the first period, and it is implicitly assumed that in both periods, any finalized premium that is not within the set can be subjected to legal measures. The only requirement so far is that the set be incentive compatible.

3.4.4 Equilibrium Definition and the Optimization Problem

We model the premium schedule as a mapping from the cost shock realizations announced by the insurance company to the dollar premium amount. We consider a sequential equilibrium of this model. The *sequential* equilibrium has the insurance company choosing the premium schedule in each period while taking into account the future premiums chosen by the future insurance company (a future-self). In the language of mechanism design, a sequential equilibrium has the insurance company choosing a *static* premium that maximizes profit at every history given its future-self will do the same. In essence, a sequential equilibrium results from the solution of two separate static maximization problems in our model.

The sequential equilibrium is a contract $\{P_1, P_2(c_2)\}_{c_2 \in [\underline{c}, \bar{c}]}$ that solves the second period static optimization problem

$$\Pi_2(P_1) \equiv \max \int_{\underline{c}}^{\bar{c}} (P_2(c_2) - c_2) D_2(P_1, P_2(c_2)) dG_2(c_2), \quad (3.8)$$

subject to (3.7) and

$$P_2(c_2) \geq 0, \forall c_2 \in [\underline{c}, \bar{c}], \quad (3.9)$$

and the first period static optimization problem

$$\max(P_1 - c_1)D_1(P_1, P_2(c_2)) + \Pi_2(P_1), \quad (3.10)$$

$$P_1 \geq 0. \quad (3.11)$$

Inequalities (3.9) and (3.11) ensure that the premium is non-negative. In equilibrium, (3.9) and (3.11) will not bind. However, with $\mu > 0$, it is conceivable that the life insurance company may want to charge a negative premium in the first period to increase the demand for life insurance as described in Klemperer (1987).

We solve for the sequential equilibrium premium schedules by the backward induction method.

3.5 Characterization of the Optimal Premium Schedule

In this section, we characterize the optimal premium schedule. In section 5.1, we obtain some preliminary properties of the life insurance premium. We show that it is rigid for certain realizations of the cost shock, and the optimal premium schedule contains a jump in the premium levels. In section 5.2, we provide a cookbook method for computing the optimal premium schedule using standard mechanism design tools.

3.5.1 Qualitative Features of the Optimal Premium Schedule

To characterize the equilibrium premium schedule, we begin by examining the incentive compatibility constraints. We will first show that the only incentive compatible premiums for cost realizations within $\underline{\mathcal{C}}_2$ are constant.

Lemma 4 *For $c_2 \in \underline{\mathcal{C}}_2$, premiums do not vary with cost.*

Proof: See Appendix.

Lemma 4 tells us that the premium schedules are not sensitive to cost shocks within the set $\underline{\mathcal{C}}_2$. This result comes from the incentive compatibility constraints (3.7) when $i = j = l$. This result gives us a glimpse of the rigidity result. However, it could still be the case that the set $\underline{\mathcal{C}}_2$ is of measure zero, and incentive compatible premium schedules can depend on

price for almost all cost shocks. The first part of the following proposition rules this out, and provides a full description of the set of incentive compatible premiums.

Proposition 7 *The set of incentive compatible premiums for the second period has the following properties*

- i. $\underline{\mathcal{C}}_2$ has strictly positive measure.*
- ii. For $c_2 \in \bar{\mathcal{C}}_2$, the company charges monopoly premiums.*
- iii. There exists c_2^h and c_2^m with $c_2^h > c_2^m$ such that for all $c_2 \in [c_2^m, c_2^h)$ we have $P_2(c_2) = \bar{r}_2^N$.*
- iv. The optimal premium schedule for the second period $P_2(c_2)$ is weakly increasing in c_2 .*
- v. For all $c_2 \in \underline{\mathcal{C}}_2$ and for all $c'_2 \in \bar{\mathcal{C}}_2$, we have $c_2 < c'_2$.*

Proof: See Appendix.

Part (i) combined with Lemma 4 delivers the rigidity result. Therefore, it is incentive compatible for premiums to be unresponsive to certain cost shocks. We will denote $\bar{P}_2 = P_2(c_2)$ and the demand $\bar{D}_2 = D_2(P_1, P_2(c_2))$ for all $c_2 \in \underline{\mathcal{C}}_2$.

Part (ii) shows that the insurance company would choose flexible premiums in the cost region $\bar{\mathcal{C}}_2$. In particular, the insurance company would charge a monopoly premium in the second period for cost realizations in $\bar{\mathcal{C}}_2$. Since the company is setting a monopoly premium for the second period within the cost region $\bar{\mathcal{C}}_2$, then it is not profitable to deviate from a cost realization within $\bar{\mathcal{C}}_2$ and report a different cost that is also in the same set. (More specifically, we can refer to the incentive compatibility constraints in the appendix and see that (C.2), (C.3), (C.5) and (C.7) will always hold for the second period.)

Part (ii) also shows that the analysis can be simplified by consolidating \mathcal{C}_2^m and \mathcal{C}_2^h . Let $P_2^*(c_2)$ denote the monopoly premium for a given cost shock c_2 . This allows us to rewrite the downward deviating incentive compatibility constraints ((C.6) and (C.8)) into the following single incentive compatibility constraint

$$[P_2^*(c_2) - c_2] D_2(P_1, P_2^*(c_2)) \geq (\bar{P}_2 - c_2) \bar{D}_2, \forall c_2 \in \bar{\mathcal{C}}_2, \forall \tilde{c}_2 \in \underline{\mathcal{C}}_2. \quad (3.12)$$

Similarly, we can also rewrite the upward deviating incentive compatibility constraints ((C.9) and (C.10)) into the following incentive compatibility constraint

$$(\bar{P}_2 - c_2) \bar{D}_2 \geq [P_2^*(\tilde{c}_2) - c_2] D_2(P_1, P_2^*(\tilde{c}_2)), \forall c_2 \in \underline{\mathcal{C}}_2, \forall \tilde{c}_2 \in \bar{\mathcal{C}}_2. \quad (3.13)$$

Part (iii) says that the transaction cost creates a kink in the demand for life insurance in the second period. This is due to the fact that in the second period, the firm needs to consider the willingness to purchase for both the newcomers and the existing policyholders above the premium \bar{r}_2^N , but only the existing policyholders for any premium below \bar{r}_2^N . This generates a region of strictly positive measure of cost realizations for which the premium does not change according to cost.

Part (iv) says that the shape of the optimal premium schedule is monotonically increasing, but only weakly since the premium is rigid for a strictly positive section of cost realizations.

Part (v) combined with part (iv) says that there exists a cutoff c_2^T such that $\underline{\mathcal{C}}_2 = [c, c_2^T]$ and $\bar{\mathcal{C}}_2 = (c_2^T, \bar{c}]$. In particular, for all low cost realizations where $c_2 \leq c_2^T$, the optimal premium does not change according to cost. However, for high cost realizations where $c_2 > c_2^T$, the optimal premium is weakly increasing in cost.

Proposition 7 provides us with a description of the optimal premium schedule. To summarize, we have so far showed that the incentive compatibility constraint is binding for realizations and deviations within $\underline{\mathcal{C}}_2$, and non-binding for all others except for (3.12) and (3.13). We can further simplify the set of incentive compatible contracts with the following lemma.

Lemma 5 *The incentive compatibility constraints (3.12) and (3.13) only bind when the cost is c_2^T .*

Proof: See Appendix.

Lemma 5 allows us to reduce the incentive compatibility constraints (3.12) and (3.13) to a single binding constraint

$$[\bar{P}_2 - c_2^T] \bar{D}_2 = [P_2^*(c_2^T) - c_2^T] D_2(P_1, P_2^*(c_2^T)). \quad (3.14)$$

Another important observation derived from Lemma 5 is that there is an upward discrete jump in the optimal premium schedule for the second period: $P_2^*(c_2^T) > \bar{P}_2$. This result is summarized in the following proposition.

Proposition 8 *The optimal premium schedule has an upward discrete jump at c_2^T such that $P_2^*(c_2^T) > \bar{r}_2^E > \bar{P}_2$.*

Proof: See Appendix.

The insurance company can increase its premiums within a certain range ex-post without changing the demand of the consumers. This is similar to the kinked demand literature in industrial organization. The logic presented above shows us that the life insurance company can convince the consumers that it is increasing the premiums by announcing a premium schedule that is flexible only when the cost shock is sufficiently large. When the cost shock is large, the life insurance company will charge a high price, and Proposition 8 tells us that this price must be high enough so that it would decrease the demand of the consumers. In other words, flexibility is attainable for cost shocks beyond c_2^T because the premium above the threshold induces lapsation since $P_2^*(c_2^T) > \bar{r}_2^E$. When the decline in demand due to high lapsation rates is significant, the life insurance company will not be able to profit from the hold-up problem. Therefore, the life insurance company can credibly respond to cost shocks only when it causes a simultaneous decrease in the demand for insurance. Figure 3.8 presents a stylized illustration of the model's main insights using the second period premium as an example.

Figure 3.8 also demonstrates how the premium rigidity in cost region \underline{C}_2 is not driven by the same mechanics as the premium rigidity generated by the kinked demand in cost region $[c_2^m, c_2^h)$. The premium rigidity for costs below c_2^T are borne out of the insurance company's need for commitment to avoid holding-up the consumers when the actual cost is realized. This type of rigidity, as we have shown in Proposition 8, can generate a discontinuity in the optimal premium schedule. The premium rigidity for costs within $[c_2^m, c_2^h)$ follows exclusively from the non-differentiability of the demand function at premium \bar{r}_2^N . As we can see, the premium schedule is continuous at cost realizations c_2^m and c_2^h . Our paper focuses on the former mechanism for generating premium rigidity.

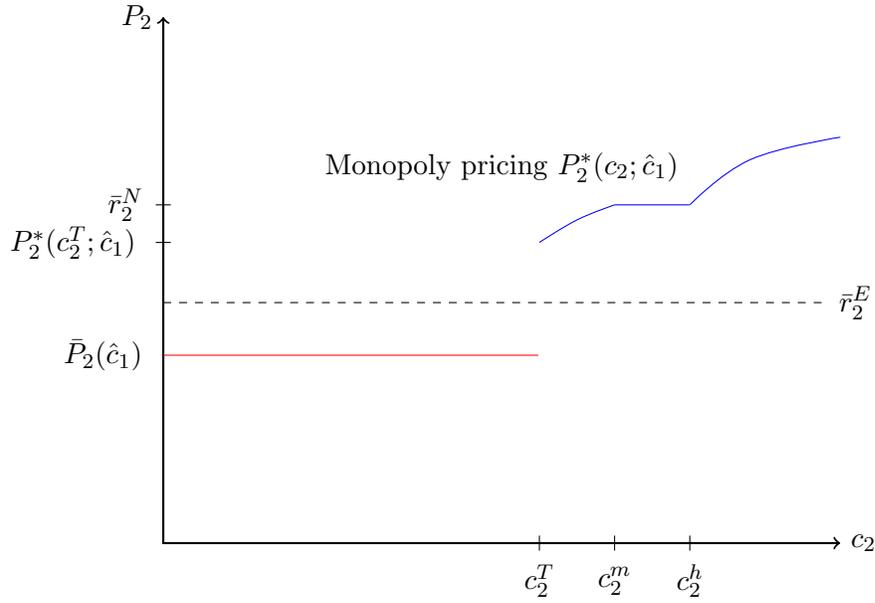


Figure 3.8: Second period incentive compatible premium profile for a given \hat{c}_1

3.5.2 Computing the Optimal Premium Schedule

We will take a step back and rewrite the incentive compatibility constraints. First, we introduce some helpful notation. We can rewrite the incentive compatibility constraints as

$$\pi(c_2) \geq \pi(\hat{c}_2, c_2), \forall c_2, \hat{c}_2 \in [c, \bar{c}],$$

where $\pi(\hat{c}_2, c_2)$ is the profit of the insurance company when it announces a cost of \hat{c}_2 while the true cost is c_2 , and $\pi(c_2) = \pi(c_2, c_2)$ is when the company reports truthfully.

Notice that the usual monotonicity of $D_2(P_1, P_2(c_2))$ with respect to second period cost c_2 necessary for incentive compatibility is trivially satisfied. Since the demand function is non-increasing in second period premiums and by part (iv) of Proposition 7, the second period demand is non-increasing in second period costs.

By Proposition 7, the expected profit conditional on the cost realization being greater than the threshold c_2^T is

$$E[\pi(c_2) | c_2 \geq c_2^T] = \int_{c_2^T}^{\bar{c}} \pi^*(c_2) dc_2,$$

where $\pi^*(c_2)$ denotes the profit under monopoly pricing. It is also straight forward to show

that

$$E[\pi(c_2) | c_2 \leq c_2^T] = G_2(c_2^T)\pi^*(c_2^T) + \bar{D}_2 \int_{\underline{c}}^{c_2^T} G_2(c_2)dc_2$$

We can now formulate the optimization problem of the life insurance company. By (3.14), the insurance company chooses the threshold c_2^T and the rigid premium \bar{P}_2 that maximizes

$$G_2(c_2^T)\pi^*(c_2^T) + \bar{D}_2 \int_{\underline{c}}^{c_2^T} G_2(c_2)dc_2 + \int_{c_2^T}^{\bar{c}} \pi^*(c_2)dc_2$$

subject to (3.14).

Notice that the optimization problem of the insurance company highlights the trade-off between commitment to a single rigid premium and retaining a flexible premium schedule. First, we discuss why the insurance company would like to choose a single rigid premium. If the premium schedule is a singleton, both the last term in the objective function and (3.14) disappear. In other words, the insurance company doesn't need to pay the incentive costs for having the option to change its premiums, as highlighted by Proposition 8. Finally, to see why the insurance company might also want to retain some flexibility, notice that as $c_2^T \rightarrow \underline{c}$, the first term in the insurance company's expected profit increases. Also, since $\pi(c_2)$ is a decreasing convex function and by the envelope theorem, $\pi'(c_2) = -D_2(P_1, P_2(c_2))$ for all $c_2 \in [\underline{c}, \bar{c}]$ where $\pi(c_2)$ is differentiable, we must have decreasing demand with respect to c_2 . Therefore, the life insurance company would like to retain some flexibility to increase the first three terms of the optimization problem if the increase in \bar{D}_2 is faster than the decrease in $G_2(c_2^T)$ as c_2^T decreases. However, if the gains from having flexibility is outweighed by the incentive costs, then it would be optimal for the company would choose a singleton for the premium schedule. We can easily rule out this corner solution by assuming that \bar{c} is sufficiently large and relatively high cost shocks happen with a large probability.

It is worth repeating that the optimal solution for the cutoff c_2^T is not \underline{c} . To see why this is the case, notice that

$$\bar{D}_2 = (1 - m_1) \left[\int_{\bar{r}_2^N}^R \int_0^R h(r_1, r_2)dr_1dr_2 + \int_{\bar{r}_2^E}^{\bar{r}_2^N} \int_{\chi(r_2)}^R h(r_1, r_2)dr_1dr_2 \right],$$

where by the definition of \bar{r}_2^E and Proposition 7,

$$\bar{r}_2^E = \frac{V_2^S(\bar{r}_2^E)}{G(c_2^T)} + \bar{P}_2.$$

When $c_2^T \rightarrow \underline{c}$, we have $\bar{r}_2^E \rightarrow \infty$, since $V_2^S(r_2) > 0$ for all values of r_2 . As a result, if c_2^T is too small, (3.14) would be violated and \bar{D}_2 would be negative.

Finally, to solve for the optimal cutoff c_2^T and the rigid premium \bar{P}_2 , we derive the first order conditions and use (3.14) to express \bar{P}_2 as a function of c_2^T . The result is summarized by the next proposition.

Proposition 9 *The cutoff c_2^T and the rigid premium \bar{P}_2 in the optimal second period premium schedule are given by (3.14) and the following equation:*

$$\begin{aligned} \frac{G_2(c_2^T)}{g_2(c_2^T)} [1 + (1 - G_2(c_2^T))] D_2(P_1, P_2^*(c_2^T)) - \int_{c_2^T}^{\bar{c}} [1 - G_2(c_2)] D_2(P_1, P_2^*(c_2)) dc_2 \quad (3.15) \\ = \Phi(c_2^T) \bar{D}_2 + \frac{G_2(c_2^T)}{g_2(c_2^T)} [1 - G_2(c_2^T)] c_2^T \frac{\partial \bar{D}_2}{\partial c_2^T}, \end{aligned}$$

where $\Phi(c_2^T) = [1 - 2G_2(c_2^T)] c_2^T + \frac{G_2(c_2^T)}{g_2(c_2^T)} + \int_{\underline{c}}^{c_2^T} G_2(c_2) dc_2$.

Proof: See Appendix.

Proposition 9 gives the two equations that pin down the optimal premium schedule (c_2^T, \bar{P}_2) for the second period. For any given optimal $c_2^T \in (\underline{c}, \bar{c})$, (3.14) gives us the the optimal second period rigid premium level \bar{P}_2 . Notice that (3.15) gives us the optimal second period cutoff c_2^T . First off, the left hand side of (3.15) is negative if $c_2^T = \underline{c}$ while the right hand side is strictly positive. Therefore, the optimal second period premium schedule is never fully flexible, confirming part (i) of Proposition 7.

Finally, once we have the optimal second period premium schedule, we are now ready to characterize the optimal first period premium using backward induction. This is presented in our next proposition.

Proposition 10 *Given the optimal second period premium schedule $P_2(c_2)$, the optimal first period premium is characterized by the following equation*

$$D_1(P_1, P_2(c_2)) + (P_1 - c_1) \frac{\partial D_1}{\partial P_1} + \frac{\partial \Pi_2}{\partial P_1} = 0. \quad (3.16)$$

Furthermore, let P_1^{pNR} be the optimal premium for the non-renewable life insurance contract without transaction cost μ and P_1^{pR} be the optimal premium for the renewable life insurance contract without transaction cost. We have $P_1^{pNR} = P_1^{pR} > P_1$.

Proof: See Appendix.

The second part of Proposition 10 states that the life insurance company will charge a lower first period premium than compared to an environment without transaction costs. To see why this is the case, we can imagine a static market. The demand for the contract would be $D(P) = 1 - H(P + \mu)$ and the optimal premium would be described by $P = \frac{1 - H(P + \mu)}{h(P + \mu)} + c$. Since the hazard rate for the distribution of private values is assumed to be increasing, then $P(\mu > 0) < P(\mu = 0)$. The same logic applies to Proposition 10. This shows that the transaction cost μ generates a first period “teaser” rate to attract consumers. This is similar to results exhibited in switching cost models like Klemperer (1987).

3.6 Alternative Explanations

In this section, we attempt to briefly describe some alternative theories that could potentially explain our empirical findings and then attempt to rule them out.

3.6.1 Menu Cost

We begin by investigating the relationship between premium duration and the relative size of adjustment. A plausible hypothesis addressing the observed phenomena is that life insurance companies face a heterogeneous menu cost to changing their premiums, in response to an industry-wide cost shock. In such a world, many companies would postpone any changes until the deviation of the listed price from marginal cost is large enough, and then adjust correspondingly. In that case, the life insurance market should exhibit a positive relationship between the duration of premiums and the size of changes. This relationship is depicted in Figure 3.9, where each point represents a single incidence of a price adjustment. Consequently, the magnitude of change is plotted against the duration of premium. As can be noticed, all points are scattered without a clear pattern and the correlation between the two variables (captured by the regression line) is about 0.13. We conclude that the menu cost is not a relevant theory to explain price rigidities observed in the life insurance market.

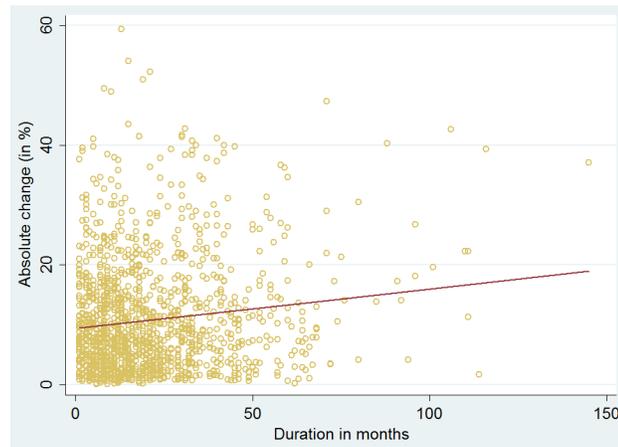


Figure 3.9: Premium duration and adjustment size in the life insurance market

3.6.2 Adverse Selection and Avoiding the ‘Death Spiral’

A potential explanation that could account for the rigid premiums and the infrequent premium changes observed in the data would be the presence of adverse selection. One way for adverse selection to cause such a phenomenon in the premiums observed would be for the following to concurrently happen. If policyholders possess private information on their health status, high risk individuals may increase their coverage needs relative to low risk individuals. This creates a need for the insurance company to increase their premiums to cover the higher expenses. However, if the price elasticity of demand for life insurance is higher for low risk individuals than high risk individuals, then a small increase in prices may trigger a ‘death spiral’ that could severely deteriorate the quality of the pool of policyholders. As a result, as long as the increase in coverage demanded by the high risk individuals do not burden the operation of the insurance company, it is reluctant to change its premiums.

This theory is appealing since insurance markets are believed to be plagued by adverse selection. However, Cawley and Philipson (1999) found no evidence of adverse selection in term life insurance. Without the presence of adverse selection, it is difficult to explain the pricing pattern exhibited by the insurance companies as an attempt to avoid the death spiral.

Even if we assumed the presence of adverse selection in the term life insurance market, Pauly et al. (2003) showed that the elasticity of demand for ART with respect to changes

in risk is relatively small. In other words, it is unlikely a high risk individuals would be signing a substantially larger amount of insurance coverage compared to low risk individuals. Furthermore, they showed that the price elasticity of demand for ART is also sufficiently small and would require a severe adverse selection problems (significant portion of the policyholders to acquire a substantial difference in information about health status) to trigger a death spiral.

3.7 Conclusion

We show that the market for life insurance has exhibited a remarkable degree of price rigidity over the past two decades. Firms that changed premiums in the analyzed sample did so on average every 39 months, preferring one-time jumps of large magnitude to more frequent and gradual price adjustments. We build a theoretical model to explain this phenomenon, based on the fundamental assumption that consumers are locked-in due to a relationship-specific investment. In line with what we find in the data, the model predicts that premiums will remain constant for a wide range of cost shock realizations, while any potential changes take the form of discrete jumps.

The theory in this paper can be applied more generally to environments plagued by the hold-up problem. This may include worker compensations, trading between upstream and downstream firms, international trade and other settings where the hold-up problem has been documented. The key insight of the model is that there is a cost in indulging in one's temptation, and it is optimal to give in to this temptation only when the cost is sufficiently large. The theory could also be used in models with time inconsistencies. For example, it could potentially shed light on the optimal degree of discretion delegated to a monetary authority with private information and temptation to stimulate the economy through surprise inflation.

