Optimal Social Insurance, Private Information, and Non-exclusive Contracting

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Pricila Maziero

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Larry E. Jones, Patrick J. Kehoe

July 2009
Acknowledgments

I am greatly indebted to Larry Jones, Patrick Kehoe, and V.V. Chari for their continuous help, guidance and support. I would like to thank Arpad Abraham, Rajesh Aggarwal, Andy Atkeson, Francesca Carapella, Simona Cociuba, Kim-Sau Chung, Mike Golosov, Roozbeh Hosseini, Narayana Kocherlakota, Erzo Luttmer, Ellen McGrattan, Lee Ohanian, Fabrizio Perri, Chris Phelan, Anderson Schneider, Martin Schneider, Chris Sleet, Aleh Tsyvinski, Gianluca Violante, Pierre Yared, and Sevin Yeltekin and Warren Weber for comments and suggestions. I am grateful to Art Rolnick and Warren Weber for allowing me to spend valuable years at the Minneapolis Fed. I thank Fabrizio Perri and Dirk Krueger for providing the CEX data.

Special thanks for Laurence Ales for the years of friendship and inumerous discussions. This thesis is based on joint work with Laurence.

My greatest thanks are to my parents, siblings and Anderson for the unconditional love and support.
Dedication

This dissertation is dedicated to my parents.
## Contents

Acknowledgements

Dedication

List of Tables

List of Figures

I Introduction

II Accounting For Private Information

1 Introduction

2 Environment

3 Characterizing the Allocation

### 2 Environment

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Recursive formulation</td>
<td>11</td>
</tr>
<tr>
<td>2.2 Optimality conditions</td>
<td>13</td>
</tr>
<tr>
<td>2.3 The role of publicly observed shocks</td>
<td>15</td>
</tr>
</tbody>
</table>

### 3 Characterizing the Allocation

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 The static allocation</td>
<td>18</td>
</tr>
<tr>
<td>3.2 The multi-period allocation</td>
<td>20</td>
</tr>
</tbody>
</table>
16 Characterization of Equilibrium

16.1 Implementation of Equilibrium ........................................ 150

16.2 The Rothschild-Stiglitz Equilibrium ................................. 152

17 Appendix

17.1 Proofs of Section 16 ................................................... 155

References ................................................................. 160
List of Tables

1  Population statistics for the static environment . . . . . . . . . . . . 19
6  Consumption response to income shocks . . . . . . . . . . . . . . . 50
2  Sample selection for PSID and CEX . . . . . . . . . . . . . . . . . 65
3  Benchmark estimation . . . . . . . . . . . . . . . . . . . . . . . . . 66
4  Summary of moments . . . . . . . . . . . . . . . . . . . . . . . . . . 67
5  Parameter estimates, robustness checks . . . . . . . . . . . . . . . . 68
7  Summary statistics for the PSID and CEX samples used. . . . . . . 69
8  Estimation results for risk aversion . . . . . . . . . . . . . . . . . . 110
9  Estimation results for risk aversion: multiple education groups . . . 111
10 Estimation results for risk aversion . . . . . . . . . . . . . . . . . . 112
11 Estimation results for risk aversion . . . . . . . . . . . . . . . . . . 113
12 Intra temporal distortion: singles . . . . . . . . . . . . . . . . . . . 115
13 Intra temporal distortion: singles . . . . . . . . . . . . . . . . . . . 116
14 Intra temporal distortion: couples . . . . . . . . . . . . . . . . . . . 117
List of Figures

1. Policy functions for consumption ............................................. 17
2. Support of promised utility by age ........................................... 21
3. Panel (a) changes in individual expected intratemporal distortions;
   panel (b) changes in the individual variance of continuation utility .... 23
4. Panel (a) variance of hours by size of labor productivity shocks; panel
   (b) variance of consumption by size of labor productivity shocks. ... 24
5. Covariances for different amount of private information. ............... 26
6. Data: Life-cycle profiles ...................................................... 30
7. Data: Life-cycle profiles ...................................................... 31
8. Benchmark environment fit on matched moments. ......................... 36
9. Evolution of the average distortion on the marginal of substitution between con-
   sumption and leisure over age. ............................................. 38
10. Transfers over age. ............................................................ 39
11. Score function and criterion function ...................................... 43
12. Estimation result with alternate utility function. ......................... 48
13. Estimation result with alternate utility function. ......................... 49
14. Estimation of consumption response to income shocks. ................. 51
15. Testing for randomization .................................................... 62
<table>
<thead>
<tr>
<th>16</th>
<th>Separating Equilibrium</th>
<th>149</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>Pooling Equilibrium</td>
<td>154</td>
</tr>
<tr>
<td>18</td>
<td>Pooling Equilibrium and Latent Contracts</td>
<td>154</td>
</tr>
</tbody>
</table>
Part I

Introduction

This thesis studies the optimal provision of social insurance in the presence of asymmetric information. Agents are subject to idiosyncratic shocks and sign contracts to insure against them. The main friction present in all the environments studied in this thesis is that these shocks are privately observed by agents. Each chapter of the thesis considers a particular environment with asymmetric information and for each case we characterize the optimal contract and study its quantitative implications.

The focus of chapter 2 is to study the quantitative properties of constrained efficient allocations in an environment where risk sharing is limited by the presence of private information. We consider a life cycle version of a standard Mirrlees economy where shocks to labor productivity have a component that is public information and one that is private information. Agents sign an exclusive contract with a planner that prescribes the consumption and output allocations. The presence of private shocks has important implications for the age profiles of consumption and income implied by the optimal contract. First, they introduce an endogenous dispersion of continuation utilities. As a result, consumption inequality rises with age even if the variance of the shocks does not. Second, they introduce an endogenous rise of the distortion on the marginal rate of substitution between consumption and leisure over the life cycle. This is because, as agents age, the ability to properly provide
incentives for work must become less and less tied to promises of benefits (through either increased leisure or consumption) in future periods. Both of these features are also present in the US survey data. We look at the data through the lens of our model and estimate the fraction of labor productivity that is private information. We find that for the model and data to be consistent, a large fraction of shocks to labor productivity must be private information.

In chapter 3 we study how the presence of non-exclusive contracts limits the amount of insurance provided in a decentralized economy. We consider a dynamic Mirrleesian economy in which agents are privately informed about idiosyncratic labor productivity shocks. Agents sign privately observable insurance contracts with multiple firms (i.e., they are non-exclusive), which include both labor supply and savings aspects. Firms have no restriction on the contracts they can offer, interact strategically. In equilibrium, contrary to the case with exclusive contracts, a standard Euler equation holds, the marginal rate of substitution between consumption and leisure is equated to the worker’s marginal productivity. Also, each agent receives zero net present value of transfers. To sustain this equilibrium, more than one firm must be active and must also offer latent contracts to deter deviations to more profitable contingent contracts. In this environment, the non-observability of contracts removes the possibility of additional insurance beyond self-insurance. To test the model, we allow firms to observe contracts at a cost. The model endogenously divides the population into agents that are not monitored and have access to
non-exclusive contracts and agents that have access to exclusive contracts. We use US survey data and find that high school graduates satisfy the optimality conditions implied by the non-exclusive contracts while college graduates behave according to the second group.

Chapter 4 considers the Rothschild and Stiglitz (1976) insurance environment relaxing the assumption of exclusivity of insurance contracts. Agents are privately informed about the probability of their income realization, which is publicly observed. Agents can engage in multiple insurance contract simultaneously and the terms of these contracts are not observed by other firms. Insurance providers behave non-cooperatively and compete offering menus of insurance contracts from an unrestricted contract space. We derive conditions under which a separating equilibrium exists and fully characterize it. The equilibrium allocation consists of agents with a lower probability of accident purchasing no insurance and agents with higher accident probability buying the actuarially fair competitive level of insurance. The equilibrium allocation also constitutes a linear price schedule for insurance. To sustain this allocation firms must offer latent contracts. We show that latent menus can prevent cream-skimming strategies, however pooling equilibrium still fails to exists.
Part II

Accounting For Private Information\textsuperscript{1}

1 Introduction

How well are workers able to smooth consumption and hours over their working life? Several studies have shown that, at best, the level of insurance available to workers is imperfect.\textsuperscript{2} Given that the efficient level of insurance is incompatible with the data, in this paper we ask whether the observed data can be rationalized as the outcome of a constrained efficient allocation.

To answer this question, we study the quantitative properties of constrained efficient allocations in an environment where risk sharing is limited by the presence of private information. In our environment, workers are subject to idiosyncratic labor productivity risk through their working lives. We assume that shocks to labor productivity have a component that is public information and one that is private information of the worker. Depending on the fraction of these shocks that is private information, the optimal contract features different degrees of insurance against income shocks. This enables us to draw a link between the amount of insurance we observe in the data and the amount of private information in our model. Looking

\textsuperscript{1}This chapter is coauthored with Laurence Ales.

\textsuperscript{2}See, for example, Cochrane (1991), Townsend (1994), Storesletten, Telmer, and Yaron (2001), and Attanasio and Davis (1996).
at the data through the lens of our model, we calibrate the amount of private information needed for the model to be consistent with the data. Our findings show that a calibrated version of a dynamic Mirrlees economy, like the one studied in this paper, with all of the uncertainty on labor productivity being private information, is consistent with the evolution of inequality of consumption and hours over the working life.

Household data for the U.S. show that workers are subject to large income fluctuations over the working life and that these fluctuations transmit only partially to consumption.\(^3\) Looking at the cross section, we observe that inequality in consumption is increasing over age.\(^4\) At the same time, the profile for the cross-sectional variance in hours worked is slightly decreasing over the working life. As shown in Cochrane (1991) and Storesletten, Telmer, and Yaron (2001), these facts suggest that workers are partially insured against idiosyncratic shocks.

The study of contractual arrangements that can explain the lack of full insurance is the underlying motivation for Green (1987), Thomas and Worrall (1990), Atkeson and Lucas (1992). These papers show that a repeated moral hazard environment with privately observed taste shocks or endowments can qualitatively account for two key features observed in the data: consumption responding to income shocks and the cross-sectional distribution of consumption increasing over time. Our interest is

\(^3\)See, for example, Cochrane (1991), Dynarski and Gruber (1997), and Gervais and Klein (2006).

\(^4\)See, for example, Deaton and Paxson (1994) and Heathcote, Storesletten, and Violante (2005).
in studying jointly the behavior of consumption and hours; for this reason we focus on an environment where the source of asymmetric information is the worker’s labor productivity, as in Mirrlees (1971) and Golosov, Kocherlakota, and Tsyvinski (2003).

In our model, the allocation of consumption and hours along the working life is described by an optimal incentive compatible contract. To prevent misreporting realized productivity shocks, skilled workers are rewarded with higher current consumption and higher continuation utility. The provision of incentives within the period (intratemporal distortion) translates to an increase in the covariance between consumption and labor productivity and a decrease in the covariance between hours and labor productivity with respect to the unconstrained optimum. This reflects the basic trade-off between efficiency and incentives faced in the optimal contract. As a consequence, the variance of consumption increases and the variance of hours decreases as the intratemporal distortion increases.

As originally shown in Green (1987), the provision of incentives between periods introduces an endogenous dispersion of continuation utilities. As a result, consumption inequality rises with age even if the variance of the shocks does not. A key difference in our environment is the presence of a finite horizon in the optimal contract. This implies that the increase in the dispersion of promised utility will be large early in life and will progressively slow down. This is because, as workers age, the ability to properly provide incentives for work must become less tied to promises of benefits (through either increased leisure or consumption) in future periods. As a
consequence, the provision of incentives will progressively rely more on the intratemporal distortion. This is the key mechanism that allows us to reconcile the private information environment with the data: as the intratemporal distortion increases over the working life, the cross-sectional variance of hours will remain flat or decreasing while the variance of consumption will continue to increase. This is in stark contrast to the case where labor productivity is entirely public information. In this case, as shown in Storesletten, Telmer, and Yaron (2001), any increase in the cross-sectional variance of consumption is followed by an increase in the cross-sectional variance of hours.

We solve the model numerically and use the simulated method of moments to determine parameter values. Our targets are the variances of consumption and hours along the life cycle. Our baseline estimated model can account for the increase in consumption inequality over the working life and the slight decrease in the inequality in hours that we observe in the data. In the calibrated model, 99% of the labor productivity shock is private information. The result is robust to different specifications of the utility function, different target moments, heterogeneity in initial promised utility levels, and persistence of the publicly observable component of labor productivity.

This paper is related to a growing literature that studies the distortions implied by the optimal contract in dynamic versions of the original Mirrlees environment. The focus of most of this literature is normative, looking at decentralization through
taxes in environments where the government is the sole provider of insurance. Few papers have looked at the empirical implications of the allocations of such constrained efficient problems.\textsuperscript{5} Our contribution with respect to this literature is to quantitatively characterize the allocation and the distortions along the working life, highlighting the role of observables such as age and the public component of labor productivity in the implied intratemporal distortions. In addition, we show that the data display characteristics that we would expect to originate from the optimal contract. This result raises the question, left for future research, of which existing institutional arrangements implement the constrained efficient allocation.

Papers similar to ours are Phelan (1994) and Attanasio and Pavoni (2007). The first studies how the evolution of consumption inequality generated in a standard agency problem (as in Phelan and Townsend (1991)) relates to US data.\textsuperscript{6} The key difference of our paper is the focus on a Mirrlees environment, which generates jointly the behavior of consumption and hours worked and using both series allows us to identify the amount of private information. The second focuses on a moral hazard problem with hidden savings and shows, analyzing equilibrium restrictions, how private information can explain the excess smoothness in consumption in data from the United Kingdom.


\textsuperscript{6}Also, Ai and Yang (2007) study an environment with private information and limited commitment that can account for the elasticity of consumption growth to income growth found in U.S data.
This paper is also related to a recent literature that studies an environment where workers have access to insurance that is in addition to what is available through precautionary savings. Blundell, Pistaferri, and Preston (2008) and Heathcote, Storesletten, and Violante (2007) study environments where workers are subject to two types of shocks – some that are completely insured, some are entirely uninsured. With respect to these papers, the assets available in our environment, and hence the level of insurance provided at different ages, are determined endogenously.

The paper is structured as follows: section 2 describes the environment, section 3 studies the qualitative implications of the environment, section 4 presents the data, section 5 presents our estimation strategy and results, and section 6 concludes.

2 Environment

In this section we describe the main features of the environment and define the optimal insurance contract between a planner and the workers.

Our environment is a standard dynamic Mirrlees economy similar to Golosov, Kocherlakota, and Tsyvinski (2003) and Albaris and Sleet (2006). Consider an infinite horizon economy. In every period \( t \) a new generation is born and is composed of a continuum of measure 1 of workers. Each generation lives for a finite number of periods \( N \) and every worker works for \( T \) periods, with \( T < N \). Given our focus on the effects of the incentive mechanisms during the working life, we
constrain the analysis to the ages 1 to $T$. Throughout the paper, we consider the optimal contract signed by a worker and a planner during working age.\footnote{In our environment, there is no moral hazard problem after retirement; hence, retirement can be fully characterized by the continuation utility assigned at time $T$ denoted by $w_T$. Our approach is to assume that the planner assigns to each worker the same level of $w_T$. There might be welfare gains from allowing the planner to choose $w_T$ optimally as an additional instrument for providing incentives to agents at time $T$.} A large literature on dynamic optimal contracts considers contracts with infinite length. In our environment, solving a contract with finite length has important implications for the allocations of consumption, hours, and income, which will be explained in the next section.

In addition to the standard dynamic Mirrlees environment, our environment features the presence of idiosyncratic public shocks together with idiosyncratic private shocks. This allows us to study the interaction between the two shocks and, in the quantitative analysis, the relative importance of each.

Each worker has utility defined over consumption and leisure. Assume that utility is additively separable over time, and let the period utility function be denoted by

$$u(c, l) : \mathbb{R}^2_+ \rightarrow \mathbb{R}. \tag{1}$$

Assume that $u$ is twice continuously differentiable, increasing, and concave in both arguments. Agents discount future utility at the constant rate $\beta < 1$. Given a sequence of consumption and leisure $\{c_t, l_t\}_{t=1}^T$, the expected discounted utility over
the working life is given by
\[
W \{c_t, l_t\}_{t=1}^T = \mathbb{E}_0 \sum_{t=1}^T \beta^{t-1} [u(c_t, l_t)],
\]
(2)
where \(\mathbb{E}_0\) denotes the expectation with respect to the information available at age \(t = 0\).

Uncertainty in ages 1, \ldots, \(T\) is in the form of labor productivity shocks. At every age, a worker is subject to two idiosyncratic labor productivity shocks, \(\theta_t \in \Theta_t\) and \(\eta_t \in H_t\). Let \(\theta^t \equiv (\theta_1, \ldots, \theta_t) \in \Theta^t\) and \(\eta^t \equiv (\eta_1, \ldots, \eta_t) \in H^t\) denote the histories of the shocks up to age \(t\). For a given realization of the labor productivity shocks, a worker can produce \(y\) units of effective output according to the following relation:
\[
y_t = f(\theta_t, \eta_t) \cdot l_t,
\]
(3)
where \(l_t\) denotes his labor input. Assume \(f\), the total labor productivity, is increasing in each argument. We assume the labor input is private information of the worker.
At every age \(t\), the worker learns the realizations of his labor productivity shocks \(\theta_t\) and \(\eta_t\). The shock \(\theta\) is publicly observed by all workers (from now on, we will call it the public shock), while the shock \(\eta\) is privately observed by the worker (the private shock). Let \(\pi(\theta^T, \eta^T)\) denote the probability of drawing a particular sequence of productivity shocks \(\theta^T\) and \(\eta^T\). We assume the following

**Assumption 1.** (a) For every age the public and private shocks are identically and independently distributed across workers.
(b) The realization of the private shock is independent of the realization of the public shock: \( \pi(\theta^T, \eta^T) = \pi_\theta(\theta^T) \pi_\eta(\eta^T) \).

(c) The shocks are independent over age: \( \pi_\theta(\theta^t|\theta^{t-1}) = \pi_\theta(\theta^t) \) and \( \pi_\eta(\eta^t|\eta^{t-1}) = \pi_\eta(\eta^t) \).

The purpose of the second assumption is to isolate the private information nature of the private shocks, so that nothing can be inferred from the realization of the public shock. Assumption 3-(c) is for tractability purposes. The contribution of private information to labor productivity uncertainty is summarized by \( \Omega \), the fraction of the variance of labor productivity due to private information,

\[
\Omega_t = \frac{\sigma^2_t(\eta)}{\sigma^2_t(\eta) + \sigma^2_t(\theta)} \in [0, 1].
\]

If \( \Omega = 1 \), all of the shocks to labor productivity are private information; if \( \Omega = 0 \), all of the shocks are public information.

\(^8\)In section 5.3 we relax this assumption by looking at the effects of a persistent public shock. Adding serial correlation to the privately observed shock is left for future work. Extending the model along this direction introduces several obstacles. From a computational point of view, the difficulty is in characterizing the optimal contract given the history dependent reporting strategies that must be considered by the planner and the lack of common prior on the type of the agent (some of these issues have been addressed in Fernandes and Phelan (2000) and Doepke and Townsend (2006)). Also, the presence of persistent private information introduces an additional age varying component in the provision of incentives (besides the one emphasized in this paper) making the identification of the amount of private information less transparent.
At age $t = 1$, before any uncertainty is realized, a worker signs an exclusive contract with a planner that provides insurance against labor productivity shocks over his working life. We solve for the optimal contract in this environment. Due to the revelation principle, we can restrict our study to direct mechanisms in which workers report truthfully the realization of the productivity shocks to the planner.

The contract specifies, conditional on the realized history of public shock $\theta_t$ and the reported history of private shock $\eta^t$, a level of required effective output and a level for consumption. Denote the contract by $\{c, y\} = \{c_t(\theta_t, \eta^t), y_t(\theta_t, \eta^t)\}_{t=1}^T$. Note that the planner’s problem is not subject to any aggregate uncertainty.

A contract $\{c, y\}$ is incentive compatible if it satisfies the following:

\begin{align*}
\sum_{t=1}^T \sum_{\theta_t, \eta^t} \pi_{\theta}(\theta^t) \pi_{\eta}(\eta^t) \beta^{t-1} u \left( c_t(\theta^t, \eta^t), \frac{y_t(\theta^t, \eta^t)}{f(\theta_t, \eta_t)} \right) \geq \sum_{t=1}^T \sum_{\theta_t, \tilde{\eta}^t} \pi_{\theta}(\theta^t) \pi_{\eta}(\eta^t) \beta^{t-1} u \left( c_t(\theta^t, \tilde{\eta}^t), \frac{y_t(\theta^t, \tilde{\eta}^t)}{f(\theta_t, \eta_t)} \right), \quad \forall \tilde{\eta}^t \in \tilde{H}^t.
\end{align*}

Note that in our environment, full insurance against productivity shocks is not incentive compatible. The intuition for this is straightforward if we assume that the period utility is separable in consumption and leisure. Efficiency implies that under full information, highly skilled workers should work more hours while at the same time all workers should receive the same consumption allocation independent of the realization of the productivity shocks. This contract is clearly not incentive compatible in the presence of private information, since an agent with high productivity shock is better off reporting a low productivity shock. In appendix 7.1 we extend
this argument to the case with a nonseparable (Cobb-Douglas) utility function.

The planner has access to a technology that allows transferring resources linearly over time at the constant rate $1/q$. A contract $\{c, y\}$ is feasible if it satisfies the following:  

$$
\sum_{t=1}^{T} \sum_{\theta^t, \eta^t} \pi_{\theta^t}(\theta^t) \pi_{\eta^t}(\eta^t) q^{t-1} \left( c_t (\theta^t, \eta^t) - y_t (\theta^t, \eta^t) \right) = 0. \quad (6)
$$

In this environment, the planner offers a contract that solves the following problem:

$$
\max_{\{c, y\}_{t=0}^T} \sum_{t=1}^{T} \sum_{\theta^t, \eta^t} \pi_{\theta^t}(\theta^t) \pi_{\eta^t}(\eta^t) \beta^{t-1} u \left( c_t (\theta^t, \eta^t), y_t (\theta^t, \eta^t) \right) \left( \frac{f(\theta_t, \eta_t)}{f(\theta_t, \eta_t)} \right) \quad (7)
$$

s.t. (5) and (6)

### 2.1 Recursive formulation

To compute the solution to the planner’s problem, it is convenient to rewrite the above problem recursively. We write the problem using as a state variable the continuation lifetime utility, as in Spear and Srivastava (1987) and Green (1987).

In addition, instead of solving the above utility maximization problem, we solve its dual cost minimization problem. To allow for ex-ante heterogeneity, before any uncertainty is realized, each worker is associated with a number $w_0$, which denotes his entitlement of discounted lifetime utility. As in Atkeson and Lucas (1992), we solve the correspondent planner’s problem for each level of promised utility $w_0$ for each worker of generation $t$.

In the recursive formulation we need to distinguish between the problem faced

---

$^9$This feasibility constraint abstracts from inter-generational transfers.
in period $T$, when the planner chooses current consumption and output, and all other periods $t < T$ when the planner chooses current consumption, output, and continuation utility. We refer to the problem for any $t < T$ as the $T - 1$ problem.

From here onward we make the additional assumption that the private information labor productivity shock can take only two values per period $\eta_t \in \{\eta_{H,t}, \eta_{L,t}\}$ with $\eta_{H,t} > \eta_{L,t}$ for all $t$.\footnote{The model can be extended to multiple shock values without any effect to the mechanism described in the next section. This assumption is added for computational reasons, however note that since the quantitative analysis focus on the variance of the allocation consumption and hours this assumption is not too restrictive.} We also consider the relaxed problem, only considering incentive compatibility constraints for the agent that draws $\eta_H$. In appendix 7.2 we show that the relaxed problem is equivalent to the original if the utility function is separable over consumption and leisure.\footnote{In our numerical simulations with nonseparable utility functions, we solve the relaxed problem and verify that the solution for this problem satisfies the constraints of the original problem.}

The period $T$ problem is

$$
S_T(w) = \min_{c,y} \sum_{\theta_T} \pi(\theta_T) \sum_{\eta_T} \pi(\eta_T) \left[ c_T(\theta_T, \eta_T) - y_T(\theta_T, \eta_T) \right],
$$

(8)

$$
\text{s.t.} \quad \sum_{\theta_T} \pi(\theta_T) \sum_{\eta_T} \pi(\eta_T) u \left( c_T(\theta_T, \eta_T), \frac{y_T(\theta_T, \eta_T)}{f(\theta_T, \eta_T)} \right) = w, \quad (9)
$$

$$
u \left( c_T(\theta_T, \eta_H), \frac{y_T(\theta_T, \eta_H)}{f(\theta_T, \eta_H)} \right) \geq u \left( c_T(\theta_T, \eta_L), \frac{y_T(\theta_T, \eta_L)}{f(\theta_T, \eta_H)} \right), \quad \forall \theta_T. \quad (10)
$$

10 The model can be extended to multiple shock values without any effect to the mechanism described in the next section. This assumption is added for computational reasons, however note that since the quantitative analysis focus on the variance of the allocation consumption and hours this assumption is not too restrictive.

11 In our numerical simulations with nonseparable utility functions, we solve the relaxed problem and verify that the solution for this problem satisfies the constraints of the original problem.
The time $T - 1$ problem is

$$S_{T-1}(w) = \min_{c,y,w'} \sum_{\theta} \pi(\theta) \sum_{\eta} \pi(\eta) \left[ c_{T-1}(\theta, \eta) - y_{T-1}(\theta, \eta) + qS_T(w'_{T-1}(\theta, \eta)) \right]$$

s.t. \hspace{1cm} \sum_{\theta} \pi(\theta) \sum_{\eta} \pi(\eta) \left[ u \left( c_{T-1}(\theta, \eta), \frac{y_{T-1}(\theta, \eta)}{f(\theta, \eta)} \right) + \beta w'_{T-1}(\theta, \eta) \right] = \Psi_{12}

$$u \left( c_{T-1}(\theta, \eta_H), \frac{y_{T-1}(\theta, \eta_H)}{f(\theta, \eta_H)} \right) + \beta w'_{T-1}(\theta, \eta_H) \geq 0$$

$$u \left( c_{T-1}(\theta, \eta_L), \frac{y_{T-1}(\theta, \eta_L)}{f(\theta, \eta_H)} \right) + \beta w'_{T-1}(\theta, \eta_L), \forall \theta_{T-1}. \hspace{1cm} (13)$$

At time 0, when the contract is signed, each individual is characterized by an initial level of promised utility $w_0$. The value for the planner of delivering the optimal contract is then given by $S_1(w_0)$. In our simulations the distribution of $w_0$, denoted by $\pi_w(w_0)$, is chosen so that $\sum_{w_0} \pi_w(w_0) S_1(w_0) = 0$.

### 2.2 Optimality conditions

The presence of private information, together with the nonstationarity of the problem, limits the ability to characterize analytically the optimal allocation. One of the few analytical results that can be derived relies on applying variational methods to the planner problem. This approach has been used by Rogerson (1985) and in an environment similar to ours by Golosov, Kocherlakota, and Tsyvinski (2003). The key result is that it is optimal for the planner to equate expected marginal cost whenever possible. Equating marginal costs requires the planner to be able to transfer resources between different nodes of the contract in an incentive feasible way (by a node we refer a particular history of labor productivity shocks at a given age). For
example, the Euler equation for marginal cost derived by Golosov, Kocherlakota, and Tsyvinski (2003) requires the planner at every period \( t \) to be able to transfer resources between all of the states at time \( t + 1 \) and the current period. Since time is observable, this transfer can be performed in an incentive feasible way.

The presence of a public shock in our environment enables the planner to make transfers not only between time but also between nodes that are made observable by the presence of the public shock itself. For example, the planner can equate marginal cost between periods for every realization of the publicly observable shock and within periods across different realizations of the public shock. The following proposition states this result. The additional assumption needed is separability between consumption and leisure.