Chapter 4

References

- Nageeb Ali. Learning self-control. *Quarterly Journal of Economics*, 126:857–893, 2011.
- Manuel Amador, Ivan Werning, and George-Marios Angeletos. Commitment vs. flexibility. *Econometrica*, 74(2):365–396, 2006.
- Attila Ambrus and Georgy Egorov. A comment on ‘commitment vs. flexibility’. *Econometrica*, 81(5):2113–2124, 2013.
- George-Marios Angeletos, David Laibson, Andrea Repetto, Jeremy Tobacman, and Stephen Weinberg. The hyperbolic consumption model: Calibration, simulation, and empirical evaluation. *Journal of Economic Perspectives*, 15(3):47–68, 2001.
- Robert Argento, Victoria Bryant, and John Sabelhaus. Early withdrawals from retirement accounts during the great recession. *Contemporary Economic Policy*, 33(1):1–16, 2015.
- Mark Armstrong and Jean-Charles Rochet. Multi-dimensional screening: A user’s guide. *European Economic Review*, 43:959–979, 1999.
- Nava Ashraf, Dean Karlan, and Wesley Yin. Tying odysseus to the mast: Evidence from a commitment savings product in the philippines. *Quarterly Journal of Economics*, 121(2): 635–672, 2006.
- Susan Athey, Andrew Atkeson, and Patrick Kehoe. The optimal degree of monetary policy discretion. *Econometrica*, 73, 2005.
- Lawrence Ausubel. Adverse selection in the credit card market. *Unpublished*, 1999.

- Matteo Bassi. Mirrlees meets laibson: Optimal income taxation with bounded rationality. *CSEF Working Paper*, (266), 2010.
- Paul Beaudry, Charles Blackorby, and Dezso Szalay. Taxes and employment subsidies in optimal redistribution programs. *American Economic Review*, 99(1):216–242, 2009.
- Shlomo Benartzi and Richard Thaler. Naive diversification strategies in defined contribution saving plans. *American Economic Review*, 91(1):79–98, 2001.
- Douglas Bernheim and Antonio Rangel. Addiction and cue-triggered decision processes. *American Economic Review*, 94(5):1558–1590, 2004.
- John Beshears, James Choi, Christopher Harris, David Laibson, Brigitte Madrian, and Jung Sakong. Self control and commitment: Can decreasing the liquidity of a savings account increase deposits? *NBER Working Paper*, (21474), 2015a.
- John Beshears, James Choi, Joshua Hurwitz, David Laibson, and Brigitte Madrian. Retirement savings and household decisions. *American Economic Review, Papers and Proceedings*, 105(5):420–425, 2015b.
- Mark Bils and Peter J. Klenow. Some evidence on the importance of sticky prices. *Journal of Political Economy*, 112(5), 2004.
- Alberto Bisin, Alessandro Lizzeri, and Leeat Yariv. Government policy with time inconsistent voters. *American Economic Review*, 105(6):1711–1737, 2015.
- Philip Bond and Gustav Sigurdsson. Commitment contracts. *Working Paper*, 2015.
- Tilman Borgers. An introduction to the theory of mechanism design. *Oxford University Press*, 2015.
- Barbara Butrica, Howard Iams, Karen Smith, and Eric Toder. The disappearing defined benefit pension and its potential impact on the retirement incomes of baby boomers. *Social Security Bulletin*, 69(3), 2009.
- Gabriel Carroll, James Choi, David Laibson, Brigitte Madrian, and Andrew Metrick. Optimal defaults and active decisions. *Quarterly Journal of Economics*, 124(4):1639–1674, 2009.

- John Cawley and Tomas Philipson. An empirical examination of information barriers to trade in insurance. *American Economic Review*, 89(4), 1999.
- James Choi, David Laibson, Brigitte Madrian, and Andrew Metrick. Defined contribution pensions: Plan rules, participant decisions, and the path of least resistance. *Tax Policy and the Economy* ed. James Poterba, 2002.
- Kim-Sau Chung. How i learned to love being dynamically inconsistent. *Working Paper*, 2015.
- Courtney Coile, Peter Diamond, Jonathan Gruber, and Alan Jousten. Delays in claiming social security benefits. *Journal of Public Economics*, 84, 2002.
- National Research Council. Aging and the macroeconomy: Long-term implications of an older population. *National Academies Press*, 2012.
- Pascal Courty and Hao Li. Sequential screening. *Review of Economic Studies*, 67, 2000.
- Helmuth Cremer, Pierre Pestieau, and Jean-Charles Rochet. Direct versus indirect taxation: The design of the tax structure revisited. *International Economic Review*, 42:781–799, 2001.
- Helmuth Cremer, Pierre Pestieau, and Jean-Charles Rochet. Capital income taxation when inherited wealth is not observable. *Journal of Public Economics*, 87:2475–2490, 2003.
- Igal Hendel Daily, Glenn and Alessandro Lizzeri. Does the secondary life insurance market threaten dynamic insurance? *American Economic Review Papers and Proceedings*, 98(2), 2008.
- Stefano DellaVigna. Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2), 2009.
- Stefano DellaVigna and Ulrike Malmendier. Contract design and self-control: Theory and evidence. *Quarterly Journal of Economics*, 119(2), 2004.
- Stefano DellaVigna and Ulrike Malmendier. Paying not to go to the gym. *American Economic Review*, 96(3), 2006.

- Peter Diamond and Johannes Spinnewijn. Capital income taxes with heterogeneous discount rates. *American Economic Journal: Economic Policy*, 3, 2011.
- Kfir Eliaz and Ran Spiegler. Contracting with diversely naive agents. *Review of Economic Studies*, 73, 2006.
- Peter Eso and Balasz Szentes. Optimal information disclosure in auctions and the handicap auction. *Review of Economic Studies*, 74, 2007a.
- Peter Eso and Balasz Szentes. The price of advise. *RAND Journal of Economics*, 38, 2007b.
- Susanna Esteban and Eiichi Miyagawa. Optimal menu of menus with self-control preferences. *NAJ Economics, Peer Reviews of Economics Publications*, 12, 2005.
- Hanming Fang and Edward Kung. How does life settlement affect the primary life insurance market? *NBER Working Paper*, (15761), 2010.
- Emmanuel Farhi and Xavier Gabaix. Optimal taxation with behavioral agents. *Working Paper*, 2015.
- Emmanuel Farhi and Ivan Werning. Capital taxation: Quantitative explorations of the inverse euler equation. *Journal of Political Economy*, 120(3), 2012.
- Simone Galperti. Commitment, flexibility, and optimal screening of time inconsistency. *Econometrica*, 83(4), 2015.
- Daniel Gottlieb and Kent Smetters. Lapse-based insurance. *Working Paper*, 2013.
- Faruk Gul and Wolfgang Pendorfer. Temptation and self-control. *Econometrica*, 69(6), 2001.
- Jang-Ting Guo and Alan Krause. Dynamic nonlinear income taxation with quasi-hyperbolic discounting and no commitment. *Journal of Economic Behavior and Organization*, 109, 2015.
- Refet S. Gurkaynak, Brian Sack, and Jonathan H. Wright. The u.s. treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54(8), 2007.
- Marina Halac and Pierre Yared. Fiscal rules and discretion under persistent shocks. *Econometrica*, 82(5), 2014.

- Paul Heidhues and Botond Koszegi. Futile attempts at self-control. *Journal of the European Economic Association*, 7(2-3), 2009.
- Paul Heidhues and Botond Koszegi. Exploiting naivete about self-control in the credit market. *American Economic Review*, 100(5), 2010.
- Martin F. Hellwig. A contribution to the theory of optimal utilitarian income taxation. *Journal of Public Economics*, 91, 2007.
- Igal Hendel and Alessandro Lizzeri. The role of commitment in dynamic contracts: Evidence from life insurance. *Quarterly Journal of Economics*, 118(1), 2003.
- Ali Hortacsu and Chad Syverson. Product differentiation, search costs, and competition in the mutual fund industry: A case study of s&p 500 index funds. *Quarterly Journal of Economics*, 119, 2004.
- Paul Klemperer. Markets with consumer switching costs. *Quarterly Journal of Economics*, 102(2), 1987.
- Melissa Knoll and Anya Olsen. Incentivizing delayed claiming of social security retirement benefits before reaching the full retirement age. *Social Security Bulletin*, 74(4), 2014.
- Melissa Knoll, Kirstin Appelt, Eric Johnson, and Jonathan Westfall. Time to retire: Why americans claim benefits early and how to encourage delay. *Behavioral Science & Policy*, 1(1), 2015.
- Narayana Kocherlakota. Zero expected wealth taxes: A mirrlees approach to dynamic optimal taxation. *Econometrica*, 73(5), 2005.
- Ralph S. J. Koijen and Motohiro Yogo. The cost of financial frictions for life insurers. *American Economic Review*, 105(1), 2015.
- Ralph S. J. Koijen and Motohiro Yogo. Shadow insurance. *Econometrica*, 84(3), 2016.
- Botond Koszegi. Behavioral contract theory. *Journal of Economic Literature*, 52(4), 2014.
- Daniel Krahmer and Roland Strausz. Ex post information rents in sequential screening. *Games and Economic Behavior*, 90, 2015.

- Per Krusell, Burhanettin Kuruscu, and Anthony Smith Jr. Temptation and taxation. *Econometrica*, 78(6), 2010.
- David Laibson. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112(2), 1997.
- David Laibson, Andrea Repetto, and Jeremy Tobacman. Estimating discount functions with consumption choices over the lifecycle. *NBER Working Paper*, (13314), 2007.
- LIMRA. Long-term ordinary lapse survey - u.s. *Connecticut*, 1996.
- Benjamin Lockwood. Optimal income taxation with present bias. *Working Paper*, 2015.
- George Loewenstein, Ted O'Donoghue, and Matthew Rabin. Projection bias in predicting future utility. *Quarterly Journal of Economics*, 118, 2003.
- Brigitte Madrian and Dennis Shea. The power of suggestion: Inertia in 401(k) participation and savings behavior. *Quarterly Journal of Economics*, 116(4), 2001.
- James Mirrlees. The theory of moral hazard and unobservable behaviour: Part i. *Review of Economic Studies*, 66(1), 1999.
- Christian Moser and Pedro Olea de Souza e Silva. Paternalism vs redistribution: Designing retirement savings policies with behavioral agents. *Working Paper*, 2015.
- Alicia Munnell. The average retirement age - an update. *Briefs from Center for Retirement Research at Boston College*, 15-4, 2015.
- Ted O'Donoghue and Matthew Rabin. Choice and procrastination. *Quarterly Journal of Economics*, 116, 2001.
- National Commission on Fiscal Responsibility and Reform. The moment of truth: Report of the national commission on fiscal responsibility and reform. 2010.
- Mark Pauly, Kate Withers, Krupa Subramanian-Viswanathan, Jean Lemaire, John Hershey, Katrina Armstrong, and David Asch. Price elasticity of demand for term life insurance and adverse selection. *NBER Working Paper*, 9925, 2003.
- Edward Prescott and Robert Townsend. Pareto optima and competitive equilibria with adverse selection and moral hazard. *Econometrica*, 52(1), 1984.

- John Karl Scholz, Ananth Seshadri, and Surachai Khitatrakun. Are americans saving ‘optimally’ for retirement? *Journal of Political Economy*, 114(4), 2006.
- John B. Shoven and Sita Nataraj Slavov. Does it pay to delay social security? *Journal of Pension Economics and Finance*, 13(2), 2014.
- Haiyan Shui and Lawrence Ausubel. Time inconsistency in the credit market. *Working Paper*, 2005.
- Jonathan Skinner. Are you sure you’re saving enough for retirement? *Journal of Economic Perspectives*, 21(3), 2007.
- Rani Spiegler. Bounded rationality and industrial organization. *Oxford University Press*, 2011.
- Xuan S. Tam. Long-term contracts in unsecured credit markets. *Unpublished Manuscript*, 2011.
- Richard Thaler and Shlomo Benartzi. Save more tomorrow: Using behavioral economics to increase employee savings. *Journal of Political Economy*, 112, 2004.
- Richard Thaler and H.M. Shefrin. An economic theory of self-control. *Journal of Political Economy*, 89(2), 1981.
- Elizabeth Warren. The card act: One year later. *Remarks by Elizabeth Warren on The CARD Act*, U.S. Department of the Treasury, 2011.

Appendix A

Appendix for Chapter 1

A.1 General Model

The model in this section will outline a general form of dynamic inconsistency. I will examine a less restrictive utility function and discuss both types of partial naïveté. The results presented on the savings problem with present biased agents can then be treated as corollaries to the results presented in this section. The proofs for the results in the paper before Section 1.5 are included here.

A.1.1 Setup of Model

Consider an economy with $|N| \geq 2$ goods produced with labor or other goods and a continuum of agents of measure one. There are $|M| \geq 2$ types of agents denoted by the set $\Theta = \{\theta_1, \theta_2, \dots, \theta_M\}$. The types are distributed according to $\Pr(\theta = \theta_m) = \pi_m > 0$, for all $\theta_m \in \Theta$ with $\sum_{m=1}^M \pi_m = 1$.

The production of good n depends on the labor input l_n and the vector amount of other inputs $\mathbf{x}_n \in \mathbb{R}_+^N$. Let $y_n = F_n(\mathbf{x}_n, l_n; \theta_m) \in \mathbb{R}_+$ denote a continuous and differentiable production process strictly increasing in l_n and \mathbf{x}_n of a type θ_m agent for good n . With a slight abuse of notation, let $l_n = G_n(y_n, \mathbf{x}_n; \theta_j)$ denote the inverse of $F_n(\mathbf{x}_n, l_n; \theta_j)$ with fixed input \mathbf{x}_n . Each type of agent differs in their labor production efficiency: $F_n(\mathbf{x}_n, l_n; \theta_j) > F_n(\mathbf{x}_n, l_n; \theta_k)$, for any labor input $l_n > 0$ and \mathbf{x}_n and good n with $\theta_j > \theta_k$. If a good does not depend on labor input, then it does not depend on θ . The production of at least one of the goods requires labor. Both the production efficiency of each agent and their

labor input $l = (l_1, l_2, \dots, l_N)$ are not observable by the government. The government can only observe output $y = (y_1, y_2, \dots, y_N)$ and input \mathbf{x} . With a slight abuse of notation, I will define $G(y(\theta_{\hat{m}}), \mathbf{x}(\theta_{\hat{m}}); \theta_m) \in \mathbb{R}_+^N$ as the labor input vector for a type θ_m agent with reporting strategy $\sigma(\theta_m) = \theta_{\hat{m}}$.

Consumer Utility

The agents' have the following ex-ante utility before consumption

$$U(c, l),$$

where $c = (c_1, c_2, \dots, c_N)$. The agents' utility changes when they are consuming to the ex-post utility

$$V(c, l).$$

Let $U : \mathbb{R}_+^{2N} \mapsto \mathbb{R}$ and $V : \mathbb{R}_+^{2N} \mapsto \mathbb{R}$ be continuously differentiable and let them be strictly increasing and concave in consumption: $\frac{\partial U}{\partial c_n} > 0$, $\frac{\partial^2 U}{\partial c_n^2} < 0$ and $\frac{\partial V}{\partial c_n} > 0$, $\frac{\partial^2 V}{\partial c_n^2} < 0$. Let U and V be strictly decreasing and convex in labor: $\frac{\partial U}{\partial l_n} < 0$, $\frac{\partial^2 U}{\partial l_n^2} < 0$. Also, there exists $\epsilon > 0$ such that $\frac{\partial U}{\partial c_n}, \frac{\partial V}{\partial c_n} > \epsilon$ for all goods. Finally, for any good $n \in N$, the utility from consumption for both ex-ante and ex-post utility is unbounded below, so $\lim_{c_n \rightarrow 0} \frac{\partial U}{\partial c_n} = +\infty$ and $\lim_{c_n \rightarrow 0} \frac{\partial V}{\partial c_n} = +\infty$. Hence, an interior solution for consumption is ensured. Also, for any good $n \in N$, let $\lim_{l_n \rightarrow 0} \frac{\partial U}{\partial l_n} = 0$ and $\lim_{l_n \rightarrow +\infty} \frac{\partial U}{\partial l_n} = +\infty$ so the labor supply is always strictly positive and finite. Finally, I assume additive separability of consumption and leisure, so that the labor decision does not affect the marginal utility or marginal rates of substitution in goods consumption.

I will assume that the utility from consumption is different. More precisely, I assume the marginal rate of substitution for some consumption goods is different for U than for V .

Assumption 1 *There exist $j, k \in N$ such that $\frac{\frac{\partial U}{\partial c_k}}{\frac{\partial U}{\partial c_j}} \neq \frac{\frac{\partial V}{\partial c_k}}{\frac{\partial V}{\partial c_j}}$,*

First notice that since the utility is separable in consumption and labor, Assumption 1 is independent of the agents' labor choice. Assumption 1 along with strictly increasing and concave utility implies a *single crossing condition* on the indifference curves for the ex-ante and ex-post utilities of the two goods, j and k . If the ex-post and ex-ante utilities satisfy Assumption 1, then agents exhibit *preference reversal*. I will impose an additional standard assumption on the agents' preferences.

Assumption 2 For any good $n \in N$ that depends on labor for production, the ex-ante preferences satisfy the single crossing property: $\frac{\partial}{\partial \theta} \left(-\frac{\partial U}{\partial y_n} / \frac{\partial U}{\partial c_n} \right) < 0$ and $\frac{\partial}{\partial \theta} \left(-\frac{\partial V}{\partial y_n} / \frac{\partial V}{\partial c_n} \right) < 0$

Types of Non-sophistication

Following Spiegel (2011), I will analyze the optimal allocation under two types of partial naïveté: magnitude naïveté and frequency naïveté.

Definition 12 For some $\hat{\alpha} \in (0, 1]$, agents are partially naïve in magnitude if, with probability one, they perceive their ex-post utility to be

$$W(c, l) = \hat{\alpha}U(c, l) + (1 - \hat{\alpha})V(c, l).$$

Definition 12 defines magnitude naïveté. If $\hat{\alpha} < 1$, then agents are certain that their preferences will change. However, since $\hat{\alpha}$ is bounded away from 0, agents underestimate the degree of their preference reversal.

Definition 13 Agents are partially naïve in frequency if they believe their ex-post utility to be $V(c, l)$ with probability $1 - \hat{\alpha}$, where $\hat{\alpha} \in (0, 1]$. In other words, let $W(c, l)$ denote the expected ex-post utility of the agent:

$$W(c, l) = \hat{\alpha}U(c, l) + (1 - \hat{\alpha})V(c, l).$$

Definition 13 defines frequency naïveté. In essence, if $\hat{\alpha} < 1$, the agents attach a positive probability to the likelihood of a change in the preference. However, since $\hat{\alpha}$ is bounded away from 0, they underestimate the probability of their preferences changing.

Under both definitions, if $\hat{\alpha} = 1$, the agents are fully naïve, and if $\hat{\alpha} = 0$, the agents are sophisticated. I will refer to $\hat{\alpha}$ as describing the *sophistication level* of an agent.

Timing, Welfare Criterion and Other Assumptions

The timing of the model is shown in Figure A.1 and is the same as in the previous sections.

Similar to the savings problem, the government has full commitment so the revelation principle applies and the analysis proceeds by using a direct mechanism.

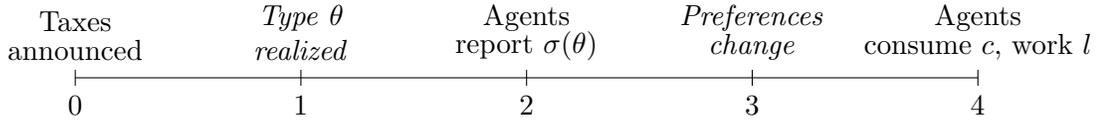


Figure A.1: Timing of Events

The government evaluates allocations at date 0 using the welfare criterion presented in Section 1.2,

$$\sum_{m=1}^M \pi_m [\kappa U(c(\theta_m), l(\theta_m)) + (1 - \kappa)V(c(\theta_m), l(\theta_m))], \quad (\text{A.1})$$

where $(c(\theta_m), l(\theta_m))$ denotes the allocation type m agent consumes, and let $\kappa \in (0, 1]$.

I will continue to assume that the agents do not have access to an insurance market for skill realizations nor does a market for commitment device exists.

The No Private Information Case

Without private information, the government maximizes social welfare (A.1) subject to the feasibility constraint

$$\sum_{m=1}^M \pi_m [F_n(\mathbf{x}_n(\theta_m), l_n(\theta_m); \theta_m) - c_n(\theta_m)] = 0, \forall n \in N. \quad (\text{A.2})$$

As was the case in the previous section, with complete information, the agents work according to their skill type: more productive agents work more than the less productive agents. The government is also able to achieve full insurance without distortions. For consumption smoothing, the government chooses an appropriate linear tax to correct the distortion caused by the preference reversal. The implementation does not change with regards to the agents' sophistication level.

A.1.2 The Effects of Non-sophistication

With non-sophisticated agents, the government issues the following menu

$$\{(c^R(\theta_m), y^R(\theta_m), \mathbf{x}^R(\theta_m)), (c^I(\theta_m), y^I(\theta_m), \mathbf{x}^I(\theta_m))\}_{\theta_m \in \Theta}.$$

Set $y^R(\theta_m) = y^I(\theta_m) = y(\theta_m)$ and $\mathbf{x}^R(\theta_m) = \mathbf{x}^I(\theta_m) = \mathbf{x}(\theta_m)$, since $c^R(\theta) \neq c^I(\theta)$ is enough to exploit non-sophistication. Denote $l(\theta_m) = G(y(\theta_m), \mathbf{x}(\theta_m); \theta_m)$ as the labor supply under truth-telling.

For magnitude naïveté, agents make their reporting decision based on $U(c, l)$ while anticipating a taste change of $W(c, l)$. Therefore, they require $c^I(\theta)$ to be more appealing than $c^R(\theta)$ under $W(c, l)$, and the reporting strategy is evaluated using $U(c, l)$. While for frequency naïveté, agents make their reporting decision based on their expected ex-post utility $W(c, l)$. More specifically, they require $c^I(\theta)$ to be more appealing than $c^R(\theta)$ under $U(c, l)$, and the reporting strategy is evaluated at the expected utility of $U(c, l)$.

Planning Problem with Magnitude Naïveté

Under magnitude naïveté, the government's problem is to choose an allocation menu $\{c^R(\theta_m), c^I(\theta_m), y(\theta_m), \mathbf{x}(\theta_m)\}_{m=1}^M$ to maximize (A.1) subject to the feasibility constraint (A.2) evaluated at the real allocations and

$$U(c^I(\theta_m), l(\theta_m)) \geq U[c^I(\theta_{\hat{m}}), G(y(\theta_{\hat{m}}), \mathbf{x}(\theta_{\hat{m}}); \theta_m)], \forall \theta_m, \theta_{\hat{m}} \in \Theta, \theta_m \neq \theta_{\hat{m}}, \quad (\text{A.3})$$

$$W(c^I(\theta_m), l(\theta_m)) \geq W(c^R(\theta_m), l(\theta_m)), \forall \theta_m \in \Theta, \quad (\text{A.4})$$

$$V(c^R(\theta_m), l(\theta_m)) \geq V(c^I(\theta_m), l(\theta_m)), \forall \theta_m \in \Theta. \quad (\text{A.5})$$

Constraint (A.3) is the incentive compatibility constraint. Constraints (A.4) and (A.5) are the fooling and executability constraints.

Planning Problem with Frequency Naïveté

For frequency naïveté, the government chooses $\{c^R(\theta_m), c^I(\theta_m), y(\theta_m), \mathbf{x}(\theta_m)\}_{m=1}^M$ to maximize (A.1) subject to the feasibility constraint (A.2) evaluated at the real allocations and the fooling constraint (A.5) with the following incentive compatibility constraint $\forall \theta_m, \theta_{\hat{m}} \in \Theta, \theta_m \neq \theta_{\hat{m}}$,

$$\begin{aligned} & \hat{\alpha}U(c^I(\theta_m), l(\theta_m)) + (1 - \hat{\alpha})U(c^R(\theta_m), l(\theta_m)) \\ & \geq \hat{\alpha}U[c^I(\theta_{\hat{m}}), G(y(\theta_{\hat{m}}), \mathbf{x}(\theta_{\hat{m}}); \theta_m)] + (1 - \hat{\alpha})U[c^R(\theta_{\hat{m}}), G(y(\theta_{\hat{m}}), \mathbf{x}(\theta_{\hat{m}}); \theta_m)], \end{aligned} \quad (\text{A.6})$$

and the fooling constraint for the imaginary allocation

$$U(c^I(\theta_m), l(\theta_m)) \geq U(c^R(\theta_m), l(\theta_m)), \forall \theta_m \in \Theta. \quad (\text{A.7})$$

The difference between frequency naïveté and magnitude naïveté lies in the beliefs of the future preference. In frequency naïveté, the agents believe with some probability $\hat{\alpha}$ that they will choose the imaginary allocation evaluated at the ex-ante preference $U(c, l)$, which is represented in (A.7). This is different from the fooling constraint (A.4) for magnitude naïveté, where the agents are certain that their preferences would change, but underestimate the extent of this shift.

Result for Non-sophisticated Agents

By Assumption 1 and Assumption 2, it can be shown that the government can achieve the efficient allocation.

Proposition 11 *The optimal allocation for the environment where agents have private information on productivity and are fully naïve or partially naïve in magnitude or frequency about their preference changes is the efficient allocation.*

Proof Let $\{c^*(\theta_m), y^*(\theta_m), \mathbf{x}^*(\theta_m)\}_{\theta_m \in \Theta}$ denote the set of efficient allocations, and

$$U^*(\theta_m) = U(c^*(\theta_m), l^*(\theta_m)),$$

$$V^*(\theta_m) = V(c^*(\theta_m), l^*(\theta_m)),$$

where $l^*(\theta_m)$ denotes the efficient labor supply when the agents are truthful. Let the real allocation be the efficient allocation. It suffices to show that the efficient allocation can be supported by imaginary allocations that satisfy the fooling and incentive compatibility constraints. I will first prove the case for magnitude naïveté.

The set of incentive compatibility constraints can be rewritten as, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

$$U(c^I(\theta_m), l^*(\theta_m)) \geq U(c^I(\theta_{\hat{m}}), l^*(\theta_{\hat{m}})) + \Delta(\theta_{\hat{m}}; \theta_m),$$

where $\Delta(\theta_{\hat{m}}; \theta_m) \geq 0$ denotes the minimum information rent that prevents a θ_m type from misreporting to be a $\theta_{\hat{m}}$ type. Since labor is additively separable from consumption in utility and output is fixed at the efficient level, $\Delta(\theta_{\hat{m}}; \theta_m)$ is only a function of types θ_m and $\theta_{\hat{m}}$. Since there exists $\epsilon > 0$ such that $\frac{\partial U}{\partial c_n} > \epsilon$ for all goods, it is possible to find $c^I(\theta_m)$ such that the incentive compatibility constraint is satisfied for any $\Delta(\theta_{\hat{m}}; \theta_m)$. By Assumption 2, since the efficient labor supply is strictly increasing with productivity, the ex-ante utility evaluated at the imaginary allocations also need to be strictly increasing,

which can be constructed. This implies that it is sufficient to focus on local downward incentive compatibility constraints.

Let $c^I(\theta_m)$ be chosen so that $\forall \theta_m \in \Theta$,

$$V(c^I(\theta_m), l^*(\theta_m)) = V^*(\theta_m). \quad (\text{A.8})$$

Hence, for the magnitude naïveté case, if the incentive compatibility constraints are satisfied, then the fooling constraints are immediately satisfied as well.

I will now show that for any given $c^I(\theta_{m-1})$ and for an arbitrary $\Delta(\theta_{m-1}; \theta_m) \in \mathbb{R}_+$, it is possible for the efficient allocations to satisfy the incentive compatibility constraints. This is akin to showing that there does not exist a finite solution to the following programming problem

$$\max_{c^I(\theta_m)} U(c^I(\theta_m), l^*(\theta_m))$$

subject to (A.8). The proof will proceed by contradiction.

Suppose there exists a finite $\hat{c}^I(\theta_m)$ such that it solves the programming problem above. Since the utility functions are unbounded below and $\hat{c}^I(\theta_m)$ is finite, it must be the case that $\hat{c}^I(\theta_m) \in \mathbb{R}_{++}^N$. By Assumption 1 and without loss of generality, there exists $j, k \in N$ such that

$$\frac{\frac{\partial U}{\partial \hat{c}_k^I(\theta_m)}}{\frac{\partial U}{\partial \hat{c}_j^I(\theta_m)}} > \frac{\frac{\partial V}{\partial \hat{c}_k^I(\theta_m)}}{\frac{\partial V}{\partial \hat{c}_j^I(\theta_m)}}.$$

For any $\epsilon > 0$, choose a new imaginary consumption allocation $\hat{c}^{I'}(\theta_m)$, where

$$\hat{c}_k^{I'}(\theta_m) = \hat{c}_k^I(\theta_m) + \epsilon$$

and

$$\hat{c}_j^{I'}(\theta_m) = \hat{c}_j^I(\theta_m) - \frac{\frac{\partial V}{\partial \hat{c}_k^I(\theta_m)} \epsilon}{\frac{\partial V}{\partial \hat{c}_j^I(\theta_m)}}.$$

This leaves the ex-post utility unchanged: $V(c^{I'}(\theta_m), l^*(\theta_m)) = V^*(\theta_m)$. However, this new imaginary allocation strictly increases the ex-ante utility:

$$U(c^{I'}(\theta_m), l^*(\theta_m)) \approx U(c^I(\theta_m), l^*(\theta_m)) + \frac{\partial U}{\partial \hat{c}_k^I(\theta_m)} \epsilon - \frac{\partial U}{\partial \hat{c}_j^I(\theta_m)} \left(\frac{\frac{\partial V}{\partial \hat{c}_k^I(\theta_m)} \epsilon}{\frac{\partial V}{\partial \hat{c}_j^I(\theta_m)}} \right),$$

which due to preference reversal,

$$\frac{\partial U}{\partial \hat{c}_k^I(\theta_m)} \epsilon - \frac{\partial U}{\partial \hat{c}_j^I(\theta_m)} \left(\frac{\frac{\partial V}{\partial \hat{c}_k^I(\theta_m)} \epsilon}{\frac{\partial V}{\partial \hat{c}_j^I(\theta_m)}} \right) > 0.$$

This is a contradiction. Therefore, there does not exist a finite solution to the programming problem above. As a result, the efficient allocation is implementable for magnitude naïveté regardless of the amount of information rent $\Delta(\theta_{m-1}; \theta_m)$.

It is now straightforward to prove for the case of frequency naïveté. Again, choose the imaginary consumption such that $V(c^I(\theta_m), l^*(\theta_m)) = V^*(\theta_m)$. By Assumption 2, the labor supply is strictly increasing with productivity, so the ex-ante utility evaluated at the imaginary consumption can be constructed to be strictly increasing as well. Thus, it suffices to focus on local downward incentive compatibility constraints, which can be rewritten as $\forall \theta_m \in \Theta \setminus \theta_1$,

$$\hat{\alpha}U(c^I(\theta_m), l^*(\theta_m)) + (1 - \hat{\alpha})U(c^*(\theta_m), l^*(\theta_m)) \geq U(c^*(\theta_m), l^*(\theta_m)) + \Delta(\theta_{m-1}; \theta_m).$$

Notice that if it is possible to choose the imaginary consumption so that the incentive compatibility constraints are satisfied, then it immediately follows that the fooling constraints are satisfied. Hence, it suffices to show that the same programming problem shown above does not have any finite solutions. This completes the proof. ■

Proposition 11 states that the private information problem can be alleviated if the agents are not sophisticated regardless of the type of naïveté. This is accomplished by using a fooling mechanism.

Corollary 5 *If $\hat{\alpha} > 0$, it is optimal for the government to implement a fooling mechanism.*

Proof By Proposition 11, the government can implement the efficient allocation. Let $\{(c^R(\theta_m), y(\theta_m)), \mathbf{x}(\theta_m)\}_{\theta_m \in \Theta}$ be the efficient allocation. Suppose the government does not fool the agents, then by definition, for all productivity types, agents evaluate the incentive compatibility constraints at the real allocation. This violates incentive compatibility if the real allocations are the efficient allocations. It follows that for some types, the agents must use the imaginary allocations to evaluate incentive compatibility and thus the government must implement a fooling mechanism to achieve the efficient allocation. ■

To exploit the agents' naïveté, the government loads the information rent on goods that they value during the reporting stage, but would not value as much relative to other goods after the preference change. By Assumption 1, suppose $\frac{\partial U}{\partial c_k} / \frac{\partial U}{\partial c_j} > \frac{\partial V}{\partial c_k} / \frac{\partial V}{\partial c_j}$, then the agents

value good k more than good j at the reporting stage. The government can then promise more of good k than good j for the imaginary allocations to elicit truthful reports as long as the agents hold the wrong beliefs. However, after the preference change, the promise of more good k is less appealing, and the agents would no longer choose the imaginary allocations but the real allocations, with less of good k .

With magnitude naïveté, the government is able to achieve full information efficient welfare for any sophistication level $\hat{\alpha} \in (0, 1]$. However, with fully sophisticated agents ($\hat{\alpha} = 0$), the fooling mechanism can only implement the constrained efficient optimum. Similar discontinuities in welfare can also be found in analysis involving magnitude naïveté, as in Heidhues and Koszegi (2010). A more surprising result for partial naïveté is that this discontinuity is present for naïveté in frequency as well. Spiegler (2011) has shown the optimal contract to be continuous with respect to cognitive limitations in a second-degree price discrimination setting for frequency naïveté. However, Proposition 11 shows that this continuity result does not hold in the optimal taxation setting even with naïveté in frequency.

A.1.3 The Effects of Sophistication

The government introduces the following menu for sophisticated agents

$$\{(c^R(\theta_m), y^R(\theta_m), \mathbf{x}^R(\theta_m)), (c^T(\theta_m), y^T(\theta_m), \mathbf{x}^T(\theta_m))\}_{\theta_m \in \Theta}.$$

Denote $l^R(\theta_m) = G(y^R(\theta_m), \mathbf{x}^R(\theta_m); \theta_m)$ and $l^T(\theta_m) = G(y^T(\theta_m), \mathbf{x}^T(\theta_m); \theta_m)$ as the labor supply under truth-telling for the real and threat allocations respectively.

The government maximizes (A.1) subject to the feasibility constraint (A.2) evaluated at the real allocations and the following incentive compatibility constraints, $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

$$U(c^R(\theta_m), l^R(\theta_m)) \geq U[c^T(\theta_{\hat{m}}), G(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m)], \quad (\text{A.9})$$

and the credible threat constraints $\forall \theta_m, \theta_{\hat{m}} \in \Theta, \theta_m \neq \theta_{\hat{m}}$,

$$V(c^R(\theta_m), l^R(\theta_m)) \geq V(c^T(\theta_m), l^T(\theta_m)), \quad (\text{A.10})$$

$$V[c^T(\theta_{\hat{m}}), G(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m)] \geq V[c^R(\theta_{\hat{m}}), G(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m)]. \quad (\text{A.11})$$

Result for Sophisticated Agents

Using Assumption 1 and Assumption 2, the following proposition shows that private information does not matter in an environment with sophisticated agents.

Proposition 12 *The optimal allocation for the environment where agents have private information on productivity and are sophisticated is the efficient allocation.*

Proof Let $\{c^*(\theta_m), y^*(\theta_m), \mathbf{x}^*(\theta_m)\}_{\theta_m \in \Theta}$ denote the set of efficient allocations. Let the real allocations be the efficient allocations. Since the utility functions are additively separable in consumption and labor, let

$$V(c, l) = v(c) - h(l),$$

and

$$U(c, l) = u(c) - h(l).$$

I will first deter local downward misreporting for type θ_2 , and then show how global downward misreports can be prevented by deterring local downward misreports. Note that type θ_1 does not need to be deterred from misreporting when the real allocation is the efficient allocation. Choose the threat allocation so that the following holds

$$v(c^T(\theta_1)) - h[G(y^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2)] = v(c^*(\theta_1)) - h[G(y^*(\theta_1), \mathbf{x}^*(\theta_1); \theta_2)]. \quad (\text{A.12})$$

By Assumption 2, labor supply for the threat allocation can be increased with a corresponding increase in utility from threat consumption such that (A.12) holds and productivity type θ_1 agent would not choose the threat allocation. To see this, notice that the following problem has no finite solution (the government is choosing a utility level $v(c^T(\theta_1))$, not the consumption bundle $c^T(\theta_1)$)

$$\min_{v(c^T(\theta_1)), l^T(\theta_1)} v(c^T(\theta_1)) - h(l^T(\theta_1)) \quad \text{s.t. (A.12)}.$$

Hence, it is possible to find $(v(c^T(\theta_1)), l^T(\theta_1))$ so that, along with (A.12), the following holds

$$v(c^*(\theta_1)) - h(l^*(\theta_1)) \geq v(c^T(\theta_1)) - h(l^T(\theta_1)).$$

Thus, by construction, the credible threat constraints are satisfied with $v(c^T(\theta_1)) > v(c^*(\theta_1))$ and $l^T(\theta_1) > l^*(\theta_1)$.

Next, fix the utility derived from the threat allocation $v(c^T(\theta_1))$ and inputs $l^T(\theta_1)$ and $\mathbf{x}^T(\theta_1)$ such that all credible threat constraints are satisfied. Hence, all that remains is to show the threat consumption can be chosen to satisfy the incentive compatibility constraint. It is always possible to find such a threat consumption if the following programming problem yields no finite solutions for any given $\Omega(\theta_1; \theta_2) > 0$,

$$\min_{c^T(\theta_1)} u(c^T(\theta_1))$$

subject to

$$v(c^T(\theta_1)) = v(c^*(\theta_1)) + \Omega(\theta_1; \theta_2),$$

where the constraint is a rewriting of (A.12), in essence, $\Omega(\theta_1; \theta_2) = h[G(y^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2)] - h[G(y^*(\theta_1), \mathbf{x}^*(\theta_1); \theta_2)]$.

By Assumption 1, it can be shown that the programming problem does not have a finite solution using the arguments established for the proof of Proposition 11. Hence, it is possible to choose a set of threat allocation in the menu for type θ_1 such that local downward incentive constraints and credible threat constraints hold.

By the convexity of $h(\cdot)$ and the fact that $l^T(\theta_1) > l^*(\theta_1)$, for any $\theta_m > \theta_2$ the following relationship holds: $\Omega(\theta_1; \theta_2) > \Omega(\theta_1, \theta_m)$. Hence, more productive agents misreporting as a θ_1 agent would also choose the threat allocation. As a result, if $c^T(\theta_1)$ is chosen so that

$$u(c^T(\theta_1)) \leq \min_{\theta_m > \theta_1} [u(c^*(\theta_m)) + h[G(y^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_m)] - h(l^*(\theta_m))],$$

then all types are deterred from misreporting as θ_1 . Hence, to deter type θ_3 from local downward misreporting, a similar process can be used to find the threat allocation. Choose the threat allocation $(c^T(\theta_2), y^T(\theta_2), \mathbf{x}^T(\theta_2))$ so that the following holds

$$\begin{aligned} & v(c^T(\theta_2)) - h[G(y^T(\theta_2), \mathbf{x}^T(\theta_2); \theta_3)] \\ & = \max \{v(c^*(\theta_2)) - h[G(y^*(\theta_2), \mathbf{x}^*(\theta_2); \theta_3)], v(c^T(\theta_1)) - h[G(y^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_3)]\}. \end{aligned}$$

Using a similar argument, the threat allocation can be chosen so that the credible threat constraints hold and the incentive compatibility constraints for all types are satisfied.

Finally, by induction, it is possible to achieve global incentive compatibility with credible threats. Hence, the efficient allocation is implementable. ■

The efficient allocation is achieved by using a threat mechanism for sophisticated agents. Since as long as agents are not fully naïve, they are either fully or partially aware of their time-inconsistency problem ($\hat{\alpha} < 1$) and a threat mechanism can potentially be utilized.

For magnitude naïveté, the threats are evaluated using the erroneous ex-post utility W , so the credible threat constraints $\forall \theta_m, \theta_{\hat{m}} \in \Theta, \theta_m \neq \theta_{\hat{m}}$ are

$$W(c^R(\theta_m), l^R(\theta_m)) \geq W(c^T(\theta_m), l^T(\theta_m)),$$

$$W[c^T(\theta_{\hat{m}}), G(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m)] \geq W[c^R(\theta_{\hat{m}}), G(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m)],$$

and the executability constraint

$$V(c^R(\theta_m), l^R(\theta_m)) \geq V(c^T(\theta_m), l^T(\theta_m)), \forall \theta_m \in \Theta. \quad (\text{A.13})$$

For frequency naïveté, the credible threat constraints remain the same.¹ However, the incentive compatibility constraints $\forall \theta_m, \theta_{\hat{m}} \in \Theta, \theta_m \neq \theta_{\hat{m}}$ are

$$U(c^R(\theta_m), l^R(\theta_m)) \geq \hat{\alpha} U[c^R(\theta_{\hat{m}}), G(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m)] \\ + (1 - \hat{\alpha}) U[c^T(\theta_{\hat{m}}), G(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m)].$$

The threat allocation makes sure that the partially naïve agent does not think misreporting is worth the risk of consuming the threat allocation. The following corollary shows that a threat mechanism works on partially naïve agents.

Corollary 6 *It is optimal to threaten the agents when $\hat{\alpha} < 1$.*

Proof Let $\{c^*(\theta_m), y^*(\theta_m), \mathbf{x}^*(\theta_m)\}_{\theta_m \in \Theta}$ denote the set of efficient allocations, and set the real allocations to be the efficient allocations. Set both the threat and real inputs be at the efficient level. Also, let $l^*(\theta_m)$ and $l^T(\theta_m)$ denote the efficient labor supply under truth-telling for real and threat allocations respectively. I will first show the result for magnitude naïveté.

First, examine downward misreporting for type θ_2 . Choose the threat allocation so that the following holds

$$\min_{\theta_m} \{u(c^*(\theta_m)) - h(l^*(\theta_m))\} = u(c^T(\theta_1)) - h[G(y^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2)]. \quad (\text{A.14})$$

¹I require that a truthful agent would prefer the real allocation regardless of which utility it is evaluated at, while a misreporting agent would prefer the threat allocation regardless of the utility it is evaluated at.

Notice that the following programming problem has no finite solution:

$$\min_{c^T(\theta_1), y^T(\theta_1)} v(c^T(\theta_1)) - h[G(\hat{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_1)]$$

subject to

$$\begin{aligned} \hat{\alpha}u(c^T(\theta_1)) + (1 - \hat{\alpha})v(c^T(\theta_1)) - h[G(y^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2)] & \quad (\text{A.15}) \\ \geq \hat{\alpha}u(c^*(\theta_1)) + (1 - \hat{\alpha})v(c^*(\theta_1)) - h[G(y^*(\theta_1), \mathbf{x}^*(\theta_1); \theta_2)]. \end{aligned}$$

and (A.14). To see why this is the case, suppose there exists a finite solution $(\hat{c}^T(\theta_1), \hat{y}^T(\theta_1))$, then (A.15) can be rewritten as

$$v(\hat{c}^T(\theta_1)) - h[G(\hat{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2)] = \Psi(\theta_1, \theta_2),$$

where Ψ only depends on the efficient allocation for types θ_1 and θ_2 and the sophistication level $\hat{\alpha}$. This implies that

$$\hat{y}^T(\theta_1) = \arg \min h[G(\hat{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2)] - h[G(\hat{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_1)] \quad \text{s.t. (A.14)},$$

which is a contradiction.² Hence, it is possible to choose the threat allocations to satisfy the executability, local downward incentive compatibility and credible threat constraints.

Once it is established that type θ_2 is deterred from choosing $(c^*(\theta_1), y^*(\theta_1), \mathbf{x}^*(\theta_1))$, higher types can be deterred. This can be done by choosing $c^T(\theta_1)$ so that the ex-ante utility is sufficiently low while the ex-post utility is sufficiently high, and then adjusting the output so the credible threat and executability constraints hold. Using a similar argument above, by induction, the efficient allocation is implementable using a threat mechanism for magnitude naïveté.

Now, I will analyze the case for frequency naïveté. For frequency naïveté, all constraints are the same as in the magnitude naïveté case, except for the incentive compatibility constraints. However, since it is possible to arbitrarily decrease the ex-ante utility evaluated at the threat allocation, the argument for the magnitude naïveté holds for frequency naïveté as well. This completes the proof. ■

²Since increasing the threat output such that $h[G(\tilde{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2)] = h[G(\hat{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2)] + \epsilon$, would further decrease $h[G(\hat{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_2)] - h[G(\hat{y}^T(\theta_1), \mathbf{x}^T(\theta_1); \theta_1)]$. By Assumption 1, it is possible to choose $\tilde{c}^T(\theta_1)$ such that (A.14) holds and $v(\tilde{c}^T(\theta_1))$ satisfies (A.14).

A.1.4 Model with Diversely Naïve Agents

When agents are heterogeneous in productivity and sophistication level, the government is still capable of achieving the efficient allocation. The following lemma demonstrates a monotonicity result for fooling and threat mechanisms similar to the one in Section 1.5.

Lemma 6 *A fooling mechanism that is effective for agents of sophistication level $\acute{\alpha}$ is also effective for more naïve agents. A threat mechanism that is effective for agents of sophistication level $\grave{\alpha}$ is also effective for more sophisticated agents.*

Proof First, look at fooling mechanisms for magnitude naïveté. If a fooling mechanism

$$\{(c^R(\theta_m), y(\theta_m), \mathbf{x}(\theta_m)), (c^I(\theta_m), y(\theta_m), \mathbf{x}(\theta_m))\}_{\theta_m \in \Theta}$$

is effective for an agent with sophistication level $\acute{\alpha}$ of magnitude naïveté, then the following fooling constraint must be satisfied

$$\begin{aligned} & \acute{\alpha}U(c^I(\theta_m), l(\theta_m)) + (1 - \acute{\alpha})V(c^I(\theta_m), l(\theta_m)) \\ & \geq \acute{\alpha}U(c^R(\theta_m), l(\theta_m)) + (1 - \acute{\alpha})V(c^R(\theta_m), l(\theta_m)), \end{aligned}$$

where $l(\theta)$ denotes the efficient labor supply when the agent is truthful. Since the executability constraint is also satisfied, for any $\hat{\alpha} > \acute{\alpha}$, the fooling constraint is relaxed and all other constraints are unchanged. Hence, a fooling mechanism that is effective for $\acute{\alpha}$ is effective for more naïve agents.