**Proposition 1.** Let \( U(c,l) = u(c) - v(l) \). Necessary conditions for an interior optimal contract are

\[
\frac{1}{u_c(c(\theta^t, \eta^t))} = \frac{q}{\beta} \sum_{\eta^{t+1}} \pi(\eta^{t+1}|\eta^t) \frac{1}{u_c(c(\theta^{t+1}, \eta^{t+1}))}, \quad \forall \theta^{t+1}, \eta^t, t, \quad (14)
\]

\[
\sum_{\eta^t} \frac{\pi(\eta^t|\eta^{t-1})}{u_c(c([\theta^{t-1}, \theta^t], \eta^t))} = \sum_{\eta^t} \frac{\pi(\eta^t|\eta^{t-1})}{u_c(c([\theta^{t-1}, \theta^t], \eta^t))}, \quad \forall \theta^t, \theta_t, \theta^{t-1}, \eta^{t-1}. \quad (15)
\]

**Proof.** In appendix 7.3. □

A direct implication of (14) is the standard inverse Euler derived by Golosov, Kocherlakota, and Tsyvinski (2003). This equation implies that current marginal cost is
equated to the expected future marginal cost:

\[
\frac{1}{u_c(c(\theta^t, \eta^t))} = \frac{q}{\beta} \sum_{\theta^{t+1}, \eta^{t+1}} \pi(\theta^{t+1}|\theta^t) \pi(\eta^{t+1}|\eta^t) \frac{1}{u_c(c(\theta^{t+1}, \eta^{t+1}))}, \quad \forall \theta^t, \eta^t.
\]  

Equation (15) is a novel feature of this environment. It implies that, within a period, the planner equates the inverse of marginal utility of consumption across different realizations of the public shock. If \( \Omega = 0 \), full insurance is incentive feasible, and equation (15) implies that marginal utility of consumption (and hence consumption) is constant across all states.

### 2.3 The role of publicly observed shocks

In this section we determine how consumption is affected by the realization of the public shock. If \( \Omega \neq 0 \), from equation (15) it is not clear whether the worker is fully insured against the realization of the public shock. This is of particular interest, since one of the tests that can be used to reject Pareto optimal allocations (see, for example, Attanasio and Davis (1996)) is based on detecting a covariance different from zero between consumption and a publicly observable characteristic.

In an environment with separable utility and without private information, consumption does not depend on the realization of the idiosyncratic productivity shock. The following proposition shows that when \( f(\theta, \eta) = \theta \cdot \eta \), in the presence of private information, consumption depends on \( \theta \).

**Proposition 2.** Assume \( u(c, l) = u(c) - v(l) \). Let \( v(l) = \frac{\phi}{1+\gamma} l^{1+\gamma} \) and \( f(\theta, \eta) = \theta \cdot \eta \).

Then for any allocation \( \{c, y\} \) that solves the relaxed problem, we have \( c(\theta, \eta) \neq \)
\[ c(\hat{\theta}, \eta) \text{ for all } \theta, \hat{\theta}, \eta. \]

**Proof.** In appendix 7.4.

The key intuition for this result is how different realizations of \( \theta \) can affect the severity of the incentive problem. Define the following variable:

\[ \Delta(\theta) = f(\theta, \eta_H) - f(\theta, \eta_L), \quad \forall \theta. \tag{17} \]

For a given value of \( \theta \), \( \Delta(\theta) \) denotes the effective amount of labor productivity that the worker with realization \( \eta_H \) can misreport. This implies that if \( \Delta(\theta) \) varies with \( \theta \), after a given realization of the public shock, the planner faces a different incentive problem. In the proof of the proposition we show that as a consequence, the multiplier on the incentive compatibility constraint, and hence the level of consumption, depends on \( \theta \). In figure 1-(a) we illustrate the results of the proposition. The plot displays typical policy function for the case with \( f(\theta, \eta) = \theta \cdot \eta \).

On the other hand, if \( f(\theta, \eta) = \theta + \eta \), \( \Delta(\theta) \) is independent of \( \theta \). The policy functions for consumption under this specification are displayed in figure 1-(b). In this case, the allocation for consumption does not depend on the realization of the public shock.

### 3 Characterizing the Allocation

In this section we characterize the properties of the cross-sectional moments for consumption and hours implied by the optimal allocation. Our benchmark parametric
Figure 1: Policy functions for consumption with $\Omega = 0.5$. In panel (a) $f(\theta, \eta) = \theta \eta$; in panel (b) $f(\theta, \eta) = \theta + \eta$.

The Cobb-Douglas utility function is:

$$u(c, l) = \frac{c^\alpha (1 - l)^{1-\alpha}}{1 - \sigma},$$

where the consumption share is $\alpha \in (0, 1)$ and the curvature parameter is $\sigma > 1$. The Cobb-Douglas utility function implies a constant elasticity of substitution between consumption and leisure equal to 1. In section 5.3 we also look at utility functions with different values for the elasticity of substitution, and with nonconstant elasticity of substitution.

We first look at the static environment. This will be the starting point in drawing a connection between the distortions induced by the incentive constraint and the properties of the allocation for consumption and hours.
3.1 The static allocation

We start with the static Mirrleesian benchmark, setting $T = 1$. With only one period, the planner can provide incentives only by distorting consumption and hours with respect to the first best allocation, what we refer to as the intratemporal margin. The distortion on the marginal rate of substitution between consumption and leisure is summarized by the following:\(^{13}\)

$$
\tau_{cl}(\theta, \eta) = 1 + \frac{1}{\theta \eta} \frac{u_l(c(\theta, \eta), l(\theta, \eta))}{u_c(c(\theta, \eta), l(\theta, \eta))}.
$$

(19)

In the full information case efficiency implies that $\tau_{cl} = 0$ and hours are set according to current labor productivity, which induces a volatility of hours directly related to the volatility of labor productivity. Also consumption is determined equating marginal utility of consumption across workers. The only source of volatility of consumption depends on the cross partial derivative of the utility between consumption and leisure.

\(^{12}\)For a detailed review of the literature on the static and dynamic Mirrleesian environment, we refer the reader to Tuomala (1990) and Golosov, Tsyvinski, and Werning (2006).

\(^{13}\)The distortion is not independent of the realization of the public shock. In particular, an agent with a high realization of the public shock will be subject to a higher average distortion.
Table 1: Population statistics for the static environment

<table>
<thead>
<tr>
<th>( \Omega )</th>
<th>( var(c) )</th>
<th>( var(l) )</th>
<th>( E[\tau_d] )</th>
<th>( cov(c, \theta \eta) )</th>
<th>( cov(l, \theta \eta) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.067</td>
<td>0.017</td>
<td>0.069</td>
<td>0.060</td>
<td>0.030</td>
</tr>
<tr>
<td>0.75</td>
<td>0.052</td>
<td>0.030</td>
<td>0.054</td>
<td>0.050</td>
<td>0.036</td>
</tr>
<tr>
<td>0.25</td>
<td>0.023</td>
<td>0.059</td>
<td>0.021</td>
<td>0.032</td>
<td>0.052</td>
</tr>
<tr>
<td>0</td>
<td>0.011</td>
<td>0.080</td>
<td>0</td>
<td>0.024</td>
<td>0.065</td>
</tr>
</tbody>
</table>

Cobb-Douglas utility with \( \sigma = 3 \), \( \alpha = 1/3 \).

When \( \Omega \neq 0 \), due to the cost of providing incentives, it is not optimal for the planner to induce the full information level of hours. This fact is illustrated in table 1. As \( \Omega \) increases from zero, the variance of hours decreases. At the same time the additional rewards to the skilled agent in implementing the desired level of hours cause the variance of consumption to increase. The role of incentives is also illustrated in the last two columns of table 1. Looking at the covariance between consumption and hours with labor productivity, we observe that the response of consumption increases as \( \Omega \) increases, while the response of hours decreases.

In this example, different distortions are achieved by varying \( \Omega \). In a dynamic environment, for any fixed level of \( \Omega \neq 0 \), we observe that the intratemporal distortion changes with age and, in particular, increases endogenously over the life cycle.
3.2 The multi-period allocation

We now look at the dynamic environment and set $T = 6$.\textsuperscript{14} From the incentive constraint (13), we observe that in every period $t < T$, the planner has at its disposal two instruments to induce truthful revelation of the high productivity shock: a worker can be rewarded with high current consumption and leisure and with high future continuation utility. Whenever possible, it is always optimal to provide incentives using the two instruments. We begin by looking at the behavior of continuation utility. From the first order-conditions of the planner problem, we have the following equations that relate current and future marginal cost for the planner:

\begin{align}
\lambda_t + \frac{\mu_t(\theta)}{\pi(\eta_H)\pi(\theta)} &= \frac{q}{\beta} S_{t+1}(w'_t(\theta, \eta_H)), \quad \forall \theta, \quad (20) \\
\lambda_t - \frac{\mu_t(\theta)}{\pi(\eta_L)\pi(\theta)} &= \frac{q}{\beta} S_{t+1}(w'_t(\theta, \eta_L)), \quad \forall \theta, \quad (21)
\end{align}

with $\lambda_t$ the multiplier on the promised utility constraint (9) and $\mu_t(\theta)$ the multiplier on the incentive constraint (10). Equations (20) and (21) determine the evolution of promised utility. A positive multiplier on the incentive constraint, together with the cost function of the planner being increasing and convex imply a spreading out of continuation utilities. In addition, the convexity of the cost function implies that this spreading out is asymmetric.\textsuperscript{15} The evolution of promised utility for an exante

\textsuperscript{14}From here onwards, we assume that a period represents a five-year interval, with the initial period set at age 25.

\textsuperscript{15}The spreading out of continuation utility has been shown numerically by Phelan and Townsend (1991). Asymptotic limit results have also been studied by Thomas and Worrall (1990), Atkeson and Lucas (1992), Aiyagari and Alvarez (1995), and Phelan (1998).
homogeneous population is plotted in figure 2. We observe that the support of promised utility for the population increases over age. Unlike previous results, due to the nonstationarity of the value function, the spread is fast in early periods and slows down as the worker ages.

The mechanism in play is the following: since the worker values smoothing of consumption across time, as he ages, the planner provides insurance substituting progressively from incentives provided on the intertemporal margin (rewarding by varying continuation utility) to incentives provided using the intratemporal (rewarding by varying current consumption and leisure). This is particularly stark in the last period where only current consumption and leisure can be used to provide incentives. To illustrate the implications of the finite horizon effect and differentiate them from the ex post heterogeneity induced in the population, we consider the evolu-

Figure 2: Support of promised utility by age. In panel (a) $\Omega = 1$; in panel (b) $\Omega = \frac{1}{2}$. 
tion of the allocation and continuation utility for the "average" individual. That is, for every age we look at a worker with the mean value of promised utility of the population. We then compute the expected distortion of the intratemporal margin faced by the worker, as well as the conditional variance of continuation utility for the following period. Figure 3 illustrates this result. In this particular example, the intratemporal distortion monotonically increases over age by a factor of 3, while the individual variance of continuation utility monotonically decreases (the same result holds averaging across the population and qualitatively holds for different parameter specifications). The conditional variance decreasing over time explains why the total variance of continuation utility grows at a progressively slower rate. This can be seen by decomposing the total variance by conditioning on current continuation utility:

\[
\text{var}(w_{t+1}) = E[\text{var}(w_{t+1}|w_t)] + \text{var}(E(w_{t+1}|w_t))
\]

\[
= E[\text{var}(w'_t)] + \text{var}(E(w'_t)) .
\]

The first term, as stated, is decreasing while the second one is increasing over age. The behavior of promised utility affects directly the allocations of consumption and hours. The increasing variance of promised utility contributes to an increase in the variance of both over age. However the way incentives are provided has different implications for the variance of consumption and hours. As was noted in the static environment, a high distortion on the intratemporal margin causes hours to vary less
Figure 3: Panel (a) changes in individual expected intratemporal distortions; panel (b) changes in the individual variance of continuation utility. $\Omega = 1$. Values are normalized to 1 for age group 25-30.

with changes in labor productivity. This implies a reduction in the variance of hours as the intratemporal distortion increases. Overall the finite horizon effect, together with the spreading out of continuation utility, makes the evolution of the variance of hours a quantitative question. In figure 4-(a) we observe that for small variances of the labor productivity hours, the incentive effect dominates and variance of hours tends to decrease. For large values of the productivity shock, the spreading out of continuation utility dominates and variance of hours increases.

The effect on the variance of consumption is unambiguous. As the intratemporal distortion increases, so does its effect on the variance of consumption.\textsuperscript{16} The variance

\textsuperscript{16}The increase in the variance of consumption and the effect on the variance of hours depend on the assumption that consumption and leisure are complements ($\sigma > 1$). Under this assumption, the
Figure 4: Panel (a) variance of hours by size of labor productivity shocks; panel (b) variance of consumption by size of labor productivity shocks. Values are normalized to 1 for age group 25-30.

of consumption is increasing over the life cycle due to an increase in the variance of promised utility and an increase in the intratemporal distortion. In figure 4-(b) we observe that variance of consumption increases over the working life and the increase is convex. The convex increase in the variance of consumption (which is also robust once we introduce persistence in the publicly observable component of labor productivity shocks) is a specific prediction of this environment which is different from other models of consumption insurance.\(^{17}\)

\(^{17}\)In an environment with self-insurance with a single bond, if the income process is persistent, the increase in the variance of consumption is concave. This comes from the fact that the realization of uncertainty early in life generates a large heterogeneity in consumption paths early in life. Although the US data displays a roughly linear increase of variance of consumption, Deaton and Paxson (1994)
Finally, we look at the relationship between consumption and output. From equation (3) when $f(\theta, \eta) = \theta \eta$, we have that

$$\log y = \log \theta + \log \eta + \log l. \quad (22)$$

When $\Omega = 1$, using the above we obtain

$$\Delta \text{cov}(\log c, \log y) = \Delta \text{cov}(\log c, \log \eta) + \Delta \text{cov}(\log c, \log l). \quad (23)$$

The increasing distortion in the intratemporal margin will cause (as described in section 3.1) an increase of the covariance between consumption and the privately observed productivity shock (the first term on the right side of equation (23)). In our numerical simulations we observe that the second term in (23) is flat and slightly decreasing with age; overall the first term dominates, increasing the covariance between consumption and output over the working life. The covariance between $c$ and $y$ over age is plotted in figure 5 for different values of the curvature parameter and different amounts of private information.

From figure 5 we observe how without private information the covariance between $c$ and $y$ remains flat over age (when $\Omega = 0$ the level of the covariance is set by the total variance of the uncertainty and by the cross-partial between consumption and leisure in the utility function). Increasing $\sigma$, through its effect on the cross-partial, increases the level of the covariance while increasing $\Omega$ increases its growth rate. This increase is convex for the United Kingdom and Taiwan.
3.3 Implementation of the optimal allocation

In this paper we focus on the optimal contract derived from a constrained efficient problem subject to an information friction. By analyzing the data through the lens of our model, our goal is to verify if the allocation in the data is compatible with the predictions of the model without taking a stand on how this allocation is actually implemented. Several papers have proposed decentralizations for environments similar to ours. Prescott and Townsend (1984) show, for a general class of economies, that a competitive equilibrium in which firms are allowed to offer history-dependent contracts is Pareto optimal. Following the seminal work of Mirrlees (1971), the public finance literature has focused on implementing the constrained optimal al-
location as a competitive equilibrium with taxes. In most of the papers following this approach the optimal tax schedule used by the government is the only instrument that provides insurance to the worker. Recent papers in this tradition are Kocherlakota (2005) and Albanesi and Sleet (2006), which show that in a dynamic environment the optimal allocation can be implemented with a nonlinear income tax that depends on the entire history of productivity shocks (the former) or on the current productivity shock and wealth level (the latter). In a similar environment in which the worker’s disability is unobservable and permanent, Golosov and Tsyvinski (2006) show that the constrained efficient allocation can be decentralized as a competitive equilibrium with an asset-tested disability policy. Grochulski (2007) shows that the informational constrained allocation can be implemented using an institutional arrangement that resembles the US personal bankruptcy code. Following Kocherlakota (1998) and Prescott and Townsend (1984), Kapicka (2007) shows that the optimal allocation can be decentralized with workers sequentially trading one-period income-contingent assets.

4 The Data

We use two different data sources, the Michigan Panel Study of Income Dynamics (PSID) and the Consumer Expenditure Survey (CEX). Our main source for consumption expenditures and hours is the CEX; labor income is taken from the PSID. In order to make the data from both surveys as comparable as possible, we apply
the same sample selection to both. We consider household heads (reference person in the CEX) as those between ages 25 and 55 who worked more than 520 hours and less than 5096 hours per year and with positive labor income.\textsuperscript{18} We exclude households with wage less than half of the minimum wage in any given year. Table 2 describes the number of households in each stage of the sample selection. All the nominal data are deflated using the consumer price index calculated by the Bureau of Labor Statistics with base 1982-84$=$100.

In Table 7 (in appendix 7.6) we present some descriptive statistics from both surveys.\textsuperscript{19} All the earnings variables and hours refer to the household head, while the expenditure variables are total household expenditure per adult equivalent.\textsuperscript{20} The earnings and hours data are from the 1968-1993 waves of the PSID, corresponding to income earned in the years 1967-1992. The measure of earnings used includes head’s labor part of farm income and business income, wages, bonuses, overtime, commissions, professional practice, labor part of income from roomers and boarders or business income. In our benchmark experiment we use hours worked from CEX.

The consumption data is from the Krueger and Perri CEX dataset for the period

\textsuperscript{18}By stopping at age 55 we also minimize the discrepancy between consumption expenditure and actual consumption (due to the progressive larger use of leisure in both preparation and shopping time) highlighted in Aguiar and Hurst (2005).

\textsuperscript{19}From table 7 we observe that, in the period during which both surveys overlap, they have similar characteristics. Workers in the CEX sample are on average older and more educated than the PSID sample. Overall, we conclude that the two data sets are consistent.

\textsuperscript{20}We use the Census definition of adult equivalence.
from 1980 to 2003. In the CEX data our baseline sample is limited to households who
responded to all four interviews and with no missing consumption data. Since
the earnings data are annual and consumption data are measured every quarter
during one year, we sum the expenditures reported in the four quarterly interviews.
The consumption measure used includes the sum of expenditures on nondurable
consumption goods, services, and small durable goods, plus the imputed services
from housing and vehicles. The earnings data correspond to total labor income.

Our focus is on the life cycle moments of consumption, hours, and earnings
distribution. Due to data availability, we construct for each data set a synthetic
panel of repeated cross sections. To derive the life cycle moments of interest, we
first calculate each moment for a particular year/age cell. We include the worker on
a "cell" of age \(a\) on year \(t\) if his reported age in year \(t\) is between \(a - 2\) and \(a + 2\). A
typical cell constructed with this procedure contains a few hundred observations with
average size of 225 households in the CEX and 318 in the PSID. Following Heathcote,
Storesletten, and Violante (2005), we control for time effects when calculating the life
cycle moments. Specifically we run a linear regression of each moment in dummies
for age and time. The moments used in the estimation, reported in the graphs that
follow, are the coefficients on the age dummies normalized to match the average
value of the moment in the total sample.

In figure 6 we report the cross-sectional variances for consumption, hours, and
earnings over the working life. The first fact to be noted is the large increase in the
variance of income over the working life (14 log points), consumption increases less (3 log points), while the variance of hours is roughly constant with a slight decrease over the working life. In order to compare the two data sets used we plot the cross-sectional variance of hours from both; we observe that this moment is very similar in both datasets over the ages considered.

In figure 7 we observe that the covariance of hours and consumption does not display any particular trend, remaining essentially flat across the life cycle. We also observe a significant increase in the covariance of consumption and earnings.

In figure 7 we observe that the covariance of hours and consumption does not display any particular trend, remaining essentially flat across the life cycle. We also observe a significant increase in the covariance of consumption and earnings.
Figure 7: Life-cycle profiles for covariances, source: CEX and PSID. Panel (a) displays the covariance between consumption expenditure and hours, panel (b) the covariance between consumption and income. The dotted lines denote two standard errors introduced when controlling for time effect.

5 Estimation Strategy and Results

In this section we quantitatively assess how the constrained efficient environment described can account for the working life profiles of consumption, hours, and earnings that we observe in the data. In doing so, we also determine how much private information on labor productivity we need to introduce to make the model and the data consistent.

Due to the nonlinearity of the optimal contract, it is not possible to separately determine from equilibrium conditions the size of the private information and the preference parameters. On the other hand, for any combination of parameters we
can solve for the optimal contract and simulate a population. This enables us to determine preference parameters and Ω using a minimum distance estimator, minimizing the distance between moments generated by the model and moments observed in the data. We follow the procedure described in Gourinchas and Parker (2002) and estimate our model using the method of simulated moments.

Denote by Γ the vector of parameters to be estimated. From our model we determine the individual values of consumption, hours, and income as functions of the parameters and promised utility, denoted respectively by $c_{it}(w_{it}, \Gamma)$, $l_{it}(w_{it}, \Gamma)$, and $y_{it}(w_{it}, \Gamma)$. Our target moments are the cross-sectional variances of consumption hours and labor income by age.\(^{21}\) We denote the cross-sectional variances in the model by $\sigma^2_{c,t}(\Gamma), \sigma^2_{l,t}(\Gamma), \sigma^2_{y,t}(\Gamma)$. From the data we compute the equivalent moments denoted by $\hat{\sigma}^2_{c,t}, \hat{\sigma}^2_{l,t}, \hat{\sigma}^2_{y,t}$. For a given moment generated from the model we calculate the distance from its empirical counterpart, $g_x(\Gamma) = \sigma_x^2(\Gamma) - \hat{\sigma}_x^2$. Let $g(\Gamma)$ be the vector of length $J$, where $J$ values of $g_x$ are stacked. The minimum distance estimator for the parameter vector $\Gamma$ will be given by

$$\Gamma^* \equiv \arg\min_{\Gamma} g(\Gamma) \cdot W \cdot g(\Gamma)' ,$$

where $W$ is a $J \times J$ positive semi-definite weighting matrix. Once we obtain the value of $\Gamma^*$ we compute the properties of the gradient at the minimum determining

\(^{21}\)Targeting effective hours worked is consistent with our informational assumption. Although in our model hours are assumed to be non-observable by the planner, since we consider direct truth-telling mechanisms, the number of hours worked is actually known along the equilibrium path.
if any two parameters are linearly substitutes (or close to). For our benchmark estimation we set \( W \) equal to the identity matrix.

To make the data and the values generated by the model compatible, we scale dollar-denominated quantities so that the model matches the average consumption value in the data. Also, the total feasible number of hours of work is set at 5200 per year (approximately 14 hours of work for every day of the year). Throughout the quantitative analysis we fix \( \beta = q = 0.9 \). Our benchmark utility function will be Cobb-Douglas as in equation (18), of this utility function we estimate the curvature parameter (\( \sigma \)) and the share of consumption (\( \alpha \)). We restrict shock to two realizations per period: \( \theta_{t,H} > \theta_{t,L} \) for the public shock and \( \eta_H > \eta_L \) for the private shock. The average value of labor productivity is held constant (normalized to 1) along the life cycle, but we do allow for an increasing variance for both private and public shocks. The two shocks are parameterized by the following, for every age \( t \)

\[
\theta_{t,H} - \theta_{t,L} = 2h_\theta[1 + g_v(t - 1)],
\]
\[
\eta_{H} - \eta_{L} = 2h_\eta[1 + g_v(t - 1)].
\]

With this formulation we can determine the evolution of the two shocks with only three parameters: \( h_\theta, h_\eta \) the magnitude of the shocks at age 25 and \( g_v \) that denotes how the variances of the two shocks increase (or decrease if negative) in every age period. This specification for the shocks enables us also to maintain a constant \( \Omega \)
across the working life, from equation (4) we have that

\[ \Omega_t = \Omega = \frac{h^2}{h^2 + h^2}. \]  \hspace{1cm} (27)

We introduce heterogeneity at age 25 in the form of heterogeneity in continuation utilities. We consider at age 25 two distinct groups: the \( w \) "rich" group with initial promised utility given by \( w_H \) and the \( w \) "poor" group with initial promised utility given by \( w_L \). These two values of \( w \) are determined by the parameter \( \delta \) as follows

\[
\begin{align*}
\log(w_H) &= \log(w_0) + \log(\delta) \\
\log(w_L) &= \log(w_0) - \log(\delta)
\end{align*}
\]

where for a given \( \delta \) the value \( w_0 \) is determined so that the feasibility condition holds

\[ \sum_{i=H,L} S_1(w_i) = 0. \]  \hspace{1cm} (28)

In our benchmark estimation we will estimate the following 6 parameters \{\( \sigma, \alpha, g_v, h_\theta, h_\eta, \delta \}\).

5.1 Results

In our first set of results, we focus on the profiles of consumption and hours. In particular, we look at the cross-sectional variance of consumption and at the cross-sectional variance of hours from ages 25 to 55. In the previous section it was shown how introducing private information enables us to have an increase in the variance of consumption without increasing the variance of hours. Hence, looking at these moments, directly exploits the mechanism induced by private information. Column (1) of table 3 displays the results.
The model is locally identified (the gradient of the score function (24) at the minimum has full rank). The value of $\Omega$ needed to generate the observed increase in inequality in consumption is 0.99; all of the labor productivity shocks are private information of the worker.\footnote{The difference in the value of the score function for values of $\Omega=0.99$ and $\Omega=1$ is within numerical rounding.} Estimation results are further discussed in section 5.2. The curvature parameter ($\sigma$) and the share of consumption ($\alpha$) in the utility function are, respectively, 1.46 and 0.69. This implies a value of risk aversion equal to 1.32 and a value for the Frisch elasticity of leisure equal to 0.90.\footnote{The implied coefficient of risk aversion is $\rho = 1 - \alpha + \alpha\sigma$ and the Frisch elasticity of leisure $\phi = \rho/\sigma$.} With the value of elasticity of leisure $\phi$, we can approximate the Frisch elasticity of labor supply by multiplying $\phi$ by $\frac{1-\alpha}{\alpha}$, the resulting value is then equal to 0.40. This value is well within the common estimates in the labor literature (refer to Browning, Hansen, and Heckman (1999) table 3.3). In addition, the value of $g_v$ is close to zero and negative; thus, the total variance of labor productivity decreases (slightly) over age. Figure 8 displays the fit of the model with respect to the targeted moments.
Figure 8: Benchmark environment fit on matched moments.

The benchmark environment successfully accounts for the level and the increase in the variance of consumption during the working life. In figure 8-(a) we plot the profile for the variance of consumption; the increase of the profile, as described in the previous section, is convex. Figure 8-(b) displays the level for the variance of hours. The model also captures the level and the slightly decreasing pattern in the cross-sectional variance (in the data, hours decrease by 0.004 and by 0.006 in the model). As shown in Bound, Brown, Duncan, and Rodgers (1994), hours are subject to a large measurement error. This can introduce an upward bias in the estimate of the magnitude of labor productivity shocks. In section 5.3, we try and control for the measurement error for the variance of hours.

In section 3 we showed how the finite horizon nature of the problem induces an
increase in the distortion of the marginal rate of substitution between consumption and leisure over the working life. We now look at how this quantity evolves in the data. From (19) for the Cobb-Douglas utility function we have

\[ \tau_{cl} = 1 - \frac{1}{\theta \eta} \frac{1 - \frac{\alpha c}{L - l}}{\alpha} \]

(29)

where \( L = 5200 \). In the data we calculate this quantity under the assumption that imputed hourly wages are equal to the marginal productivity of the worker (the product of the two skill shocks).\(^{24}\) When calculating how the value of \( \tau_{cl} \) evolves over the ages 25-55 we observe that two factors are important: family composition and the definition of consumption (including or not housing). Our benchmark calculation of \( \tau_{cl} \), plotted in figure 9, considers only single households and does not include services imputed from housing in the definition of consumption.\(^{25}\) We observe that \( \tau_{cl} \) clearly displays an increasing trend over age in the data.

The growth rate of \( \tau_{cl} \) is decreasing in \( \alpha \). For the benchmark estimation (\( \alpha = 0.69 \)), for ages up to 40-45, \( \tau_{cl} \) in the model increases at the same rate as \( \tau_{cl} \) in the data. After that, the model overestimates the increase as shown in figure 9-(b). Moving to a lower value of \( \alpha \) increases the growth rate of \( \tau_{cl} \) in the data as shown in figure 9-(a). This moment is of particular interest since any market setting, where workers equate the marginal utility of consumption to the marginal

\(^{24}\)By doing so, we are not taking into account incentives provided through wages.