Next, fooling mechanisms for frequency naïveté also have the same form as the mechanism above. Suppose the fooling mechanism is effective for an agent with sophistication level $\acute{\alpha}$ of frequency naïveté, then the following incentive compatibility constraint must be satisfied

$$\begin{aligned} & \acute{\alpha}U(c^I(\theta_m), l(\theta_m)) + (1 - \acute{\alpha})U(c^R(\theta_m), l(\theta_m)) \\ & \geq \acute{\alpha}U[c^I(\theta_{\hat{m}}), G(y(\theta_{\hat{m}}), \mathbf{x}(\theta_{\hat{m}}); \theta_m)] + (1 - \acute{\alpha})U[c^R(\theta_{\hat{m}}), G(y(\theta_{\hat{m}}), \mathbf{x}(\theta_{\hat{m}}); \theta_m)]. \end{aligned}$$

Due to separability between consumption and effort, the utility function can be written as

$$U(c, l) = u(c) - h(l),$$

so the incentive compatibility constraint becomes

$$\begin{aligned} & \acute{\alpha} [u(c^I(\theta_m)) - u(c^I(\theta_{\hat{m}}))] + (1 - \acute{\alpha}) [u(c^R(\theta_m)) - u(c^R(\theta_{\hat{m}}))] \\ & \geq h(l(\theta_m)) - h[G(y(\theta_{\hat{m}}), \mathbf{x}(\theta_{\hat{m}}); \theta_m)]. \end{aligned}$$

Since the social welfare function is strictly concave, the efficient allocation would provide full insurance: $c^R(\theta_m) = c^R(\theta_{\hat{m}})$ for all $\theta_m \neq \theta_{\hat{m}}$. As a result, for any $\theta_m, \theta_{\hat{m}} \in \Theta$ such that $\theta_m > \theta_{\hat{m}}$ it must be that $c^I(\theta_m) > c^I(\theta_{\hat{m}})$ for the incentive constraint to hold. The incentive compatibility constraint holds trivially for $\theta_m < \theta_{\hat{m}}$ of any sophistication level. It is easy to see that for any agent with $\hat{\alpha} > \acute{\alpha}$, the incentive compatibility constraint holds as well. Since all other constraints are invariant to the sophistication level, a fooling mechanism that is effective for sophistication level $\acute{\alpha}$ is effective for more naïve agents.

Finally, for threat mechanisms, the analysis will begin by analyzing frequency naïveté. If a threat mechanism

$$\{(c^R(\theta_m), y^R(\theta_m), \mathbf{x}^R(\theta_m)), (c^T(\theta_m), y^T(\theta_m), \mathbf{x}^T(\theta_m))\}_{\theta_m \in \Theta}$$

is effective for an agent with sophistication level $\grave{\alpha}$ of frequency naïveté, then the incentive compatibility constraint must be satisfied

$$\begin{aligned} U(c^R(\theta_m), l^R(\theta_m)) & \geq \grave{\alpha} U[c^R(\theta_{\hat{m}}), G(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m)] \\ & + (1 - \grave{\alpha}) U[c^T(\theta_{\hat{m}}), G(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m)], \end{aligned}$$

where $l^R(\theta)$ denotes the efficient labor supply under the real allocation when the agent is truthful. Note that for $\theta_m > \theta_{\hat{m}}$ to implement the efficient allocation, it must be that

$$U[c^R(\theta_{\hat{m}}), G(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m)] > U[c^T(\theta_{\hat{m}}), G(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m)].$$

If not, then the incentive compatibility constraint would not hold.³ As a result, for any agent with $\hat{\alpha} < \grave{\alpha}$, the incentive compatibility constraint would hold as well. The incentive compatibility constraint would hold trivially for $\theta_m < \theta_{\hat{m}}$ of any sophistication level. The other constraints do not depend on $\hat{\alpha}$, so a threat mechanism that is effective for sophistication level $\grave{\alpha}$ would also be effective for more sophisticated agents.

A threat mechanism for magnitude naïveté agents with sophistication level $\grave{\alpha}$ would also have the same policy menu as the threat mechanism for frequency naïveté. The credible

³This is because $U(c^R(\theta_m), l^R(\theta_m)) < U[c^R(\theta_{\hat{m}}), G(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m)]$ at the efficient allocation.

threat constraints are

$$\begin{aligned} & \hat{\alpha}U(c^R(\theta_m), l^R(\theta_m)) + (1 - \hat{\alpha})V(c^R(\theta_m), l^R(\theta_m)) \\ & \geq \hat{\alpha}U(c^T(\theta_m), l^T(\theta_m)) + (1 - \hat{\alpha})V(c^T(\theta_m), l^T(\theta_m)), \end{aligned}$$

$$\begin{aligned} & \hat{\alpha}U[c^T(\theta_{\hat{m}}), G(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m)] + (1 - \hat{\alpha})V[c^T(\theta_{\hat{m}}), G(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m)] \\ & \geq \hat{\alpha}U[c^R(\theta_{\hat{m}}), G(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m)] + (1 - \hat{\alpha})V[c^R(\theta_{\hat{m}}), G(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m)], \end{aligned}$$

where $l^T(\theta)$ denotes the efficient labor supply under the threat allocation when the agent is truthful. The executability constraint ensures that for any $\hat{\alpha} < \hat{\alpha}$, a truth-telling agent would always prefer the real allocations over the threat allocations. Separability between consumption and effort helps check that the other credible threat constraint (a misreporting agent would always choose the threat allocation over the real allocation) works for more sophisticated agents as well. Let

$$V(c, l) = v(c) - h(l),$$

so the credible threat constraint can be rewritten as

$$\begin{aligned} & \hat{\alpha}[u(c^R(\theta_{\hat{m}})) - u(c^T(\theta_{\hat{m}}))] + (1 - \hat{\alpha})[v(c^R(\theta_{\hat{m}})) - v(c^T(\theta_{\hat{m}}))] \\ & \leq h[G(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m)] - h[G(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m)]. \end{aligned}$$

Notice that the incentive compatibility implies

$$u(c^R(\theta_m)) - u(c^T(\theta_{\hat{m}})) \geq h(l^R(\theta_m)) - h[G(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m)].$$

The mechanism is effective which implies that it has full insurance, $c^R(\theta_m) = c^R(\theta_{\hat{m}})$, and for any $\theta_m > \theta_{\hat{m}}$,

$$h(l^R(\theta_m)) > h[G(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m)].$$

Hence, $u(c^R(\theta_m)) - u(c^T(\theta_{\hat{m}})) > h[G(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m)] - h[G(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m)]$, so it must be the case that

$$v(c^R(\theta_{\hat{m}})) - v(c^T(\theta_{\hat{m}})) < h[G(y^R(\theta_{\hat{m}}), \mathbf{x}^R(\theta_{\hat{m}}); \theta_m)] - h[G(y^T(\theta_{\hat{m}}), \mathbf{x}^T(\theta_{\hat{m}}); \theta_m)].$$

As a result, for any agent with $\hat{\alpha} < \hat{\alpha}$, a misreporting agent would also choose the threat allocation over the real allocation. This proves that a threat mechanism that is effective for

sophistication level $\hat{\alpha}$ is also effective for more sophisticated agents. ■

With Lemma 6, the government can target sophistication level, $\tilde{\alpha}$, such that all agents who are more sophisticated than $\tilde{\alpha}$ are threatened by using the same threat mechanism and those who are more naïve are fooled by using the same fooling mechanism. The government targets sophistication level $\tilde{\alpha} \in (0, 1)$ and introduce the following hybrid mechanism

$$\{(c^R(\theta_m), y^R(\theta_m), \mathbf{x}^R(\theta_m)), (c^I(\theta_m), y^I(\theta_m), \mathbf{x}^I(\theta_m)), (\mathbf{c}^T(\theta_m), \mathbf{y}^T(\theta_m), \mathbf{x}^T(\theta_m))\}_{\theta_m \in \Theta}.$$

The imaginary and threat allocations are chosen such that agents with sophistication level $\tilde{\alpha}$ are fooled and threatened with effective mechanisms. It can be shown that the fooling and the threat components do not interact in a hybrid mechanism and full efficiency is achievable.

Proposition 13 *The optimal allocation for the environment with diversely naïve agents where agents have private information on productivity and sophistication level is the efficient allocation.*

Proof The result follows immediately from Lemma 6. ■

Model with Diversely Time-Inconsistent Agents

To model agents with varying degrees of time-inconsistency, let $\alpha \in [\underline{\alpha}, \bar{\alpha}] \subset [0, 1]$ denote the degree of preference reversal (temptation level) of the agents, where $\underline{\alpha} = 0$ and $\bar{\alpha} < 1$.

Definition 14 *An agent with preference reversal $\alpha \in [\underline{\alpha}, \bar{\alpha}]$ has ex-post utility*

$$V(c, l; \alpha) = \alpha U(c, l) + (1 - \alpha) \underline{V}(c, l),$$

where $\underline{V}(c, l) \neq U(c, l)$ is the ex-post utility of the agents with the largest degree of preference reversal.

Since $\bar{\alpha} < 1$, all agents in the economy are time-inconsistent. In this setting, the definitions of magnitude or frequency naiveté follow. In other words, a partially naïve agents has misperception $\hat{\alpha}$ in magnitude or frequency within the support $(\alpha, 1]$. If $\hat{\alpha} = \alpha$, then the agents are sophisticated. The following proposition shows that a hybrid mechanism can implement full efficiency in an economy with diversely time-inconsistent and naïve agents.

Proposition 14 *The optimal allocation for the environment with diversely time-inconsistent and naïve agents where agents have private information on productivity and sophistication level is the efficient allocation.*

Proof I will first show how the mechanism can be constructed for magnitude naiveté. First, choose a target $\tilde{\alpha} \in (\bar{\alpha}, 1)$, and construct a fooling mechanism that implements the full information efficient optimum for $\hat{\alpha} = \tilde{\alpha}$ and $\alpha = \bar{\alpha}$. By Proposition 11, this can be done. The executability and fooling constraints are satisfied $\forall \theta_m \in \Theta$,

$$\begin{aligned} V(c^*(\theta_m), l^*(\theta_m); \bar{\alpha}) &\geq V(c^I(\theta_m), l^*(\theta_m); \bar{\alpha}), \\ \tilde{\alpha}U(c^I(\theta_m), l^*(\theta_m)) + (1 - \tilde{\alpha})\underline{V}(c^I(\theta_m), l^*(\theta_m)) \\ &\geq \tilde{\alpha}U(c^*(\theta_m), l^*(\theta_m)) + (1 - \tilde{\alpha})\underline{V}(c^*(\theta_m), l^*(\theta_m)). \end{aligned}$$

The executability constraint can be rewritten as

$$\begin{aligned} \bar{\alpha}U(c^*(\theta_m), l^*(\theta_m)) + (1 - \bar{\alpha})\underline{V}(c^*(\theta_m), l^*(\theta_m)) \\ \geq \bar{\alpha}U(c^I(\theta_m), l^*(\theta_m)) + (1 - \bar{\alpha})\underline{V}(c^I(\theta_m), l^*(\theta_m)). \end{aligned}$$

The executability and fooling constraints imply the following:

$$U(c^I(\theta_m), l^*(\theta_m)) - U(c^*(\theta_m), l^*(\theta_m)) > \underline{V}(c^I(\theta_m), l^*(\theta_m)) - \underline{V}(c^*(\theta_m), l^*(\theta_m)). \quad (\text{A.16})$$

I will now show that the efficient allocation is implementable for all agents more naïve ($\hat{\alpha} > \tilde{\alpha}$) and with larger degrees of preference reversals ($\alpha < \bar{\alpha}$). To see this, first notice that by Lemma 6, all agents with $\alpha = \bar{\alpha}$ and $\hat{\alpha} > \tilde{\alpha}$ would be fooled into choosing the efficient allocation in equilibrium. Notice for agents with $\alpha < \bar{\alpha}$, the executability constraint is

$$\begin{aligned} \alpha U(c^*(\theta_m), l^*(\theta_m)) + (1 - \alpha)\underline{V}(c^*(\theta_m), l^*(\theta_m)) \\ \geq \alpha U(c^I(\theta_m), l^*(\theta_m)) + (1 - \alpha)\underline{V}(c^I(\theta_m), l^*(\theta_m)), \end{aligned}$$

which holds trivially by (A.16). Since the fooling constraints do not depend on α and the incentive compatibility constraints do not depend on α and $\hat{\alpha}$, a fooling mechanism for $\alpha = \bar{\alpha}$ and $\hat{\alpha} = \tilde{\alpha}$ is able to deceive all agents in the economy more naïve than $\tilde{\alpha}$.

Next, consider a threat mechanism that implements the full information efficient optimum for $\hat{\alpha} = \tilde{\alpha} \in (\bar{\alpha}, 1)$ and $\alpha = \underline{\alpha}$. This can be accomplished by Corollary 6. I will

demonstrate that this mechanism achieves the full information optimum for more sophisticated agents of any degree of preference reversal.

By Lemma 6, the efficient allocation is implementable for agents with $\alpha = \underline{\alpha}$ and $\hat{\alpha} < \tilde{\alpha}$. Similar to the argument for the fooling mechanism, only the executability constraint depends on α . However, for magnitude naiveté, if the credible threat constraints hold for all $\hat{\alpha} \in [\underline{\alpha}, \tilde{\alpha}]$, then the executability constraint would be satisfied for any $\alpha \in [\underline{\alpha}, \bar{\alpha}]$. This is because one of the credible threat constraints is $\forall \hat{\alpha} \in [\underline{\alpha}, \tilde{\alpha}]$,

$$\begin{aligned} & \hat{\alpha}U(c^*(\theta_m), l^*(\theta_m)) + (1 - \hat{\alpha})\underline{V}(c^*(\theta_m), l^*(\theta_m)) \\ & \geq \hat{\alpha}U(c^T(\theta_m), l^T(\theta_m)) + (1 - \hat{\alpha})\underline{V}(c^T(\theta_m), l^T(\theta_m)). \end{aligned}$$

Hence, the set allocations satisfying it is a superset for the set of allocations satisfying the executability constraint for any $\alpha \in [\underline{\alpha}, \bar{\alpha}]$. Therefore, the threat mechanism for agents with $(\hat{\alpha}, \alpha) = (\tilde{\alpha}, \underline{\alpha})$ would be able to implement the efficient allocation for more sophisticated agents of any degree of preference reversal. Furthermore, the relatively sophisticated agents would weakly prefer the threat mechanism over the fooling mechanism, and the relatively naïve would strictly prefer the fooling mechanism over the threat mechanism. This completes the proof for magnitude naiveté.

I will now prove the case for frequency naiveté. First, target a sophistication level $\tilde{\alpha} \in (\bar{\alpha}, 1)$ and degree of preference reversal $\alpha = \bar{\alpha}$ with an effective fooling mechanism, which can be done by Proposition 11. By Lemma 6, the efficient allocation is implemented for more naïve agents and $\alpha = \bar{\alpha}$. Hence, the incentive compatibility and fooling constraints would also hold for all agents with $\hat{\alpha} > \tilde{\alpha}$ and $\alpha \in [\underline{\alpha}, \bar{\alpha}]$. Only the executability constraint varies with α . By construction, both the executability and fooling constraints hold at $\alpha = \bar{\alpha}$, for any $\theta_m \in \Theta$,

$$\begin{aligned} & \bar{\alpha}U(c^*(\theta_m), l^*(\theta_m)) + (1 - \bar{\alpha})\underline{V}(c^*(\theta_m), l^*(\theta_m)) \\ & \geq \bar{\alpha}U(c^I(\theta_m), l^*(\theta_m)) + (1 - \bar{\alpha})\underline{V}(c^I(\theta_m), l^*(\theta_m)), \end{aligned}$$

and $U(c^I(\theta_m), l^*(\theta_m)) \geq U(c^*(\theta_m), l^*(\theta_m))$. This implies that $\underline{V}(c^*(\theta_m), l^*(\theta_m)) \geq \underline{V}(c^I(\theta_m), l^*(\theta_m))$. Hence, the executability constraints for agents with $\alpha < \bar{\alpha}$ would also hold trivially. Therefore, the efficient allocation can be implemented for all agents more naïve than $\tilde{\alpha}$ regardless of their degree of preference reversal.

Finally, construct an effective threat mechanism for agents with the same sophistication level $\tilde{\alpha}$ and degree of preference reversal $\alpha = \underline{\alpha}$, which can be done by Corollary 6. By

Lemma 6, agents with $\hat{\alpha} < \tilde{\alpha}$ and $\alpha = \underline{\alpha}$ would be threatened into choosing the efficient allocation. Notice the only difference between magnitude and frequency naiveté for a threat mechanism are the incentive compatibility constraints, which do not depend on α . Hence, the same argument used to show that the executability constraint holds for all degrees of preference reversal in the magnitude naiveté case also works for frequency naiveté. As a result, the efficient allocation can be implemented for all agents more sophisticated than $\tilde{\alpha}$ regardless of the degree of preference reversal. This completes the proof. ■

A.2 Proofs for Section 1.6 and Beyond

Proof of Theorem 5:

Notice that the deterrent and threat allocations can deter TI agents of any productivity type from misreporting. Hence, only the incentive compatibility constraints for the TC agents need to be considered. The proof will focus on a relaxed problem, where for within consistency deviations, only the local downward incentive compatibility constraints are considered along with a monotonicity constraints on $y^P : y_m^P \leq y_{m+1}^P$ for all $\theta_m \in \Theta$.

The incentive compatibility constraints are non-smooth. The problem can be reformulated by introducing a vector of new variables $x = (x_2, \dots, x_M)$ so the constraints are smooth. In essence, the incentive compatibility constraints are rewritten as $\forall \theta_m, \theta_{\hat{m}} \in \Theta$

$$u(c_m^P) - h\left(\frac{y_m^P}{\theta_m}\right) + w(k_m^P) \geq x_m \quad (\text{A.17})$$

$$x_m - u(c_{m-1}^P) - h\left(\frac{y_{m-1}^P}{\theta_m}\right) + w(k_{m-1}^P) \geq 0 \quad (\text{A.18})$$

$$x_m - u(c_{\hat{m}}^R) - h\left(\frac{y_{\hat{m}}^R}{\theta_m}\right) + w(k_{\hat{m}}^R) \geq 0. \quad (\text{A.19})$$

Let λ_m denote the Lagrange multiplier associated with inequality (A.17). Let λ_m^{TC} denote the Lagrange multiplier associated with inequality (A.18). Let $\lambda_{\hat{m}|m}^{TI}$ denote the Lagrange multiplier associated with inequality (A.19). Let μ be the Lagrange multiplier associated with the feasibility constraint. Finally, denote γ_m as the Lagrange multiplier on

the monotonicity constraint. The Lagrangian for the problem is

$$\begin{aligned}
\mathcal{L} = & \sum_{\theta_m} \pi_m [\phi U(c_m^R, k_m^R, y_m^R; \theta_m) + (1 - \phi) U(c_m^P, k_m^P, y_m^P; \theta_m)] \\
& + \sum_{\theta_m} \lambda_m [U(c_m^P, k_m^P, y_m^P; \theta_m) - x_m] + \sum_{\theta_m} \lambda_m^{TC} [x_m - U(c_{m-1}^P, k_{m-1}^P, y_{m-1}^P; \theta_m)] \\
& + \sum_{\theta_m} \sum_{\theta_{\hat{m}}} \lambda_{\hat{m}|m}^{TI} [x_m - U(c_{\hat{m}}^R, k_{\hat{m}}^R, y_{\hat{m}}^R; \theta_m)] \\
& + \mu \left\{ \sum_{\theta_m} \pi_m [\phi (y_m^R - c_m^R - k_m^R) + (1 - \phi) (y_m^P - c_m^P - k_m^P)] \right\}.
\end{aligned}$$

The first order conditions for the real allocations are

$$\begin{aligned}
\left(\phi \pi_m - \sum_{\theta_{\hat{m}}} \lambda_{\hat{m}|m}^{TI} \right) u'(c_m^R) &= \left(\phi \pi_m - \sum_{\theta_{\hat{m}}} \lambda_{\hat{m}|m}^{TI} \right) w'(k_m^R) = \phi \pi_m \mu, \\
\phi \pi_m \frac{1}{\theta_m} h' \left(\frac{y_m^R}{\theta_m} \right) - \sum_{\theta_{\hat{m}}} \lambda_{\hat{m}|m}^{TI} \frac{1}{\theta_{\hat{m}}} h' \left(\frac{y_m^R}{\theta_{\hat{m}}} \right) &= \phi \pi_m \mu.
\end{aligned}$$

The first order conditions for the persistent allocations are

$$\begin{aligned}
[(1 - \phi) \pi_m + \lambda_m - \lambda_{m+1}^{TC}] u'(c_m^P) &= [(1 - \phi) \pi_m + \lambda_m - \lambda_{m+1}^{TC}] w'(k_m^P) = (1 - \phi) \pi_m \mu, \\
[(1 - \phi) \pi_m + \lambda_m] \frac{1}{\theta_m} h' \left(\frac{y_m^P}{\theta_m} \right) - \lambda_{m+1}^{TC} \frac{1}{\theta_{m+1}} h' \left(\frac{y_m^P}{\theta_{m+1}} \right) &= (1 - \phi) \pi_m \mu + \gamma_m - \gamma_{m+1}.
\end{aligned}$$

The first order condition on x_m gives the following relationship for Lagrange multipliers

$$\lambda_m = \lambda_m^{TC} + \sum_{\theta_{\hat{m}}} \lambda_{\hat{m}|m}^{TI}.$$

The Kuhn-Tucker conditions for a solution are the first order conditions above and the complementary slackness conditions. By Hellwig (2007), at the optimum, it must be the case that at least one of the multipliers λ_m or γ_m is strictly positive. Also, notice that perfect consumption smoothing is implemented: $c_m^R = k_m^R$ and $c_m^P = k_m^P$ for all $\theta_m \in \Theta$.

Suppose for some θ_m , $(c_m^R, k_m^R, y_m^R) \neq (c_m^P, k_m^P, y_m^P)$. There are three possible cases. The first case is $c_m^R < c_m^P$ and $y_m^R < y_m^P$ with $\lambda_{m|m}^{TI} > 0$ and $\lambda_{\hat{m}|m}^{TI} = 0$ for all $\theta_{\hat{m}} \neq \theta_m$. The second case is $c_m^R > c_m^P$ and $y_m^P < y_m^R$ with some $\theta_{m+k} > \theta_m$ such that $\lambda_{m|m+k}^{TI} \geq 0$ and $\lambda_{\hat{m}|m}^{TI} = 0$ for all $\theta_{\hat{m}} \neq \theta_{m+k}$. The third case is $\lambda_{\hat{m}|m}^{TI} = 0$ for all $\theta_{\hat{m}} \in \Theta$.

For the first case, the first order conditions on consumption imply

$$\frac{\lambda_m - \lambda_{m+1}^{TC}}{1 - \phi} > -\frac{\lambda_{m|m}^{TI}}{\phi}.$$

From the first order conditions on output and by the fact that $h(\cdot)$ is strictly convex with $\theta_{m+1} > \theta_m$, the following inequality must hold at the optimum

$$\left(1 - \frac{\lambda_{m|m}^{TI}}{\phi \pi_m}\right) \frac{1}{\theta_m} h' \left(\frac{y_m^R}{\theta_m} \right) \geq \left[1 + \frac{\lambda_m - \lambda_{m+1}^{TC}}{(1 - \phi) \pi_m}\right] \frac{1}{\theta_m} h' \left(\frac{y_m^P}{\theta_m} \right) + \frac{\gamma_{m+1} - \gamma_m}{(1 - \phi) \pi_m},$$

Thus, for $y_m^P > y_m^R$, it must be that $\gamma_m > \gamma_{m+1}$. Thus, $y_m = y_{m-1}$, which implies $c_m = c_{m-1}$ and $k_m = k_{m-1}$, or else the upward incentive compatibility constraints would be violated. However, if $c_m^R < c_m^P$, $k_m^R < k_m^P$ and $y_m^R < y_m^P$ with $\lambda_{m|m}^{TI} > 0$, then the TC agent with productivity θ_{m-1} would be strictly better off misreporting as a TI agent with productivity θ_m . The first case is not possible.

For the second case, notice that if for some $\theta_m > \theta_1$, $(c_m^R, k_m^R, y_m^R) = (c_m^P, k_m^P, y_m^P)$, then the government has a welfare improving deviation for the real allocations. If $u'(c_m^P) < \mu$, then consider the following new deviation in real allocation: $\tilde{c}_m^R = c_m^R + \alpha \epsilon$, $\tilde{y}_m^R = y_m^R + \epsilon$ and $\tilde{k}_m^R = k_m^R$, with $\alpha = [u'(c_m^P)]^{-1} \frac{1}{\theta_{m+1}} h' \left(\frac{y_m^P}{\theta_{m+1}} \right)$. The gain in welfare from a higher net output more than offsets the loss in utility for the type θ_m TI agent. Notice that this deviation is not possible for the lowest type, so $(c_1^R, k_1^R, y_1^R) = (c_1^P, k_1^P, y_1^P)$.

It is easy to see that the third case is satisfied as long as $\lambda_m^{TC} \geq \lambda_{m+1}^{TC}$, in essence, $u'(c_m^P) > \mu$. Note the highest productivity, θ_M , belongs to the third case.

Finally, due to the co-monotonicity of c and y (Hellwig (2007)) for the persistent allocation, it must be the case that only high types belong to the third case, and lower types belong to the second case. This establishes the existence of a cutoff $\bar{\theta}$. ■

Proof of Theorem 6:

Part (i) of the theorem is immediate from the first order conditions presented in the proof for Theorem 5.

For part (ii), if $\theta_m \geq \bar{\theta}$, then by Theorem 5 the TI agents are strictly worse off than TC agents of the same productivity. Hence, from the first order conditions, $\sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI} = 0$ and the result follows. For TC agents with θ_M productivity, by Hellwig (2007), it must be that $U_c \left(c_M^P, k_M^P, \frac{y_M^P}{\theta_M} \right) + U_y \left(c_M^P, k_M^P, \frac{y_M^P}{\theta_M} \right) = 0$. The result follows.

For part (iii), if $\theta_m < \bar{\theta}$, then by Theorem 5 the the TC agents are either separated from the TI agents or bunched with the TI agents of the same productivity. If the TI agents are separated from the TC agents, then from the first order conditions it must have $\sum_{\theta_m} \lambda_{m|\hat{m}}^{TI} \geq 0$. It is thus immediate that $U_c(c_m^R, k_m^R, \frac{y_m^R}{\theta_m}) + U_y(c_m^R, k_m^R, \frac{y_m^R}{\theta_m}) \geq 0$. If the TC agents are bunched with TI agents, then by Hellwig (2007), the distortion away from efficiency must be downward: $U_c(c_m^P, k_m^P, \frac{y_m^P}{\theta_m}) + U_y(c_m^P, k_m^P, \frac{y_m^P}{\theta_m}) > 0$, which is true also for any TC agent with $\theta_m < \theta_M$. ■

Proof of Theorem 7:

By the envelope theorem,

$$\begin{aligned} \frac{\partial W^T}{\partial \phi} &= \sum_{\theta_m} \pi_m [U(c_m^R, k_m^R, y_m^R; \theta_m) - U(c_m^P, k_m^P, y_m^P; \theta_m)] \\ &\quad + \mu \sum_{\theta_m} \pi_m [(y_m^R - c_m^R - k_m^R) - (y_m^P - c_m^P - k_m^P)], \end{aligned}$$

where μ is the Lagrange multiplier on the feasibility constraint. By Theorem 6, for all $\theta_m \geq \bar{\theta}$, $c_m^R < c_m^P$ and $k_m^R < k_m^P$. Since $u(\cdot)$ and $w(\cdot)$ are strictly concave, it follows that $\theta_m \geq \bar{\theta}$,

$$u(c_m^P) - u(c_m^R) < u'(c_m^R)(c_m^P - c_m^R) = \mu(c_m^P - c_m^R),$$

$$w(k_m^P) - w(k_m^R) < w'(k_m^R)(k_m^P - k_m^R) = \mu(k_m^P - k_m^R),$$

where the last equality follows from the first order necessary conditions on c_m^R and k_m^R . Following a similar argument, since $y_m^P < y_m^R$ and $h(\cdot)$ is strictly convex, the following inequality holds

$$h\left(\frac{y_m^R}{\theta_m}\right) - h\left(\frac{y_m^P}{\theta_m}\right) < \frac{1}{\theta_m} h'\left(\frac{y_m^R}{\theta_m}\right) (y_m^R - y_m^P) = \mu(y_m^R - y_m^P).$$

Substituting the inequalities above, the following relationship holds

$$\begin{aligned} \frac{\partial W^F}{\partial \phi} &> \sum_{\theta_m < \bar{\theta}} \pi_m \left\{ [u(c_m^R) - u(c_m^P)] - \mu(c_m^R - c_m^P) - \left[h\left(\frac{y_m^R}{\theta_m}\right) - h\left(\frac{y_m^P}{\theta_m}\right) \right] + \mu(y_m^R - y_m^P) \right. \\ &\quad \left. + [w(k_m^R) - w(k_m^P)] - \mu(k_m^R - k_m^P) \right\}. \end{aligned}$$

For $\theta_m < \bar{\theta}$, by Theorem 5, either $(c_m^R, k_m^R, y_m^R) = (c_m^P, k_m^P, y_m^P)$ or $(c_m^R, k_m^R, y_m^R) \gg (c_m^P, k_m^P, y_m^P)$. It is sufficient to focus on the latter case.

When $(c_m^R, k_m^R, y_m^R) \gg (c_m^P, k_m^P, y_m^P)$, either $\lambda_{m|\hat{m}}^{TI} = 0$ for all $\theta_{\hat{m}}$ or there exists $\theta_{\hat{m}}$ such that $\lambda_{m|\hat{m}}^{TI} > 0$. For the former case, the same procedure delineated above can be applied. For the latter, suppose there exists $\theta_m \in \Theta$ such that

$$H(c_m^R, k_m^R, y_m^R) \equiv U(c_m^R, k_m^R, y_m^R; \theta_m) + \mu(y_m^R - c_m^R - k_m^R) - U(c_m^P, k_m^P, y_m^P; \theta_m) - \mu(y_m^P - c_m^P - k_m^P) < 0.$$

Let $\Gamma = u(c) + w(k)$. From the proof of Theorem 5, if $(c_m^P, k_m^P, y_m^P) = (c_m^R, k_m^R, y_m^R)$, then there exists perturbations $d\Gamma$ and dy such that the gain in welfare from the higher net output outweighs the loss in utility for type θ_m TI agent. Also, from the first order conditions,

$$\mu = \frac{1}{\sum_{\theta_m} \frac{\pi_m}{u'(c_m^R)}}.$$

Thus, by the continuity of $H(c_m^R, k_m^R, y_m^R)$, there exists $(\bar{c}_m^R, \bar{k}_m^R, \bar{y}_m^R)$ such that $H(\bar{c}_m^R, \bar{k}_m^R, \bar{y}_m^R) = 0$ and $(c_m^R, k_m^R, y_m^R) \gg (\bar{c}_m^R, \bar{k}_m^R, \bar{y}_m^R) \gg (c_m^P, k_m^P, y_m^P)$. Let $\bar{\mu}$ denote the multiplier on the feasibility constraint associated with $(\bar{c}_m^R, \bar{k}_m^R, \bar{y}_m^R)$. Since $(c_m^R, k_m^R, y_m^R) \neq (\bar{c}_m^R, \bar{k}_m^R, \bar{y}_m^R)$, then consider $\tilde{c}_m^R = \bar{c}_m^R + \alpha\epsilon$ and $\tilde{y}_m^R = \bar{y}_m^R + \epsilon$ for $\epsilon > 0$ and $\alpha = [u'(c_m^P)]^{-1} \frac{1}{\theta_{m+1}} h' \left(\frac{y_m^P}{\theta_{m+1}} \right)$. This perturbation improves welfare, so it must be the case that $H(\tilde{c}_m^R, \bar{k}_m^R, \tilde{y}_m^R) > 0$. Thus, the neighborhood around any (c, k, y) such that $H(c, k, y) = 0$ has a perturbation $d\Gamma$ and dy that improves welfare. Therefore, it can never be the case that $H(c_m^R, k_m^R, y_m^R) < 0$.

Hence, $\frac{\partial W^F}{\partial \phi} > 0$. Furthermore, by the maximum theorem, $W^T(\phi)$ is a continuous function over $[0, 1]$. Thus, welfare increases continuously from the constrained efficient level to the efficient level as ϕ increases. ■

Proof of Theorem 8:

The first part of the proof will establish the implementability of the optimal allocation from the threat mechanism. Choose the persistent and real allocations to be the same as the optimal persistent and real allocations from the threat mechanism. The deterrent allocations can be the same, or different as long as it is capable of deterring TI agents from misreporting to be TC agents. For the imaginary allocations, set $y_m^I = y_m^P$ for all $\theta_m \in \Theta$, and

$$u(c_m^I) + w(k_m^I) = u(c_m^P) + w(k_m^P).$$

In particular, choose $c_1^I = c_1^P = c_1^R$ and $k_1^I = k_1^P = k_1^R$. Notice that with this construction, all of the incentive compatibility constraints are satisfied, since the persistent allocations from the threat mechanism are incentive compatible. Next, choose $(c_m^I, k_m^I)_{\theta_m \in \Theta}$ such that the fooling and executability constraints are satisfied: $\forall \theta_m \in \Theta$

$$u(c_m^I) - h\left(\frac{y_m^I}{\theta_m}\right) + \hat{\beta}w(k_m^I) \geq u(c_m^R) - h\left(\frac{y_m^R}{\theta_m}\right) + \hat{\beta}w(k_m^R),$$

$$u(c_m^R) - h\left(\frac{y_m^R}{\theta_m}\right) + \beta w(k_m^R) \geq u(c_m^I) - h\left(\frac{y_m^I}{\theta_m}\right) + \beta w(k_m^I).$$

By the construction of the imaginary allocations, the fooling and executability constraints can be rewritten as

$$w(k_m^I) \leq \frac{U(c_m^P, k_m^P, y_m^P; \theta_m) - U(c_m^R, k_m^R, y_m^R; \theta_m)}{1 - \hat{\beta}} + w(k_m^R),$$

$$w(k_m^I) \geq \frac{U(c_m^P, k_m^P, y_m^P; \theta_m) - U(c_m^R, k_m^R, y_m^R; \theta_m)}{1 - \beta} + w(k_m^R).$$

By the incentive compatibility of the persistent and real allocations and by Theorem 5, $U(c_m^P, k_m^P, y_m^P; \theta_m) > U(c_m^R, k_m^R, y_m^R; \theta_m)$ for all $\theta_m \in \Theta \setminus \theta_1$. Therefore, it is possible to find (c_m^I, k_m^I) for all productivity types to satisfy the fooling constraints when $1 > \hat{\beta} > \beta$.

The second part of the proof will argue the optimality of the optimal threat allocation in a fooling mechanism. Note that a direct truth-telling fooling mechanism has extra potentially binding constraints that a direct truth-telling threat mechanism does not have: $\forall \theta_m, \theta_{\hat{m}} \in \Theta$,

$$U(c_m^P, k_m^P, y_m^P; \theta_m) \geq U(c_{\hat{m}}^I, k_{\hat{m}}^I, y_{\hat{m}}^I; \theta_m).$$

For a threat mechanism, a TC agent would never choose the threat allocations. However, in a fooling mechanism, the government needs to deter the TC agent from pretending to be a TI agent and choose the imaginary allocations. Hence, the government's problem in a fooling mechanism has more constraints than a threat mechanism. Since the optimal allocation from a direct threat mechanism is implementable in a direct fooling mechanism, then the optimal allocations in both mechanisms must be equivalent. ■

Proof of Theorem 9:

Similar to the proof of Theorem 5, the proof will focus on a relaxed problem. In essence, only the local downward incentive compatibility constraints are considered along with a

monotonicity constraints on y^P . The incentive compatibility constraints are (A.17), (A.18) and (A.19). Notice the fooling constraint is included in (A.19), when $\theta_{\hat{m}} = \theta_m$. The executability constraint is

$$u(c_m^R) - h\left(\frac{y_m^R}{\theta_m}\right) + \beta w(k_m^R) \geq u(c_m^P) - h\left(\frac{y_m^P}{\theta_m}\right) + \beta w(k_m^P).$$

Notice the fooling and executability constraints imply that

$$k_m^P \geq k_m^R, \tag{A.20}$$

and

$$u(c_m^R) - h\left(\frac{y_m^R}{\theta_m}\right) \geq u(c_m^P) - h\left(\frac{y_m^P}{\theta_m}\right). \tag{A.21}$$

Let ω_m be the Lagrange multiplier associated with the executability constraint for TC agents of type θ_m . The rest of the Lagrange multipliers are the same as in the proof for Theorem 5.

The first order conditions for the real allocations are

$$\begin{aligned} \left(\phi\pi_m - \sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI} + \omega_m \right) u'(c_m^R) &= \phi\pi_m\mu, \\ \left(\phi\pi_m - \sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI} + \beta\omega_m \right) w'(k_m^R) &= \phi\pi_m\mu \\ (\phi\pi_m + \omega_m) \frac{1}{\theta_m} h'\left(\frac{y_m^R}{\theta_m}\right) - \sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI} \frac{1}{\theta_{\hat{m}}} h'\left(\frac{y_m^R}{\theta_{\hat{m}}}\right) &= \phi\pi_m\mu. \end{aligned}$$

The first order conditions for the persistent allocations are

$$\begin{aligned} [(1 - \phi)\pi_m + \lambda_m - \lambda_{m+1}^{TC} - \omega_m] u'(c_m^P) &= (1 - \phi)\pi_m\mu, \\ [(1 - \phi)\pi_m + \lambda_m - \lambda_{m+1}^{TC} - \beta\omega_m] w'(k_m^P) &= (1 - \phi)\pi_m\mu, \\ [(1 - \phi)\pi_m + \lambda_m - \omega_m] \frac{1}{\theta_m} h'\left(\frac{y_m^P}{\theta_m}\right) - \lambda_{m+1}^{TC} \frac{1}{\theta_{m+1}} h'\left(\frac{y_m^P}{\theta_{m+1}}\right) &= (1 - \phi)\pi_m\mu + \gamma_m - \gamma_{m+1}. \end{aligned}$$

The first order condition on x_m gives the following relationship for Lagrange multipliers

$$\lambda_m = \lambda_m^{TC} + \sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI}.$$

By the Kuhn-Tucker Theorem, the solution is characterized by the first order conditions above and the complementary slackness conditions. Notice that if $\omega_m > 0$, then there will

be intertemporal distortions for type θ_m . Next, the proof will proceed to show how $\omega_m > 0$ for all $\theta_m \in \Theta \setminus \theta_1$, there by demonstrating part (i).

Suppose there exists $\theta_m \in \Theta \setminus \theta_1$ such that $\omega_m = 0$. By the first order conditions and (A.20), it must be the case that $c_m^P \geq c_m^R$ and with strict inequality only if $k_m^P > k_m^R$. There are two cases: if $k_m^P > k_m^R$ or $k_m^P = k_m^R$.

If $k_m^P > k_m^R$, then $c_m^P > c_m^R$ and

$$1 + \frac{\lambda_m - \lambda_{m+1}^{TC}}{(1 - \phi)\pi_m} > 1 - \frac{\sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI}}{\phi\pi_m}.$$

Thus by (A.21), it must be the case that $y_m^P > y_m^R$. If $\sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI} = 0$, then

$$\left(1 + \frac{\lambda_m - \lambda_{m+1}^{TC}}{(1 - \phi)\pi_m}\right) \frac{1}{\theta_m} h' \left(\frac{y_m^P}{\theta_m}\right) + \frac{\gamma_{m+1} - \gamma_m}{(1 - \phi)\pi_m} < \left(1 - \frac{\sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI}}{\phi\pi_m}\right) \frac{1}{\theta_m} h' \left(\frac{y_m^R}{\theta_m}\right), \quad (\text{A.22})$$

which implies $\gamma_m > 0$ if $y_m^P > y_m^R$ is true. Thus, it follows that $y_m^P = y_{m-1}^P$ with $c_m^P = c_{m-1}^P$ and $k_m^P = k_{m-1}^P$. Hence, with $\sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI} = 0$, then either a θ_{m-1} agent has an incentive to misreport to be a θ_m agent, or vice versa, which is a contradiction, so $\sum_{\theta_{\hat{m}}} \lambda_{m|\hat{m}}^{TI} > 0$. If this is the case, then it must be that $\lambda_{m|m}^{TI} > 0$ with $\lambda_{m|\hat{m}}^{TI} = 0$ for all $\theta_{\hat{m}} \in \Theta \setminus \theta_m$. This is because if $\lambda_{m|\hat{m}}^{TI} > 0$ for $\theta_{\hat{m}} > \theta_m$, then $y_m^R \geq y_m^P$ and if $\lambda_{m|\hat{m}}^{TI} > 0$ for $\theta_{\hat{m}} < \theta_m$, then there is a welfare enhancing perturbation by increasing the output and consumption. However, if $\lambda_{m|m}^{TI} > 0$, then (A.22) must be true, which is a contradiction. Hence, if $\omega_m = 0$ for some $\theta_m \in \Theta \setminus \theta_1$, then it can not be the case that $k_m^P > k_m^R$.

Now, examine the case when $k_m^P = k_m^R$, then $c_m^P = c_m^R$. By the fooling and executability constraints, this means that $y_m^P = y_m^R$. However, this is not optimal. Consider the perturbation: $\tilde{c}_m^R = c_m^R + \alpha\epsilon$, $\tilde{y}_m^R = y_m^R + \epsilon$ and $\tilde{k}_m^R = k_m^R$, with $\alpha = [u'(c_m^P)]^{-1} \frac{1}{\theta_{m+1}} h' \left(\frac{y_m^P}{\theta_{m+1}}\right)$. As was demonstrated in the proof for Theorem 5, the gain in welfare from the higher net output more than offsets the loss in utility for the type θ_m TI agent. Hence, for all $\theta_m \in \Theta_m \setminus \theta_1$, it must be that $\omega_m > 0$. This implies the TI agents under-save, and the TC agents over-save.

For part (ii), notice part (i) implies $\omega_M > 0$. Suppose $\lambda_{M|M}^{TI} > 0$, then both the fooling and executability constraints for type θ_M are binding by the complementary slackness condition. The binding constraints imply $k_M^P = k_M^R$ with $c_M^P = c_M^R$ and $y_M^P = y_M^R$. From the first order condition on c , it implies that

$$\frac{\lambda_M - \omega_M}{1 - \phi} = \frac{\omega_M - \lambda_{M|M}^{TI}}{\phi},$$

which implies $k_m^P > k_m^R$ from the first order conditions on k . Thus, $\lambda_{M|M}^{TI} = 0$. If $\lambda_{M|M}^{TI} = 0$, notice that it can not be the case that $k_M^P = k_M^R$, since by the first order conditions, this would imply $c_M^R > c_M^P$ and $y_M^R < y_M^P$ which violates the fooling (incentive compatibility) constraint. Hence, $k_M^P > k_M^R$. The case with $c_m^P \geq c_m^R$ and $y_m^P \geq y_m^R$ can be ruled out since this violates the first order conditions. It follows that $c_M^P < c_M^R$ and $y_M^P > y_M^R$. This proves part (ii).

For part (iii), first notice that if $\omega_1 = 0$, then by the same argument in the proof of Theorem 5, $(c_1^R, k_1^R, y_1^R) = (c_1^P, k_1^P, y_1^P)$.

Suppose $\omega_1 > 0$, then by (A.20) and the first order condition on savings, it must be that

$$\frac{-\lambda_2^{TC} - \beta\omega_1}{1 - \phi} \geq \frac{\beta\omega_1 - \sum_{\theta_{\hat{m}}} \lambda_{1|\hat{m}}^{TI}}{\phi}. \quad (\text{A.23})$$

Hence, there exists $\theta_{\hat{m}}$ such that $\lambda_{1|\hat{m}}^{TI} > 0$.

If $c_1^P \geq c_1^R$, then $y_1^P \geq y_1^R$. by (A.21). Thus, it must be the case that $\lambda_{1|1}^{TI} > 0$. By complementary slackness condition, both the executability and fooling constraints for type θ_1 must bind. It follows that $k_1^R = k_1^P$, which implies that (A.23) holds with equality. However, from the first order conditions on consumption, $c_1^P \geq c_1^R$ is not possible.

Now check $c_1^R > c_1^P$, which implies

$$\frac{-\lambda_2^{TC} - \omega_1}{1 - \phi} < \frac{\omega_1 - \sum_{\theta_{\hat{m}}} \lambda_{1|\hat{m}}^{TI}}{\phi}. \quad (\text{A.24})$$

If $y_1^R < y_1^P$, then it has to be the case that $\lambda_{1|1}^{TI} > 0$. Hence, it must be the case that $k_1^R = k_1^P$ by the complementary slackness conditions. This violates the executability constraint. Finally, it remains to check $(c_1^R, y_1^R) \gg (c_1^P, y_1^P)$. Since $y_1^R > y_1^P$, from the first order conditions, it must be that

$$\frac{\omega_1}{\phi\pi_1} \frac{1}{\theta_1} h' \left(\frac{y_1^R}{\theta_1} \right) + \frac{\omega_1}{(1 - \phi)\pi_1} \frac{1}{\theta_1} h' \left(\frac{y_1^P}{\theta_1} \right) < \frac{\lambda_{1|\hat{m}}^{TI}}{\phi\pi_1} \frac{1}{\theta_{\hat{m}}} h' \left(\frac{y_1^R}{\theta_{\hat{m}}} \right) - \frac{\lambda_2^{TC}}{(1 - \phi)\pi_1} \frac{1}{\theta_2} h' \left(\frac{y_1^P}{\theta_2} \right),$$

which contradicts (A.24). Thus, at the optimum, $\omega_1 = 0$. This implies perfect consumption smoothing for type θ_1 . ■

Proof of Theorem 10:

By Theorem 8, it is without loss of generality to focus on a fooling mechanism for $W(\hat{\beta} < 1)$. Let $\mathcal{A}(\hat{\beta} < 1)$ and $\mathcal{A}(\hat{\beta} = 1)$ denote the sets of real and persistent allocations

that are implementable under a fooling mechanism with partially naïve and fully naïve agents respectively.