\(^{25}\)By restricting the sample to single households we do not have to consider the joint decision of hours worked within a household. This joint decision might have significant effect when calculating \( \tau_{cl} \) since it includes ratios of marginal utility.
Figure 9: Evolution of the average distortion on the marginal of substitution between consumption and leisure over age.

disutility of leisure as an exogenously incomplete markets model with endogenous labor, displays a flat profile for $\tau_{cl}$. The only way to induce an increase in this quantity is by introducing individual taste shocks in the value of $\alpha$ that increase in variance as the worker age (as for example in Badel and Huggett (2006)).

Finally, we look at how large are the transfers needed to implement the constrained efficient allocation. In the model this quantity can be calculated directly by looking at the differences between output produced and the consumption level. This transfer has no direct equivalent in the data, being the sum of multiple observable (change in asset position, transfer income) and unobservable quantities (transfers within the firm). We can, however, get a measure that approximates these transfers
Figure 10: Transfers over age.

from the PSID.\textsuperscript{26} Figure 10 shows the relation between these two variables. Overall, the transfers needed to implement are higher but not too distant from what can be measured in the data, particularly considering that the measure constructed from the data is a lower bound of the actual transfers taking place between workers.

In table 3 we report some initial robustness checks. We first look at the effect

\textsuperscript{26}The PSID for the years 1969 to 1985 continuously reports any additional transfer income the household received during the previous year. This variable includes transfers from publicly funded programs (food stamps, child nutrition programs, supplemental feeding programs, supplemental social security income, AFDC, earned income tax credit) and transfers received by family and nonfamily members. In our sample, 24\% of the household-year observation received a transfer, and in total 67\% of the households received a transfer at some time. These transfers are significant, averaging $1930 \ (1983\ \text{dollars}) \ and\ account\ for\ 70\% \ to\ 90\% \ of\ total\ food\ expenditures.
of setting the initial heterogeneity in \( w \) equal to zero (\( \delta = 0 \)) (column (2)). The environment without initial heterogeneity cannot account for the entire level in the heterogeneity in consumption, so in this test we only look at the increase in the variance of consumption over age.\(^{27}\) For the remaining cases, we keep \( \delta = 0 \).

We next look at the effect of fixing a lower share of consumption in the utility function (column (3)). We target average hours worked (column (4)) and restrict to a stationary process for labor productivity, fixing \( g_v = 0 \) (column (5)). All the cases confirm that shocks to labor productivity are entirely private information. In column (3) of table 4, we observe that for a value of \( \alpha = 1/3 \), only 60% of the increase in the cross-sectional variance for consumption is accounted for. The value of \( \alpha \) is important for its effect on the average hours worked; targeting this additional moment determines a level of \( \alpha = 0.46 \). With this additional restriction we account for 75% of the cross-sectional increase in the variance of consumption (column (4) in tables 3 and 4). Section 5.3 considers additional robustness checks.

### 5.2 Discussion

To understand why the minimum distance estimator returns a high value for \( \Omega \), we first consider the implications for the profiles of the variances of consumption and hours when \( \Omega \) is equal to zero. In this case, we can solve directly for these moments.

\(^{27}\)A similar limitation is also discussed in Phelan (1994).
The problem is characterized by the following first-order conditions:

\[
\begin{align*}
    u_c(c(\theta), l(\theta)) &= -\frac{1}{\theta} u_t(c(\theta), l(\theta)), \quad \forall \theta, \quad (30) \\
    u_c(c(\theta), l(\theta)) &= \frac{1}{\lambda}, \quad \forall \theta, \quad (31)
\end{align*}
\]

where \(\lambda\) is the multiplier on the promise-keeping constraint. In the Cobb-Douglas case from (30) and taking logs,

\[
\ln c_\theta = \ln \theta + \ln \frac{\alpha}{1-\alpha} + \ln \left(1 - \frac{y\theta}{\theta}\right), \quad (32)
\]

similarly, from (31) we have

\[
\ln c_\theta = \ln \lambda + \ln \alpha + \alpha (1 - \sigma) \ln c_\theta + (1 - \alpha) (1 - \sigma) \ln \left(1 - \frac{y\theta}{\theta}\right). \quad (33)
\]

Combining the previous two equations, we get

\[
\begin{align*}
    Var[\ln c_\theta] &= (\phi - 1)^2 Var[\ln \theta], \quad (34) \\
    \Delta Var[\ln c_\theta] &= (\phi - 1)^2 \Delta Var[\ln \theta], \quad (35)
\end{align*}
\]

where \(\phi = \frac{1-\alpha + \alpha \sigma}{\sigma}\) is the Frisch elasticity of leisure. From the same set of equations we can solve for leisure, obtaining

\[
\begin{align*}
    Var[\ln (1-l_\theta)] &= \phi^2 Var[\ln \theta], \quad (36) \\
    \Delta Var[\ln (1-l_\theta)] &= \phi^2 \Delta Var[\ln \theta]. \quad (37)
\end{align*}
\]

From equations (35) and (37), we observe that any increase in the variance of consumption is followed by an increase in the variance of hours.\(^{28}\) This feature highlights the relationship between leisure and consumption variance.
the difficulty of a full information insurance environment in describing the profile of consumption and hours.\textsuperscript{29} In our environment, private information is necessary to provide an increasing variance in consumption while at the same time keeping the variance of hours constant.

We now provide some intuition on the values obtained for the preference parameters. In the minimization procedure starting, for example, from an initial guess of \((\sigma = 3, \alpha = \frac{1}{3})\), we observe that the minimization path ultimately progresses, decreasing risk aversion and increasing the elasticity of leisure (decreasing \(\sigma\) and increasing \(\alpha\)), for the following reasons. For a given level of uncertainty, a high elasticity of leisure makes the spread in hours at the optimum larger. This translates, in the presence of private information, to a more severe moral hazard problem; which implies larger distortions on both the intratemporal and inter-temporal margin causing a larger spreading out of consumption and continuation utility. Also, a low value of risk aversion, although reducing the need to provide insurance, increases the elasticity of intertemporal substitution, making it less costly to the planner to provide incentives intertemporally. The additional tension that determines the value of the risk aversion and elasticity is given by the cross-partial derivative between consumption and leisure. As \(\sigma\) approaches 1 the cross-partial tends to zero, and as the complementarity between consumption and leisure decreases, it becomes more costly for the planner to induce variation in consumption, since now the first best

\textsuperscript{29}This result is also robust to different values of the elasticity of substitution between consumption and leisure, as shown in Storesletten, Telmer, and Yaron (2001).
level of consumption is constant.

We now want to determine how precisely the moments chosen for the estimation procedure can estimate the amount of private information.

![Figure 11: Surface plot of score function for benchmark estimation (a); Criterion function for benchmark estimation with respect to $\Omega$ (b).](image)

In figure 11-(a) we plot the values of the score function (24) as a function of $\sigma(\theta)$ and $\sigma(\eta)$. The minimum is obtained at the lower right corner marked by the "x". The lines are isocurves denoting how the function increases from the minimum. Lines closer together denote a steeper change. For each point in the graph, the distance from the origin denotes the total variance of the shock, while the angular distance from the horizontal axis denotes the amount of public information. The "x" being close to the horizontal axis denotes an estimate of almost all private information for the skill shock. What we observe is that the total variance is estimated more
precisely than the value of \( \Omega \): the score function displays a semi circular ridge at a constant distance from the origin, this determines the variance of the skill shock; within this ridge the score function is more flat (fewer isocurves) although it displays a minimum at the estimated value of \( \Omega \) close to 1.

In figure 11-(b) we plot the criterion function with respect to \( \Omega \). Each point in this curve is generated by keeping the value of \( \Omega \) fixed and estimating all of the remaining parameters. What we observe is that, as expected, the minimum is close to 1. However the criterion function rises slowly as we move away from 1. This indicates that the moments chosen are sensitive to increases of \( \Omega \) as we move away from the full information case (\( \Omega = 1 \)) but become less responsive as we move to higher values of \( \Omega \). These results suggest that a large value for \( \Omega \) is necessary for the model to be consistent with data, although a point estimate is likely to be imprecisely estimated.

5.3 Robustness checks

In this section we look at additional robustness checks.

Optimal weighting matrix

The estimation in the previous section was performed using the identity matrix as a weighting matrix in the minimization criterion. We performed the same estimation using an optimal weighting matrix. We adopt a continuously updated optimal weighting matrix as described in Hansen, Heaton, and Yaron (1996). In this case the
weighting matrix is evaluated at each iteration during the minimization procedure from the variance-covariance matrix of the simulated moments. The parameters are now determined by

$$\Gamma^* \equiv \arg\min_{\Gamma} g(\Gamma) \cdot W(\Gamma)^{-1} \cdot g(\Gamma)'$$

s.t. $$W(\Gamma) = E \left[ g(\Gamma) \cdot g(\Gamma)' \right].$$

The parameter results are reported in column (6) of table 5. A summary of the fit of the model is displayed in table 4. With respect to the benchmark estimation, the optimal matrix puts more weight on moments early in life than on moments later in life. Also the variance of consumption is weighted more than the variance of hours. Overall, the differences with respect to the benchmark estimation are small.

**General CES utility function**

The Cobb-Douglas utility function restricts the elasticity of substitution between consumption and leisure ($\epsilon$) to 1. We relax this implicit constraint by looking at the more general CES utility function,

$$u(c,l) = \left[ \alpha c^{\nu} + (1 - \alpha) (1 - l)^{\nu} \right]^{\frac{1-\sigma}{\nu(1-\sigma)}},$$

(38)

where now $\epsilon = \frac{1}{1-\rho}$. The results are in column (7) of tables 4 and 5. In the estimation, given the difficulty in crossing the value corresponding to $\epsilon = 1$, we estimate starting from each region with $\epsilon > 1$ and $\epsilon < 1$. The point estimate for the elasticity of substitution is $\epsilon = 1.08$. Since this value is close to 1, there are no significant changes with respect to the Cobb-Douglas utility function.
Controlling for measurement error

In section 5 we used the level of the cross-sectional variance of hours as a target moment. However, the presence of measurement error can bias the level upward. To control for this effect, we reestimate the benchmark environment by cutting the cross-sectional variance of hours by 30%. The results are in column (8) of tables 4 and 5.

Allowing persistence of the public shock

So far we have assumed that the labor productivity process is independent over age. We relax this assumption by introducing persistence in the public component of labor productivity. We model the public shock with a two state Markov chain. The transition matrix is bistochastic, and the probability of remaining in the same state is given by $\rho$. We also introduce ex-ante heterogeneity in the population by differentiating workers by their initial seed. As in the previous section, we target the cross-sectional variance of consumption and hours. The results are shown in table 5, column (5). Introducing persistence on the public shock has a large effect on the estimated composition of the labor productivity shocks: the point estimate for the value of $\Omega$ is now equal to .79, the estimated value for $\rho$ is 0.99, indicating that public shocks to labor productivity are permanent. We interpret this result as a supportive argument for the importance of private information shocks: even when

---

30 Using the PSID validation study, Bound, Brown, Duncan, and Rodgers (1994) find a signal to noise ratio for the variance of hours ranging from .2 to .3.
the only shock allowed to have serial correlation is the public one, we still require a significant fraction of the shock to be privately observed.

**Separable utility function and income variance**

The Cobb-Douglas utility function used in the benchmark estimation limits the ability to independently vary risk aversion and the Frisch elasticity of labor supply. We also performed the estimation with the following utility function:

$$u(c, l) = \alpha \frac{c^{1-\sigma}}{1-\sigma} - \frac{l^{1+\alpha_t}}{1+\alpha_t}. \quad (39)$$

This specification is commonly used in the labor literature.\(^{31}\) The coefficient of risk aversion is given by \(\sigma\) and \(1/\alpha_t\) is the Frisch elasticity of labor supply. With this specification we also target the cross-sectional variance of income. In order to interpret effective output in the model as labor income, we need to assume that labor markets are perfectly competitive and workers are paid their marginal productivity. The parameters’ estimates are given in table 5, column (9). The fit of the model is displayed in figures 12 and 13.

Overall, the model captures the evolution of the cross-sectional moments over the life cycle. The high value of the growth rate of the variance of labor productivity \((g_v = 0.117)\), needed to match the increase in the variance of income, causes the model to overshoot the variance of consumption at age 55. Also the variance of hours is slightly increasing and underestimated early in the working life.

\(^{31}\)See Browning, Hansen, and Heckman (1999).
5.4 The response of consumption to income shocks

Under the assumption that workers are paid at their marginal productivity, we can also study how consumption responds to income changes. This measure, which has been extensively studied, reflects the insurance possibilities of workers against income fluctuations and has important policy implications. In particular, Kaplan and Violante (2008) show that a calibrated version of a Bewley economy is not compatible with the degree of consumption insurance observed in the US data. We compute the response of consumption growth to income growth ($\alpha_2$) from the

---

32See Kaplan and Violante (2008) and references therein.
following regression:\footnote{As controls we use: change in family composition (including: marital status, number of babies, kids and number of adults in the households), a quartic in age and dummies for the month and for quarter of the interview.}

\[
\Delta \log c_i^t = \alpha_1 + \alpha_2 \Delta \log y_i^t + \text{controls}. \tag{40}
\]

We compute the value of \(\alpha_2\) from the CEX data using an OLS and instrumental variable estimation as in Dynarski and Gruber (1997). Results are in table 6. We perform the same estimation on panel data generated by the model. For our baseline environment with non separable utility, we find a value of \(\alpha_2\) equal to 0.107 which falls within our estimates using OLS and IV on CEX data.\footnote{Gervais and Klein (2006) show that the standard IV estimates overstates the true value \(\alpha_2\). Using a projection method they estimate in the CEX a value of \(\alpha_2 = 0.1\). Also note, Ai and}
can be derived in an environment with separable utility. In this case, if workers are fully insured against income shocks, the value of \( \alpha_2 \) is 0 (marginal utility of consumption is held constant). In our environment with private information and separable preferences \( \alpha_2 \) is equal to 0.067, which emphasizes the limited insurance possibilities available to workers.

Table 6: Consumption response to income shocks

<table>
<thead>
<tr>
<th>Source</th>
<th>( \alpha_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data - OLS</td>
<td>0.028 (.004)</td>
</tr>
<tr>
<td>Data - IV</td>
<td>0.177 (.021)</td>
</tr>
<tr>
<td>Data - IV-20%</td>
<td>0.134 (.018)</td>
</tr>
<tr>
<td>Model - separable</td>
<td>0.067</td>
</tr>
<tr>
<td>Model - non separable</td>
<td>0.107</td>
</tr>
</tbody>
</table>

Estimation of \( \alpha_2 \) using OLS, instrumental variables (IV) and instrumental variables removing changes in income smaller than 20% (IV-20%). Source: CEX data and authors calculations.

Finally we look at a particular prediction of our environment. The model predicts that the value of \( \alpha_2 \) should be increasing in age due to the progressive importance given to within period incentives. We calculate this statistic with the same restric-

Yang (2007), in an environment with private information and limited commitment, find a value of \( \alpha_2 = 0.269 \).
tion imposed to generate picture 9 (restricting to single household and removing services from housing from consumption). Figure 14 displays the result. The value of $\alpha_2$ is increasing in age up to age 40-45 as predicted by the model. The pattern is less clear (and with large standard errors) as we approach the retirement age.

6 Concluding Remarks

In this paper we show that household data for the U.S. can be rationalized as the outcome of an environment where risk sharing is efficient but limited by the presence of private information. We estimate a dynamic Mirrleesian economy and show that it can account for the evolution of inequality of consumption and hours over the working life when labor productivity shocks are entirely private information of the
worker. We characterize the finite horizon optimal contract and show how the provision of incentives differs along the life cycle: early in life continuation utility plays an important role in providing incentives, later in life intratemporal distortions on the marginal rate of substitution between consumption and leisure become more important.

The result of this paper suggests that private information is quantitatively an important friction when studying risk sharing. This provides strong supporting evidence for recent papers that study the asset pricing implications of constrained efficient allocation with private information as in Kocherlakota and Pistaferri (2009) and Kocherlakota and Pistaferri (2007). Accounting for the presence of private information in the data can have strong implications for designing policies that address inequality, redistribution and insurance. For example, the welfare gains from policies that reduce inequality in an economy in which workers can trade a single bond can be quite large. However in our environment, any policy that addresses inequality without recognizing the role of incentives introduced by the presence of private information can potentially be welfare decreasing.
7 Appendix

7.1 Incentives and Nonseparability

In this section we show that the full information allocation is not incentive compatible for the environment with Cobb-Douglas utility:

\[ u(c, l) = \left[ c^\alpha (1 - l)^{1-\alpha} \right]^{1-\sigma}. \]

For separable utility functions the result is straightforward given that the first best allocation requires constant consumption but not constant output across individuals with different skills. With a Cobb-Douglas utility function if \( \sigma > 1 \), consumption and labor are Frisch complements (the cross-partial derivative \( u_{cl} > 0 \)). This implies that in the first best allocation, a worker with high productivity works more but also consumes more. We show that if faced with the full information allocation, a high-skill worker is better off lying and receiving the allocation of a low-skill agent. That is \( u(c(\theta_H), \frac{u(\theta_H)}{\sigma_H}) < u(c(\theta_L), \frac{u(\theta_L)}{\sigma_H}) \). Recall that for the Cobb-Douglas utility we have

\[ u_c(c, l) = \frac{\alpha}{c} (1 - \sigma) u(c, l), \quad (41) \]
\[ u_l(c, l) = -\frac{1 - \alpha}{1 - l} (1 - \sigma) u(c, l). \quad (42) \]

From the first-order conditions (30) and (31) we have

\[ (1 - l(\theta)) = \frac{(1 - \alpha)}{\alpha} \frac{c(\theta)}{\theta}, \quad \forall \theta, \quad (43) \]
\[ c(\theta_L) = c(\theta_H) \left( \frac{\theta_L}{\theta_H} \right)^{(1-\alpha)(1-\sigma)} \quad (44). \]
Using (43) we can rewrite the utility function as
\[
  u(c(\theta), l(\theta)) = \frac{c(\theta) (\frac{(1-\alpha)}{\alpha})^{1-\alpha} (1-\sigma)^{1-\sigma}}{1-\sigma} = \frac{c(\theta)^{(1-\alpha)(1-\sigma)}}{1-\sigma}.
\]
Substituting (44) in the above for \( \theta = \theta_L \),
\[
  u(c(\theta_L), l(\theta_L)) = \frac{c(\theta_H)^{(1-\alpha)(1-\sigma)}}{1-\sigma} \left( \frac{(1-\alpha)}{\alpha} \right)^{\frac{(1-\sigma)^{1-\sigma}}{1-\sigma}} \frac{1}{1-\sigma} = \frac{u(c(\theta_H), l(\theta_H))}{\theta_H \theta_L} \left( \frac{(1-\alpha)(1-\sigma)}{\sigma} \right)^{\frac{(1-\sigma)^{1-\sigma}}{1-\sigma}}.
\]
By assumption \( 0 < \left( \frac{\theta_H}{\theta_L} \right)^{\frac{(1-\alpha)(1-\sigma)}{\sigma}} < 1 \). This implies that \( u(c(\theta_H), \frac{y(\theta_H)}{\theta_H}) < u(c(\theta_L), \frac{y(\theta_L)}{\theta_L}) \). Given that \( \sigma > 1 \), \( u(c(\theta_H), l(\theta_H)) \) and \( u(c(\theta_L), l(\theta_L)) \) are both negative. From this result it follows that
\[
  u(c(\theta_H), \frac{y(\theta_H)}{\theta_H}) < u(c(\theta_L), \frac{y(\theta_L)}{\theta_L}) < u(c(\theta_L), \frac{y(\theta_L)}{\theta_H}).
\]
Hence, the first best allocation is not incentive-compatible for the worker with high productivity shock.

7.2 Relaxed Recursive Problem

In this section we justify our use of the relaxed recursive formulation described in section 2.1. Denote the original maximization problem by \((P1)\).
Proposition 3. Assume $u(c, l) = u(c) - v(l)$ with $v$ a convex function, then any allocation $\{c, y\}$ that solves (P2) also solves (P1).

Proof. Let the allocation $\{c, y\}$ be a solution to (P2) and suppose it does not satisfy (P1). Let the relaxed maximization problem be the original problem without constraints (50) and (54). Denote it by (P2).

The time $T-1$ problem is

$$S_{T-1}(w) = \min_{c,y,w'} \sum_{\theta} \pi(\theta) \sum_{\eta} \pi(\eta) \left[ c_{T-1}(\theta, \eta) - y_{T-1}(\theta, \eta) + qS_T(w_{T-1}^\prime(\theta, \eta)) \right],$$

s.t.

$$\sum_{\theta} \pi(\theta) \sum_{\eta} \pi(\eta) u \left( c_{T-1}(\theta, \eta), \frac{y_{T-1}(\theta, \eta)}{f(\theta, \eta)} \right) + \beta w_{T-1}^\prime(\theta, \eta) \geq 0,$$

$$u \left( c_{T-1}(\theta, \eta_H), \frac{y_{T-1}(\theta, \eta_H)}{f(\theta, \eta_H)} \right) + \beta w_{T-1}^\prime(\theta, \eta_H) \geq 0,$$

$$u \left( c_{T-1}(\theta, \eta_L), \frac{y_{T-1}(\theta, \eta_L)}{f(\theta, \eta_L)} \right) + \beta w_{T-1}^\prime(\theta, \eta_L), \ \forall \theta_{T-1},$$

$$u \left( c_{T-1}(\theta, \eta_H), \frac{y_{T-1}(\theta, \eta_H)}{f(\theta, \eta_H)} \right) + \beta w_{T-1}^\prime(\theta, \eta_H), \forall \theta_{T-1}.$$ (53)

Let the relaxed maximization problem be the original problem without constraints (50) and (54). Denote it by (P2).
(54) for some \( \theta_{T-1} \). Then

\[
\begin{align*}
&u(c_{T-1}(\theta, \eta_H)) - v \left( \frac{y_{T-1}(\theta, \eta_H)}{f(\theta, \eta_L)} \right) + \beta w'_{T-1}(\theta, \eta_H) > \\
&u(c_{T-1}(\theta, \eta_L)) - v \left( \frac{y_{T-1}(\theta, \eta_L)}{f(\theta, \eta_L)} \right) + \beta w'_{T-1}(\theta, \eta_L).
\end{align*}
\]

(55)

We know that any allocation that solves \((P2)\) must have (13) holding with an equality. Substituting this constraint in the previous equation we get

\[
\begin{align*}
v \left( \frac{y_{T-1}(\theta, \eta_H)}{f(\theta, \eta_H)} \right) - v \left( \frac{y_{T-1}(\theta, \eta_L)}{f(\theta, \eta_L)} \right) > v \left( \frac{y_{T-1}(\theta, \eta_L)}{f(\theta, \eta_H)} \right) - v \left( \frac{y_{T-1}(\theta, \eta_L)}{f(\theta, \eta_L)} \right),
\end{align*}
\]

(56)

Since \( v \) is convex, then for any \( \varepsilon > 0, x > \hat{x} \) we have

\[
\begin{align*}
v(x) - v(x - \varepsilon) &> v(\hat{x}) - v(\hat{x} - \varepsilon), \\
v(x) - v(x - \varepsilon) &> v(\hat{x}) - v(\hat{x} - \varepsilon f(\theta, \eta_H)/f(\theta, \eta_H)).
\end{align*}
\]

Let \( x \equiv \frac{y_{T-1}(\theta, \eta_H)}{f(\theta, \eta_L)}, \hat{x} \equiv \frac{y_{T-1}(\theta, \eta_H)}{f(\theta, \eta_H)} \) and \( \varepsilon \equiv \frac{y_{T-1}(\theta, \eta_H) - y_{T-1}(\theta, \eta_L)}{f(\theta, \eta_L)} \) then:

\[
\begin{align*}
v \left( \frac{y_{T-1}(\theta, \eta_H)}{f(\theta, \eta_L)} \right) - v \left( \frac{y_{T-1}(\theta, \eta_L)}{f(\theta, \eta_L)} \right) &> v \left( \frac{y_{T-1}(\theta, \eta_H)}{f(\theta, \eta_L)} \right) - v \left( \frac{y_{T-1}(\theta, \eta_L)}{f(\theta, \eta_L)} \right).
\end{align*}
\]

Contradicting (56). Hence any allocation that solves \((P2)\) also solves the original problem \((P1)\). The same proof holds for the time \( T \) problem. \( \blacksquare \)

### 7.3 Proof of Proposition 1

**Proof.** The proof follows Rogerson (1985) closely. Considering the planner’s problem as allocating utility levels to workers, let \( \bar{u}(\theta, \eta) = u(c(\theta, \eta)) \) be the utility derived from consumption in state \((\theta, \eta)\). Let \( C(\bar{u}) \) be the cost for the planner of
providing utility level \( \bar{u} \). To show (14), consider the following perturbation of the optimal contract \( \bar{u}^* \). For some \( \eta_t \in H \) and some \( \theta_{t+1} \in \Theta \), let \( \bar{u}(\theta^t, [\eta^{t-1}, \eta_t]) = \bar{u}^*(\theta^t, [\eta^{t-1}, \eta_t]) - \Delta \) and \( \bar{u}([\theta^t, \theta_{t+1}], \eta^{t+1}) = u^*([\theta^t, \theta_{t+1}], \eta^{t+1}) + \Delta/\beta \) for all \( \theta^t \) and \( u(\theta^t, [\eta^{t-1}, \tilde{\eta}_t]) = u^*(\theta^t, [\eta^{t-1}, \tilde{\eta}_t]) \) and \( u([\theta^t, \tilde{\theta}_{t+1}], \eta^{t+1}) = u^*([\theta^t, \tilde{\theta}_{t+1}], \eta^{t+1}) \) for \( \eta_t \neq \tilde{\eta}_t \) and for \( \theta_{t+1} \neq \tilde{\theta}_{t+1} \). The labor allocations are left unchanged. This contract is still incentive compatible given that the time and the shock \( \theta \) is publicly observed. This contract minimizes the cost for the planner if the following holds:

\[
\lim_{\Delta \to 0} \frac{\partial}{\partial \Delta} \left[ C' \left( u(\theta^t, \eta^t) - \Delta \right) + q \sum_{\eta^{t+1}} \pi(\eta^{t+1}|\eta^t) C \left( u(\theta^{t+1}, \eta^{t+1}) + \frac{\Delta}{\beta} \right) \right] = 0, \tag{57}
\]

which implies

\[
C'(u(\theta^t, \eta^t)) = \frac{q}{\beta} \sum_{\eta^{t+1}} \pi(\eta^{t+1}|\eta^t) C \left( u(\theta^{t+1}, \eta^{t+1}) + \frac{\Delta}{\beta} \right). \tag{58}
\]

Equation (14) then follows given that \( C'(u(\theta^t, \eta^t)) = \frac{1}{u_c([\theta^t, \eta^t])} \). To show (15) we proceed in a similar way. At any given period \( t \), consider two \( \theta_t, \tilde{\theta}_t \in \Theta \); for all \( \eta_t \) and \( \theta^{t-1} \) let \( u([\theta^{t-1}, \theta_t], \eta^t) = u^*([\theta^{t-1}, \theta_t], \eta^t) - \Delta \) and \( u([\theta^{t-1}, \tilde{\theta}_t], \eta^t) = u^*([\theta^{t-1}, \tilde{\theta}_t], \eta^t) + \Delta \). For all the remaining histories, the labor allocations are unchanged. This perturbation of the optimal contract does not affect incentives of the worker, since the transfers \( \Delta \) are contingent on observables and the total utility of the worker is unchanged. Optimality of this contract requires

\[
\lim_{\Delta \to 0} \frac{\partial}{\partial \Delta} \left[ \sum_{\eta^t} \pi(\eta^t|\eta^{t-1}) C(u([\theta^{t-1}, \tilde{\theta}_t], \eta^t) - \Delta) + \sum_{\eta^{t}} \pi(\eta^{t}|\eta^{t-1}) C'(u([\theta^{t-1}, \theta_t], \eta^t) + \Delta) \right] = 0. \tag{59}
\]
Equation (15) then follows from

\[ \sum_{\eta'} \pi(\eta'|\eta^{t-1}) C'(u([\theta^{t-1}, \hat{\theta}_t], \eta')) = \sum_{\eta'} \pi(\eta'|\eta^{t-1}) C'(u([\theta^{t-1}, \theta_t], \eta')), \forall \hat{\theta}_t, \theta_t. \]  