First, I will show that $\mathcal{A}(\hat{\beta} = 1) \subseteq \mathcal{A}(\hat{\beta} < 1)$. Note that for any $\mathbf{a} \in \mathcal{A}(\hat{\beta} = 1)$, the persistent allocation is equivalent to the imaginary allocation and the incentive compatibility constraints for the persistent allocation is trivially satisfied. Also, the incentive compatibility constraints of the real allocations hold. Hence, $\mathbf{a} \in \mathcal{A}(\hat{\beta} < 1)$, which implies $W(\hat{\beta} < 1) \geq W(\hat{\beta} = 1)$.

Finally, I will show that the optimal allocation $a^* \in \mathcal{A}(\hat{\beta} < 1)$ is not implementable when agents are fully naïve, in essence, $a^* \notin \mathcal{A}(\hat{\beta} = 1)$. To implement a^* with fully naïve agents, the government has to set the imaginary allocations equal to the persistent allocations. By Theorem 5 and Theorem 8, a^* has the following property for $\theta_M : c_M^P > c_M^R, k_M^P > k_M^R$ and $y_M^P < y_M^R$. Hence, the following fooling constraint is violated

$$u(c_M^R) - h\left(\frac{y_M^R}{\theta_M}\right) + \beta w(k_M^R) \geq u(c_M^P) - h\left(\frac{y_M^P}{\theta_M}\right) + \beta w(k_M^P).$$

Therefore, the real allocation in a^* is not implementable when agents are fully naïve. This implies that $W(\hat{\beta} < 1) > W(\hat{\beta} = 1)$. ■

Proof of Proposition 3:

To see how P^n implements the efficient allocation, first choose $\tau^* = 1 - \beta$ so, on the equilibrium path, consumption is smooth: $u'(c(y)) = w'(k(y))$. Also, set the income tax to be

$$T^e(y) = \begin{cases} y - c^* - \beta k^* & \text{if } y \in [y_1^*, \infty) \\ y & \text{if } y \in [0, y_1^*) \end{cases}$$

where (c^*, k^*) is the efficient consumption and savings and y_1^* is the efficient output for the least productive agents. By construction, no agent will ever produce $y < y_1^*$.

Let $Y = \{y_1^*, \dots, y_M^*\}$ denote the set of efficient outputs. Next, define $\tau(y)$ as an increasing step function, so $\tau(y) = \tau(y_m^*)$ if $y \in [y_m^*, y_{m+1}^*)$ with $\tau(y_1^*) = \tau^*$ and $\tau(y_m^*) < \tau(y_{m+1}^*)$. Also, define $T^s(y)$ as an increasing step function, so $T^s(y) = T^s(y_m^*)$ if $y \in (y_{m-1}^*, y_m^*]$ with $T^s(y_1^*) = 0$ and $T^s(y_m^*) > T^s(y_{m-1}^*)$. By the construction of $\tau(y)$, $T^e(y)$ and $T^s(y)$, agents would only choose $y \in Y$. If the planner expects the doer to switch from the default savings subsidy τ^* , for a given y , the doer is perceived to choose consumption

and savings by solving

$$\max_{c,k} u(c) + \hat{\beta}w(k) \text{ s.t. } c + (1 - \tau(y))k = y - T^e(y) - T^s(y).$$

Hence, the perceived optimal consumption and savings choices are functions of $\tau(y)$ and $T^s(y) : c(\tau(y), T^s(y))$ and $k(\tau(y), T^s(y))$.

An agent with productivity θ_m selects y by solving the following problem:

$$\max_{y \in Y} u[c(\tau(y), T^s(y))] - h\left(\frac{y}{\theta_m}\right) + w[k(\tau(y), T^s(y))].$$

By the setup, it is sufficient to choose policies that deter adjacent downward deviations. For type θ_m agents, the government can construct $\tau(y_m^*)$ for given $T^s(y_m^*) > T^s(y_{m-1}^*)$ and $\tau(y_{m-1}^*)$ to satisfy

$$u[c(\tau(y_m^*), T^s(y_m^*))] + w[k(\tau(y_m^*), T^s(y_m^*))] - [u[c(\tau(y_{m-1}^*), T^s(y_{m-1}^*))] + w[k(\tau(y_{m-1}^*), T^s(y_{m-1}^*))]] = \Delta_m,$$

where $\Delta_m = h\left(\frac{y_m^*}{\theta_m}\right) - h\left(\frac{y_{m-1}^*}{\theta_m}\right)$. Note that the least productive agent would choose the efficient allocation for θ_1 . Therefore, by the incentive compatibility constraints, $\tau(y_m^*)$ is a function of $(T^s(y_2^*), \dots, T^s(y_m^*))$.⁴

The government can choose $T^s(y_2^*)$ so the following holds

$$u[c(\tau(y_2^*), T^s(y_2^*))] + \hat{\beta}w[k(\tau(y_2^*), T^s(y_2^*))] = u(c^*) + \hat{\beta}w(k^*).$$

Since $\tau(y_2^*) > \tau^*$, then it must be that $T^s(y_2^*) > T^s(y_1^*)$ for the above fooling constraint to hold with equality. By iteration, the government chooses $T^s = (T^s(y_2^*), \dots, T^s(y_M^*))$ so that the fooling constraints bind and both $T^s(y)$ and $\tau(y)$ are increasing step functions.

To check that the construction above implements the efficient allocation, notice that the executability constraints are trivially satisfied, because $\tau(y) \geq \tau^*$ for any y and the fooling constraints bind. As a result, the agents would choose the efficient allocation if they produced the efficient output. Also, notice that both the fooling and incentive compatibility constraints hold by construction, so the agents would produce efficiently. Finally, the government budget constraint holds, since $\sum \pi_m T^e(y_m^*) + \tau^* k^* = 0$. ■

⁴This is due to the fact that from the incentive compatibility constraint for type θ_2 agents, $\tau(y_2^*)$ is a function of $T^s(y_2^*)$. By induction, $\tau(y_m^*)$ is a function of $(T^s(y_2^*), \dots, T^s(y_m^*))$.

Proof of Proposition 4:

Consider the following repayment plan for a projection $y^e \in [y_m^*, y_{m+1}^*)$,

$$R(b, y^e, y) = \begin{cases} \eta(y^e) - \rho^* b & \text{if } y_m^* \leq y < \bar{y}(y^e) \\ \hat{\eta}(y^e, y) - \hat{\rho}(y^e) b & \text{if } y \geq \bar{y}(y^e) \\ b & \text{if } y < y_m^* \end{cases}$$

and $\bar{y}(y^e) = \infty$ if $y^e \geq y_M^*$. The repayment plan consists of the principal, η , and a variable payment contingent on savings, ρb . The construction of the repayment plan ensures the agent would not produce less than y_m^* if the initial projection was $y^e \in [y_m^*, y_{m+1}^*)$. The government sets $\rho^* = \frac{1}{\beta} - 1$, so consumption is smoothed on the equilibrium path. To see this, notice that the sequential budget constraints (1.13) and (1.14) can be consolidated, and the budget constraint on the equilibrium path is

$$c + \frac{k}{1 + \rho^*} = y + L(y^e) - T(y) - \eta(y^e)/_{1+\rho^*}.$$

Consequently, with $\rho^* = \frac{1}{\beta} - 1$, the doer would choose the appropriate savings amount on the equilibrium path. In essence, ρ^* is a discount on the repayment contingent on savings. Also, the government can set $L(y^e) = \beta\eta(y^e)$, which means the agents do not need to pay any interest on the commitment loan if they choose the appropriate savings b and income $y < \bar{y}(y^e)$. The income tax is

$$T(y) = \begin{cases} y - c^* - \beta k^* & \text{if } y \in [y_1^*, \infty) \\ y & \text{if } y \in [0, y_1^*) \end{cases}$$

so agents would never produce less than y_1^* . The agents would choose the efficient allocation, (c^*, k^*, y_m^*) , on the equilibrium path. To see why this is the case, first notice the doer solves the following problem on the equilibrium path:

$$\max V(c, k, y; \theta) \text{ s.t. } c + \frac{k}{1+\rho^*} = y + L(y^e) - T(y) - \eta(y^e)/_{1+\rho^*}.$$

With the constructed policies, an agent would never produce more than y_m^* on the equilibrium path, and the on-equilibrium budget constraint is $(c - c^*) + \beta(k - k^*) = 0$. Thus, for the first order conditions and the budget constraint to hold, the unique solution to the doer's problem is the efficient allocation.

Next, to make sure a type θ_m agent would choose a commitment loan for projection $y^e \in [y_m^*, y_{m+1}^*]$, construction of the off-equilibrium path policy, $\{(\hat{\eta}(y^e, y), \hat{\rho}(y^e), \bar{y}(y^e))\}_{m=1}^M$, along with the fixed principal payment $\eta(y^e)$ will be demonstrated.

To prevent deviation, the off-equilibrium path policy needs to deter agents from choosing $L(y^e)$ that is not intended for their productivity. First, for any $y_1^e, y_2^e \in [y_m^*, y_{m+1}^*]$, set $\eta(y_1^e) = \eta(y_2^e)$, $\hat{\rho}(y_1^e) = \hat{\rho}(y_2^e)$, $\bar{y}(y_1^e) = \bar{y}(y_2^e)$ and $\hat{\eta}(y_1^e, y) = \hat{\eta}(y_2^e, y)$. Next, set $\hat{\eta}(y^e, y) = (y - \bar{y}(y^e))(1 + \hat{\rho}(y^e))$ so the agents would never choose to produce more than $\bar{y}(y^e)$. In essence, if the agent has output $y = \bar{y}(y^e)$, then the agent does not need to pay the principal η , but will instead repay an amount proportional to his savings b .

Substituting in the policies, a type θ_{m+1} agent contemplating a commitment loan for $y^e \in [y_m^*, y_{m+1}^*]$ predicts his doer would solve the following program if $y \in [y_m^*, \bar{y}(y^e)]$

$$\max_{c, y, k} u(c) - h\left(\frac{y}{\theta_{m+1}}\right) + \hat{\beta}w(k) \text{ s.t. } c + \beta k = (c^* + \beta k^*).$$

It is obvious the optimal output is y_m^* , and let (\tilde{c}, \tilde{k}) denote the solution for consumption and savings. Notice the solution to the problem does not depend on the agent's actual productivity θ_{m+1} nor θ_m . If $y = \bar{y}(y^e)$, the type θ_{m+1} agent would solve the following problem

$$\max_{c, k} u(c) + \hat{\beta}w(k) \text{ s.t. } c + \frac{k}{1 + \hat{\rho}(y^e)} = c^* + \beta k^* + \beta \eta(y^e),$$

and let (\hat{c}_m, \hat{k}_m) denote the solution, which also does not depend on θ_{m+1} .

Let $\bar{y}(y^e) = y_m^* + \alpha_m$, with $\alpha_m > 0$. The government constructs $\hat{\rho}(y^e) < \rho^*$ for all productivity types to tempt the doer to increase c and decrease k , which is undesirable to the planner. With $\hat{\rho}(y^e) < \rho^*$, the government can construct $\eta(y^e) > 0$ to ensure the credible threat constraint holds, given $\alpha_m > 0$:

$$u(\hat{c}_m) - h\left(\frac{y_m^* + \alpha_m}{\theta_{m+1}}\right) + \hat{\beta}w(\hat{k}_m) \geq u(\tilde{c}) - h\left(\frac{y_m^*}{\theta_{m+1}}\right) + \hat{\beta}w(\tilde{k}). \quad (\text{A.25})$$

For any given $\hat{\rho}(y^e) < \rho^*$ and $\alpha_m > 0$, construct $\eta(y^e)$ such that (A.25) binds. Let $\eta_m(\hat{\rho}(y^e), \alpha_m)$ denote the solution for $y^e \in [y_m^*, y_{m+1}^*]$. Also, with (A.25) binding, higher productivity types would strictly prefer producing output $y = y_m^* + \alpha_m$ if they chose the loan for projected income $y^e \in [y_m^*, y_{m+1}^*]$.

Next, the government chooses $\hat{\rho}(y^e)$ to satisfy the incentive compatibility constraint:

$$u(\tilde{c}) - h\left(\frac{y_M^*}{\theta_M}\right) + w(\tilde{k}) \geq u(\hat{c}_m) - h\left(\frac{y_m^* + \alpha_m}{\theta_M}\right) + w(\hat{k}_m),$$

where $(\tilde{c}, \tilde{k}, y_M^*)$ is the solution to the following problem for taking out a loan with an initial projection of $y^e \in [y_M^*, \infty)$,

$$\max_{c, y, k} u(c) - h\left(\frac{y}{\theta_M}\right) + \hat{\beta}w(k) \quad \text{s.t.} \quad c + \frac{k}{1 + \rho^*} = (c^* + \beta k^*).$$

If the θ_M agent can be deterred from picking the loan for θ_m , then all other agents would be deterred as well. Since (A.25) binds, the incentive compatibility constraint can be rewritten as

$$(1 - \hat{\beta}) \left[w(\tilde{k}) - w(\hat{k}_m) \right] \geq h\left(\frac{y_M^*}{\theta_M}\right) - h\left(\frac{y_m^*}{\theta_M}\right).$$

Notice that, for a given $\alpha_m > 0$, \hat{k}_m is an increasing function of $\hat{\rho}(y^e)$ and is the only object dependent on $\hat{\rho}(y^e)$. The government can then decrease $\hat{\rho}(y^e)$ till the incentive compatibility constraint holds for any $\alpha_m > 0$. Let $\hat{\rho}_m(\alpha_m)$ denote the level of $\hat{\rho}(y^e)$ such that the incentive compatibility constraint holds. Hence, the government can set $\eta_m(\hat{\rho}_m(\alpha_m), \alpha_m)$ and $\hat{\rho}_m(\alpha_m)$ for (A.25) and the incentive compatibility constraints to hold.

Finally, to pin down α_m , the government can increase α_m till the executability constraint for θ_m type agent holds. The executability constraint is

$$u(c^*) - h\left(\frac{y_m^*}{\theta_m}\right) + \beta w(k^*) \geq u(c'_m) - h\left(\frac{y_m^* + \alpha_m}{\theta_m}\right) + \beta w(k'_m).$$

where (c'_m, k'_m) is the solution to the following problem

$$\max u(c) + \beta w(k) \quad \text{s.t.} \quad c + \frac{k}{1 + \hat{\rho}_m(\alpha_m)} = c^* + \beta k^* + \beta \eta_m(\hat{\rho}_m(\alpha_m), \alpha_m).$$

By construction, the government budget is balanced. Also, since (A.25) binds, the actual θ_m agent would predict an output of $y < \bar{y}(y^e)$ so the credible threat constraints for θ_m hold. To see this, notice the following $h(y_m^* + \alpha_m) - h\left(\frac{y_m^*}{\theta_m}\right) > h\left(\frac{y_m^* + \alpha_m}{\theta_{m+1}}\right) - h\left(\frac{y_m^*}{\theta_{m+1}}\right)$, which implies $u(\hat{c}_m) - h\left(\frac{y_m^*}{\theta_m}\right) + \hat{\beta}w(\hat{k}_m) < u(\tilde{c}) - h\left(\frac{y_m^*}{\theta_m}\right) + \hat{\beta}w(\tilde{k})$. The process above can be repeated for all productivity types to ensure global incentive compatibility. ■

Appendix B

Appendix for Chapter 2

B.1 Proofs

Proof of Theorem 11: By (2.1), it has to be the case that $(q_\theta^R, p_\theta^R) \in C_\theta^*$. Set $(q_\theta^R, p_\theta^R) = (q_\theta^*, p_\theta^*)$ for all $\theta \in \Theta$, so (2.3) automatically holds. For a threat mechanism, it remains to be shown that $(q_{\theta'|\theta}^T, p_{\theta'|\theta}^T)$ can be constructed so that $(q_{\theta'|\theta}^T, p_{\theta'|\theta}^T) \in C_{\theta'|\theta}^*$ and (2.2) is satisfied, while (2.1) holds. Since $(q_\theta^R, p_\theta^R) = (q_\theta^*, p_\theta^*)$ for all $\theta \in \Theta$, the threat option only needs to deter downward misreports.

First, consider the lowest type $\underline{\theta}$. Next, notice the following programming problem has no finite solution for any $\theta > \underline{\theta}$

$$\min_{q_{\underline{\theta}|\theta}^T, p_{\underline{\theta}|\theta}^T} U(q_{\underline{\theta}|\theta}^T, p_{\underline{\theta}|\theta}^T; \theta)$$

subject to

$$V(q_{\underline{\theta}|\theta}^T, p_{\underline{\theta}|\theta}^T; \theta) = V(q_{\underline{\theta}}^R, p_{\underline{\theta}}^R; \theta).$$

This is true because $\beta < 1$, so $q_{\underline{\theta}|\theta}^T$ can always be increased with a corresponding increase in $p_{\underline{\theta}|\theta}^T$ such that the equality constraint, $V(q_{\underline{\theta}|\theta}^T, p_{\underline{\theta}|\theta}^T; \theta) = V(q_{\underline{\theta}}^R, p_{\underline{\theta}}^R; \theta)$ holds, while $U(q_{\underline{\theta}|\theta}^T, p_{\underline{\theta}|\theta}^T; \theta)$ is lowered.

Next, to see that (2.1) is automatically satisfied, single-crossing ensures that if $V(q_{\underline{\theta}|\theta}^T, p_{\underline{\theta}|\theta}^T; \theta) = V(q_{\underline{\theta}}^R, p_{\underline{\theta}}^R; \theta)$ and $q_{\underline{\theta}|\theta}^T > q_{\underline{\theta}}^R$ for $\theta > \underline{\theta}$, then $V(q_{\underline{\theta}|\theta}^T, p_{\underline{\theta}|\theta}^T; \underline{\theta}) < V(q_{\underline{\theta}}^R, p_{\underline{\theta}}^R; \underline{\theta})$. Hence, $(q_{\underline{\theta}}^R, p_{\underline{\theta}}^R)$ is implementable.

Finally, the process above can be repeated for all types $\theta > \underline{\theta}$, which completes the proof. ■

Proof of Corollary 4: The proof follows from the proofs for Theorem 11. First, set $(q_\sigma^R(\mu), p_\sigma^R(\mu)) = (q_\sigma^*(\mu), p_\sigma^*(\mu))$ for all $\sigma \in \Sigma$ and $\mu \in [0, 1]$, so (2.3) automatically holds. Hence, the threat needs to deter downward misreports.

Since μ is observable, choose $(q_{\sigma|\sigma'}^T(\mu), p_{\sigma|\sigma'}^T(\mu))$ with $\sigma < \sigma'$ such that for all $\mu \in [0, 1]$ and $\sigma, \sigma' \in \Sigma$ with $\sigma < \sigma'$,

$$\theta(\sigma', \mu) q_{\sigma|\sigma'}^T(\mu) - \beta p_{\sigma|\sigma'}^T(\mu) = \theta(\sigma', \mu) q_\sigma^R(\mu) - \beta p_\sigma^R(\mu). \quad (\text{B.1})$$

This ensures that $(q_{\sigma|\sigma'}^T(\mu), p_{\sigma|\sigma'}^T(\mu)) \in C_{\sigma|\sigma'}^*(\mu)$.

Next, consider the lowest signal σ_1 and notice the following programming problem has no finite solution for any $\sigma > \sigma_1$

$$\min_{q_{\sigma_1|\sigma}^T, p_{\sigma_1|\sigma}^T} \int_0^1 \theta(\sigma, \mu) q_{\sigma_1|\sigma}^T(\mu) - p_{\sigma_1|\sigma}^T(\mu) d\mu$$

subject to (B.1). This is because if there were a finite solution, say $(q_{\sigma_1|\sigma}^{T'}(\mu), p_{\sigma_1|\sigma}^{T'}(\mu))$, then a new allocation $(q_{\sigma_1|\sigma}^{T''}(\mu), p_{\sigma_1|\sigma}^{T''}(\mu))$ can be constructed where

$$q_{\sigma_1|\sigma}^{T''}(\mu) = q_{\sigma_1|\sigma}^{T'}(\mu) + \frac{\Delta}{\theta(\sigma, \mu)}, \text{ and } p_{\sigma_1|\sigma}^{T''}(\mu) = p_{\sigma_1|\sigma}^{T'}(\mu) + \frac{\Delta}{\beta}.$$

With this new allocation, (B.1) is still satisfied, but $U(q_{\sigma_1|\sigma}^{T'}, p_{\sigma_1|\sigma}^{T'}; \sigma)$ can be lowered due to $\beta < 1$. Therefore, the threat allocation can be chosen to deter all $\sigma > \sigma_1$ from misreporting as σ_1 .

Next, observe that if (B.1) is satisfied and $q_{\sigma_1|\sigma}^T(\mu) > q_{\sigma_1}^R(\mu)$ for all σ and μ , then a truthful σ_1 consumer would always choose the real allocation at $t = 1$ for any realization of μ .

Finally, the process above can be repeated for all signals $\sigma > \sigma_1$. This completes the proof. ■

Proof of Lemma 2: If $\left\{ (q_\sigma^R(\mu), p_\sigma^R(\mu))_{\mu \in [0, \bar{\mu}_\sigma]} \right\}_{\sigma \in \Sigma}$ satisfies (2.1), then it must be the case that for any $\sigma \in \Sigma$ and $\mu \in [0, \bar{\mu}]$,

$$\mu = \arg \max_{\hat{\mu}} \theta(\sigma, \mu) q_\sigma^R(\hat{\mu}) - \beta p_\sigma^R(\hat{\mu}).$$

Therefore, $V(q_\sigma^R(\mu), p_\sigma^R(\mu); \mu) = \max_{\hat{\mu}} \theta(\sigma, \mu) q_\sigma^R(\hat{\mu}) - \beta p_\sigma^R(\hat{\mu})$ is the maximum of a family of affine functions, so V is convex, which implies absolute continuity and differentiability almost everywhere. It must be the case that the first order condition

$$\theta(\sigma, \mu) \frac{\partial q_\sigma^R(\mu)}{\partial \mu} = \beta \frac{\partial p_\sigma^R(\mu)}{\partial \mu}. \quad (\text{B.2})$$

and the second order condition

$$\theta(\sigma, \mu) \frac{\partial^2 q_\sigma^R(\mu)}{\partial \mu^2} - \beta \frac{\partial^2 p_\sigma^R(\mu)}{\partial \mu^2} \leq 0. \quad (\text{B.3})$$

are satisfied. Further differentiating (B.2), we have

$$\frac{\partial \theta(\sigma, \mu)}{\partial \mu} \frac{\partial q_\sigma^R(\mu)}{\partial \mu} + \theta(\sigma, \mu) \frac{\partial^2 q_\sigma^R(\mu)}{\partial \mu^2} - \beta \frac{\partial^2 p_\sigma^R(\mu)}{\partial \mu^2} = 0.$$

By (B.3) and the fact that θ is increasing in μ , we have that q_σ^R must be weakly increasing in μ . This shows necessity.