(60)

7.4 Proof of Proposition 2

Proof. Suppose not. Then there is a \( \eta \) so that \( c(\theta, \eta) = c(\hat{\theta}, \eta) \). Let \( \eta = \eta_H \), from the first order conditions for \( c \)

\[ w'(c(\theta, \eta_H)) [\lambda \pi(\theta) \pi(\eta_H) + \mu(\theta)] = \pi(\theta) \pi(\eta_H), \forall \theta. \]

This implies that \( \mu(\theta) = \mu(\hat{\theta}) = \mu \) and \( c(\theta, \eta_L) = c(\hat{\theta}, \eta_L) \) (If we assume in the contradicting assumption that \( c(\theta, \eta_L) = c(\hat{\theta}, \eta_L) \), we also get that \( \mu(\theta) = \mu(\hat{\theta}) = \mu \) and \( c(\theta, \eta_H) = c(\hat{\theta}, \eta_H) \).) From the first-order conditions (FOCs) for \( w'(\theta, \eta) \)

\[ \pi(\theta) \pi(\eta_H) \beta \lambda + \beta \mu = \pi(\theta) \pi(\eta_H) q_{\tau \theta}(w'(\theta, \eta_H)), \forall \theta, \]

\[ \pi(\theta) \pi(\eta_L) \beta \lambda - \beta \mu = \pi(\theta) \pi(\eta_L) q_{\tau \theta}(w'(\theta, \eta_L)), \forall \theta. \]

This implies

\[ w'(\theta, \eta_H) = w'(\hat{\theta}, \eta_H), \quad w'(\theta, \eta_L) = w'(\hat{\theta}, \eta_L). \]  

(61)

From the FOCs for \( y(\theta, \eta_H) \)

\[ \frac{1}{\theta_{\eta H}} v'(y(\theta, \eta_H)) \frac{\partial \pi(\theta) \pi(\eta_H) + \mu}{\partial \eta_H} = -\pi(\theta) \pi(\eta_H), \forall \theta, \]

\[ \frac{1}{\theta_{\eta H}} v'(y(\theta, \eta_H)) = \frac{1}{\theta_{\eta H}} v'(y(\hat{\theta}, \eta_H)). \]
Using the parametric form for the utility function, this equation implies

\[
\frac{y(\theta, \eta_H)^\gamma}{(\theta \eta_H)^{1+\gamma}} = \frac{y(\hat{\theta}, \eta_H)^\gamma}{(\hat{\theta} \eta_H)^{1+\gamma}}. \tag{62}
\]

From the FOCs for \(y(\theta, \eta_L)\)

\[
\frac{\lambda \pi(\theta) \pi(\eta_L)}{\theta \eta_L} v' \left( \frac{y(\theta, \eta_L)}{\theta \eta_L} \right) - \frac{\mu}{\theta \eta_H} v' \left( \frac{y(\theta, \eta_L)}{\theta \eta_H} \right) = -\pi(\theta) \pi(\eta_L), \quad \forall \theta,
\]

\[
\frac{\lambda \pi(\hat{\theta}) \pi(\eta_L)}{\hat{\theta} \eta_L} v' \left( \frac{\hat{y}(\theta, \eta_L)}{\hat{\theta} \eta_L} \right) - \frac{\mu}{\hat{\theta} \eta_H} v' \left( \frac{\hat{y}(\theta, \eta_L)}{\hat{\theta} \eta_H} \right),
\]

\[
\frac{y(\theta, \eta_L)^\gamma}{\theta^{1+\gamma}} \left( \frac{\lambda \pi(\theta) \pi(\eta_L)}{\eta_L^{1+\gamma}} - \frac{\mu}{\eta_H^{1+\gamma}} \right) = \frac{\hat{y}(\theta, \eta_L)^\gamma}{\theta^{1+\gamma}} \left( \frac{\lambda \pi(\hat{\theta}) \pi(\eta_L)}{\eta_L^{1+\gamma}} - \frac{\mu}{\eta_H^{1+\gamma}} \right).
\]

This implies

\[
\frac{y(\theta, \eta_L)^\gamma}{(\theta \eta_H)^{1+\gamma}} = \frac{y(\hat{\theta}, \eta_L)^\gamma}{(\hat{\theta} \eta_H)^{1+\gamma}}. \tag{63}
\]

Since the multiplier on the incentive-compatibility constraint is strictly positive, this equation holds with equality for each \(\theta\). Summing the incentive-compatibility constraint for both \(\theta\), using (61) and the fact that the consumption is independent of \(\theta\) we have

\[
v \left( \frac{y(\hat{\theta}, \eta_H)}{\theta \eta_H} \right) - v \left( \frac{y(\theta, \eta_H)}{\theta \eta_H} \right) = v \left( \frac{y(\hat{\theta}, \eta_L)}{\theta \eta_H} \right) - v \left( \frac{y(\theta, \eta_L)}{\theta \eta_H} \right),
\]

\[
\frac{y(\hat{\theta}, \eta_L, w)^{\gamma+1}}{(\theta \eta_H)^{1+\gamma}} - \frac{y(\theta, \eta_H, w)^{\gamma+1}}{(\theta \eta_H)^{1+\gamma}} = \frac{y(\hat{\theta}, \eta_L, w)^{\gamma+1}}{(\theta \eta_H)^{1+\gamma}} - \frac{y(\theta, \eta_L, w)^{\gamma+1}}{(\theta \eta_H)^{1+\gamma}}.
\]
Substituting in this expression equations (62) and (63)

\[
\frac{y(\hat{\theta}, \eta_H)^{\gamma+1}}{(\hat{\theta} \eta_H)^{1+\gamma}} - \frac{y(\hat{\theta}, \eta_H)^{\gamma+1}}{(\hat{\theta} \eta_H / \theta \eta_H)^{(1+\gamma)^2}(\theta \eta_H)^{1+\gamma}} = \frac{1}{(\hat{\theta} \eta_H)^{1+\gamma}} - \frac{1}{(\hat{\theta} \eta_H / \theta \eta_H)^{(1+\gamma)^2}(\theta \eta_H)^{1+\gamma}}.
\]

So that

\[
y(\hat{\theta}, \eta_H) = y(\hat{\theta}, \eta_L).
\] (64)

Note that the above, together with \(c(\eta_H) > c(\eta_L), w'(\eta_H) > w'(\eta_L)\) (these relations come from the FOCs and \(\mu > 0\)), implies:

\[
u(c(\eta_H)) - v\left(\frac{y(\theta, \eta_H)}{\theta \eta_L}\right) + \beta w'(\eta_H) > u(c(\eta_L)) - v\left(\frac{y(\theta, \eta_L)}{\theta \eta_L}\right) + \beta w'(\eta_L).
\]

Hence, the allocation does not satisfy the incentive-compatibility constraint for an agent with a low private shock. This implies that there is some allocation \(\{c, y\}\) that solves the relaxed problem \((P2)\) and violates the incentive-compatibility constraint for the low agent. This is a contradiction to Proposition 3. A similar proof holds for the time \(T\) problem.

### 7.5 Numerical Procedure

Computing the solution to the dynamic moral hazard environment described in this paper presents two difficulties: the problem is nonstationary and the incentive
constraints introduce a nonconvexity in the programming problem. We adopt a computation procedure similar to the procedure developed in Phelan and Townsend (1991). A key difference is in how the possible nonconvexities in the problem are dealt with. Phelan and Townsend (1991) confine the allocation on a grid and allow the planner to choose lotteries on such allocations. This procedure transforms the dynamic program in a linear programming problem. The use of lotteries in our environment makes the computing problem quickly intractable due to the presence of nonseparable preferences (the lottery in this case has to be defined on the joint distribution of consumption and leisure) and due to the heterogeneity of individuals, so that a lottery has to be computed not only for every age but also for every realization of the public and private shock. Our approach does not rely on lotteries. We do not impose any grid on the allocation and restrict the planner to only choose degenerate lotteries. This restriction is not binding. Our theoretical justification is based on the works of Arnott and Stiglitz (1988) and Kehoe, Levine, and Prescott (2002), which show that in many moral hazard environments under the assumption of nonincreasing risk aversion, the use of lotteries is not optimal. Our environment with separable utility (or inelastic labor) falls directly in this category. When the utility is nonseparable, we cannot show that lotteries will not be optimal. In this case we verify ex post if the allocation can be improved with the use of lotteries.
Figure 15: Results allowing for randomization in the Cobb-Douglas case with $\Omega = 1$; panel (a) shows the probability distribution for a single allocation (effective output for $\eta_l$); panel (b) displays the values of the joint probability distribution on all the allocation for a given age.

To determine if the use of lotteries is optimal we first compute our solution without lotteries, then include the solution found on an equally spaced grid of consumption, hours and effective output. The results are shown in figure 15. We observe that the probability chosen is single peaked at the optimum allocation found without lotteries and quickly (the graphs are in log scale) falls into numerical noise.

We now describe the steps taken to solve the environment and to compute the moments used in the estimation.

1. *(Obtain the policy functions)* The first step is to derive the policy functions of the problem described in section 2.1. We solve the problem iterating backward, starting from the last period $T$, in our case $T = 7$. Given that we do not know
the ex post evolution of the state variable $w$ (promised utility), we solve the problem for time $T$ on a grid of possible $w$ for each value of $w$. The system of equations given by the first-order condition is solved using the Newton method and using as a guess the solution of the equivalent full information problem (this improves efficiency of the computation and stability over a wide range of parameters of the utility). Having solved for the optimal policy function, we compute the value function of the planner $S_T$ and its numerical derivatives. The first derivatives are computed using a two-sided difference formula, second derivatives using a three-point formula. Moving backward in time in period $T - 1$, we repeat the above procedure using the computed values of $S_T$ to determine the allocation for time $T - 1$. Whenever necessary, we interpolate $S_T$ using a cubic spline interpolation. The procedure described is repeated for all the periods $T - 1, ..., 1$.

2. *(Simulate the population)* With the policy functions we can simulate our panel. For each age we determine the value of consumption, hours, and effective labor using a cubic spline on the policy functions. In our benchmark ($T = 6$) we simulate every possible history of labor productivity for the individuals. Given the possibility of four different realizations of the uncertainty for every age, the panel generated contains a total of $4^T$ individuals. When we allow for initial heterogeneity in $w_0$, we construct a panel for each value of $w_0$. We set $w_0$ so that aggregate feasibility holds, that is $S_1(w_0) = 0$. In the case of time
zero heterogeneity in $w_0$, the previous condition becomes $\sum_{w_0} S_1 (w_0) = 0$.

3. *(Estimate)* In the final step we compute the same statistics on the artificial panel as in the data. The estimation procedure requires minimizing the distance between data moments and artificial moments. The minimization is performed using the Nelder-Mead simplex algorithm. As described in Lagarias, Reeds, Wright, and Wright (1998) this method does not guarantee convergence to the minimum. Our heuristic approach in assuring that we have in fact reached a minimum is the following: restarting the minimization procedure from the minimum found and starting from a different initial point in the simplex.

7.6 Data
<table>
<thead>
<tr>
<th></th>
<th>PSID</th>
<th>CEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline sample</td>
<td>192,897</td>
<td>69,816</td>
</tr>
<tr>
<td>Exclude SEO sample</td>
<td>109,342</td>
<td>NA</td>
</tr>
<tr>
<td>Hours restriction</td>
<td>85,811</td>
<td>46,559</td>
</tr>
<tr>
<td>Earnings$\leq$0</td>
<td>NA</td>
<td>46,002</td>
</tr>
<tr>
<td>Labor income$\leq$0</td>
<td>76,633</td>
<td>45,745</td>
</tr>
<tr>
<td>Minimum wage restriction</td>
<td>67,023</td>
<td>43,802</td>
</tr>
<tr>
<td>Age $&gt;$21 and $\leq$55</td>
<td>56,628</td>
<td>36,871</td>
</tr>
<tr>
<td>Food$\leq$0</td>
<td>47,757</td>
<td>NA</td>
</tr>
<tr>
<td>Final sample</td>
<td>47,757</td>
<td>36,871</td>
</tr>
</tbody>
</table>

Numbers indicate total observations remaining at each stage of the sample selection.
### Table 3: Benchmark estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta, q$</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.46</td>
<td>1.27</td>
<td>3.88</td>
<td>1.59</td>
<td>1.42</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.69</td>
<td>0.67</td>
<td>$\frac{1}{3}$</td>
<td>0.46</td>
<td>0.49</td>
</tr>
<tr>
<td>$g_v$</td>
<td>$-0.0073$</td>
<td>$-0.014$</td>
<td>$-0.001$</td>
<td>$-0.006$</td>
<td>0</td>
</tr>
<tr>
<td>$\delta$</td>
<td>1.22</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>0.99</td>
<td>0.99</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Results: benchmark estimation (1), results without heterogeneity in $w$ (2), fixing $\alpha = 1/3$ (3), targeting mean hours (4), fixing $g_v = 0$ (5). Note: values of $\beta$ and $q$ are held fixed. For columns (2) to (5) only the increase in the variance of consumption is targeted.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \text{var}(c)$</td>
<td>0.0287</td>
<td>0.017</td>
<td>0.0214</td>
<td>0.0238</td>
<td>0.0285</td>
</tr>
<tr>
<td>$\Delta \text{var}(l)$</td>
<td>−0.0062</td>
<td>0.0058</td>
<td>−0.0036</td>
<td>0.0022</td>
<td>−0.004</td>
</tr>
<tr>
<td>$\Delta \text{cov}(c,l)$</td>
<td>0.0045</td>
<td>0.002</td>
<td>−0.0031</td>
<td>−0.001</td>
<td>−0.0004</td>
</tr>
<tr>
<td>$E[l]$</td>
<td>2960</td>
<td>1540</td>
<td>2124</td>
<td>2244</td>
<td>2123</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \text{var}(c)$</td>
<td>0.0284</td>
<td>0.0281</td>
<td>0.025</td>
<td>0.0258</td>
<td>0.0285</td>
</tr>
<tr>
<td>$\Delta \text{var}(l)$</td>
<td>−0.01</td>
<td>−0.0013</td>
<td>−0.003</td>
<td>−0.006</td>
<td>−0.004</td>
</tr>
<tr>
<td>$\Delta \text{cov}(c,l)$</td>
<td>0.0049</td>
<td>0.0048</td>
<td>0.0021</td>
<td>0.0035</td>
<td>−0.0004</td>
</tr>
<tr>
<td>$E[l]$</td>
<td>2132</td>
<td>2831</td>
<td>2564</td>
<td>3119</td>
<td>2123</td>
</tr>
</tbody>
</table>

Summary moments: benchmark estimation (1), results fixing $\alpha = 1/3$ (2), result targeting mean hours (3), results fixing $g_v = 0$ (4), with persistence of the public shock (5), using optimal weighting matrix (6), using general CES utility function (7), controlling for measurement error in hours (8).
### Table 5: Parameter estimates, robustness checks

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.47</td>
<td>0.64</td>
<td>0.56</td>
<td>1.27</td>
<td>0.83</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>2.84</td>
<td>1.46</td>
<td>1.77</td>
<td>0.681</td>
<td>0.82</td>
</tr>
<tr>
<td>$\alpha_l$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.66</td>
</tr>
<tr>
<td>$g_v$</td>
<td>0.007</td>
<td>–0.005</td>
<td>–0.007</td>
<td>–0.008</td>
<td>0.12</td>
</tr>
<tr>
<td>$\delta$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.13</td>
</tr>
<tr>
<td>$\nu$</td>
<td>–</td>
<td>–</td>
<td>0.0725</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.99</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>0.79</td>
<td>1</td>
<td>0.98</td>
<td>1</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Parameter estimates with persistence of public shock (5), using optimal weighting matrix (6), using general CES utility function (7), controlling for measurement error in hours (8), alternative utility function (9).
Table 7: Summary statistics for the PSID and CEX samples used.

<table>
<thead>
<tr>
<th></th>
<th>PSID (68-91)</th>
<th>CEX (80-04)</th>
<th>PSID (80-91)</th>
<th>CEX (80-91)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>35.71 (9.47)</td>
<td>39.17 (8.74)</td>
<td>35.52 (8.92)</td>
<td>38.14 (8.79)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school dropout</td>
<td>11.96</td>
<td>6.99</td>
<td>10.34</td>
<td>8.27</td>
</tr>
<tr>
<td>High school graduate</td>
<td>36.61</td>
<td>29.26</td>
<td>35.52</td>
<td>31.92</td>
</tr>
<tr>
<td>College</td>
<td>44.99</td>
<td>60.46</td>
<td>51.11</td>
<td>55.93</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>90.42</td>
<td>86.95</td>
<td>91.13</td>
<td>88.18</td>
</tr>
<tr>
<td>Black</td>
<td>7.42</td>
<td>9</td>
<td>7.06</td>
<td>8.55</td>
</tr>
<tr>
<td>Family composition</td>
<td>3.19 (1.56)</td>
<td>3.07 (1.58)</td>
<td>3.02 (1.42)</td>
<td>3.15 (1.62)</td>
</tr>
<tr>
<td>Average earnings ($)</td>
<td>26,594 (18,168)</td>
<td>30,340 (20,406)</td>
<td>26,519 (20,474)</td>
<td>28,491 (16,908)</td>
</tr>
<tr>
<td>Average consumption ($)</td>
<td>NA</td>
<td>13,542 (6,842)</td>
<td>NA</td>
<td>13,166 (6,541)</td>
</tr>
<tr>
<td>Food ($)</td>
<td>4,493 (2,344)</td>
<td>3,791 (1,965)</td>
<td>4,218 (2,340)</td>
<td>3,998 (2,036)</td>
</tr>
<tr>
<td>Rent ($)</td>
<td>935 (1,890)</td>
<td>262 (487)</td>
<td>1,007 (2,028)</td>
<td>258 (469)</td>
</tr>
<tr>
<td>Hours</td>
<td>2,203 (588)</td>
<td>2,123 (567)</td>
<td>2,191 (580)</td>
<td>2,178 (559)</td>
</tr>
</tbody>
</table>

Note - All dollar amounts in 1983 dollars.
Part III

Non-exclusive Dynamic Contracts, Competition, and the Limits of Insurance

8 Introduction

What type of contractual arrangements are available to workers in a decentralized economy when firms compete for the provision of social insurance? In this paper, we study how, in a decentralized economy, the presence of non-exclusive contracts endogenously limits the contracts offered and hence the amount of insurance. We find that competition and non-observability of insurance contracts significantly reduce the amount of insurance provided: the equilibrium allocation in our environment is equivalent to a self-insurance economy and only linear contracts are offered.

Multiple credit and labor relations are an important aspect of everyday life. Survey data shows that individuals and households receive insurance against idiosyncratic risk from a multitude of sources: publicly provided insurance (unemployment, Medicare, Medicaid, disability, food stamps, progressive income taxation); privately

---

35This chapter is coauthored with Laurence Ales.
provided insurance (employer, between and within family transfers); financial instruments in credit markets; and housing and other large durable goods. The same consideration is true for labor relationships. Paxson and Sicherman (1994) look at the number of concurrent labor relationships held by survey respondents of the Panel Study of Income Dynamics (PSID) between 1977 and 1990 and the Current Population Survey (CPS) of 1991. They find that for any given year, 20% of working males held at least a second job, and during their working life there is at least a 50% probability of holding a second job. However, monitoring all the transactions an agent might engage in with other firms is very costly for an individual firm, especially if these relationships include activities in the informal labor market, private savings, and the ability to transfer leisure into consumption through either home production or shopping time (see Aguiar and Hurst (2005)). Motivated by these considerations, the key friction addressed in this paper is the non-exclusivity and non-observability of contractual relations. In the first part of the paper, we characterize the optimal contract under the assumption that none of the labor and credit relations an agent engages in can be observed by an individual firm.\footnote{The Panel Study of Income Dynamics reports for the years 1969 to 1985 a measure of income transfer received by households. We find that, in a given year, 24% of the households report receiving a transfer and 67% of the households received a transfer at some stage. These transfers are significant, averaging $1,930 (1983 dollars) and represent between 70% to 90% of total food expenditures.} We interpret \footnote{The characterization under exclusive contracts is well understood, see Prescott and Townsend (1984).}
this friction as reflecting both the costs that a firm might incur when monitoring the transactions agents engage in and the inability of firms to offer contracts contingent on the agents’ actions with other firms in the economy. In the second part of the paper, we endogeneize the observability of contracts by allowing firms to costly monitor contracts and take the model to the data.

The environment studied is a finite horizon dynamic Mirrleesian economy in which agents are privately informed about idiosyncratic labor productivity shocks that evolve over time. Agents wish to insure this risk by signing contracts with insurance providers (firms). Agents are not limited to a single insurance/labor relationship and can sign contracts with multiple firms. The contracting arrangements are private information of the contracting parties. Given this friction, in general, the communication between agent and firms cannot be limited to the exogenous private shock of agents (firms might also seek information about the other relations the agent has engaged in), as in the case with observable contracts. We extend the results in the common agency literature to our dynamic environment and characterize equilibrium using a menu game. In this game, each firm offers collections of payoff relevant alternatives – menus – and delegates to the agent the choice within these menus. The choice of the agent from a menu can reveal information about his type and the other contractual arrangements in which he might be involved. We impose no restriction on the contracts that firms can offer. A firm can, for example, offer a spot labor contract, a linear intertemporal borrowing and saving contract, a

---

state contingent dynamic insurance contract, and so on.

The non-observability of contracts removes the possibility of additional insurance beyond self-insurance and only linear contracts arise in equilibrium. We find that three optimality conditions must hold in equilibrium. First, the intertemporal marginal rate of substitution between consumption at time \( t \) and consumption at \( t + 1 \) is equal to the marginal rate of transformation (a standard Euler equation holds).\(^{39}\) Second, the marginal rate of substitution between consumption and leisure is equated to the marginal productivity for any time and any history.\(^{40}\) Third, the net present value of the transfers received in equilibrium is equal zero for every agent in the economy. These optimality conditions imply that the unique equilibrium allocation is equivalent to an economy in which agents can trade non-contingent bonds and are paid their marginal productivity and in which there is no redistribution. The intuition for this result is the following. If, for example, a firm offers an intertemporal contract at an implicit rate of return lower than the marginal rate of transformation, it would provide a profitable opportunity for an entrant: it can offer a contract with a return slightly higher and make profits.\(^{41}\) Such entry cannot be

\(^{39}\)If contracts are exclusive, the Euler equation does not hold and agents are savings constrained (see Golosov, Kocherlakota, and Tsyvinski (2003)).

\(^{40}\)This is also different with respect to the exclusive contracting environment (see, for example, Mirrlees (1971) and Golosov, Tsyvinski, and Werning (2006)), where this relation holds only for the highest skill type, while all of the remaining types face a distortion on the intratemporal margin that discourages consumption and hours provided.

\(^{41}\)Or similarly, offering a labor contract at an implicit wage lower than marginal productivity.
prevented by the first firm by also offering latent contracts because it cannot induce negative profits to the entrant.

These results, linking side trading and linear contracts, are reminiscent of Allen (1985), Hammond (1987), Cole and Kocharlakota (2001). We contribute to this literature by explicitly modeling competition between firms and determining endogenously the market structure. To sustain the equilibrium allocation we show that an incumbent firm must offer latent contracts to deter deviations of other incumbent firms. Moreover, in equilibrium more than one firm must offer the equilibrium allocation. The intuition for this result is that the equilibrium allocation is the most profitable non-contingent contract; however some contingent contracts deliver higher profits. If there is a unique incumbent or no latent contracts, a firm will deviate and offer one of these contracts.

To derive testable implications between non-exclusivity of contracts and the availability of insurance in the data, we generalize the model, relaxing the assumption about the observability of contracts. We assume that at time 0, a firm can pay a cost for each agent which allows the firm to observe all the contracts the agent signs. We assume that agents are heterogeneous with respect to the probability distribution of the productivity shock: some agents draw the productivity shock from a low mean distribution, while others draw from a distribution with higher

\[\text{In our equilibrium characterization, restricting to direct mechanisms, while not restrictive in the previous papers, results in non-existence of equilibrium.}\]
mean. If the cost is paid, a firm offers the optimal contract under exclusivity (as in Golosov, Kocherlakota, and Tsyvinski (2003) and Albanesi and Sleet (2006)). If the cost is not paid, firms offer the contract described in this paper, which implements the self-insurance allocation. With this extension, the model endogenously partitions the population into groups with access to different insurance contracts. Agents with lower average productivity have access to non-exclusive contracts while agents with higher productivity have access to exclusive contracts. We use US survey data to test whether agents’ consumptions and hours allocations, when grouped by education attainment, satisfy the optimality conditions under exclusive or non-exclusive contracts. We find that the consumption of college graduates evolves according to the inverse Euler equation, while for individuals with less than college, the consumption satisfies the standard Euler equation. Looking at the static consumption-leisure distortion calculated in the data, we investigate how it evolves as agents age. The model prescribes a constant distortion over age if workers have access to non-exclusive contracts while an increasing distortion in the other case. We find that also in this dimension, we cannot reject the hypothesis that high school graduates behave according to the linear contracts whereas the other group is closer to the constrained efficient contract.

**Related Literature**

This paper is related the literature on optimal social insurance contracts and its implementation through taxation, commonly referred to as *new dynamic public fi-*
nance. In general, the environment studied in these papers assumes that insurance is provided by a unique provider - the government - who perfectly controls both consumption and labor decision of the agents. With respect to this literature, this paper has two distinct implications. Our main result suggests that the constrained efficient allocation cannot be implemented in decentralized environments unless every aspect of the contracting is observable, thus making necessary the provision of insurance via taxes or a centralized institution that makes information public. However, our results also highlight that the presence of hidden and self-enforcing activities (for both consumption and labor) might undo any incentives the government provides through taxes. Related to this last point, our work is also related to a literature on optimal contract in the presence of hidden trades. In particular, Cole and Kocherlakota (2001) show that, in an private information endowment economy, equilibrium is equivalent to self-insurance when agents can secretly save in a storage technology. In an environment similar to ours, Golosov and Tsyvinski (2007) characterize equilibrium when agents can engage in hidden trades of Arrow-Debreu securities. They show that a standard Euler equation holds and that the decentralized equilibrium is not efficient, since firms do not internalize the effects of the contracts offered on the market rate of return. This paper can be seen as a generalization of the previous two papers, in the sense in that those the recontracting possibilities are assumed

\textsuperscript{43} For a review, refer to Kocherlakota (2006) and Albanesi (2008).

\textsuperscript{44} For example Cole and Kocherlakota (2001), Golosov and Tsyvinski (2007) and Abraham and Pavoni (2005).
exogenously (a market with linear prices or a storage technology) while in this paper the recontracting market is a result of an equilibrium game between insurance providers.

This paper also relates to Bisin and Guaitoli (2004), who analyze a static moral hazard environment under non-exclusive contracting. Their main result shows that latent contracts are used to sustain the equilibrium. However, the nature of the moral hazard environment, differently from our environment, enables latent contracts to prevent any profitable entry by additional insurance providers, thus delivering a positive profit equilibrium to the incumbents.

The quantitative analysis in this paper is related to Townsend (1995) and Ligon (1998). These papers investigate whether the consumption patterns in villages in Thailand and India, respectively, are consistent with the predictions of a constrained efficient allocation or the full information model. Ligon (1998) estimates the inverse Euler equation and the Euler equation for three villages in India. He finds that in two villages the consumption behavior is consistent with the Euler equation while in one village it is consistent with the constrained efficient allocation. Townsend (1995) investigates the consumption in Thai villages and finds that for some the constrained efficient allocation describes accurately the fluctuations while for others the full information model is a good benchmark. The study also emphasizes how villages differ in information flows between households (including assets and transactions) and how this could be responsible for the different insurance regimes observed.
The paper is organized as follows. In Section 9, we describe the environment and show that any equilibrium can be implemented by a menu game. Section 16 characterizes the equilibrium of our benchmark environment and shows that it is equivalent to self-insurance. We also show that latent contracts are necessary to implement the equilibrium allocation. Section 11 extends the model, allowing firms to observe contracts, and analysis its implications using US survey data. Section 12 is the conclusion.

9 Environment

Consider an economy populated by a continuum of measure one of ex ante identical agents and $I$ firms (insurance providers), with $I \geq 2$. The economy lasts for a finite number $T$ of periods. Agents’ period utility is defined over consumption $c$ and labor $l$ and is given by $u(c) - v(l)$. Agents discount future utility at rate $0 < \beta < 1$. Assume $u : \mathbb{R}_+ \rightarrow \mathbb{R}$ is twice continuously differentiable, increasing and strictly concave function, $\lim_{c \to 0} u'(c) = \infty$ and $\lim_{c \to \infty} u'(c) = 0$; and $v : \mathbb{R}_+ \rightarrow \mathbb{R}$ is twice continuously differentiable, increasing and strictly convex function and $\lim_{l \to \bar{L}} v'(l) = \infty$, where $\bar{L}$ is the maximum feasible number of hours in a period. At every time $t = 1, 2, \ldots, T$, each agent draws a privately observed productivity shock $\theta_t \in \Theta$, where $\Theta$ is a finite set and its smallest element is strictly positive.

We assume the law of large numbers holds. The shock is distributed according

\[ \text{Assume } \min_{\theta \in \Theta} \theta > 0. \]
to probability distribution $\pi(\cdot)$ and is independent and identically distributed over time and across agents. Let $\theta^t = (\theta_1, ..., \theta_t)$ denote the history of uncertainty of an agent up to time $t$. Given a sequence of consumption and leisure $\{c, l\} = \{c_t, l_t\}_{t=1}^T$, the expected discounted utility of an agent is given by

$$U(\{c, l\}) = \mathbb{E}_0 \sum_{t=1}^T \beta^{t-1}[u(c_t) - v(l_t)].$$

For a given realization of the labor productivity shock $\theta$, an agent can produce $y$ units of effective output according to $y = \theta l$, where $l$ denotes his labor input. We assume the labor input is private information of the agent while output $y$ is publicly observable to each firm for which the agent is producing output $y$.

Each firm $i \in \{1, ..., I\}$ offers labor and credit contracts to agents to insure against productivity shocks. A contract prescribes, at every time $t$, output requirement $y^i_t$ and consumption transfer $y^i_t + b^i_t$. The period profit of firm $i$ is given by $V^i(b^i) = -b^i$. Firms can transfer resources over time at constant rate $q$.\(^{46}\)

An important feature of our environment is that agents can sign contracts simultaneously with more than one firm, and the terms of the contract between an agent and a firm $i$ are not observed by other firms.\(^{47}\) We do not impose any restriction on the contracts offered by each firm. For example, a firm can offer a contract for the entire time horizon $t = 1, ..., T$; for a particular set of dates; only credit contracts ($y_t = 0, \forall t$); only labor contracts, or both. We also do not impose any specific contingency on the contracts; in particular, we do not restrict to linear contracts.

\(^{46}\)This fixed interest can be interpreted as the firm having access to external credit markets.

\(^{47}\)We assume that each agent is atomless and no interaction between agents is allowed.
At time 0, before any uncertainty is realized, agents sign a contract with each firm $i$.\textsuperscript{48} To take into account the voluntary participation of agents, every firm is required to offer at time 0 a null contract that determines no output requirement and no consumption transfers in every period. The contracts offered by a firm at time 0 are contingent on the future communication between that firm and the agent. We assume that contracts must be honored and neither firms nor agents can renege on them.\textsuperscript{49}

**Communication and Menu Games**

*Communication*

Firms and agents communicate according to communication mechanism,\textsuperscript{50} which

\textsuperscript{48}The ability of a firm to offer a contract prescribing transfers and output requirement at any future date $t$ and giving the agents the option of not entering the contract (offering the $(0,0)$ pair at time $t$) is, for the agent, a costless option of entering into that contract at time $t$. This, together with the fact that, for a firm, not contracting with an agent between time 0 and time $t$ does not reveal any additional, makes our analysis equivalent to the case where agents decide to enter or not a contract at any future date, not only at time 0.

\textsuperscript{49}We interpret contracts as self-enforcing in the following way. Both agents and firms have access to an enforcement mechanism (“court”) upon the payment of a cost, whenever one of the parties reneges on a contract. If this cost is paid, the terms of the contract between the two parties in consideration become public, and this court can enforce a punishment to the party that reneged on the contract. If either firms or agents falsely report a breach on the contracts, they can also be punished by court. We assume this punishment can be made large enough so that in equilibrium neither firms nor agents will renege the contracts signed.

\textsuperscript{50}No communication between firms is allowed.
consists of message spaces $R_i^t$ for time 0 and message spaces $M_i^t$ for each $t \in \{1, ..., T\}$, for each firm $i \in \{1, ..., I\}$. Denote the set of all possible messages that can be exchanged by an agent and firm $i$ up to time $t$ by $M_i^{t,t} = M_i^1 \times ... \times M_i^t$.

Each firm chooses allocation functions $g_i^t : M_i^{t,t} \rightarrow \mathbb{R}^2$, which specify transfers of consumption and output at time $t$, and $\phi_i : R_i^i \rightarrow G_i^1(M_i^{i,1}) \times ... \times G_i^T(M_i^{i,T})$, where $G_i^j(M_i^{i,t})$ is the set of all measurable mappings from message space $M_i^{i,t}$ to the allocation space $\mathbb{R}^2$. Let $(b(m_i^{i,t}), y(m_i^{i,t})) = g_i^t(m_i^{i,t})$ denote the allocation received by an agent who sends messages $m_i^{i,t} = (m_i^1, ..., m_i^t)$ to firm $i$. The function $\phi_i$ determines the contracts an agent will face in all subsequent periods. Denote by $G_i^t(M_i^t) = G_i^1(M_i^{i,1}) \times ... \times G_i^T(M_i^{i,T})$ and $M_i^t = M_i^1 \times ... \times M_i^t$. Let $\Phi_i^t(R_i^i, M_i^t)$ be the set of all measurable mappings from message space $R_i^t$ to the set $G_i^t$ and note that $\phi_i^t \in \Phi_i^t(R_i^i, M_i^t)$. Let $M = \times_{i=1}^I M_i^i$ and $R = \times_{i=1}^I R_i^i$. Denote the game associated with the communication mechanism $(M, R)$ by $\Gamma_{M,R}$.

At time 0, before any uncertainty is realized, each firm $i$ simultaneously offers a collection of allocation functions $\phi_i$, and agents communicate with firms sending a message $r_i$. This message determines, through $\phi_i$, the functions $g_i^t$ at every period $t$. The timing of the game $\Gamma_{M,R}$ is the following:

- At time 0:
  1. Each firm $i$ simultaneously offers contract $\phi_i : R_i^i \rightarrow G_i^t(M_i^i)$;
  2. Agents send a report $r_i^t \in R_i^i$ to each firm $i$. 

81
At time $t$:

1. Agent learns his private type $\theta_t$;
2. Firm offers allocation rule $g^i_t : \mathcal{M}^{i,t} \rightarrow \mathbb{R}^2$ according to $\phi^i(r^i)$;
3. Agent sends a message $m^i_t \in \mathcal{M}^i_t$ to each firm $i$;
4. Payoffs are realized.

Given messages $(\mathcal{M}, \mathcal{R})$, we consider a static Nash equilibrium played by firms at time 0 when choosing the contracts that are offered in future periods. Given these contracts, agents optimize choosing the report at time 0 and messages in every period $t = 1, \ldots, T$.

**Definition 1 (Equilibrium of Communication Game).** A pure strategy equilibrium of $\Gamma_{\mathcal{M}, \mathcal{R}}$ is $(r^*, m^*, \phi^*, g^*)$ such that:\footnote{We do not allow random strategies.}

1. **Agent's message** $m^*_t : G^1_t \times \cdots \times G^I_t \times \Theta^t \rightarrow \mathcal{M}_t$ solves for each $t \in \{1, \ldots, T\}$:

   \[
   U_t (m^{t-1}, \theta_t | g^*) = \max_{m_t \in \mathcal{M}_t} u \left( \sum_{i=1}^I (b^i(m^{i,t}) + y(m^{i,t})) \right) - v \left( \frac{\sum_{i=1}^I y(m^{i,t})}{\theta_t} \right) + \beta \sum_{\theta_{t+1}} \pi(\theta_{t+1}) U_{t+1} (m_t, \theta_{t+1} | g^*) ,
   \]

   subject to $\sum_{i=1}^I (b^i(m^{i,t}) + y(m^{i,t})) \geq 0$, $\sum_{i=1}^I y(m^{i,t}) \geq 0$, $\forall t$

   where $(b(m^{i,t}), y(m^{i,t})) = g^*_{t-i}(m^{i,t})$.

2. **Agent's reporting strategy at time 0**, $r^*: G^1 \times \cdots \times G^I \rightarrow \mathcal{R}$ solves:

   \[
   \max_{r \in \mathcal{R}} \sum_{\theta_1} \pi(\theta_1) U_1 (m^0, \theta_1 | r)
   \]
where \( g^i = \phi^{i,*}(r_i) \).

3. For each \( i \in \{1,...,I\} \), taking as given the choices of the other firms and the agents’ choices, firm’s allocation function \( \phi^{i,*} \) solves:

\[
V^i(\phi^{i,*}, \phi^{-i,*}) = \min \sum_{t=0}^{T} \sum_{\theta^t} \pi(\theta^t)q^t b^i_t(\theta^t),
\]

\[
b^i_t(\theta^t) = b(m^t, \theta^t), \quad g^i = \phi^i(r^{i,*}) \quad \text{and} \quad g^{-i,*} = \phi^{-i,*}(r^{-i,*}).
\]

Denote the equilibrium allocation of a general communication game by \((b^*, y^*)\).

**Menu Games**

If contracts are exclusive (or equivalently observable), the environment is equivalent to a standard dynamic Mirrleesian environment as in Golosov, Kocherlakota, and Tsyvinski (2003). In this case, the revelation principle guarantees that without loss of generality, firms can restrict to direct mechanisms that are incentive compatible.

However, under non-exclusive contracting, the preference ordering of the agents is influenced not only by their exogenous private information, but also by the set of contracts offered. In particular, the choice of an agent in the contracts offered by firm \( i \) depends on the contracts offered by other firms. This implies that restricting to a direct mechanism may not allow a firm to have a rich enough communication with the agent in order to obtain information on the other contracts.

In order to characterize the contracts offered by each firm, we extend the delegation principle proved by Peters (2001) and Martimort and Stole (2002) to our environment. This principle states that, without loss of generality, the equilibrium
outcomes of any communication game can be implemented as an equilibrium of a menu game. The key idea is that any communication in the original communication mechanism can be replaced by firms offering menus of payoff-relevant alternatives and delegating to the agents the choice within this menu. To incorporate a richer communication between firms and agents, firms might offer menus with elements that are not chosen in equilibrium (latent contracts). As highlighted by Arnott and Stiglitz (1991), offering latent contracts might be necessary to sustain particular equilibria by deterring entry of additional insurance providers and by preventing deviation of the incumbent insurance providers.\footnote{Our environment differs from the previous literature along two dimensions. First, the environment is dynamic in the sense that the exogenous uncertainty is realized in every period. Second, agents choose a communication-contingent contract from each firm \( i \) before any uncertainty is realized. This is important since at time 0, agents are identical thus might be possible to extract more information about the contracts being offered by other firms.}

A communication mechanism induces allocation functions and, hence, distribution over allocations. This means that to prove the equivalence between the equilibrium allocation of a given communication mechanism and the equilibrium of a menu game, it is essential that the menus offered are rich enough to capture the strategies used to implement equilibrium in a communication mechanism. In our environment, a menu is a sequence of sets, where each set is a subset of the allocation space \( \mathbb{R} \times \mathbb{R}_+ \). For a message space \((\mathcal{M}, \mathcal{R})\), define, for each firm \( i \), the set
\[
C_i^t(m^{i,t-1}, \mathcal{M}_i^t|G_i^t)
\]
as the menu that can be implemented through a message space.
\( \mathcal{M}_i^t \) at time \( t \) given a history of messages \( m^{i,t-1} \) and a set of allocation functions \( G_i^t \).

Formally, a menu at time \( t \) is the following set:

\[
C_i^t(m^{i,t-1}, \mathcal{M}_i^t|G_i^t) \equiv \{ C_i^t \subseteq \mathbb{R} \times \mathbb{R}_+ : \exists g_i^t \in G_i^t(\mathcal{M}_i^{i,t}) : C_i^t = \text{Im}(g_i^t|m^{i,t-1}) \} \quad \forall t, \forall i
\]

(66)

where

\[
\text{Im}(g_i^t|m^{i,t-1}) = \{ x \in \mathbb{R} \times \mathbb{R}_+ : \exists m_i^t \in \mathcal{M}_i^t : x = g_i^t(m^{i,t-1}, m_i^t) \} \quad \forall t, \forall i.
\]

(67)

Each set defined in (66) contains all subsets of \( \mathbb{R}^2 \) with cardinality at most \( \mathcal{M}_i^t \).

For any subset \( G_i^t \subseteq G_i^t(\mathcal{M}_i^{i,t}) \), let \( G_i^t = G_i^1 \times ... \times G_i^t \) and define a sequence of menus offered by firms at time 0 as:

\[
C(G_i^t) = \{ C_i^t \subseteq \mathcal{M}_i^t : \exists g_i^t \in G_i^t(\mathcal{M}_i^{i,t}) : C_i^t = \text{Im}(g_i^t|m^{i,t-1}) \} \quad \forall m^{i,t-1} \in \mathcal{M}_i^{i,t-1}, m_i^t \in \mathcal{M}_i^t \}
\]

(68)

At time 0, each agent chooses a sequence of menus in the collection offered by firm \( i \). Define \( C_i \) as the collection of menus that are consistent with a communication system \((\mathcal{M}, \mathcal{R})\).

\[
C_i(\mathcal{R}^i, \mathcal{M}^i) \equiv \{ C_i \subseteq \mathcal{M}^i(\mathcal{R}^i) : \exists \phi_i \in \Phi^i(\mathcal{R}^i, \mathcal{M}^i) : G_i = \text{Im}(\phi_i) \}.
\]

(69)

This set contains all the collections of sets \( C_i \) with cardinality less than or equal to the cardinality of \( \mathcal{R}^i \). Without explicitly writing the dependence on the message spaces, let \( C = C_i(\mathcal{R}^i, \mathcal{M}^i) \) and let \( C_i \) be a generic element of \( C_i \). Let \( C = \prod_i C_i \) and \( \mathcal{C} = \prod_i C_i \) be the collection of all menus. Let \( \Gamma_{C,\mathcal{C}} \) be the game associated with menus \((C, \mathcal{C})\).
Definition 2 (Equilibrium of Menu Games). A pure strategy equilibrium of a menu game is a collection of menus \( \hat{C} \) and agents’ choices \( \hat{C} \in \hat{C} \) and \((\hat{b}_i, \hat{y}_t)\) ∈ \( C_i^t(\hat{b}_{i,t-1}, \hat{y}_{i,t-1}|\hat{C}^n) \) ∀\( t \in \{1,...,T\} \), ∀\( i \in \{1,...,I\} \).\(^{53}\)

1. Agents’ choice at time \( t \), \((\hat{b}_t, \hat{y}_t) : C_t(\hat{b}_{t-1}, \hat{y}_{t-1}|\hat{C}) \times \Theta^t \rightarrow C_t(\hat{b}_{t-1}, \hat{y}_{t-1}|\hat{C})\) solves:

\[
U_t \left( b^{t-1}, y^{t-1}, \theta_t | \hat{C} \right) = \max_{(b_t,y_t) \in C_t(\hat{b}_{t-1}, \hat{y}_{t-1}|\hat{C})} u \left( \sum_{i=1}^{I} (b^i_t + y^i_t) \right) - v \left( \sum_{i=1}^{I} y^i_t \right) + \beta \sum_{\theta_{t+1}} \pi(\theta_{t+1}) U_{t+1} \left( b^t, y^t, \theta_{t+1} | \hat{C} \right),
\]

subject to \( \sum_{i=1}^{I} (b^i_t + y^i_t) \geq 0, \sum_{i=1}^{I} y^i_t \geq 0 \) ∀\( t \).

2. Agents’ choice at time 0, \( \hat{C} : \hat{C} \rightarrow \hat{C} \) solves:

\[
\max_{C \in \hat{C}} \sum_{\theta_1} \pi(\theta_1) U_1 (b^0, y^0, \theta_1 | C).
\]

3. For each \( i \in \{1,...,I\} \), \( C^i \) solves, taking as given \( \hat{C}_{-i} \) chosen by firms \( -i \) and the agents’ choice \( \hat{C}_{-i}, \{\hat{b}_{t}(\theta^t), \hat{y}_{t}(\theta^t)\}_{t=1}^{T} \):

\[
V^i(\hat{C}^i, \hat{C}_{-i}) \equiv \min_{\theta^t} \sum_{t=0}^{T} \sum_{\theta^t} \pi(\theta^t) q^t \hat{b}^i_t(\theta^t)
\]

\( \hat{b}^i_t(\theta^t) \in \hat{C}_i^t(\hat{b}_{i,t-1}, \hat{y}_{i,t-1}|\hat{C}^i), \hat{C}_i^t(\hat{b}_{i,t-1}, \hat{y}_{i,t-1}|\hat{C}^i) \in \hat{C}^i \)

\( \hat{b}^{-i}_t(\theta^t) \in \hat{C}_{t-i}^{-i}(\hat{b}_{-i,t-1}, \hat{y}_{-i,t-1}|\hat{C}_{-i}), \hat{C}_{t-i}^{-i}(\hat{b}_{-i,t-1}, \hat{y}_{-i,t-1}|\hat{C}_{-i}) \in \hat{C}_{-i} \).

Denote the equilibrium allocation of a menu game by \((\hat{b}, \hat{y})\).

Note that a menu might contain more alternatives than the cardinality of the type space, implying that some alternatives are not chosen in equilibrium. Similarly,

\(^{53}\)We do not allow for random menus.
at time 0 a firm might offer more than one set of contracts, also implying that some contracts are offered and not chosen by agents in equilibrium. We denote a contract as latent if it is offered in equilibrium by a firm but is not chosen in equilibrium by any agent. As we show in this paper, latent contracts have an important role in sustaining equilibrium allocations by preventing other firms from deviating to other contracts.

The following proposition shows that an equilibrium in a general communication system can be implemented as an equilibrium of a menu game. In this menu game, the collection of menus offered by each firm must be compatible with the general communication mechanism as defined above.

**Proposition 4** (Delegation Principle). Let \((b^*, y^*)\) be an equilibrium allocation of a general communication game \(\Gamma_{\mathcal{M}, \mathcal{R}}\). Then there exists \((\hat{b}, \hat{y})\) that is an equilibrium allocation of a menu game \(\Gamma_{\mathcal{C}, \mathcal{C}}\) and \((b^*, y^*) = (\hat{b}, \hat{y})\).

**Proof.** In Appendix 13.1.

Proposition 4 states that for given message spaces \((\mathcal{M}, \mathcal{R})\), there exists a menu game that implements the same equilibrium allocation. It is important to note that message spaces restrict the menus that can be offered in a menu game. From the previous result, if firms are allowed to use unrestricted message spaces, the same equilibrium can be implemented if firms can offer unrestricted menus as stated in Corollary 1 in Martimort and Stole (2002). From now on, we focus on unrestricted menu games.
The presence of two rounds of communication (at time 0 and at every time \( t \)) allows to further simplify the unrestricted menu game. We show that any time \( t \) menu that contains latent points (allocations not chosen in equilibrium) can alternatively be replaced by a time \( t \) menu with the same number of elements as the type space and latent menus at time 0. This implies that, without loss of generality, we can restrict firms to offering time \( t \) menus that have the same cardinality of the type space, which we call minimal menus.

**Definition 3** (Minimal Menus). A menu \( C^i \in C^i \) is minimal if for all \( C^{-i} \in C^{-i} \) and \((b^i_t, y^i_t) \in C^i_t\), for all \( C^i_t \in C^i \), there exists \( \theta_t \in \Theta \), such that \((b^i_t, y^i_t) = (b^{*,i}(\theta^t), y^{*,i}(\theta^t))\).

Intuitively, a menu is minimal if all of its elements are chosen by some agent in equilibrium.

**Proposition 5.** Let \( C = \{C^i, C^{-i}\} \) be an equilibrium of a menu game. There exists a payoff equivalent equilibrium \( \tilde{C} \), such that every \( \tilde{C}^i \in \tilde{C}^i \) is a minimal menu for all \( i \).

**Proof.** In Appendix 13.1.

10 Equilibrium Characterization

An important message of the previous section is that direct mechanisms might not be sufficient when characterizing the optimal contract. This means that firms
might offer latent (off-equilibrium) contracts. In this section, the use of latent contracts plays an important role, in particular to show that an equilibrium exists. We show that equilibrium would fail to exist if firms were restricted to offer direct mechanisms.\textsuperscript{54}

Characterization under Exclusive Contracts

Before characterizing the optimality conditions in our environment, we review two robust equilibrium conditions in an environment in which there is competition between insurance providers and \textit{contracts are exclusive}.\textsuperscript{55} The seminal paper of Prescott and Townsend (1984) shows that in a general class of private information economy, the first welfare theorem holds. The decentralized economy is equivalent to a planning problem that maximizes the ex ante lifetime utility of the agents subject to feasibility and incentive compatibility constraints (in every period for every realization agents weakly prefer the allocation designed for them).

In an environment similar to ours, and in the presence of exclusive contracting,

\textsuperscript{54}Throughout the paper, an incumbent refers to a firm that offers a menu that contains transfers and/or output recommendations other than the null contract and some agent chooses some of these contracts in equilibrium. An entrant refers to an insurance provider that, at all times, every agent chooses the null contract from the menus offered by this firm. We assume the number of firms \( I \) is large enough so that an entrant always exists.

\textsuperscript{55}Alternatively, the allocation can be implemented in an economy in which all the contracts an agent signs are observable and a firm can offer a contract contingent on agent’s actions with all other firms.
the equilibrium allocation has the following features:\textsuperscript{56}

1. The marginal rate of substitution between consumption and leisure is equated to the marginal productivity only for the highest type, originally shown by Mirrlees (1971):

\begin{equation}
    u'(c(\bar{\theta})) = \frac{1}{\bar{\theta}} v' \left( \frac{y(\bar{\theta})}{\bar{\theta}} \right),
\end{equation}

\begin{equation}
    u'(c(\theta)) > \frac{1}{\bar{\theta}} v' \left( \frac{y(\theta)}{\theta} \right), \quad \forall \theta \neq \bar{\theta}, \theta \in \Theta,
\end{equation}

where \( \bar{\theta} \equiv \max_{\theta \in \Theta} \theta \). The intuition for this result is the following: in order to separate types, it is optimal to discourage less productive agents to work. This implies that all but the most productive agents work and consume less than they would in a competitive environment.

2. If preferences are separable in consumption and leisure, the marginal rate of substitution of consumption between any two periods differs from the intertemporal rate of transformation for all types (the standard Euler equation does not hold):

\begin{equation}
    \frac{1}{u'(c(\theta^t))} = \frac{1}{\beta R} E \left[ \frac{1}{u'(c(\theta^{t+1}))} | \theta^t \right], \quad \forall t, \theta^t.
\end{equation}

This equation, derived originally by Rogerson (1985) and generalized in Golosov, Kocherlakota, and Tsyvinski (2003), implies that for all periods \( u'(c(\theta^t)) \) <

\textsuperscript{56} For a review of the results of constrained efficient allocation in dynamic Mirrleesian environments, refer to Golosov, Tsyvinski, and Werning (2006).
This means that it is optimal to make any type of agent saving constrained in order to encourage the truthful revelation of productivity in future periods.

10.1 Optimality Conditions under Non-exclusivity

We now derive the equilibrium conditions in the presence of non-exclusive contracting. This friction implies that the above equilibrium conditions cannot be implemented.

Under exclusivity, the optimal contract provides incentives to more skilled workers by discouraging less skilled agents to work (with respect to the full information allocation). The next lemma shows that this distortion cannot be implemented when contracts are not exclusive, since agents can work an extra amount to other firms.

**Lemma 1.** In any equilibrium for every $\theta^t \in \Theta^t$, for all $t$ the following holds:

$$u'(b(\theta^t) + y(\theta^t)) \leq v' \left( \frac{y(\theta^t)}{\theta_t} \right) \frac{1}{\theta_t},$$

(73)

where $b(\theta^t) = \sum_i b^i(\theta^t)$ and $y(\theta^t) = \sum_i y^i(\theta^t)$ and where $(b^i(\theta^t), y^i(\theta^t))$ are the contracts chosen by an agent with history $\theta^t$ from firm $i$ at time $t$.

**Proof.** Suppose that for some history $\theta^t$ equation (73) does not hold:

$$u'(b(\theta^t) + y(\theta^t)) > v' \left( \frac{y(\theta^t)}{\theta_t} \right) \frac{1}{\theta_t},$$

(74)
In this case, the agent would like to consume and work more than the equilibrium contract. An entrant can make strictly positive profits offering a supplemental contract with more consumption and output. Consider an entrant that offers the contract at time $t$, $C^E_t = \{(-\varepsilon, \delta^*(\varepsilon)), (0, 0)\}$ where $\delta^*$ and $\varepsilon$ are constructed as follows. Let $\delta^*(\varepsilon|\theta_t)$ be the solution of the following problem:

$$U(\varepsilon|\theta_t) \equiv \max_{\delta \geq 0} u(b(\theta^t) + y(\theta^t) + \delta - \varepsilon) - v\left(\frac{y(\theta^t) + \delta}{\theta^t}\right). \quad (75)$$

A necessary first order condition for this problem is:

$$u'(b(\theta^t) + y(\theta^t) + \delta^*(\varepsilon|\theta_t) - \varepsilon) \leq v'\left(\frac{y(\theta^t) + \delta^*(\varepsilon|\theta_t)}{\theta^t}\right) \frac{1}{\theta^t}. \quad (76)$$

If $\varepsilon = 0$, the solution for the above problem is $\delta^*(0|\theta_t) > 0$ given that (117) holds. From the Theorem of the Maximum, the solution $\delta^*(\varepsilon)$ is continuous on $\varepsilon$. Fix $\epsilon_1 > 0$ such that $|\delta^*(0) - 0| > \epsilon_1$. There exists $\epsilon_2 > 0$ such that if $|\varepsilon - 0| < \epsilon_2$ then $|\delta^*(\varepsilon) - \delta^*(0)| < \epsilon_1$. Let $\varepsilon$ be such that $0 < \varepsilon < \epsilon_2$.

An entrant offering this contract makes strictly positive profits, proportional to $\varepsilon$, and the agent is strictly better off given that his utility is higher in some history with positive probability. This contract is always profitable for the entrant even if other type $\tilde{\theta}_t$ accepts the deviating contract. The only way to deter this deviation is to have some latent contract that makes no agent willing to choose it. However, if such a contract existed, it would have been chosen in the original equilibrium, contradicting the fact that it is a latent contract.

When contracts are exclusive, the provision of incentives imply that agents
are savings constrained. The following lemma shows that this fails under non-exclusivity.

**Lemma 2.** In any equilibrium for every \( \theta^t \in \Theta^t \), for all \( t \), the following holds:

\[
    u'(c_t(\theta^t)) = \frac{\beta}{q} \sum_{\theta_{t+1}} u'(c_{t+1}(\theta^{t+1})) \pi(\theta_{t+1}),
\]

where \( c_t(\theta^t) = \sum_{i=1}^{I} (b^i_t(\theta^t) + y^i_t(\theta^t)) \).