To show sufficiency, suppose both monotonicity and local incentive compatibility constraints are satisfied. However, there exists a $\sigma \in \Sigma$ and $\mu, \mu' \in [0, \bar{\mu}_\sigma]$ such that (2.1) is violated:

$$\theta(\sigma, \mu) q_\sigma^R(\mu') - \beta p_\sigma^R(\mu') > \theta(\sigma, \mu) q_\sigma^R(\mu) - \beta p_\sigma^R(\mu).$$

This implies that

$$\int_\mu^{\mu'} \theta(\sigma, \mu) \frac{\partial q_\sigma^R(x)}{\partial x} - \beta \frac{\partial p_\sigma^R(x)}{\partial x} dx > 0.$$

Hence, if $\mu' > \mu$ there exists a $\tilde{\mu} \in (\mu, \mu')$ such that

$$\theta(\sigma, \mu) \frac{\partial q_\sigma^R(\tilde{\mu})}{\partial \tilde{\mu}} - \beta \frac{\partial p_\sigma^R(\tilde{\mu})}{\partial \tilde{\mu}} > 0.$$

Since q_σ^R is non-decreasing and $\theta(\sigma, \mu) < \theta(\sigma, \tilde{\mu})$, then

$$\theta(\sigma, \tilde{\mu}) \frac{\partial q_\sigma^R(\tilde{\mu})}{\partial \tilde{\mu}} - \beta \frac{\partial p_\sigma^R(\tilde{\mu})}{\partial \tilde{\mu}} > 0,$$

which violates local incentive compatibility. With a similar argument, a contradiction can be obtained when $\mu' < \mu$. This shows sufficiency. ■

Proof of Lemma 3: If the real allocations are ex-post incentive compatible, then by Lemma 2, for any $\mu \in [0, \bar{\mu}_\sigma]$,

$$\theta(\sigma, \bar{\mu}_\sigma) q_\sigma^R(\bar{\mu}_\sigma) - \beta p_\sigma^R(\bar{\mu}_\sigma) \geq \theta(\sigma, \bar{\mu}_\sigma) q_\sigma^R(\mu) - \beta p_\sigma^R(\mu).$$

Hence, if $\theta(\sigma, \bar{\mu}_\sigma) q_\sigma^T - \beta p_\sigma^T = \theta(\sigma, \bar{\mu}_\sigma) q_\sigma^R(\bar{\mu}_\sigma) - \beta p_\sigma^R(\bar{\mu}_\sigma)$, then

$$\theta(\sigma, \bar{\mu}_\sigma) [q_\sigma^T - q_\sigma^R(\mu)] \geq \beta [p_\sigma^T - p_\sigma^R(\mu)].$$

Since θ is increasing in μ and $q_\sigma^T \geq q_\sigma^R(\mu)$ and $p_\sigma^T \geq p_\sigma^R(\mu)$ for all $\mu \in [0, \bar{\mu}_\sigma]$, then for any $\mu' \in (\bar{\mu}_\sigma, 1]$ and $\mu \in [0, \bar{\mu}_\sigma]$,

$$\theta(\sigma, \mu') [q_\sigma^T - q_\sigma^R(\mu)] > \beta [p_\sigma^T - p_\sigma^R(\mu)].$$

As a result, consumers with shock $\mu' \in (\bar{\mu}_\sigma, 1]$ would choose the threat allocation.

Next, suppose there exists some $\mu \in [0, \bar{\mu}_\sigma]$ such that

$$\theta(\sigma, \mu) q_\sigma^T - p_\sigma^T > \theta(\sigma, \mu) q_\sigma^R(\mu) - \beta p_\sigma^R(\mu).$$

Since the real allocations are ex-post incentive compatible, this implies

$$\theta(\sigma, \mu) q_\sigma^T - p_\sigma^T > \theta(\sigma, \mu) q_\sigma^R(\bar{\mu}_\sigma) - \beta p_\sigma^R(\bar{\mu}_\sigma).$$

Hence, $\theta(\sigma, \mu) [q_\sigma^T - q_\sigma^R(\bar{\mu}_\sigma)] > \beta [p_\sigma^T - p_\sigma^R(\bar{\mu}_\sigma)]$. Finally, because θ is increasing in μ , the following must be true:

$$\theta(\sigma, \bar{\mu}_\sigma) q_\sigma^T - \beta p_\sigma^T > \theta(\sigma, \bar{\mu}_\sigma) q_\sigma^R(\bar{\mu}_\sigma) - \beta p_\sigma^R(\bar{\mu}_\sigma).$$

This is a contradiction, which completes the necessity part of the proof.

To show sufficiency, suppose that real allocations and $\left\{ (q_\sigma^T(\mu), p_\sigma^T(\mu))_{\mu \in (\bar{\mu}_\sigma, 1]} \right\}_{\sigma \in \Sigma}$ are ex-post incentive compatible. It has to be the case that

$$\theta(\sigma, \bar{\mu}_\sigma) q_\sigma^T(\bar{\mu}_\sigma) - \beta p_\sigma^T(\bar{\mu}_\sigma) = \theta(\sigma, \bar{\mu}_\sigma) q_\sigma^R(\bar{\mu}_\sigma) - \beta p_\sigma^R(\bar{\mu}_\sigma). \quad (\text{B.4})$$

A violation of (B.4) would violate ex-post incentive compatibility. Also, Lemma 2 holds for the threat allocations too if it satisfies ex-post incentive compatibility, which implies $q_\sigma^T(\mu)$ is non-decreasing in μ for all σ and $\mu \in (\bar{\mu}_\sigma, 1]$ and

$$\theta(\sigma, \mu) \frac{\partial q_\sigma^T(\mu)}{\partial \mu} = \beta \frac{\partial p_\sigma^T(\mu)}{\partial \mu}.$$

Suppose $q_\sigma^T(\mu)$ varies with μ , then by revenue equivalence, $q_\sigma^T(\mu) = q_\sigma^R(\mu)$ for all $\mu \in (\bar{\mu}_\sigma, 1]$. Therefore, for the threat to satisfy monotonicity, local incentive compatibility and (B.4), it has to be the case that $q_\sigma^T(\mu) = q_\sigma^T(\mu') = q_\sigma^T$ and $p_\sigma^T(\mu) = p_\sigma^T(\mu') = p_\sigma^T$ for all

$\mu, \mu' \in (\bar{\mu}_\sigma, 1]$. Finally, suppose $q_\sigma^T < q_\sigma^R(\bar{\mu}_\sigma)$ and $p_\sigma^T < p_\sigma^R(\bar{\mu}_\sigma)$ for some $\sigma \in \Sigma$. If (B.4) holds, then for any $\mu > \bar{\mu}_\sigma$, consumers would prefer the real allocations over the threat allocations. Hence, $q_\sigma^T \geq q_\sigma^R(\bar{\mu}_\sigma)$ and $p_\sigma^T \geq p_\sigma^R(\bar{\mu}_\sigma)$ for all $\sigma \in \Sigma$. This completes the proof.

■

Proof of Theorem 12: Set $q_\sigma^R(\mu) = q_\sigma^*(\mu)$ for any $\sigma \in \Sigma$ and $\mu \in [0, \bar{\mu}_\sigma]$ with $\bar{\mu}_\sigma \in (0, 1)$, so that the first best quantity is consumed for real allocations. Therefore, by Lemma 2, for any $\sigma \in \Sigma$ and $\mu \in [0, \bar{\mu}_\sigma]$, write $p_\sigma^R(\mu) = \Psi_\sigma(\mu) + \Omega_\sigma(\bar{\mu}_\sigma)$, where $\Psi_\sigma(\mu)$ is pinned down by $q_\sigma^R(\mu)$ and $\Omega_\sigma(\bar{\mu}_\sigma)$ will be determined later.

First, I will show that a threat can be constructed to deter deviations. Notice that only immediate downward deviations need to be deterred. In essence, threats need to deter σ_{n+1} from misreporting as σ_n for global ex-ante incentive compatibility to hold.

Let $\mu_{n+1}(\mu)$ be a function that satisfies $\theta(\sigma_{n+1}, \mu_{n+1}(\mu)) = \theta(\sigma_n, \mu)$ for any $\mu \in [0, 1]$. If a consumer with σ_{n+1} signal misreports as σ_n at $t = 0$, then he chooses the real allocations for any $\mu \in [0, \mu_{n+1}(\bar{\mu}_n)]$ and the threat allocations for any $\mu \in (\mu_{n+1}(\bar{\mu}_n), 1]$. Since μ is private information, for any $\mu \in [0, \mu_{n+1}(\bar{\mu}_n)]$, at $t = 1$ the consumer solves

$$\begin{aligned} \max_{\hat{\mu}} \theta(\sigma_{n+1}, \mu) q_n^R(\hat{\mu}) - \beta p_n^R(\hat{\mu}) &= \max_{\hat{\mu}} \theta(\sigma_n, \mu_{n+1}^{-1}(\mu)) q_n^R(\hat{\mu}) - \beta p_n^R(\hat{\mu}) \\ &= \theta(\sigma_n, \mu_{n+1}^{-1}(\mu)) q_n^R(\mu_{n+1}^{-1}(\mu)) - \beta p_n^R(\mu_{n+1}^{-1}(\mu)) \\ &= \theta(\sigma_{n+1}, \mu) q_n^R(\mu_{n+1}^{-1}(\mu)) - \beta p_n^R(\mu_{n+1}^{-1}(\mu)). \end{aligned}$$

The first and third equality comes from the definition of $\mu_{n+1}(\cdot)$ and the fact that the inverse is well-defined. The second equality comes from ex-post incentive compatibility. Hence, a locally misreporting σ_{n+1} consumer at $t = 1$ can expect

$$\int_0^{\mu_{n+1}(\bar{\mu}_n)} \theta(\sigma_{n+1}, \mu) q_n^R(\mu_{n+1}^{-1}(\mu)) - p_n^R(\mu_{n+1}^{-1}(\mu)) d\mu + \int_{\mu_{n+1}(\bar{\mu}_n)}^1 \theta(\sigma_{n+1}, \mu) q_n^T(\bar{\mu}_n) - p_n^T(\bar{\mu}_n) d\mu.$$

Let $\Phi(\bar{\mu}_n; \Omega(\bar{\mu}_n)) = \int_0^{\mu_{n+1}(\bar{\mu}_n)} \theta(\sigma_{n+1}, \mu) q_n^R(\mu_{n+1}^{-1}(\mu)) - p_n^R(\mu_{n+1}^{-1}(\mu)) d\mu$. Choose the threat allocation $(q_n^T(\bar{\mu}_n), p_n^T(\bar{\mu}_n))$ so that

$$\int_{\mu_{n+1}(\bar{\mu}_n)}^1 \theta(\sigma_{n+1}, \mu) q_n^T(\bar{\mu}_n) - p_n^T(\bar{\mu}_n) d\mu = -\Phi(\bar{\mu}_n; \Omega(\bar{\mu}_n)), \quad (\text{B.5})$$

and

$$\theta(\sigma_n, \bar{\mu}_n) q_n^T(\bar{\mu}_n) - \beta p_n^T(\bar{\mu}_n) = \theta(\sigma_n, \bar{\mu}_n) q_n^R(\bar{\mu}_n) - \beta p_n^R(\bar{\mu}_n). \quad (\text{B.6})$$

If (B.5) holds, then by the participation constraint (2.3), the ex-ante incentive compatibility constraint (2.2) of σ_{n+1} consumer would hold as well. By Lemma 3, if (B.6) is satisfied, then the threat is ex-post incentive compatible. Let $(q_n^{T*}(\bar{\mu}_n; \Omega(\bar{\mu}_n)), p_n^{T*}(\bar{\mu}_n; \Omega(\bar{\mu}_n)))$ denote the threat allocation that satisfies (B.5) and (B.6). By repeating the argument above, it is possible to find the set of threat allocations such that both the ex-post incentive compatibility (2.1) and ex-ante incentive compatibility (2.2) constraints are satisfied. By Cramer's rule,

$$q_n^{T*}(\bar{\mu}_n) = \frac{\beta \Phi(\bar{\mu}_n; \Omega(\bar{\mu}_n)) + [1 - \mu_{n+1}(\bar{\mu}_n)] \Gamma(\bar{\mu}_n; \Omega(\bar{\mu}_n))}{\int_{\mu_{n+1}(\bar{\mu}_n)}^1 \theta(\sigma_{n+1}, \mu_{n+1}(\bar{\mu}_n)) - \beta \theta(\sigma_{n+1}, \mu) d\mu},$$

$$p_n^{T*}(\bar{\mu}_n) = \frac{\Gamma(\bar{\mu}_n; \Omega(\bar{\mu}_n)) \int_{\mu_{n+1}(\bar{\mu}_n)}^1 \theta(\sigma_{n+1}, \mu) d\mu + \theta(\sigma_n, \bar{\mu}_n) \Phi(\bar{\mu}_n; \Omega(\bar{\mu}_n))}{\int_{\mu_{n+1}(\bar{\mu}_n)}^1 \theta(\sigma_{n+1}, \mu_{n+1}(\bar{\mu}_n)) - \beta \theta(\sigma_{n+1}, \mu) d\mu},$$

where $\Gamma(\bar{\mu}_n; \Omega(\bar{\mu}_n)) = \theta(\sigma_n, \bar{\mu}_n) q_n^R(\bar{\mu}_n) - \beta p_n^R(\bar{\mu}_n)$. Furthermore, it is possible to show that

$$q_n^{T*}(\bar{\mu}_n) < \frac{\bar{\theta} q_n^R(1) [1 - (1 - \beta) \mu_{n+1}(\bar{\mu}_n)] - \beta \left[\int_0^{\mu_{n+1}(\bar{\mu}_n)} p_n^R(\mu_{n+1}^{-1}(\mu)) d\mu + \int_{\mu_{n+1}(\bar{\mu}_n)}^1 p_n^R(\bar{\mu}_n) d\mu \right]}{[1 - \mu_{n+1}(\bar{\mu}_n)] [\theta(\sigma_n, \bar{\mu}_n) - \beta \bar{\theta}]}. \quad (\text{B.7})$$

Next, I will show that the participation constraint (2.3) binds for all $\bar{\mu}_\sigma$ and $\sigma \in \Sigma$. To see this, notice that for any $\bar{\mu}_\sigma$, one needs to find $\Omega(\bar{\mu}_\sigma)$ such that

$$\int_0^{\bar{\mu}_\sigma} \theta(\sigma, \mu) q_\sigma^R(\mu) - \Psi_\sigma(\mu) d\mu + \int_{\bar{\mu}_\sigma}^1 \theta(\sigma, \mu) q_\sigma^{T*}(\bar{\mu}_\sigma; \Omega(\bar{\mu}_\sigma)) - p_\sigma^{T*}(\bar{\mu}_\sigma; \Omega(\bar{\mu}_\sigma)) d\mu = \bar{\mu}_\sigma \Omega_\sigma(\bar{\mu}_\sigma).$$

Substituting in $(q_n^{T*}(\bar{\mu}_n; \Omega(\bar{\mu}_n)), p_n^{T*}(\bar{\mu}_n; \Omega(\bar{\mu}_n)))$, one gets

$$\Omega_n(\bar{\mu}_n) = \frac{a(\bar{\mu}_n) + b(\bar{\mu}_n) c(\bar{\mu}_n) \int_{\bar{\mu}_n}^1 \theta(\sigma_n, \mu) d\mu - (1 - \bar{\mu}_n) b(\bar{\mu}_n) d(\bar{\mu}_n)}{\bar{\mu}_n + \beta b(\bar{\mu}_n) \int_{\bar{\mu}_n}^1 \theta(\sigma_n, \mu) d\mu - (1 - \bar{\mu}_n) b(\bar{\mu}_n) e(\bar{\mu}_n)},$$

where

$$\begin{aligned}
a(\bar{\mu}_n) &= \int_0^{\bar{\mu}_\sigma} \theta(\sigma, \mu) q_\sigma^R(\mu) - \Psi_\sigma(\mu) d\mu, \\
b(\bar{\mu}_n) &= \left[\int_{\mu_{n+1}(\bar{\mu}_n)}^1 \theta(\sigma_{n+1}, \mu_{n+1}(\bar{\mu}_n)) - \beta \theta(\sigma_{n+1}, \mu) d\mu \right]^{-1}, \\
c(\bar{\mu}_n) &= \beta \int_0^{\mu_{n+1}(\bar{\mu}_n)} \theta(\sigma_{n+1}, \mu) q_n^R(\mu_{n+1}^{-1}(\mu)) - \Psi_n(\mu_{n+1}^{-1}(\mu)) d\mu \\
&\quad + \int_{\mu_{n+1}(\bar{\mu}_n)}^1 \theta(\sigma_n, \bar{\mu}_n) q_n^R(\bar{\mu}_n) - \beta \Psi_n(\bar{\mu}_n) d\mu, \\
d(\bar{\mu}_n) &= \theta(\sigma_n, \bar{\mu}_n) \int_0^{\mu_{n+1}(\bar{\mu}_n)} \theta(\sigma_{n+1}, \mu) q_n^R(\mu_{n+1}^{-1}(\mu)) - \Psi_n(\mu_{n+1}^{-1}(\mu)) d\mu \\
&\quad + [\theta(\sigma_n, \bar{\mu}_n) q_n^R(\bar{\mu}_n) - \beta \Psi_n(\bar{\mu}_n)] \int_{\mu_{n+1}(\bar{\mu}_n)}^1 \theta(\sigma_{n+1}, \mu) d\mu, \\
e(\bar{\mu}_n) &= \theta(\sigma_n, \bar{\mu}_n) \mu_{n+1}(\bar{\mu}_n) + \beta \int_{\mu_{n+1}(\bar{\mu}_n)}^1 \theta(\sigma_{n+1}, \mu) d\mu.
\end{aligned}$$

Since $\lim_{\theta \rightarrow \bar{\theta}} \pi(\theta|\sigma_n)/\pi(\theta|\sigma_{n+1}) = 0$, then

$$\lim_{\bar{\mu}_n \rightarrow 1} \frac{1 - \bar{\mu}_n}{1 - \mu_{n+1}(\bar{\mu}_n)} = 0. \tag{B.8}$$

Hence, $\lim_{\bar{\mu}_n \rightarrow 1} \Omega_n(\bar{\mu}_n) < \infty$.

Finally, to see how perfect price discrimination is virtually implementable, notice that

$$\begin{aligned}
W^* - W(\bar{\mu}) &= \sum_{n=1}^N \gamma_n \int_{\bar{\mu}_n}^1 \theta(\sigma_n, \mu) [q_n^*(\mu) - q_n^T(\bar{\mu}_n)] - [c(q_n^*(\mu)) - c(q_n^T(\bar{\mu}_n))] d\mu \\
&< \sum_{n=1}^N (1 - \bar{\mu}_n) \gamma_n \{ \bar{\theta} q_n^*(1) - \underline{\theta} q_n^T(\bar{\mu}_n) + [c(q_n^T(\bar{\mu}_n)) - c(q_n^*(\bar{\mu}_n))] \} \\
&< \sum_{n=1}^N (1 - \bar{\mu}_n) \gamma_n \{ [c'(q_n^T(\bar{\mu}_n)) - \underline{\theta}] q_n^T(\bar{\mu}_n) + \bar{\theta} q_n^*(1) \} \\
&< \sum_{n=1}^N (1 - \bar{\mu}_n) \gamma_n [(M - \underline{\theta}) q_n^T(\bar{\mu}_n) + \bar{\theta} q_n^*(1)].
\end{aligned}$$

By (B.7), (B.8) and the fact that $\lim_{\bar{\mu}_n \rightarrow 1} \Omega_n(\bar{\mu}_n) < \infty$, it follows $\lim_{\mu \rightarrow 1} W(\bar{\mu}) = W^*$. ■

Appendix C

Appendix for Chapter 3

C.1 Estimating the Marginal Cost of a Life Insurance Company

In what follows, let $m_{t,n,\bar{n}}$ denote the period t mortality rate of age n individuals who bought life insurance at age \bar{n} ¹, and let N be the maximum attainable age according to the corresponding mortality tables. Let $R_t(i)$ be the (annualized) interest rate on zero-coupon risk-free securities with maturity i at time t . Further, denote by $\ell_{x,\bar{n},i}$ the lapsation rate after the i th year of an x -year level-term insurance policy, where $x \in \{1, 5, 10, 20\}$, purchased at age \bar{n} (i.e. the probability that a policyholder fails to pay in time the outstanding premium). Finally, let $\gamma_{t,n,\bar{n}}$ stand for the growth rate in premium between age n and $n+1$ for a schedule posted at time t ². We define the net premium for an x level-term policy acquired at age n per dollar of death benefit as

$$V_t^x(n) \equiv \left(\sum_{i=1}^{N-n} \frac{\prod_{j=0}^{i-2} (1 - m_{t,n+j,n})(1 - \ell_{x,n,j+1})m_{t,n+i-1,n}}{R_t^i(i)} \right) \left(1 + \sum_{i=1}^{N-n-1} \frac{\prod_{j=0}^{i-1} (1 - m_{t,n+j,n})(1 - \ell_{x,n,j+1})\gamma_{t,n+j,n}}{R_t^i(i)} \right)^{-1} \quad (\text{C.1})$$

¹It is important to keep track of different cohorts of the insured due to adverse selection, i.e. individuals who have already held a policy tend to have significantly higher mortality rates than the same-age newcomers.

²Notice that, by construction, $\gamma_{t,n,\bar{n}} = 1$ for all n such that $\text{mod}(n - \bar{n}, x) \neq 0$, i.e. depending on the policy term, premiums are allowed to increase only every x years.

The net premium of a renewable term life insurance is difficult to calculate and to the best of our knowledge there is no agreed-upon way to do so. The first problem is that the premium schedule is increasing in age and different companies apply different growth schemes. The second problem relates to lapsations which are not modeled or predicted easily. In the pricing of other life insurance products (e.g. universal or whole life insurance), characterized by significant front-loading of premium schedules, lapsation rates are often disregarded since a policyholder’s incentive to lapse decreases over time. With short-term renewable policies though, lapsation is an important factor because most customers acquire them for a limited number of years only. For this reason, in equation (C.1) we propose a modified version of the standard formula (see e.g. Koijen and Yogo (2015)), in that we account for lapsation as well as the premium schedule that increases in age.

In order to obtain actuarially fair growth rates of the premium, we implement the following algorithm. First, compute formula (C.1) assuming the policy is a whole-life insurance (premiums are constant forever, i.e. $\gamma_{t,n+j,n} = 1, \forall j > 0$). Having obtained an increasing sequence of actuarial values at renewal dates across different age profiles $V_t^x(n), V_t^x(n+1), \dots, V_t^x(N-1)$, we next calculate the resulting actual growth rates, i.e. $\gamma_{t,n+j,n} = V_t^x(n+j+1)/V_t^x(n+j), \forall j > 0$. Finally, we plug the growth rates obtained in this way in formula (C.1) to compute the net premium across time. It is important to notice that a life insurance company can choose among many different premium growth patterns which are all actuarially fair (i.e. they all equate the present expected value of cash flows between the company and the consumer). The advantage of our method is that it pins down the growth rates that result purely from increased mortality rates due to aging.