**Proof.** In appendix 13.2. \( \blacksquare \)

The intuition for the result is the following. If the equilibrium allocation does not satisfy the Euler equation, an entrant firm can offer a savings (borrowing) contract at time \( t \) with an implicit interest rate lower (higher) than the marginal rate of transformation. As long as this contract is accepted, the entrant makes strictly positive profits and such contract can be constructed in a way that provides higher utility to the agent.

In the next proposition, we show that in equilibrium the marginal rate of substitution (MRS) between consumption and leisure is equated to the marginal productivity for every history and also that the lifetime transfer received under any history is equal to zero, so that there is no cross-subsidization between types.

**Proposition 6.** In any equilibrium the following two conditions hold:

1. Zero net present value of transfers:

\[
    \sum_{t=1}^{T} \left( \frac{1}{q} \right)^{1-t} b_t(\theta^t) = 0 \quad \forall \theta^T \in \Theta^T.
\]

93
2. **MRS equal to marginal productivity:**

\[
u'(b(\theta^t) + y(\theta^t)) = v' \left( \frac{y(\theta^t)}{\theta_t} \right) \frac{1}{\theta_t} \quad \forall \theta^t, t. \tag{79}\]

**Proof.** In appendix 13.2.

So far we characterized three necessary properties of the equilibrium allocation: (77), (78), and (79). In subsection 10.2, we show that there is a unique allocation that satisfies these conditions, which we denote by \( \{\hat{b}, \hat{y}\} = \{(b(\theta^t), y(\theta^t))_{t=1}^{T} \mid \theta^t \in \Theta^t\} \). The next proposition shows that an equilibrium exists by determining strategies of the firms (menus) that sustain this allocation as an equilibrium. A crucial element of the proof is that the equilibrium strategies must contain latent menus. These menus are similar to the ones derived in the characterization of equilibrium to show that any contract other than self-insurance is unprofitable.

**Proposition 7.** Allocation \( \{\hat{b}, \hat{y}\} \) is the unique equilibrium allocation of a menu game.

**Proof.** We construct strategies of the firms and the agents that sustain allocation \( \{\hat{b}, \hat{y}\} \) as an equilibrium. Let firm \( i \in \{1, 2\} \) offer the following menus:

\[
\hat{C}_1^i = \left\{(b_1^i, y_1^i) : b_1^i \in \mathbb{R}, \ y_1^i \in \mathbb{R}_+ \mid u'(b_1^i + y_1^i) = \frac{1}{\theta} v' \left( \frac{y_1^i}{\theta} \right) \forall \theta \in \Theta \right\},
\]

\[
\hat{C}_2^i(b_1^{i,T-1}, y_1^{i,T-1}) = \left\{(b_T^i, y_T^i) : b_T^i = 0, \ y_T^i \in \mathbb{R}_+ \mid u' \left( -\frac{1}{q} b_{T-1}^i + y_T^i \right) = \frac{1}{\theta} v' \left( \frac{y_T^i}{\theta} \right) \forall \theta \in \Theta \right\},
\]

and for periods \( t = 2, \ldots, T - 1 \):

\[
\hat{C}_2^i(b_t^{i,t-1}, y_t^{i,t-1}) = \left\{(b_t^i, y_t^i) : b_t^i \in \mathbb{R}, \ y_t^i \in \mathbb{R}_+ \mid u' \left( -\frac{1}{q} b_{t-1}^i + b_t^i + y_t^i \right) = \frac{1}{\theta} v' \left( \frac{y_t^i}{\theta} \right) \forall \theta \in \Theta \right\}.
\]
These firms also offer the following latent menus:

Dynamic Contract: for all $t = 1, \ldots, T$

$$C^i_{t,D}(b^{i,t-1}, y^{i,t-1}) = \left\{ (b^i_t, y^i_t) : b^i_t \in \mathbb{R}, \ y^i_t = 0 | b^i_t = -\frac{1}{q} b^i_{t-1} + x, \ x \in \mathbb{R} \right\}, \ b^i_T = b^i_0 = 0$$

Static Contract: for all $t = 1, \ldots, T$

$$C^i_{t,S} = \{(0, \delta) : \delta \in \mathbb{R}_+ \}$$

Remaining firms $i \in \{3, \ldots, I\}$ offer the null contract. Given these menus, the agents choose at time zero menu $\hat{C}^i$ from one of the two firms. We derive the agents’ choices by backward induction. At time $T$, an agent with history $(\theta^T-1, \theta_T)$ and past choices $(\tilde{b}(\theta^{T-1}), \tilde{y}(\theta^{T-1}))$ chooses from menu $C^i_T(\tilde{b}(\theta^{T-1}), \tilde{y}(\theta^{T-1}))$ the allocation $(\tilde{b}(\theta^T), \tilde{y}(\theta^T))$ such that

$$u' \left( -\frac{1}{q} \tilde{b}(\theta^{t-1}) + \tilde{y}(\theta^t) \right) = \frac{1}{\theta_t} v' \left( \frac{\tilde{y}(\theta^t)}{\theta_t} \right).$$

For time $t \in \{1, \ldots, T-1\}$, an agent with history $\theta^t$ and past choices $(\tilde{b}(\theta^{t-1}), \tilde{y}(\theta^{t-1}))$ chooses from menu $C^i_t(\tilde{b}(\theta^{t-1}), \tilde{y}(\theta^{t-1}))$ allocation $(\tilde{b}(\theta^t), \tilde{y}(\theta^t))$ such that

$$u' \left( -\frac{1}{q} \tilde{b}(\theta^{t-1}) + \tilde{b}(\theta^t) + \tilde{y}(\theta^t) \right) = \frac{1}{\theta_t} v' \left( \frac{\tilde{y}(\theta^t)}{\theta_t} \right).$$

Given agents’ choices, firm $i$’s profit is $\sum_{t=1}^T \sum_{\theta^t} q^i \tilde{b}(\theta^t) = 0$.

We next show that such strategies constitute an equilibrium by showing that there are no profitable deviations by firms. In particular, the latent contracts $C^i,S$ and $C^i,D$ are sufficient to deter any potential deviations.\footnote{Note that only offering menus $\tilde{C}^i$ is not an equilibrium since either an incumbent or an entrant will deviate, offering profitable welfare increasing menu, in the shape of a contingent contract. As}
As a first step, we show that it is not profitable for any firm to offer a contract that
specifies only intertemporal transfers (without any output requirements). Suppose
firm \( j \neq 1, 2 \) offers a menu \( C^j \) containing sequences of transfers \( \{b_t\}_{t=1}^T \). For each
feasible sequence in this menu, define the net present values of a sequence by:
\[
NPV(\{b_t\}_{t=1}^T) = \sum_{t=1}^T \frac{1}{q^{t-1}} b_t.
\]
The menu \( C^j \) is chosen by agents and is profitable only if contains at least one feasible sequence with \( NPV > 0 \) and one with \( NPV < 0 \).

Denote by \( \{\tilde{b}_t\}_{t=1}^T \) the feasible transfer with highest NPV. In the presence of
menu \( C^D \), all agents choose this sequence, implying that the entrant makes negative

For a sufficiently small \( \varepsilon \), the agent prefers this contract to the original, and in addition, this
deviation provides additional \( \frac{\varepsilon}{q-1} \) profits.

A sequence \( \{b_t\}_{t=1}^T \) is feasible if \( b_t \in C^j(b^{T-1}) \forall b_t, t \).

If all sequences have NPV=0, the menu is not chosen, since the equilibrium allocation \( \{\hat{b}, \hat{y}\} \) is
the allocation that maximizes agents’ welfare with no redistribution. Similarly, if all transfers are
negative, the menu is also not chosen, while if all transfer have \( NPV > 0 \) the firm makes a loss.
profits. Suppose not: there is an agent with history $\theta^T$ that chooses a sequence $\{b_t\}_{t=1}^T \neq \{\tilde{b}_t\}_{t=1}^T$. This agent is better-off by choosing the sequence $\{\tilde{b}_t\}_{t=1}^T$ and the following strategy in the menu $C^D$: $\delta_t = b_t - \tilde{b}_t$. This strategy enables him to replicate his original allocation and have extra resources, since the net present value of $\{\delta_t\}_{t=1}^T$ is negative:

$$\delta_0 = \sum_{t=1}^T \frac{1}{q^{1-t}} b_t - \sum_{t=1}^T \frac{1}{q^{1-t}} \tilde{b}_t < 0.$$ 

These additional resources can be used to increase consumption in any period, making the agent better-off.

The next step is to rule out contracts that offer jointly consumption transfers and output requirements. In appendix 13.3 we show that, in the last period, a contract that specifies transfers from productive to unproductive is unprofitable and any contract that implies redistribution from unproductive to more productive agents reduces agents’ welfare with respect to the equilibrium allocation. Thus, in the last period, a firm can only provide a contract with no redistribution.\(^{60}\)

For the dynamic case we focus on a two period example with two values of productivity shock, $\theta_H > \theta_L$. If the firm provides negative redistribution at time 1, given appendix 13.3, the contract will not be chosen at time zero. The remaining alternative is to provide, at time 1, some redistribution from the productive to the unproductive agent. To do this, a firm must offer transfers with higher net present

---

\(^{60}\)In a static environment this completes the proof since it rules out the existence of a contract that is, at the same time, profitable and preferred by the agents.
value together with higher output requirement. If not, the productive agent deviates, using both $C_{i,D}$ and $C_{i,S}$, replicating his original allocation and receiving transfers with higher NPV. Suppose now that at time 1, the $\theta_H$ agent receives transfers equal to $b_1 - \Delta$ while $\theta_L$ agent receives $b_1 + \Delta$ (with $\Delta > 0$). The best case for both agents is to receive transfers at time 2 that does not depend on the realization of the type in that period. Thus we can write transfers for the high type as $b_{2,H}$ and for the low type $b_{2,L}$. These transfers are such that $b_1 - \Delta + q b_{2,H} > b_1 + \Delta + q b_{2,L}$. This implies that a lower rate of return is charged to low productivity agents relative to high productivity agents. Since the low agent has lower consumption, this interest rate differential is welfare decreasing. Hence the benefits to the high agent are offset by the utility loss of the low agent. And, from an ex-ante perspective, the agent is better-off choosing the original equilibrium.

Finally for $\{\hat{b}, \hat{y}\}$ to be sustained as an equilibrium allocation, at least two firms must offer the equilibrium and the latent contracts. If not, the unique firm active in equilibrium will re-optimize, and offer a contract that implies some redistribution (as the example in footnote 57) since no latent contract is preventing such deviation. ❭

Summarizing, the allocation $\{\hat{b}, \hat{y}\}$ can be sustained in equilibrium by at least two incumbents simultaneously offering the menu $\hat{C}_i$ and the latent contracts $C_{i,S}$, and $C_{i,D}$. This is necessary to prevent deviations by any firm to a more profitable and ex ante welfare improving contract that features redistribution. This result highlights the importance of allowing firms to offer latent contracts. If offering such
contracts were not allowed, as in direct mechanisms, equilibrium would fail to exist
in this environment.

10.2 Equivalence to Self-Insurance

In the previous propositions we showed that the equilibrium allocation satisfies a
standard Euler equation, the marginal rate of substitution between consumption and
leisure is equated to marginal productivity in every period, and the net present value
of transfers received under any history is equal to zero (there is no redistribution).
These equilibrium conditions are the same optimality conditions in a decentralized
economy in which agents can borrow and save at rate \( R = 1/q \).

Let \( \{c^*, y^*\} = \{c^*(\theta^t), y^*(\theta^t)\}_{t=1}^T \) be the solution to the following problem:

\[
\max_{c,y \geq 0} \sum_{t=1}^{T} \sum_{\theta^t} \beta^{t-1} \pi(\theta^t) \left[ u(c(\theta^t)) - v\left(\frac{y(\theta^t)}{\theta_t}\right)\right]
\]

\[\text{s.t.} \quad \sum_{t=1}^{T} \frac{c(\theta^t) - y(\theta^t)}{R^{1-t}} = 0, \quad \forall \theta^T,\]

where \( R \) is taken as given.

**Proposition 8.** Let \( \{\hat{b}, \hat{y}\} = \{\hat{b}(\theta^t), \hat{y}(\theta^t)\}_{t=1}^T \) be the equilibrium allocation of a
menu game. Let the agents’ consumption be \( \hat{c}(\theta^t) = \hat{b}(\theta^t) + \hat{y}(\theta^t) \) for all \( \theta^t \) and for
all \( t \). If \( R = 1/q \), \( c^*(\theta^t) = \hat{c}(\theta^t) \) and \( y^*(\theta^t) = \hat{y}(\theta^t) \) for all \( \theta^t \) and for all \( t \).
Proof. The first order conditions of (83) are:

\[ u'(c(\theta^t)) = \beta R \sum_{\theta_{t+1}} u'(c(\theta^{t+1})) \pi(\theta_{t+1}), \]  
(84)

\[ u'(c(\theta^t)) = \frac{1}{\theta_t} v'\left(\frac{y(\theta^t)}{\theta^t}\right), \]  
(85)

\[ \sum_{t=1}^{T} \frac{c(\theta^t) - y(\theta^t)}{R^{1-t}} = 0, \quad \forall \theta^T. \]  
(86)

A solution to (83) exists. Also, the maximization problem (83) has a strictly concave objective function and the constraint set is convex; hence, the first order conditions are necessary and sufficient for the optimum and the optimum is unique.

The previous proposition summarizes how non-exclusivity and non-observability of contracts limit the ability to provide insurance and also the contracts that are offered in equilibrium. Our environment with firms interacting strategically and being allowed to offer any type of contracts, in equilibrium, is equivalent to an environment with competitive firms offering linear contracts with no redistribution. A immediate implication of the proposition is that the equilibrium is unique in terms of allocation.

**Corollary 1.** There is a unique equilibrium allocation of a menu game.

11 Endogenous Insurance and Quantitative Analysis

In this section, we derive a simple testable model that endogenously generates heterogeneous insurance regimes. To do this, we relax the assumption on observability
of the contracts. As in a costly state verification model, we give firms the option of paying a fixed cost, $\gamma \geq 0$, to monitor all the transactions an agent engages in.\textsuperscript{61} We assume that agents are heterogeneous with respect to the probability distribution of the productivity shock. There are two groups of agents: the first group draws the productivity shock from a low mean distribution, while the second draws from a distribution with higher mean. We show that in this modified environment, different groups of agents will have access to different insurance possibilities. Using US survey data, we show that this extension can rationalize the coexistence of multiple insurance regimes observed in the data.

11.1 Monitoring Costs

At time 0 (and only at time 0), before offering a set of contracts to an agent of type $j \in \{1, 2\}$, each firm chooses between the following two options: pay a cost $\gamma$ to observe all the contracts an agent engages in, and choose which contract to offer under full observability; or not pay the cost and offer the most profitable contract under non-exclusivity. Agents are heterogeneous with respect to the probability distribution of the productivity shock. A fraction of agents (“low mean agents”) draws, at every time $t$, a shock $\theta_t \in \Theta$, distributed according to $\pi(\cdot)$ while the a fraction of agents (“high mean” agents) draws the productivity shock $\lambda \theta_t$, where

\textsuperscript{61}Note that costly state verification models as in Townsend (1979) allow, upon paying the cost, the realization of uncertainty to be observable. Here we keep the realization of uncertainty private but allow the contracts an agent sign to be observable.
\( \theta_t \in \Theta \), distributed according to \( \pi(\cdot) \) and \( \lambda > 1 \). Let \( \bar{\theta} \) and \( \lambda \bar{\theta} \) be the average productivity of, respectively, low and high mean agents. Whether an agent is a low or high mean is publicly known by all the firms. For each group of agents, a firm decides whether to pay or not the monitoring cost and which contracts to offer in each case.

If a firm monitors an agent, the environment is equivalent to the one described in Prescott and Townsend (1984). We refer to optimal contract in this case as the “exclusive contract”. If the monitoring cost is not paid, the environment is the one studied in previous sections of this paper. From Proposition 8, this environment is equivalent in terms of allocation to a self-insurance economy, in which agents can borrow and save at fixed rate \( R \) and are paid wages equal to marginal productivity. We refer to the optimal contract in this case as the “non-exclusive” contract.

To determine which contract each group of agent will have access to, for a given value of monitoring cost \( \gamma \), firms compare the lifetime utility delivered under exclusive and non-exclusive contracts. This means that a firm finds profitable to pay the cost and offer the exclusive contract if agent’s utility is higher in this case. If firms do not find it profitable to pay the cost, an agent will receive the lifetime level of utility associated with non-exclusive contracts. We show that, under a particular assumption on the utility function, there exists a level of the monitoring cost such that low mean agents have access to the non-exclusive contract, whereas high mean agents have access to the exclusive contracts. In appendix 13.4, we show that if
\( \gamma = 0 \), the exclusive contract is always preferred over the non-exclusive, since it is cheaper to provide a given level of lifetime utility under exclusive contracts. For analytical convenience, we assume the following utility specification.

**Assumption 2.** \( u(c) = \log c; \quad v(l) = -a \log(1-l) \).

The zero profits level of lifetime utility \( \bar{w}^{NE}(\cdot) \) is defined as:

\[
\bar{w}^{NE}(x) = \max_{c,y} \sum_{t=1}^{T} \sum_{\theta \in \Theta} \beta^{t-1} \pi(\theta^t) \left[ u\left(c(\theta^t)\right) - v\left(y(\theta^t) x_{\theta_t}\right)\right] \\
\sum_{t} \left[ \frac{c(\theta_t^t) - y(\theta_t^t)}{R^{1-t}} \right] = 0, \quad \forall \theta^T,
\]

where the argument \( x = 1 \), \( \lambda \) refers to the agents’ productivity distribution.

Similarly for the exclusive contracts, define \( w^{E}(\cdot|\gamma) \) as follows. Note that for the exclusive contracts, the lifetime utility level that delivers zero profits also depends on the monitoring cost.

\[
w^{E}(x\theta|\gamma) = \max_{c,y} \sum_{t=1}^{T} \sum_{\theta \in \Theta} \beta^{t-1} \pi(\theta^t) \left[ u\left(c(\theta^t)\right) - v\left(y(\theta^t) x_{\theta_t}\right)\right] \\
\sum_{\theta^t, t} \beta^{t-1} \pi(\theta^t) \left[ u\left(c(\theta^t)\right) - v\left(y(\theta^t) x_{\theta_t}\right)\right] \geq \sum_{\theta^t, t} \beta^{t-1} \pi(\theta^t) \left[ u\left(c(\bar{\theta}^t)\right) - v\left(y(\bar{\theta}^t) x_{\theta_t}\right)\right] \forall \bar{\theta}^T,
\]

\[
\sum_{t} \left[ \frac{c(\theta_t^t) - y(\theta_t^t)}{q^{t-1}} \right] = \gamma, \quad \forall \theta^T.
\]

The following proposition states that there exists a value for the monitoring cost so that different agents have access to different insurance contracts.

**Proposition 9.** There exists \( \gamma^* > 0 \) such that:

\[
\bar{w}^{NE}(\bar{\theta}) = w^{E}(\bar{\theta}|\gamma^*) \\
\bar{w}^{NE}(\lambda \bar{\theta}) < w^{E}(\lambda \bar{\theta}|\gamma^*)
\]
Proof. In appendix 13.4. □

The steps to show the result are the following. We first show that, under non-exclusive contract, indirect lifetime utility of high mean agents is proportional to the lifetime utility of low mean agents (by a factor proportional to $\lambda$). The assumption on the utility function is crucial to show this result. Second, we show that, under exclusive contracts, the lifetime utility is scaled by a factor larger than $\lambda$. This implies that, for a given $\lambda$, there is a value of the monitoring cost so that the firms can promise a higher lifetime utility under the exclusive contract than under the non-exclusive. The same result can also be proved if $u(c,l) = \frac{c^{1-\sigma} l^{1-\sigma}}{1-\sigma}$. The general CRRA case, with $u(c,l) = \frac{c^{1-\sigma} l^{1-\sigma}}{1-\sigma} + a^{1-\sigma_l}$, is verified numerically.\textsuperscript{62}

11.2 Quantitative Implications

So far we showed that our extended model implies that different groups of agents have access to different insurance contracts. This implies that, along some dimensions, the allocations is characterized by a different set of equilibrium conditions. To test the implications of the model, we use US household survey data and divide the population by education attainment: those with less than a college degree and

\textsuperscript{62}Another way to endogenously divide the population in two different insurance regimes is to assume agents are heterogenous with respect to the monitoring cost $\gamma$. In this case, we show that there exists a cutoff value $\gamma^*$ such that if agents have cost $\gamma$, with $0 < \gamma \leq \gamma^*$, they have access to the exclusive contract and receive lifetime utility $w(\gamma)$. While agents with cost $\gamma$, with $\gamma > \gamma^*$, have access to the non-exclusive contract, receiving lifetime utility $\bar{w}^{NE}$. 

104
those who completed college or more. We estimate for each of the groups two implications of the model: an intertemporal optimality condition on consumption and an intratemporal condition on consumption and leisure. We consider the education level a proxy for a worker’s average productivity and according to our model, agents with higher average productivity (college graduates) satisfy the optimality conditions of the exclusive contracts, while agents with lower average productivity (high school graduates) satisfy the optimality conditions of non-exclusive contracts.

**Data**

We use the Krueger and Perri Consumer Expenditure Survey (CEX) dataset for the period 1980 to 2003 and divide the population by the education level of the reference person. To abstract from college and retirement decisions, we restrict our sample to households with the reference person age is between 25 and 55. We only consider reference person who worked more than 520 hours and less than 5096 hours per year and with positive labor income. We exclude households with wage less than half of the minimum wage in any given year and households who responded to all four interviews and with no missing consumption data. All the nominal data are deflated using the consumer price index calculated by the Bureau of Labor Statistics with base 1982-84=100.\(^{63}\) The consumption measure used includes the sum of expenditures on nondurable consumption goods, services, and small durable goods, plus the imputed services from housing and vehicles, as calculate by Krueger

\(^{63}\)For a more detailed description of the data and sample selection, refer to section 4.
Intertemporal Optimality Conditions

The first implication we test is an intertemporal optimality condition regarding the evolution of consumption. We showed that if agents have access to exclusive contracts, the consumption allocation satisfies the following inverse Euler equation:

\[
\frac{1}{u'(c_t(\theta^t))} \frac{\beta}{q} = E_t \left[ \frac{1}{u'(c_{t+1}(\theta^{t+1}))} \right].
\] (90)

On the other hand, if agents have access to non-exclusive contracts, the allocation must satisfy the following standard Euler equation:

\[
u'(c_t(\theta^t)) = \frac{\beta}{q} E_t \left[ u'(c_{t+1}(\theta^{t+1})) \right].
\] (91)

Assuming \( u(c) = \frac{c^{1-\sigma}}{1-\sigma} \), these equations imply, respectively:

\[
c_t(\theta^t)^\sigma \frac{\beta}{q} = E_t \left[ c_{t+1}(\theta^{t+1})^\sigma \right],
\] (92)

\[
c_t(\theta^t)^{-\sigma} = \frac{\beta}{q} E_t \left[ c_{t+1}(\theta^{t+1})^{-\sigma} \right].
\] (93)

These two equations can be nested in the following:

\[
c_t(\theta^t)^b \left( \frac{\beta}{q} \right)^{\frac{b}{b+1}} = E_t \left[ c_{t+1}(\theta^{t+1})^b \right].
\] (94)

If the inverse Euler equation (92) holds, then \( b > 0 \), whereas if the standard Euler equation (93) holds, \( b < 0 \). Taking expectation of (94) at time \( t \), we get

\[
\sum_{\theta^{t+1}} \pi(\theta^{t+1}) \left[ c_t(\theta^t)^b \left( \frac{\beta}{q} \right)^{\frac{b}{b+1}} - c_{t+1}(\theta^{t+1})^b \right] = 0, \quad \forall \theta^t.
\] (95)
For a given education group, we test whether the intertemporal consumption decision is compatible with exclusive or non-exclusive contracts by estimating the parameter $b$ in (95). If, for an education group, the value of $b$ is negative, the consumption of these agents is consistent with the predictions of (87). If the estimation of $b$ has a positive value, it implies that agents’ consumption satisfies the implications of (88). Our theory predicts that for more educated individuals the value of $b$ is positive. The analysis here closely follows Ligon (1998) and Kocherlakota and Pistaferri (2008).

\textit{Estimation Procedure and Results}

A typical household is on the sample for a total period of four quarters. For the estimation, we construct sample averages as follows. Denote by $c_{i,t}$ the consumption for household $i$ in the quarter that ends with month $t$, and let $N_t$ be the number of observations available at time $t$. We de-seasonalize consumption with dummies corresponding to the month the household was interviewed. The sample analog of equation (95) is:

$$g(b) = \frac{1}{T} \sum_{t=3}^{T} \left( \frac{\beta}{q} \right)^{\frac{b}{\pi}} \frac{1}{N_{t-3}} \sum_{i=1}^{N_{t-3}} c_{i,t-3}^b - \frac{1}{N_t} \sum_{i=1}^{N_t} c_{i,t}^b. \quad (96)$$

As shown by Kocherlakota and Pistaferri (2008), this sample analog is still valid in the presence of multiplicative classical measurement error in the consumption data.

\textsuperscript{64}In particular, Ligon (1998) tests whether the standard Euler or inverse Euler condition better describes the consumption behavior for three Indian villages. His results indicate that two out of three village provides evidence for the inverse Euler equation.
The main disadvantage of this sample analog is that, by taking means over the population, it does not take into account individual changes on consumption over time. An alternative valid sample analog is the following:

\[
\tilde{g}(b) = \frac{1}{T} \sum_{t=3}^{T} \left[ \left( \frac{\beta}{q} \right)^{\frac{b}{q}} \cdot \frac{1}{N_t} \sum_{i=1}^{N_t} \left( \frac{c_{i,t-3}}{c_{i,t}} \right)^b \right]
\]

where in this equation \(N_t\) is the total number of households with consumption data for time \(t\) and \(t-3\). The estimation of this equation, in the presence of multiplicative classical measurement error in consumption, implies inconsistent estimation of the parameter \(\beta\). A standard approach in the literature is to estimate the log-linearized version of this sample analog. Simple algebra shows that the log-linearized versions of equations (90) and (91) result in the same log-linearized equation. This means that this procedure cannot be used to test whether the consumption of a group of household satisfy (90) or (91).

In table 8 we report the estimation of parameter \(b\) for the two education groups, assuming \(\frac{\beta}{q} = 1\). We estimate the parameter \(b\) in (96) using non-linear generalized method of moments. We find that for college graduates \(b = 0.855\), which is consistent with exclusive contracts. While for individuals with education less than college the estimation indicates \(b = -1.128\), which corresponds to consumption evolving as predicted by non-exclusive contracts. Note that for agents with education less than college we reject that \(b\) is positive, while for college graduates we cannot reject a negative value for \(b\).

\textsuperscript{65}See Attanasio and Low (2004) and Ludvigson and Paxson (2001).
As a robustness check, we perform the same estimation by dividing the population into four education groups: those with less than a high school education, those who completed high school, those with some college, and those who completed college or more. The results are reported in table 9 and are consistent with the previous one: for individuals with education less than college the estimated value of \( b \) is negative, while for college graduates, this value is positive.

Another robustness check performed is to estimate the moment condition for all the households that have answered at least one of the interviews, not only for the households who have answered the four interviews. The results are reported in table 10 and are consistent with the results for the baseline sample. For both groups, the estimation of the coefficient of risk aversion is bigger than in the benchmark case and the standard errors are smaller. In this case, for both education groups, we can reject the value of \( b \) being the sign than the estimated.

We also perform our benchmark estimation by using the previous period interest rate as an instrument. The results are displayed in table 11 and are consistent with the benchmark results.

**Intratemporal distortions**

The second tested implication of the model regards the joint consumption and leisure decision at a given time. Define, for an individual \( j \), the intratemporal labor distortion as

\[
\tau_{ij}(t) = \frac{1}{\theta_i^j} \frac{\nu^j(y_i^j)}{u'(c_i^j)}
\]

(98)
Table 8: Estimation results for risk aversion

<table>
<thead>
<tr>
<th>Consumption</th>
<th>Less than College</th>
<th>College</th>
<th>All Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline de-seasonalized</td>
<td>-1.128</td>
<td>0.855</td>
<td>-1.030</td>
</tr>
<tr>
<td></td>
<td>(0.459)</td>
<td>(0.716)</td>
<td>(0.728)</td>
</tr>
<tr>
<td>Baseline truncated*</td>
<td>-1.128</td>
<td>0.864</td>
<td>-1.030</td>
</tr>
<tr>
<td></td>
<td>(0.458)</td>
<td>(0.726)</td>
<td>(0.728)</td>
</tr>
</tbody>
</table>

Estimation results for risk aversion from (95). A positive solution denotes the coefficient of risk aversion consistent with the household’s decision under the constrained efficient contract, while a negative solution denotes the estimated risk aversion consistent with the group being under a borrowing and saving contract. *We drop households with consumption changes bigger than 5 times in absolute value.
Table 9: Estimation results for risk aversion: multiple education groups

<table>
<thead>
<tr>
<th>Consumption</th>
<th>Less than HS</th>
<th>HS</th>
<th>Less than College</th>
<th>College</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline de-seasonalized</td>
<td>-0.773</td>
<td>-1.346</td>
<td>-0.962</td>
<td>0.855</td>
</tr>
<tr>
<td></td>
<td>(0.355)</td>
<td>(0.415)</td>
<td>(0.865)</td>
<td>(0.716)</td>
</tr>
<tr>
<td>Baseline truncated</td>
<td>-0.774</td>
<td>-1.348</td>
<td>-0.962</td>
<td>0.864</td>
</tr>
<tr>
<td></td>
<td>(0.355)</td>
<td>(0.414)</td>
<td>(0.864)</td>
<td>(0.726)</td>
</tr>
</tbody>
</table>
Table 10: Estimation results for risk aversion

<table>
<thead>
<tr>
<th>Consumption</th>
<th>Education Group</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Less than College</td>
<td>College</td>
<td>All Sample</td>
</tr>
<tr>
<td>Baseline de-seasonalized</td>
<td>-0.970</td>
<td>1.157</td>
<td>-0.898</td>
</tr>
<tr>
<td></td>
<td>(0.206)</td>
<td>(0.467)</td>
<td>(0.341)</td>
</tr>
<tr>
<td>Baseline truncated</td>
<td>-1.00</td>
<td>1.155</td>
<td>-0.904</td>
</tr>
<tr>
<td></td>
<td>(0.210)</td>
<td>(0.486)</td>
<td>(0.360)</td>
</tr>
</tbody>
</table>

Estimation results for risk aversion from (95) with households who answered at least one interview.
Table 11: Estimation results for risk aversion

<table>
<thead>
<tr>
<th></th>
<th>Education Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumption</td>
<td></td>
</tr>
<tr>
<td>Less than College</td>
<td>-0.972</td>
</tr>
<tr>
<td>College</td>
<td>1.167</td>
</tr>
<tr>
<td>All Sample</td>
<td>-0.890</td>
</tr>
<tr>
<td>Baseline instrumented</td>
<td></td>
</tr>
<tr>
<td>(0.213)</td>
<td>(0.450)</td>
</tr>
<tr>
<td></td>
<td>(0.365)</td>
</tr>
</tbody>
</table>

Estimation results for risk aversion from (95) using previous period interest rate as instrument.

As shown in section 3, if contracts are exclusive, an individual faces an increasing average value of $\tau_{cl}$ over the course of his working life. This result relies on the age dependent provision of incentives. As workers age (and the termination of the optimal contract gets closer) it is optimal to progressively provide more incentives using current promises of consumption and leisure (thus distorting more the static consumption-leisure condition) rather than promises of future consumption and leisure.

On the other hand, as proved in the previous section, if agents have access to non-exclusive contracts, $\tau_{cl}(t)$ is constant over age, since in this case the MRS is always equated to agent’s marginal productivity. Hence evaluating how this distortion evolves over the working life provides another testable implication of the model.
Estimation Procedure and Results

Using the following utility function $u(c, l) = \left(\frac{c^{\alpha} l^{1-\alpha}}{1-\sigma}\right)^{1-\sigma}$, equation (98) is:

$$
\tau^*_d(t) = 1 - \frac{1 - \alpha}{\alpha} \frac{1}{\theta_t^d} \frac{c^d_t}{L - l^d_t},
$$

(99)

The main advantage of using this utility is that the intratemporal distortion is not affected by the risk aversion parameter and $\alpha$ does not affect the behavior of $\tau_{cd}(t)$ over time.

To estimate the dependence of $\tau_{cd}$ on age, we regress its value on age. We run the regression on the standardized values of all variables.\footnote{Precisely: $\tau^*_d(age) = \tau_d + \delta \cdot age + \varepsilon^*_d$.} We calculate the labor distortion as follows. We use as proxy for a worker’s marginal productivity the imputed hourly wage, which is calculated dividing the total labor income by the total number of hours in a year For the measure of consumption, $c^d_t$ we use total consumption expenditure, $l^d_t$ is the yearly hours worked, and $L$ is the feasible amount of yearly working hours, set at 5200. To abstract from changes in family composition, we restrict the sample to individuals who are single.\footnote{In our baseline sample the single individuals represent 18% of the population.}

The results of this estimation are displayed in table 12. To control for heteroscedasticity and outliers, we estimate $\delta$ using a robust regression and for completeness we also report the OLS estimation. The coefficient on age for the entire sample is positive, as highlighted in Ales and Maziero (2008). We find that a zero coefficient (implying independence over age) cannot be rejected for individuals with education less than...
Table 12: Intra temporal distortion: singles

<table>
<thead>
<tr>
<th>δ</th>
<th>Less than College</th>
<th>College</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.011 (0.008)</td>
<td>0.016 (0.009)</td>
</tr>
<tr>
<td>Robust Regression</td>
<td>-0.0018 (0.005)</td>
<td>0.021 (0.005)</td>
</tr>
<tr>
<td>t-statistic</td>
<td>-0.36</td>
<td>3.93</td>
</tr>
</tbody>
</table>

college, whereas for college graduates the value of the coefficient is positive and significant, indicating that the labor distortion increases with age.

We also estimate (98) for a specification of the utility function that is separable on consumption and leisure. We assume $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$ and $v(l) = \frac{l^{1-\sigma_L}}{1-\sigma_L}$ with $\sigma_L = 2$.

Table 13 shows the results using for the coefficient of risk aversion the estimation of the Euler equation.\footnote{We also estimate the equation for different values of risk aversion, within the range estimated in the literature, and the result is qualitatively unchanged.} The result is the same as in the non-separable case: for less educated individuals, the coefficient on age is not significantly different than zero, while for college graduates this coefficient is positive. As a robustness check, we also calculate the labor distortion including married individuals in the sample. In this
Table 13: Intra temporal distortion: singles

<table>
<thead>
<tr>
<th>δ</th>
<th>Less than College</th>
<th>College</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.011 (0.008)</td>
<td>0.016 (0.009)</td>
</tr>
<tr>
<td>Robust Regression</td>
<td>-0.0078 (0.005)</td>
<td>0.05 (0.006)</td>
</tr>
<tr>
<td>t-statistic</td>
<td>-1.44</td>
<td>7.28</td>
</tr>
</tbody>
</table>

Intra temporal distortion by age and education group for singles using estimates for risk aversion derived from the estimation of the Euler equation.
Intra temporal distortion by age and education for household containing two adults.

<table>
<thead>
<tr>
<th>Education Group</th>
<th>Less than College</th>
<th>College</th>
</tr>
</thead>
<tbody>
<tr>
<td>δ</td>
<td>-0.0007 (0.0003)</td>
<td>0.0003 (0.0005)</td>
</tr>
<tr>
<td>OLS</td>
<td>-0.0004 (0.0002)</td>
<td>0.0011 (0.0003)</td>
</tr>
<tr>
<td>Robust Regression</td>
<td>-1.76</td>
<td>3.59</td>
</tr>
</tbody>
</table>

Due to data limitation, we restrict the sample to households with two or less adults, the reference person and the spouse. In this case, the results are also consistent with our benchmark estimation, as reported in table 14.

---

69The CEX records hours and earnings for the reference person and the spouse.
12 Conclusion

In this paper, we study a decentralized environment when firms compete for the provision of insurance. We focus on how the presence of non-exclusive trades endogenously limits the contracts offered, and consequently the amount of insurance implemented. We consider an environment in which consumers are privately informed about their skill shocks that evolve over time and can sign non-observable contracts with insurance providers. Our main results are that competition reduces the amount of insurance provided, the equilibrium is equivalent to a self-insurance economy, and only linear contracts are offered. Also, in equilibrium there is no redistribution.

To derive testable implications of the model, we extend the model and relax the assumption on the observability of contracts: firms can pay a cost to observe all the contracts an agent signs. Assuming agents are heterogeneous with respect to this cost, we find that agents with lower monitoring costs have access to the constrained efficient contract, while agents with higher monitoring costs have access to contracts that implement the self-insurance allocation. This implies that the first group of agents attains a higher level of lifetime utility. Considering education as a proxy for lifetime utility, we test the different intertemporal and intratemporal implications of this model using US data. We find that agents with a high level of education satisfy the optimality conditions of the constrained efficient model while the consumption and hours of agents with less education evolve according to the
self-insurance economy.
13 Appendix

13.1 Proofs of Section 9

Proof of Proposition 4

Proof. The proof is by construction. Starting from the equilibrium strategies of a
general communication game, we construct strategies for a menu game and show
that these strategies constitute an equilibrium.

Define as in (66) and (69) respectively the menus and the collection of menus
that are compatible with message spaces \((\mathcal{M}, \mathcal{R})\). Define the strategy of firm \(i\) in
this menu game as:

\[
\hat{C}_i = \{C_i \subseteq C^i (G^i) \mid G^i = \text{Im}(\phi^i,*)\}. \tag{100}
\]

The collection of menus \(\hat{C}_i\) contains all the subsets of the allocation space that are
consistent with the collection of allocation functions in the original equilibrium.

Agents’ strategies are defined as follows.

\[
\hat{C}_i = \{\hat{C}_i \subseteq \hat{C}_i : \hat{C}_i = \text{Im}(g^{i,*}_t|m^{i,t-1,*}) \text{ and } g^{i,*}_t = \phi^{i,*}(r^{i,*})\}
\]

\[
(\hat{b}^i(\theta^t), \hat{y}^i(\theta^t)) = g^{i,*}_t(m^{i,t,*}(\theta^t)).
\]

Note that by construction \(\hat{C}_i \subseteq \hat{C}_i\) and \((\hat{b}^i(\theta^t), \hat{y}^i(\theta^t)) \in \hat{C}_i, \forall \theta^t, \forall t\). The menu
\(\hat{C}_i\) is the subset of allocation space, \(\mathbb{R}^2\), that corresponds to the allocation function
chosen by the agent in the original equilibrium. Also \((\hat{b}_i, \hat{y}_i)\) corresponds to allocation
determined by the allocation function given the equilibrium message sent by each type \( \theta^t \). If agents and firms follow these strategies, the equilibrium allocation in the menu game is the same as in the original equilibrium.

First let's show that the agents' strategies are an equilibrium. Suppose that at some time \( t \), for some firm \( i \) \( \exists (b^i_t, y^i_t) \in \hat{C}^i_t \) such that:

\[
\begin{align*}
& u \left( \sum_{i=1}^{I} (b^i_t + y^i_t) - v \left( \frac{\sum_{i=1}^{I} y^i_t}{\theta_t} \right) + \beta \sum_{\theta_{t+1}} \pi(\theta_{t+1}) U_{t+1} \left( \hat{b}^{i-1}, b^i_t, \hat{y}^{i-1}, y^i_t, \theta_{t+1} | \hat{C} \right) > \\
& u \left( \sum_{i=1}^{I} (\hat{b}^i_t + \hat{y}^i_t) - v \left( \frac{\sum_{i=1}^{I} \hat{y}^i_t}{\theta_t} \right) + \beta \sum_{\theta_{t+1}} \pi(\theta_{t+1}) U_{t+1} \left( \hat{b}^i_t, \hat{y}^i_t, \theta_{t+1} | \hat{C} \right).
\end{align*}
\]

Since \((b^i_t, y^i_t) \in \hat{C}^i_t\), there exists \( m^i_t \in M^i_t \) such that \((b^i_t, y^i_t) = g^i_{t, *}(m^i_t)\). Replacing in the agents' payoff:

\[
\begin{align*}
& u \left( \sum_{i=1}^{I} (b^i_t(m^{i,t}) + y(m^{i,t})) - v \left( \frac{\sum_{i=1}^{I} y(m^{i,t})}{\theta_t} \right) + \beta \sum_{\theta_{t+1}} \pi(\theta_{t+1}) U_{t+1} \left( m^i_t, \theta_{t+1} | g^* \right) > \\
& u \left( \sum_{i=1}^{I} (\hat{b}(m^{i,t,*}) + y(m^{i,t,*})) - v \left( \frac{\sum_{i=1}^{I} y(m^{i,t,*})}{\theta_t} \right) + \beta \sum_{\theta_{t+1}} \pi(\theta_{t+1}) U_{t+1} \left( m^{i,*}, \theta_{t+1} | g^* \right).
\end{align*}
\]

But this contradicts \( m^{i,*} \) being an equilibrium in the original game. Now suppose \( \hat{C}^i \) is not an equilibrium for some \( i \). There exists some \( C^i \in \hat{C}^i \) such that:

\[
U(C^i, \hat{C}_{-i}) > U(\hat{C}).
\]

Since \( C^i \in \hat{C}^i, \exists r^i \in R^i \) such that \( C^i = \text{Im}(g^i) \) and \( g^i = \phi^{i,*}(r^i) \). Replacing in the agents' payoff:

\[
U \left( g^i, g_{-i}^* \right) > U \left( g^{i,*}, g_{-i}^* \right).
\]

But this contradicts \( r^{i,*} \) being an equilibrium in the original game.
Finally, we check that firms’ strategies constitute an equilibrium. Suppose $\exists C^i \in C^i(\mathcal{R}^i, \mathcal{M}^i)$ such that $V^i(C^i, \hat{C}^{-i}) > V^i(\hat{C}^i, \hat{C}^{-i})$.

Since $C^i \in C^i(\mathcal{R}^i, \mathcal{M}^i)$, there exists $\phi^i$ such that $g^i = \phi^i(r^i,*)$. Replacing in the firm’s payoff in the original game $V^i(\phi^i, \phi^*_c) > V^i(\phi^i, \hat{\phi}^*_c)$. But this contradicts $\phi^i$ being an equilibrium in the original game.

Proof of Proposition 5

Proof. We show the equivalence by construction. For a given firm $i$, by assumption there exists at least one $C^i \in C^i$ which is not minimal. As notation let $C^i \times C^{-i} = C^1 \times C^i \times \ldots \times C^N$, and let $U(C^i, C^{-i})$ be the lifetime utility of a sequence of menus $C$ as defined in equilibrium. Define the set

$$P(C^i) = \{C^{-i} \in C^i | U(C^i, C^{-i}) \geq U(\hat{C}^i, C^{-i}) \quad \forall \hat{C}^i \in C^i\}. \quad (101)$$

The set $P(C^i)$ contains all the menus $C^{-i}$ offered by other firms $-i$ that resulted in $C^i$ being chosen from firm $i$. Note that if $C^i$ is the unique element of $C^i$ the set $P(C^i) = C^i$.

For each $C^{-i} \in P(C^i)$, construct the following sequence of menus:

$$\hat{C}^i_t(C^{-i}|C^i) \equiv \{(b^i_t(\theta^i, C^i, C^{-i}), y^i_t(\theta^i, C^i, C^{-i})) \in C^i_t, \quad \forall \theta_t \in \Theta\}, \quad \forall C^{-i} \in C^i, \forall t. \quad (102)$$

Each set $\hat{C}^i_t(C^{-i}|C^i)$ contains the actual equilibrium choices of each type of agent and is a minimal menu. Let $\hat{C}^i(C^{-i}) \equiv \{\hat{C}^i_t(C^{-i}|C^i) \forall C^{-i} \in C^i, \forall t\}$. Finally let $\tilde{C}^i = \{\hat{C}^i(C^{-i}) \forall C^{-i} \in P(C^i)\}$. We now replace the menu $C^i \in C^i$ by $\tilde{C}^i$.
and show that the equilibrium is the same. Let $\tilde{C} = \{(\tilde{C}_i \setminus C_i), \tilde{C}_i\}$. We prove the statement in two steps. We first show that each element of $\tilde{C}_i$ is chosen by the agent if and only if $C_i$ was chosen in the original equilibrium. We then show that $\tilde{C} = \{\tilde{C}_i, C^{-i}\}$ is an equilibrium of the menu game by showing that none of the firms $-i$ deviates to any $\tilde{C}^{-i}$.

To show the first step, given that $C_i$ was chosen in the original equilibrium $U(C_i, C^{-i}) \geq U(\hat{C}_i, C^{-i}) \forall \hat{C}_i \in C_i$. By construction, we have that $U(\tilde{C}_i(C^{-i}), C^{-i}) \geq U(C_i, C^{-i})$ and $U(\tilde{C}_i(C^{-i}), C^{-i}) \geq U(\hat{C}_i, C^{-i})$, for all $\hat{C}_i \in \tilde{C}_i$, so that $U(\tilde{C}_i(C^{-i}), C^{-i}) \geq U(C_i, C^{-i})$ for all $C_i \in \tilde{C}_i$. To prove the reverse, suppose $\hat{C}_i \in \tilde{C}_i$ is chosen by the agent. By the definition of $\tilde{C}_i$, if $(b_i^t, y_i^t) \in \hat{C}_i \in \tilde{C}_i$ then $(b_i^t, y_i^t) \in C_i \in C_i$ so $U(C_i, C^{-i}) \geq U(\hat{C}_i, C^{-i})$ and $U(C_i, C^{-i}) \geq U(C', C^{-i})$ for all $C' \in \tilde{C}_i$. Given that $\hat{C}_i$ is chosen then $U(\hat{C}_i, C^{-i}) \geq U(C', C^{-i})$ for all $C' \in C_i \setminus \tilde{C}_i$. Combining these inequalities, we get that $U(C_i, C^{-i}) \geq U(C', C^{-i})$ for all $C' \in C_i$, implying that $C_i$ is chosen in the original equilibrium.

Suppose there exists a collection of menus $\tilde{C}^{-i}$ so that $V^{-i}(\tilde{C}^{-i}, \tilde{C}_i) > V^{-i}(C^{-i}, \tilde{C}_i)$. Let $C^*$ denote the equilibrium choice of the agent, such that $U\left(C^*, C^{-i}\right) \geq U(\tilde{C}_i, C^{-i})$, for all $(\tilde{C}_i, \tilde{C}^{-i}) \in \tilde{C}_i \times \tilde{C}^{-i}$. The first case is if $\tilde{C}^{-i} \cap P(C_i) = 0$. If $C^{-i,*} \in \tilde{C}^{-i} \cap C^{-i}$, we immediately reach a contradiction since $C^*$ was also chosen in the previous equilibrium so that profits must be equal. If $C^{-i,*} \notin \tilde{C}^{-i} \cap C^{-i}$, then we reach a contradiction with $C$ being an equilibrium, since firm $-i$ would have deviated from offering the menu $C^{-i} \setminus P(C_i) \cup C^{-i,*}$ and would make strictly greater
profits.

The second case is if $\tilde{C}^{-i} \cap P(C^i) \neq 0$. In this case, if $C^{-i,*} \in P(C^i)$ we immediately reach a contradiction since the agent chooses the same menu in both equilibria so profits are the same. If $C^{-i,*} \notin P(C^i)$ then we contradict $C$ being an equilibrium, since firm $-i$ would deviate from offering $C^{-i}\P(C^i) \cup C^{-i,*}$.

Repeating this procedure for every non-minimal menu $C^i$ in the original $C^i$, we construct $\tilde{C}$ where every menu is minimal.

13.2 Proofs of Section 16

13.2.1 Proof of Lemma 2

Proof. Suppose that for some history $\hat{\theta}^t$ equation (77) does not hold.

Case 1:

\[ u'(c_t(\hat{\theta}^t)) > \frac{\beta}{q} \sum_{\theta_{t+1}} u'(c_{t+1}(\hat{\theta}^t, \theta_{t+1})) \pi(\theta_{t+1}). \]  

(103)

In this case, the agent is borrowing constrained. An entrant can make strictly positive profits offering a borrowing contract at a rate higher than $1/q$, contradicting the original allocation being an equilibrium. The first step is to construct the contract to be offered by a firm. Let $\delta^*(\varepsilon)$ be the solution of the following problem:

\[ U(\varepsilon) \equiv \max_{\delta \geq 0} u(c_t(\hat{\theta}^t) + \delta) + \beta E_t u \left( c_{t+1}(\hat{\theta}^t, \theta_{t+1}) - \delta \left( \frac{1}{q} + \varepsilon \right) \right). \]  

(104)

A necessary first order condition for this problem is:

\[ u'(c_t(\hat{\theta}^t) + \delta) \leq \beta \left( \frac{1}{q} + \varepsilon \right) E_t u' \left( c_{t+1}(\hat{\theta}^t, \theta_{t+1}) - \delta \left( \frac{1}{q} + \varepsilon \right) \right). \]  

(105)
If $\varepsilon = 0$, the solution for the above problem is $\delta^*(0) > 0$ given that (103) holds. From the Theorem of the Maximum, the solution $\delta^*(\varepsilon)$ is continuous on $\varepsilon$. Fix $\varepsilon_1 > 0$ such that $|\delta^*(0) - 0| > \varepsilon_1$. There exists $\varepsilon_2 > 0$ such that if $|\varepsilon - 0| < \varepsilon_2$ then $|\delta^*(\varepsilon) - \delta^*(0)| < \varepsilon_1$. Let $\varepsilon$ be such that $0 < \varepsilon < \varepsilon_2$. Consider an entrant that offers the contract $C_t = \{(\delta^*(\varepsilon), 0), (0, 0)\}$ and $C_{t+1} = \{(-\delta^*(\varepsilon)(\frac{1}{q} + \varepsilon), 0), (0, 0)\}$ and the contract $(0, 0)$ for all other periods. This firm is making strictly positive profits, proportional to $\delta^*(\varepsilon)\varepsilon$, and the agent is strictly better off keeping the original equilibrium together with this contract since increases his utility in a history with positive probability and keeps the same utility in all other histories.

Hence, under the original equilibrium, a firm can offer a contract that makes strictly positive profits. This contradicts the allocation being an equilibrium.

The other case can be proved using a similar argument. ■

13.2.2 Proof of Lemma 6

Proof. For a history $\theta^{t-1}$, define the net present value of transfers received from time $t$ onwards by:

$$A_t(\theta^{t-1}, \theta_T^{t-1}) = \sum_{n=t}^{T} \left(\frac{1}{q}\right)^{t-n} b_n(\theta^{t-1}, \theta_n^{t-1}),$$

(106)

where $\theta_n^{t-1} = (\theta_t, \theta_{t+1}, \ldots, \theta_n)$ is the sequence of shocks following history $\theta^{t-1}$ from time $t$ to $n$ and $b_n(\theta^{t-1}, \theta_n^{t-1})$ is the equilibrium transfer chosen at time $n$ by an equilibrium transfer $\hat{\theta}_t$, $\theta_{t+1}, \ldots, \theta_n$.

$\text{Note that } u'(c_t(\hat{\theta}_t') + \delta^*(\varepsilon)) \text{ is a finite, strictly positive number, hence also is the right hand side of equation (105). This implies that } c_{t+1}(\hat{\theta}_t', \theta_{t+1}) - \delta^*(\varepsilon)\left(\frac{1}{q} + \varepsilon\right) > 0 \text{ for all } (\hat{\theta}_t', \theta_{t+1}).$
agent with history $\theta^n$. We show, using a backward induction argument, that for all $t$, $A_s(\theta^{s-1}, \theta^T_s-1)$ is independent of $\theta^T_{s-1}$ for all $s \geq t$. This implies that $A_1(\theta^T)$ is the same for all $\theta^T \in \Theta^T$. If $A_1(\theta^T) > 0$, firms make strictly negative profits in equilibrium and would be better off offering a null contract. If $A_1(\theta^T) < 0$, an entrant can offer the same sequence of transfers giving an additional transfer $\varepsilon > 0$ in the terminal period. Since the sequence of transfers is not contingent and is profitable for all types, there is no latent contract that makes it unprofitable.

1. **Equations (78) and (79) hold for $t = T$.**

We first show that at time $T$, transfers are independent of realization of time $T$ shock and then show that for time $T$ equation (79) holds.

**Equation (78) holds at $t = T$:**

Suppose that (78) does not hold and let $b(\theta^T) = \min_{b \in C(\theta^T-1, y^T-1)} b$ and $b(\theta^T-1, \hat{\theta}_T)$ the second smallest $b$. Denote by $\hat{\theta}^T = (\theta^T-1, \hat{\theta}_T)$. The contradiction argument relies on the incumbent firm deviating to an allocation that delivers higher profits. First note that it must be true that $y(\hat{\theta}^T) + b(\hat{\theta}^T) > y(\theta^T) + b(\theta^T)$. If not, given that $b(\hat{\theta}^T) > b(\theta^T)$ then $y(\hat{\theta}^T) < y(\theta^T)$, an entrant firm can offer the following contract $\tilde{C}_T = \{(-\varepsilon, y(\theta^T) - y(\hat{\theta}^T)); (0, 0)\}$, for some $\varepsilon$ small enough. An agent with type $\theta^T$ is better off by choosing allocation $(b(\hat{\theta}^T), y(\hat{\theta}^T))$ in menu $C_T$ together with $(-\varepsilon, y(\theta^T) - y(\hat{\theta}^T))$ in menu $\tilde{C}_T$. With these choices, his utility is:

$$u \left( b(\hat{\theta}^T) - \varepsilon + y(\theta^T) \right) - v \left( \frac{y(\theta^T)}{\theta^T} \right) > u \left( b(\theta^T) + y(\theta^T) \right) - v \left( \frac{y(\theta^T)}{\theta^T} \right)$$
where the inequality holds as long as \( b(\hat{\theta}^T) - \varepsilon > b(\theta^T) \). No latent contracts can prevent this deviation, since it is profitable for the entrant as long as some agent accepts it.\(^\text{71}\)

The equilibrium allocation, being optimal for the agent, must satisfy the following:

\[
\begin{align*}
    u(b(\hat{\theta}^T) + y(\hat{\theta}^T)) - v \left( \frac{y(\hat{\theta}^T)}{\hat{\theta}^T} \right) & \geq u(b(\theta^T) + y(\theta^T)) - v \left( \frac{y(\theta^T)}{\theta^T} \right), \quad (107) \\
    u(b(\theta^T) + y(\theta^T)) - v \left( \frac{y(\theta^T)}{\theta^T} \right) & \geq u(b(\hat{\theta}^T) + y(\hat{\theta}^T)) - v \left( \frac{y(\hat{\theta}^T)}{\hat{\theta}^T} \right). \quad (108)
\end{align*}
\]

**Case 1** If (107) holds with equality, an agent of type \( \hat{\theta}^T \) is indifferent between his equilibrium choice and the choice of agent \( \theta^T \). However, the insurance providers receive strictly higher profits from the allocation \( \theta^T \), since by assumption \( b(\hat{\theta}^T) > b(\theta^T) \). This incumbent can deviate to an alternative menu that differs from the original by offering at time \( T \) only the allocation chosen by agent \( \theta^T \). No latent contract can induce lower profits to deter this deviation, since now the deviating incumbent offers a subset of the allocations that were available in the original equilibrium. The argument also holds if the equilibrium allocation is divided between multiple insurance providers.

\(^\text{71}\)Note that this case arises in the solution of the constrained efficient allocation: high skilled agents work more and make positive transfers to less skilled agents. The deviation \( \tilde{C}_T \) makes this allocation unprofitable in our environment, since it induces skilled agents to choose the allocation designed for low skilled agents and working an additional amount with entrant.
**Case 2** Suppose that (107) holds with strict inequality. Following the argument in the previous case, for any type \( \theta_T \) such that \( b(\theta_T) > b_T(\theta_T) \), it must be true that:

\[
u(b(\theta_T) + y(\theta_T)) - v\left(\frac{y(\theta_T)}{\theta_T}\right) > u(b(\theta_T) + y(\theta_T)) - v\left(\frac{y(\theta_T)}{\theta_T}\right).
\]

(109)

Otherwise, the incumbent firm will offer only the contract containing \( b_T(\theta_T) \).