In our calculation of the net premium we use the mortality tables issued by the American Society of Actuaries. We apply the 1980 Commissioners Standard Ordinary (CSO) table for all years prior to January 2001, the 2001 Valuation Basic Table (VBT) prior to January 2008 and the 2008 VBT for the time period following January 2008. We use geometric averaging on the monthly basis to smooth the transition between any two vintages of the mortality tables. It is important to emphasize that these tables are created based on the actual mortality rates among the insured rather than the general population. For this reason, they account for a potential adverse selection in the market for life insurance.³ As for the lapsation data, we use the rates published for yearly renewable term products in

³Cawley and Philipson (1999) found no strong evidence of adverse selection in the term life insurance.

LIMRA (1996). As the risk-free interest rate we use the U.S. Treasury zero-coupon yield curve⁴.

C.2 Incentive Compatibility Constraints

From (3.7), we can write out all nine incentive compatibility constraints for the second period for any \hat{c}_1 :

$$[P_2(c_2) - c_2] D_2(P_1, P_2(c_2)) \geq [P_2(\tilde{c}_2) - c_2] D_2(P_1, P_2(\tilde{c}_2)), \forall c_2, \tilde{c}_2 \in \mathcal{C}_2^h \quad (\text{C.2})$$

$$[P_2(c_2) - c_2] D_2(P_1, P_2(c_2)) \geq [P_2(\tilde{c}_2) - c_2] D_2(P_1, P_2(\tilde{c}_2)), \forall c_2, \tilde{c}_2 \in \mathcal{C}_2^m \quad (\text{C.3})$$

$$[P_2(c_2) - c_2] D_2(P_1, P_2(c_2)) \geq [P_2(\tilde{c}_2) - c_2] D_2(P_1, P_2(\tilde{c}_2)), \forall c_2, \tilde{c}_2 \in \mathcal{C}_2^l \quad (\text{C.4})$$

$$[P_2(c_2) - c_2] D_2(P_1, P_2(c_2)) \geq [P_2(\tilde{c}_2) - c_2] D_2(P_1, P_2(\tilde{c}_2)), \forall c_2 \in \mathcal{C}_2^h, \forall \tilde{c}_2 \in \mathcal{C}_2^m \quad (\text{C.5})$$

$$[P_2(c_2) - c_2] D_2(P_1, P_2(c_2)) \geq [P_2(\tilde{c}_2) - c_2] D_2(P_1, P_2(\tilde{c}_2)), \forall c_2 \in \mathcal{C}_2^h, \forall \tilde{c}_2 \in \mathcal{C}_2^l \quad (\text{C.6})$$

$$[P_2(c_2) - c_2] D_2(P_1, P_2(c_2)) \geq [P_2(\tilde{c}_2) - c_2] D_2(P_1, P_2(\tilde{c}_2)), \forall c_2 \in \mathcal{C}_2^m, \forall \tilde{c}_2 \in \mathcal{C}_2^h \quad (\text{C.7})$$

$$[P_2(c_2) - c_2] D_2(P_1, P_2(c_2)) \geq [P_2(\tilde{c}_2) - c_2] D_2(P_1, P_2(\tilde{c}_2)), \forall c_2 \in \mathcal{C}_2^m, \forall \tilde{c}_2 \in \mathcal{C}_2^l \quad (\text{C.8})$$

$$[P_2(c_2) - c_2] D_2(P_1, P_2(c_2)) \geq [P_2(\tilde{c}_2) - c_2] D_2(P_1, P_2(\tilde{c}_2)), \forall c_2 \in \mathcal{C}_2^l, \forall \tilde{c}_2 \in \mathcal{C}_2^h \quad (\text{C.9})$$

$$[P_2(c_2) - c_2] D_2(P_1, P_2(c_2)) \geq [P_2(\tilde{c}_2) - c_2] D_2(P_1, P_2(\tilde{c}_2)), \forall c_2 \in \mathcal{C}_2^l, \forall \tilde{c}_2 \in \mathcal{C}_2^m \quad (\text{C.10})$$

C.3 Proofs

Proof of Lemma 4: Since we defined $\underline{\mathcal{C}}_2 = \mathcal{C}_2^l = \{c_2 \mid P_2(c_2) \leq \bar{r}_2^E\}$, by (3.6) and (3.5) we have the following demand for $c_2 \in \underline{\mathcal{C}}_2$,

$$D_2(P_1, P_2(c_2)) = (1 - m_1) \left[\int_{\bar{r}_2^N}^R \int_0^R h(r_1, r_2) dr_1 dr_2 + \int_{\bar{r}_2^E}^{\bar{r}_2^N} \int_{\chi(r_2)}^R h(r_1, r_2) dr_1 dr_2 \right].$$

The insurance company takes the threshold valuations as given, so for cost realizations in $\underline{\mathcal{C}}_2$, the demand is independent of the variations in second period premiums.

Suppose $c_2 \in \underline{\mathcal{C}}_2$ is the true cost realization and the insurance company contemplates reporting a cost of $\tilde{c}_2 \in \underline{\mathcal{C}}_2$, then by (C.4) the incentive compatible premium schedule requires $P_2(c_2) \geq P_2(\tilde{c}_2)$. Now suppose \tilde{c}_2 is the true cost realization and the insurance company contemplates reporting a cost of c_2 , then the incentive compatible premium schedule requires $P_2(c_2) \leq P_2(\tilde{c}_2)$. Therefore, we have $P_2(c_2) = P_2(\tilde{c}_2)$ for all cost realizations and

⁴Taken from Gurkaynak et al. (2007) and averaged for each month.

deviations within the set $\underline{\mathcal{C}}_2$. ■

Proof of Proposition 7: For part (i), we first assume that $\underline{\mathcal{C}}_2$ is of measure zero. This yields $\Pr [P_2(c_2) \leq \bar{r}_2^E] = 0$, then by (3.2) we have that \bar{r}_2^E is not the second period threshold valuation of the existing policyholders, because the outside option is strictly positive. By efficiency, this implies that there exists a new threshold valuation $\bar{r}_2^{E'} > \bar{r}_2^E$ such that (3.2) binds and $\Pr [P_2(c_2; \hat{c}_1) \leq \bar{r}_2^{E'}] > 0$. We can redefine $\underline{\mathcal{C}}_2$ in terms of the new threshold $\bar{r}_2^{E'}$ and it is not measure zero, which is a contradiction.

To see why part (ii) is true, we start by analyzing (C.2) and notice that in this cost region

$$D_2(P_1, P_2(c_2)) = (1 - m_1) [1 - H_2(P_2(c_2))]. \quad (\text{C.11})$$

We will assume that $P_2(c_2)$ is incentive compatible and differentiable on \mathcal{C}_2^h , then for a given $c_2 \in \mathcal{C}_2^h$ the following must be true

$$c_2 = \arg \max_{\hat{c}_2 \in \mathcal{C}_2^h} (P_2(\hat{c}_2) - c_2) D_2(P_1, P_2(\hat{c}_2)). \quad (\text{C.12})$$

Therefore, (C.12) implies that the life insurance company will set a monopoly premium in the cost region \mathcal{C}_2^h , which is represented in the following fixed point problem:

$$P_2^*(c_2) = \frac{1 - H_2(P_2^*(c_2))}{h_2(P_2^*(c_2))} + c_2, \forall c_2 \in \mathcal{C}_2^h. \quad (\text{C.13})$$

Notice that the incentive compatible second period premium schedule for cost realizations within the set \mathcal{C}_2^h and the set itself do not depend the announced first period cost realization.

Similarly, we can examine (C.3) and notice that in this cost region the demand is

$$D_2(P_1, P_2(c_2)) = (1 - m_1) \left\{ [1 - H_2(\bar{r}_2^N)] + \int_{P_2(c_2)}^{\bar{r}_2^N} \int_{\chi(r_2)}^R h(r_1, r_2) dr_1 dr_2 \right\}. \quad (\text{C.14})$$

Following a similar argument, the life insurance company will set a monopoly premium $P_2^*(c_2)$ in the cost region \mathcal{C}_2^m according to the following equation

$$\left\{ [1 - H_2(\bar{r}_2^N)] + \int_{P_2^*(c_2)}^{\bar{r}_2^N} \int_{\chi(r_2)}^R h(r_1, r_2) dr_1 dr_2 \right\} = [P_2^*(c_2) - c_2] \left\{ \int_{\chi(P_2^*(c_2))}^R h(r_1, P_2^*(c_2)) dr_1 \right\}. \quad (\text{C.15})$$

The incentive compatible second period premium for cost realizations within the set \mathcal{C}_2^m and the set itself depends on the announced first period cost realization.

Notice that (C.5) and (C.7) hold trivially since the premiums are chosen to maximize each independent segment of the demand which also maximizes the whole demand for $c_2 \in \bar{\mathcal{C}}_2$.

Part (iii) follows from the usual kinked demand result. We can equate the solution from (C.13) to equal \bar{r}_2^N to find c_2^h , and apply the same process to (C.15) to find c_2^m . We now show that $c_2^h > c_2^m$.

From (C.13) we have the following relationship between \bar{r}_2^N and c_2^h

$$\bar{r}_2^N = \frac{1 - H_2(\bar{r}_2^N)}{h_2(\bar{r}_2^N)} + c_2^h. \quad (\text{C.16})$$

Similarly, since $\chi(\bar{r}_2^N) = \delta$, then from (C.15) we have the following relationship between \bar{r}_2^N and c_2^m

$$1 - H_2(\bar{r}_2^N) = (\bar{r}_2^N - c_2^m) \int_{\delta}^R h(r_1, \bar{r}_2^N) dr_1. \quad (\text{C.17})$$

We substitute (C.16) into (C.17) and by the definition of marginal probability distribution for $h_2(\cdot)$ we can derive the following

$$(\bar{r}_2^N - c_2^h) \int_0^R h(r_1, \bar{r}_2^N) dr_1 = (\bar{r}_2^N - c_2^m) \int_{\delta}^R h(r_1, \bar{r}_2^N) dr_1.$$

Since $\delta \geq 0$, we have $c_2^h \geq c_2^m$. Next, we show that for any c_2 in $[c_2^m, c_2^h)$ the optimal premium is \bar{r}_2^N .

To show rigidity within $[c_2^h, c_2^m)$, suppose there exists a $c_2 \in (c_2^m, c_2^h)$ such that the optimal premium $P_2(c_2)$ is strictly greater than \bar{r}_2^N . Since the optimal premium has to be incentive compatible, we have the following

$$\begin{aligned} (\bar{r}_2^N - c_2^h) D_2(P_1, \bar{r}_2^N) &\geq [P_2(c_2) - c_2^h] D_2(P_1, P_2(c_2)) \\ &= [P_2(c_2) - c_2] D_2(P_1, P_2(c_2)) - (c_2^h - c_2) D_2(P_1, P_2(c_2)) \\ &\geq (\bar{r}_2^N - c_2) D_2(P_1, \bar{r}_2^N) - (c_2^h - c_2) D_2(P_1, P_2(c_2)) \\ &> (\bar{r}_2^N - c_2) D_2(P_1, \bar{r}_2^N) - (c_2^h - c_2) D_2(P_1, \bar{r}_2^N) \\ &= (\bar{r}_2^N - c_2^h) D_2(P_1, \bar{r}_2^N). \end{aligned}$$

The first and third inequality follows from the incentive compatibility constraints. The fourth inequality follows from a weakly decreasing demand and our assumption that $P_2(c_2) > \bar{r}_2^N$ and $c_2 < c_2^h$. We have a contradiction.

We now assume there exists a $c_2 \in (c_2^m, c_2^h)$ such that the optimal premium $P_2(c_2)$ is strictly smaller than \bar{r}_2^N . Similarly, with incentive compatibility, we can show

$$\begin{aligned}
(\bar{r}_2^N - c_2^m)D_2(P_1, \bar{r}_2^N) &\geq [P_2(c_2) - c_2^m] D_2(P_1, P_2(c_2)) \\
&= [P_2(c_2) - c_2] D_2(P_1, P_2(c_2)) + (c_2 - c_2^m)D_2(P_1, P_2(c_2)) \\
&\geq (\bar{r}_2^N - c_2)D_2(P_1, \bar{r}_2^N) + (c_2 - c_2^m)D_2(P_1, P_2(c_2)) \\
&> (\bar{r}_2^N - c_2)D_2(P_1, \bar{r}_2^N) + (c_2 - c_2^m)D_2(P_1, \bar{r}_2^N) \\
&= (\bar{r}_2^N - c_2^m)D_2(P_1, \bar{r}_2^N).
\end{aligned}$$

The first and third inequality follows from the incentive compatibility constraints. The fourth inequality follows from a weakly decreasing demand and our assumption that $P_2(c_2) < \bar{r}_2^N$ and $c_2 > c_2^m$. We have a contradiction.

To prove part (iv), suppose $\exists c_2$ and $\bar{\epsilon} > 0$ such that $\forall \epsilon \in (0, \bar{\epsilon})$, we have $P_2(c_2 + \epsilon) < P_2(c_2)$. Notice that with part (i) and (ii), we can simply analyze the two cases: (i) $c_2 \in \underline{\mathcal{C}}_2$ and $c_2 + \epsilon \in \bar{\mathcal{C}}_2$, and (ii) $c_2 \in \bar{\mathcal{C}}_2$ and $c_2 + \epsilon \in \underline{\mathcal{C}}_2$. We first examine case (i). Since (3.12) and (3.13) must hold, we have the following

$$\begin{aligned}
[P_2(c_2 + \epsilon) - (c_2 + \epsilon)] D_2(P_1, P_2^*(c_2)) &\geq [\bar{P}_2 - (c_2 + \epsilon)] \bar{D}_2 \\
&> [\bar{P}_2 - c_2] \bar{D}_2 \\
&\geq [P_2(c_2 + \epsilon) - c_2] D_2(P_1, P_2^*(c_2)),
\end{aligned}$$

which is a contradiction since $\epsilon > 0$. Note the first and third inequality come from (3.12) and (3.13) respectively.

Next, we examine case (ii). Since (3.13) must hold, we have the following

$$\begin{aligned}
[\bar{P}_2 - (c_2 + \epsilon)] \bar{D}_2 &\geq [P_2^*(c_2) - (c_2 + \epsilon)] D_2(P_1, P_2^*(c_2)) \\
&= [P_2^*(c_2) - c_2] D_2(P_1, P_2^*(c_2)) - \epsilon D_2(P_1, P_2^*(c_2)) \\
&\geq [\bar{P}_2 - c_2] \bar{D}_2 - \epsilon D_2(P_1, P_2^*(c_2)).
\end{aligned}$$

The second inequality follows from (3.12). The analysis implies that $D_2(P_1, P_2^*(c_2)) \geq \bar{D}_2$. Since the demand function is weakly decreasing, it must be the case that $P_2^*(c_2) \leq \bar{r}_2^E$ and $c_2 \in \underline{\mathcal{C}}_2$. Therefore, (C.4) is violated if $P_2^*(c_2) > \bar{P}_2$ and $c_2 \in \underline{\mathcal{C}}_2$, which gives us the contradiction.

Part (v) follows immediately by examining (3.12) for the second period premium. For the second period premium, we know that $P_2^*(c_2) > \bar{P}_2$ for any $c_2 \in \bar{\mathcal{C}}_2$. Furthermore, by

part (iii), we know that the optimal premium schedule is weakly increasing in c_2 . Therefore, the result follows. ■

Proof of Lemma 5: We will first claim that the following must be true:

$$(\bar{P}_2 - c_2^T) \bar{D}_2 = [P_2^*(c_2^T) - c_2^T] D_2(P_1, P_2^*(c_2^T)).$$

The expression above has to hold, since if not, then there exists $\epsilon > 0$ such that we have a $c_2 \in (c_2^T - \epsilon, c_2^T + \epsilon)$ where one of the incentive compatibility constraints would not hold.

We first show that (3.13) holds trivially. For any $c_2 \in \underline{\mathcal{C}}_2$ and $\tilde{c}_2 \in \bar{\mathcal{C}}_2$, we have the following

$$\begin{aligned} [P_2^*(\tilde{c}_2) - c_2] D_2(P_1, P_2^*(\tilde{c}_2)) &= [P_2^*(\tilde{c}_2) - c_2^T] D_2(P_1, P_2^*(\tilde{c}_2)) + (c_2^T - c_2) D_2(P_1, P_2^*(\tilde{c}_2)) \\ &\leq [P_2^*(c_2^T) - c_2^T] D_2(P_1, P_2^*(c_2^T)) + (c_2^T - c_2) D_2(P_1, P_2^*(\tilde{c}_2)) \\ &= (\bar{P}_2 - c_2^T) \bar{D}_2 + (c_2^T - c_2) D_2(P_1, P_2^*(\tilde{c}_2)) \\ &\leq (\bar{P}_2 - c_2^T) \bar{D}_2 + (c_2^T - c_2) \bar{D}_2 \\ &= (\bar{P}_2 - c_2) \bar{D}_2. \end{aligned}$$

The first inequality follows from the insurance company's optimization problem. The second equality follows from (3.14). The second inequality follows from the monotonically decreasing demand function and parts (iv) and (v) of Proposition 7 and the fact that $c_2 \leq c_2^T$.

Next, we will show that (3.12) holds. $\forall c_2 \in \bar{\mathcal{C}}_2$,

$$\begin{aligned} [P_2^*(c_2) - c_2] D_2(P_1, P_2^*(c_2)) &\geq [P_2^*(c_2^T) - c_2] D_2(P_1, P_2^*(c_2^T)) \\ &= [P_2^*(c_2^T) - c_2^T] D_2(P_1, P_2^*(c_2^T)) - (c_2 - c_2^T) D_2(P_1, P_2^*(c_2^T)) \\ &= (\bar{P}_2 - c_2^T) \bar{D}_2 - (c_2 - c_2^T) D_2(P_1, P_2^*(c_2^T)) \\ &= (\bar{P}_2 - c_2) \bar{D}_2 + (c_2 - c_2^T) [\bar{D}_2 - D_2(P_1, P_2^*(c_2^T))] \\ &\geq (\bar{P}_2 - c_2) \bar{D}_2. \end{aligned}$$

The first inequality follows from the insurance company's optimization problem. The second equality follows from (3.14). The second inequality follows from the the definition of $\bar{\mathcal{C}}_2$ and from the monotonically decreasing demand function and the fact that $c_2 \geq c_2^T$. ■

Proof of Proposition 8: We first show that there is a discrete jump at c_2^T where $\bar{P}_2 < P_2^*(c_2^T)$. Note that Proposition 7 has ruled out the case where $\bar{P}_2 > P_2^*(c_2^T)$. We proceed by ruling out the case for $\bar{P}_2 = P_2^*(c_2^T)$.

Suppose that $\bar{P}_2 = P_2^*(c_2^T)$, then by Lemma 5, we have that (3.14) holds, which implies the following

$$\bar{D}_2 = D_2(P_1, P_2^*(c_2^T)).$$

By (3.5) and (3.6), we have $\bar{r}_2^E = P_2^*(c_2^T)$. Therefore, $E[\bar{r}_2^E - P_2(c_2) \mid P_2(c_2) \leq \bar{r}_2^E] = 0$, which violates the definition of \bar{r}_2^E . We have a contradiction and this proves that there is a discontinuity in the premium schedule at c_2^T .

To show that $\bar{r}_2^E > \bar{P}_2$, notice that by the definition of \bar{r}_2^E shown in (3.2) can be expressed as

$$(\bar{r}_2^E - \bar{P}_2) G(c_2^T) = V_2^S(\bar{r}_2^E).$$

Since the expected value of switching is assumed to be strictly positive, $\bar{r}_2^E > \bar{P}_2$ follows.

Next, to show that $P_2^*(c_2^T) > \bar{r}_2^E$, first observe that by Lemma 5 and the fact that $\bar{P}_2 < P_2^*(c_2^T)$, it must be the case that $\bar{D}_2 > D_2(P_1, P_2^*(c_2^T))$. By the definition of second period demand, we have

$$\frac{\bar{D}_2}{1 - m_1} = \left[\int_{\bar{r}_2^N}^R \int_0^R h(r_1, r_2) dr_1 dr_2 + \int_{\bar{r}_2^E}^{\bar{r}_2^N} \int_{\chi(r_2)}^R h(r_1, r_2) dr_1 dr_2 \right],$$

and

$$\frac{D_2(P_1, P_2^*(c_2^T))}{1 - m_1} = \left[\int_{\max\{\bar{r}_2^N, P_2^*(c_2^T)\}}^R \int_0^R h(r_1, r_2) dr_1 dr_2 + \int_{\max\{\bar{r}_2^E, \min\{P_2^*(c_2^T), \bar{r}_2^N\}\}}^{\bar{r}_2^N} \int_{\chi(r_2)}^R h(r_1, r_2) dr_1 dr_2 \right].$$

Suppose $\bar{r}_2^E \geq P_2^*(c_2^T)$, then $\bar{D}_2 = D_2(P_1, P_2^*(c_2^T))$, which is a contradiction. As a result, at the cutoff, $P_2^*(c_2^T) > \bar{r}_2^E > \bar{P}_2$ must hold. ■

Proof of Proposition 9: Using 3.14, we can express the incentive compatible \bar{P}_2 as a function of c_2^T . Hence, we can also express \bar{D}_2 as a function of the second period cutoff c_2^T . To complete the proof, we need to take the first order condition with respect to c_2^T . ■

Proof of Proposition 10: For a given optimal second period premium, the optimal first period premium is characterized by the first order condition (10). Also, since the second period premium is optimal, $\frac{\partial \Pi_2}{\partial P_1}$ can be found using the envelope theorem.

To show the last part of the proposition, notice that P_1^{pNR} is characterized by (3.16) without the second period premium, and P_1^{pR} is characterized by (3.16) with a first period demand function without transaction cost μ . When $\mu = 0$, no initial investment is required and there is cost to switching. In other words, there are no dynamic considerations, so consumers buy if and only if $r_t \geq P_t$ for each period. Therefore, in the absence of health shocks, a renewable contract is the same as a non-renewable contract, so they are priced the same: $P_1^{pNR} = P_1^{pR}$.

Finally, when $\mu = 0$, we have $\frac{\partial \Pi_2}{\partial P_1} = 0$ due to the fact that maximizing period by period is equivalent to maximizing present discounted value. However, if $\mu > 0$, by the definition of δ and equations (3.5) and (3.6), the envelope theorem yields us $\frac{\partial \Pi_2}{\partial P_1} < 0$. More specifically,

$$\frac{\partial \Pi_2}{\partial P_1} = -(1 - m_1) \int_{\max\{\bar{r}_2^E, \min\{P_2(c_2), \bar{r}_2^N\}\}} h(\chi(r_2), r_2) dr_2.$$

Let D_1^p denote the first period demand for life insurance in an environment without transaction cost, then

$$D_1^p(P_1) = \int_{P_1}^R \int_0^R h(r_1, r_2) dr_1 dr_2,$$

then we have

$$\frac{\partial D_1^p}{\partial P_1} = - \int_0^R h(P_1, r_2) dr_2.$$

Also, by (3.4), we have

$$\frac{\partial D_1}{\partial P_1} = - \int_{\bar{r}_2^N}^R h(\delta, r_2) dr_2 - \int_{\bar{r}_2^E}^{\bar{r}_2^N} h(\chi(r_2), r_2) dr_2.$$

Thus, for (3.16) to hold, it must be the case that $P_1 < P_1^{pNR} = P_1^{pR}$. ■