Consider the following deviation by an incumbent firm \( \tilde{b}(\theta_T) = b(\theta_T) - \varepsilon \) and \( \tilde{b}(\theta_T) = b(\theta_T) + \varepsilon - \delta \) for \( \varepsilon, \delta > 0 \) and \( \varepsilon > \delta \) (to be defined explicitly below) and keeping unchanged all the other allocations.\(^{72}\) This deviation reduces the spread of transfers and increases incumbent’s profit by a factor proportional to \( \delta \).

To show that such deviation is profitable, thus reaching a contradiction, we show that there is no latent contract \( \alpha \equiv (\alpha_b, \alpha_y) \) that can induce a reduction in the profits of this firm. Suppose such contract exists. One possibility is to induce \( \theta_T \) agents, when faced with the deviating allocation \( \tilde{b} \), to choose \( \tilde{b}(\theta_T) \). This would imply a reduction of profits, since \( \tilde{b}(\theta_T) > \tilde{b}(\theta_T) \). Such latent contract has to satisfy:

\[
u(\tilde{b}(\theta_T) + y(\theta_T) + \alpha_b + \alpha_y) - v\left(\frac{y(\theta_T) + \alpha_y}{\theta_T}\right) > u(\tilde{b}(\theta_T) + y(\theta_T)) - v\left(\frac{y(\theta_T)}{\theta_T}\right).
\]

(110)

Since \( \alpha \) is not chosen in the original equilibrium, it must also be true that

\[
u(b(\theta_T) + y(\theta_T)) - v\left(\frac{y(\theta_T)}{\theta_T}\right) \geq u(b(\theta_T) + y(\theta_T) + \alpha_b + \alpha_y) - v\left(\frac{y(\theta_T) + \alpha_y}{\theta_T}\right).
\]

(111)

\(^{72}\)If there are multiple \( \theta \) with values equal to \( b(\theta_T) \) or \( b(\theta_T) \), the same deviation applies to all such transfers.
However, \( u(\tilde{b}(\theta^T) + y(\theta^T)) > u(b(\theta^T) + y(\theta^T)) \) and \( u(b(\tilde{T}) + y(\tilde{T}) + \alpha_b + \alpha_y) \geq u(\tilde{b}(\tilde{T}) + y(\tilde{T}) + \alpha_b + \alpha_y) \), which combined with (111) implies

\[
\Delta(\tilde{T}) = \min_{\alpha \in C_{\tilde{T}}^{-1}} \left\{ u(b(\tilde{T}) + y(\tilde{T})) - v \left( \frac{y(\tilde{T})}{\theta_T} \right) + u(b(\tilde{T}) + y(\tilde{T}) + \alpha_b + \alpha_y) - v \left( \frac{y(\tilde{T}) + \alpha_y}{\theta_T} \right) \right\}.
\]

This gives the minimum utility gain agent \( \tilde{T} \) receives from choosing allocation \((b(\tilde{T}), y(\tilde{T}))\) instead of \((b(\tilde{T}), y(\tilde{T}))\) combined with any other latent contract \( \alpha \). Since (114) holds with strict inequality, \( \Delta(\tilde{T}) \) is strictly positive for each \( \tilde{T} \). Let \( \tilde{\alpha} \equiv \arg \min \Delta(\tilde{T}) \). There exists \( \varepsilon(\tilde{T}) > 0 \) such that

\[
\begin{align*}
&u(b(\tilde{T}) + y(\tilde{T})) - v \left( \frac{y(\tilde{T})}{\theta_T} \right) \geq u(b(\tilde{T}) + y(\tilde{T}) + \alpha_b + \alpha_y + \varepsilon(\tilde{T})) - v \left( \frac{y(\tilde{T}) + \alpha_y}{\theta_T} \right) > u(b(\tilde{T}) + y(\tilde{T}) + \alpha_b + \alpha_y + \varepsilon(\tilde{T}) - \delta) - v \left( \frac{y(\tilde{T}) + \alpha_y}{\theta_T} \right).
\end{align*}
\]
Let $\varepsilon = \min_{\theta \neq \bar{\theta}} \varepsilon(\bar{\theta})$. Under this choice of $\varepsilon$, the above equation contradicts (114). Equation (116) also implies that for all $\bar{\theta} \neq \theta$, choice following the deviation is the same as in the original equilibrium.

The last step in the proof requires checking that the time $T - 1$ incentive constraints hold. This is necessary in order to leave the decision of the agents unchanged at time $T - 1$. Note that for a given $\varepsilon > 0$, there exists $\delta^* > 0$ that makes the utility, calculated in time $T - 1$, of the modified contract the same as in the original contract. To see this, note that if $\delta = \varepsilon$ the change in utility of the agent is negative following the proposed deviation, while if $\delta = 0$ the utility change is positive, since the agent now faces a reduction in the spread of consumption at time $T$ because $y(\hat{\theta}T) + b(\hat{\theta}T) > y(\theta^T) + b(\theta^T)$. This implies that there exists an intermediate value of $\delta^*$ such that $\varepsilon > \delta^* > 0$ so that the change is zero. Hence, the time $T - 1$ decision will be unchanged if $\delta = \delta^*$.

**Equation (79) holds at time $t = T$:**

Lemma 1 implies that there is only one case left to consider. Suppose that for some $\bar{\theta}^T = (\bar{\theta}^{T-1}, \theta_T)$

$$u'(b(\bar{\theta}^{T-1}) + y(\theta^T)) < v' \left( \frac{y(\theta^T)}{\theta_T} \right) \frac{1}{\theta_T}. \tag{117}$$

In this case, the agent would like to consume and work less than the equilibrium contract. A deviation that reduces the total output and consumption by agent $\theta^T$ cannot be provided by an entrant, since a worker cannot deliver negative hours.
However, an incumbent firm will find it optimal to deviate from the equilibrium contract, offering an allocation with lower consumption and lower output requirement and making strictly positive profits. Formally, it offers the original contract at all time $t < T$ and at time $T$, a menu that contains a null contract, the modified allocation chosen by $\theta^T$ and the original allocation chosen by the remaining types:

$$C_T(b(\theta^{T-1}), y(\theta^{T-1})) = \left\{ (b(\theta^T) + y(\theta^T) + \delta^*(\varepsilon|\theta_T) - \varepsilon, y(\theta^T) + \delta^*(\varepsilon|\theta_T) ; (0, 0) ; (y(\tilde{\theta}^T) + b(\tilde{\theta}^T), y(\tilde{\theta}^T)) \right\}_{\tilde{\theta}^T \neq \theta^T}$$

where $\delta^*$ and $\varepsilon$ are constructed in a similar fashion to the proof of Lemma 1, with the constraint $\delta \leq 0$.

With this deviation, the incumbent makes strictly positive profits, proportional to $\varepsilon$, and there exists $\varepsilon$ so that agents' utility is unchanged following this deviation. This guarantees that no deviation at time $T - 1$ takes place. This contract is always profitable for the incumbent even if another type $\tilde{\theta}_T$ accepts it. If an agent with type $\tilde{\theta}^T$ is able to choose the pair $(b(\theta^T) + y(\theta^T) + \delta^*(\varepsilon|\theta_T) - \varepsilon, y(\theta^T) + \delta^*(\varepsilon|\theta_T))$ at time $T$, it implies that he must also have chosen the allocation sequence $\{(b(\theta^n) + y(\theta^n), y(\theta^n))\}_{n=1}^{T-1}$ in previous periods. From the previous step in the proposition, transfers from any history are independent of time $T$; i.e., this agent will receive transfers with the same net present value as in the original choice. Hence, the deviation is profitable.

2. Equations (78) and (79) hold for $t < T$. 

131
As an inductive assumption, suppose (78) holds for \( t + 1 \). We now show it holds for period \( t \). Rewrite the net present value of transfers as:

\[
A_t(\theta^{t-1}, \theta^{T}_{t-1}) = \sum_{n=t}^{T} \left( \frac{1}{q} \right)^{t-n} b_n(\theta^{t-1}, \theta^{n}_{t-1}) = \\
b_t(\theta^{t-1}, \theta_t) + q \sum_{n=t+1}^{T} \left( \frac{1}{q} \right)^{t+1-n} b_n(\theta^{t-1}, \theta^{n}_{t-1}) = b_t(\theta^{t-1}, \theta_t) + q A_{t+1}(\theta^t, \theta^T_t).
\]

By way of contradiction, there exist \( \theta_t \) and \( \hat{\theta}_t \) following history \( \theta^{t-1} \) such that

\[
b_t(\theta^{t-1}, \theta_t) + q A_{t+1}(\theta^t, \theta_t) < b_t(\theta^{t-1}, \hat{\theta}_t) + q A_{t+1}(\theta^{t-1}, \hat{\theta}_t). \tag{118}
\]

By the inductive assumption \( b_t(\theta^{t-1}, \theta_t) < b_t(\theta^{t-1}, \hat{\theta}_t) \). As in the proof for time \( T \), the contradiction argument relies on deviations by entrants to guarantee that (79) holds and on deviations by entrant and incumbent firms to imply that the net present value of transfers is zero.

Under the inductive assumption, the agent faces no distortion on both his intratemporal margin and intertemporal margin (recall Lemma 2) from time \( t + 1 \) onward. This implies that the equilibrium allocation from time \( t + 1 \) onwards is equivalent to a self-insurance economy (this will be formally proved in Proposition 8). Let \( S_{t+1}(x) \) be the utility the agent receives from entering time \( t + 1 \) with a level \( x \) of net present value of assets. The value function \( S \) is monotonically increasing in the level of assets. Given this, the agents’ equilibrium choices at time \( t \) satisfy the
following:

\[
\begin{align*}
&u(b(\hat{\theta}^t) + y(\hat{\theta}^t)) - v \left( \frac{y(\hat{\theta}^t)}{\theta_t} \right) + \beta S_{t+1} \left( qA_{t+1}(\hat{\theta}^t) - b(\hat{\theta}^t) \right) \geq \\
&u(b(\theta^t) + y(\theta^t)) - v \left( \frac{y(\theta^t)}{\theta_t} \right) + \beta S_{t+1} \left( qA_{t+1}(\theta^t) - b(\theta^t) \right),
\end{align*}
\]

(119)

and

\[
\begin{align*}
&u(b(\theta^t) + y(\theta^t)) - v \left( \frac{y(\theta^t)}{\theta_t} \right) + \beta S_{t+1} \left( qA_{t+1}(\theta^t) - b(\theta^t) \right) \geq \\
&u(b(\hat{\theta}^t) + y(\hat{\theta}^t)) - v \left( \frac{y(\hat{\theta}^t)}{\theta_t} \right) + \beta S_{t+1} \left( qA_{t+1}(\hat{\theta}^t) - b(\hat{\theta}^t) \right).
\end{align*}
\]

(120)

If \( y(\theta^t) \geq y(\hat{\theta}^t) \), an entrant can offer the following menu that enables the agent to work additional hours and move resources between time \( t \) and time \( t+1 \):

\[
\tilde{C}_t = \left\{ \left( b(\theta^t) - b(\hat{\theta}^t), y(\theta^t) - y(\hat{\theta}^t) \right); (0, 0) \right\},
\]

\[
\tilde{C}_{t+1} = \left\{ \left( -\frac{1}{q}b(\theta^t) - b(\hat{\theta}^t) \right] - \epsilon, 0 \right\}.
\]

This menu generates strictly positive profits to the entrant, proportional to \( \epsilon \). If this menu is offered, agent \( \theta^t \) will deviate, accepting the allocation for \( \hat{\theta}^t \) together with the allocation specified in the entrant’s menu. This is due to the fact that the agent can now replicate his original time \( t \) level of output and have access to a strictly higher net present value of transfers at a cost equal to \( \epsilon \).

Suppose now that \( y(\theta^t) < y(\hat{\theta}^t) \). The first case we consider is when consumption at time \( t \) is higher for the agent with a higher net present value of transfer, \( y(\theta^t) + b(\theta^t) < y(\hat{\theta}^t) + b(\hat{\theta}^t) \). As in the argument for period \( T \), inequality (119) cannot hold.
with equality. This enables us to reduce the time $t$ spread of consumption between histories $\theta^t$ and $\hat{\theta}^t$. Following the same steps of time $T$, a contradiction can be reached.

The final case is $y(\theta^t) < y(\hat{\theta}^t)$ and $y(\theta^t) + b(\theta^t) \geq y(\hat{\theta}^t) + b(\hat{\theta}^t)$. This case violates the inter-temporal Euler equation for at least one of the two types, thus contradicting Lemma 2. To see this, suppose that the Euler equation (77) holds for agent $\theta^t$. We have

$$u'(y(\theta^t) + b(\theta^t)) = \frac{\beta}{q} \sum_{\theta_{t+1}} \pi(\theta_{t+1}) u'(c(\theta^{t+1}))$$

$$\Rightarrow u'(y(\hat{\theta}^t) + b(\hat{\theta}^t)) \geq \frac{\beta}{q} \sum_{\theta_{t+1}} \pi(\theta_{t+1}) u'(c(\theta^{t+1}))$$

$$\Rightarrow u'(y(\hat{\theta}^t) + b(\hat{\theta}^t)) > \frac{\beta}{q} \sum_{\theta_{t+1}} \pi(\theta_{t+1}) u'(c(\hat{\theta}^t, \theta_{t+1})),$$

where the last implication follows from the fact that an agent with higher transfer will have higher consumption at time $t + 1$, thus a lower expected marginal utility of consumption.

To conclude, given that it was shown that the net present value of transfers is independent of the time $t$ choice, we can follow the same steps as in time $T$ to show that equation (117) holds for time $t$. □

13.3 No profitable deviation with redistribution.

We show that there is no profitable deviation at time $T$ that implies redistribution between agents.
We first show that any deviation, if chosen by agents, is such that transfers \{b(\theta_{T-1}, \theta_i)\}_{\theta_i \in \Theta} satisfy the following ordering: for all \(i, j\) if \(\theta_i > \theta_j\) then \(b(\theta_{T-1}, \theta_i) > b(\theta_{T-1}, \theta_j)\). Suppose not, so there exists \(\theta_i > \theta_j\) with \(b(\theta_{T-1}, \theta_i) < b(\theta_{T-1}, \theta_j)\). Let \(\{\hat{b}_T(\theta_{T-1}), \hat{y}(\theta_T)\}\) be the allocation chosen from the contract \(\hat{C}\) at time \(T\).\(^{73}\) The agents’ choices must satisfy the following, for all \(\theta, \hat{\theta}\):

\[
W(b, \theta) = \max_{y \geq 0} u(b + y) - v\left(\frac{y}{\theta}\right).
\]

Using this equation for \(\theta_i\) and \(\theta_j\) and from convexity of \(v\), we have that \(\hat{y}(\theta_T) + y(\theta_T) > \hat{y}(\theta_{T-1}, \hat{\theta}) + y(\theta_{T-1}, \hat{\theta})\). Agent \(\theta_i\) is better-off with the following strategy: choosing the pairs \((\hat{b}_T(\theta_{T-1}), \hat{y}(\theta_{T-1}, \hat{\theta}))\) and \((b(\theta_{T-1}, \hat{\theta}), y(\theta_{T-1}, \hat{\theta}))\) and from menu \(C^S_T\) choosing \(\delta_i = \hat{y}(\theta_T) + y(\theta_T) - (\hat{y}(\theta_{T-1}, \hat{\theta}) + y(\theta_{T-1}, \hat{\theta}))\). This allows him to have the same output requirements as in the original choice but higher consumption transfers.

We now show that such intratemporal transfers (transferring from less to more productive agents) reduce agents’ welfare with respect to the original equilibrium. Let \(N = |\Theta|\) be the number of possible shock realizations. We only need to consider the case with transfers \(\{b(\theta_{T-1}, \theta_i)\}_{\theta_i \in \Theta}\) ordered so that for all \(i, j\) if \(\theta_i > \theta_j\) then \(b(\theta_{T-1}, \theta_i) > b(\theta_{T-1}, \theta_j)\). Define the time \(T\) utility of an agent with type \(\theta\) and transfers \(b\), that can optimally chose the amount to work by the following:

\[
W(b, \theta) = \max_{y \geq 0} u(b + y) - v\left(\frac{y}{\theta}\right).
\]

\(^{73}\)From proposition (6), transfers in contract \(\hat{C}\) do not depend on time \(T\) realization of the shock.
Denote by \( y^*(b, \theta) \) the solution of problem (121) characterized by:

\[
u'(b + y^*(b, \theta)) = \frac{1}{\theta} v' \left( \frac{y^*(b, \theta)}{\theta} \right) . \tag{122}\]

Note that, for a given \( b \), \( y^* \) is increasing in \( \theta \), since \( v \) is convex. The envelope condition for (121) implies:

\[
\frac{\partial W(b, \theta)}{\partial b} = u'(b + y^*(b, \theta)) > 0. \tag{123}
\]

Given the definition of \( W \), the time \( T \) utility under equilibrium menu \( \hat{C} \) can be written as:

\[
\hat{u}_T = \sum_{i=1}^{N} \pi(\theta_i) W(\hat{b}_T(\theta^{T-1}), \theta_i) . \tag{124}\]

Let \( \hat{W} \) be the time \( T \) utility level attained accepting a deviation that delivers transfers \( b \) and let \( W^N(b) \) be the following:

\[
W^N(b) = \sum_{i=1}^{N} \pi(\theta_i) W(b_i + \hat{b}_T(\theta^{T-1}), \theta_i) , \tag{125}\]

where each individual \( W \) is as in (121). Abusing notation, set \( W(b_i, \theta_i) \equiv W(b_i + \hat{b}_T(\theta^{T-1}), \theta_i) \). Consider the most favorable case for the consumer and assume that the deviation incurs zero profits, so that \( \sum_{i=1}^{N} \pi_i b_i = 0 \). To show that any deviation reduces welfare (\( \hat{W} < \hat{u}_T \)), we first show the following

\[
\sum_{i=1}^{N} \pi_i b_i W'(0, \theta_i) < 0 , \tag{126}\]

Multiplying and dividing the above by \( W'(0, \theta) \), where \( \theta \) is the smallest \( \theta_i \), implies that the sign of (126) is determined by the sign of the following

\[
\sum_{i=1}^{N} \pi_i b_i W'(0, \theta_i) = W'(0, \theta) \sum_{i=1}^{N} \pi_i b_i \frac{W'(0, \theta_i)}{W'(0, \theta)} ,
\]
which is negative given the zero profit assumption and the fact that $W'(0, \theta_i)$ is decreasing in $\theta_i$. Define a scale parameter $g \in [0, 1]$ for all the transfers, and define the following function of the scale parameter

$$G(g) = \sum_{i=1}^{N} \pi_i W(g \cdot b_i, \theta_i).$$

(127)

Note that $G(0) = \hat{u}_T$ and by definition of $W$, $\bar{W} \leq G(1)$. Also, $G$ is monotonically decreasing in $g$ since

$$\frac{\partial G'(g)}{\partial g} = \sum_{i=1}^{N} \pi_i b_i W'(g \cdot b_i, \theta_i),$$

(128)

where $W'(g \cdot b_i, \theta_i) = u'(g \cdot b_i + y^*(g \cdot b_i, \theta_i))$. As in the previous case, we also have that

$$u'(g \cdot b_i + y^*(g \cdot b_i, \theta_i)) < u'(y^*(0, \theta_i)), \quad \text{if } b_i > 0,$$

$$u'(g \cdot b_i + y^*(g \cdot b_i, \theta_i)) > u'(y^*(0, \theta_i)), \quad \text{if } b_i < 0.$$

This implies that $G'(g) < G'(0)$ for all $g > 0$ and from (126) $G'(0) < 0$ so that $G(1) < G(0) = \bar{u}_T$.

13.4 Proofs of Section 11

Optimality of Exclusive Contracts under Zero Costs.

**Lemma 3.** For all feasible utility levels $w$, $V(w) > \Pi(w)$.

\textsuperscript{74}The set of feasible initial utility levels is the open interval $\left(\frac{1-\beta^{T+1}}{1-\beta} U, \frac{1-\beta^{T+1}}{1-\beta} \bar{U}\right)$, where $U = \inf_{c,l \geq 0} u(c) - v(l)$ and $\bar{U} = \sup_{c,l \geq 0} u(c) - v(l)$. 

137
Proof. Let \( \{ c^{NE}, y^{NE} \} \) and \( \{ c^{E}, y^{E} \} \) be the solution of (87) and (88), respectively. Since \( \{ c^{NE}, y^{NE} \} \) is in the constraint set of (88), \( V(w) \geq \Pi(w) \) for all \( w \). Suppose there exists \( w \) such that \( V(w) = \Pi(w) \). This implies that \( \{ c^{NE}, y^{NE} \} \) is one of the solutions of (88) for this \( w \). Let \( \theta_l = \min_{\theta} \Theta \). A necessary first order condition for a solution of (88) is for all feasible \( w \):
\[
u'(c(\theta^{t-1}, \theta_l)) > \frac{1}{\theta_l} u'(\frac{y(\theta^{t-1}, \theta_l)}{\theta_l}), \quad \forall \theta^{t-1}.
\]
However, since \( \{ c^{NE}, y^{NE} \} \) is a solution of (87), it must satisfy the following necessary first order condition:
\[
u'(c(\theta^{t-1}, \theta_l)) = \frac{1}{\theta_l} u'(\frac{y(\theta^{t-1}, \theta_l)}{\theta_l}), \quad \forall \theta^{t-1}.
\]
This contradicts \( \{ c^{NE}, y^{NE} \} \) being a solution of (88).\(^{75}\) So no such \( w \) exists. \( \blacksquare \)

**Proof of Proposition 9**

To prove the proposition, we first show the following two lemmas. For notation, let \( U(c, y, \Theta) \) the life-time utility of any allocation \( \{ c(\theta^t), y(\theta^t) \} \) when shocks are in \( \Theta \).

**Lemma 4.** \( w^{NE}(\lambda \bar{\theta}) = \frac{1-\beta^T}{1-\beta} \log \lambda + w^L(\bar{\theta}). \)

**Proof.** Let \( \{ c(\theta^t), y(\theta^t) \} \) be the solution of (87) for low mean agents. To prove the claim, we show that \( \{ \lambda c(\theta^t), \lambda y(\theta^t) \} \) solves the above problem for high mean agents. Suppose not, then there exists an allocation \( \{ \hat{c}(\lambda \theta^t), \hat{y}(\lambda \theta^t) \} \) that delivers

\(^{75}\)Note that (129) holds with equality only for the highest realization of utility.
higher utility \( U(\hat{c}, \hat{y}, \lambda \Theta) \). Consider the allocation \( \left\{ \frac{\hat{c}(\theta^t)}{\lambda}, \frac{\hat{y}(\theta^t)}{\lambda} \right\} \). This allocation is in the constraint set of problem (87) and delivers utility

\[
U \left( \frac{\hat{c}}{\lambda}, \frac{\hat{y}}{\lambda}, \Theta \right) = U(\hat{c}, \hat{y}, \lambda \Theta) - \frac{1 - \beta^T}{1 - \beta} \log \lambda. \tag{131}
\]

By the contradicting assumption, \( U(\hat{c}, \hat{y}, \lambda \Theta) > U(\lambda c, \lambda y, \lambda \Theta) = U(c, y, \Theta) + \frac{1 - \beta^T}{1 - \beta} \log \lambda \), which implies \( U(\hat{c}, \hat{y}, \lambda \Theta) - \frac{1 - \beta^T}{1 - \beta} \log \lambda > U(c, y, \Theta) \). Using (131), we get \( U \left( \frac{\hat{c}}{\lambda}, \frac{\hat{y}}{\lambda}, \Theta \right) > U(c, y, \Theta) \), contradicting allocation \( \left\{ c(\theta^t), y(\theta^t) \right\} \) solving (87) for low mean agents. \( \blacksquare \)

**Lemma 5.** \( w^E(\lambda \bar{\theta}) > w^E(\bar{\theta}) + \frac{1 - \beta^T}{1 - \beta} \log \lambda \).

**Proof.** Let \( \{ c(\theta^t), y(\theta^t) \} \) be the solution of (88) for low mean agents. Consider the relaxed problem with the surplus constraint (89) holding as a weak inequality.

Note that the allocation \( \{ \lambda c(\lambda \theta^t), \lambda y(\lambda \theta^t) \} \) is in the constraint set of this relaxed problem when agents are high mean. Also, this constraint must hold with equality (otherwise the extra surplus can be distributed in an incentive compatible way, increasing agent’s utility). This implies that the allocation that solves the problem must deliver strictly higher utility. This implies \( w^E(\lambda \bar{\theta}) > w^E(\bar{\theta}) + \frac{1 - \beta^T}{1 - \beta} \log \lambda \). \( \blacksquare \)

**Proof of Proposition 9**

**Proof.** Let \( \gamma^* = V(w^{NE}(\bar{\theta})) \), where the function \( V \) is the solution of the following problem:

\[
V(w_0) = \max_{c,y} \sum_{\theta^t,t} q^t \pi(\theta^t)[y(\theta^t) - c(\theta^t)] \tag{132}
\]
\[
\sum_{\theta^t} \beta^{t-1} \pi(\theta^t) \left[ u(c(\theta^t)) - v \left( \frac{y(\theta^t)}{\theta^t} \right) \right] = w_0
\]

\[
\sum_{\theta^t} \beta^{t-1} \pi(\theta^t) \left[ u(c(\theta^t)) - v \left( \frac{y(\theta^t)}{\theta^t} \right) \right] \geq \sum_{\theta^t} \beta^{t-1} \pi(\theta^t) \left[ u(c(\tilde{\theta}^t)) - v \left( \frac{y(\tilde{\theta}^t)}{\theta^t} \right) \right] \quad \forall \tilde{\theta}^t
\]

When writing problems (132), we abuse notation by denoting by \( \theta \) the agents’ labor productivity for both groups of agents. This implies

\[
w^E (\bar{\theta} | \gamma^*) = w^{NE} (\bar{\theta}).
\]

Also,

\[
w^E (\bar{\theta} | \gamma^*) + \frac{1 - \beta^T}{1 - \beta} \log \lambda = w^{NE} (\bar{\theta}) + \frac{1 - \beta^T}{1 - \beta} \log \lambda.
\]

Using the previous lemmas,

\[
w^E (\lambda \bar{\theta} | \gamma^*) > w^E (\bar{\theta} | \gamma^*) + \frac{1 - \beta^T}{1 - \beta} \log \lambda = w^{NE} (\bar{\theta}) + \frac{1 - \beta^T}{1 - \beta} \log \lambda = w^{NE} (\lambda \bar{\theta}).
\]

It is possible to break the indifference of the firms with respect to low mean agents by considering the monitoring cost \( \gamma = \gamma^* + \varepsilon \) for some \( \varepsilon \) small enough. The result holds in this case, since \( w^{NE} (\bar{\theta}) > w^E (\bar{\theta} | \gamma^*) \) and \( w^E (\bar{\theta} | \gamma^*) \) is continuous on \( \gamma \), so we can replicate the same steps. 

140
Part IV

Adverse Selection and Non-Exclusive Contracts\textsuperscript{76}

14 Introduction

Insurance contracts can be written to offset the risk connected to a wide variety of events. Examples of different type of insurance contracts involve insurance against person related events (medical, life, annuities); property events (car, home); financial events (credit default swaps: CDS\textsuperscript{77}). These type of insurance share some common properties: the realization of uncertainty can be verified and subscribers might have additional private information about the \textit{probabilities} that an event realizes. However, due to different regulatory oversight, a feature that varies greatly among them is the ability of the insurer to enter in additional insurance contracts with other insurance providers. This possibility of non-exclusive non-observable insurance holding, while being a rare instance in property insurance, is a definitive possibility in the case of credit default swaps (for example for CDS, until early 2009, were issued in private bilateral trades without any intermediation by any clearing house).\textsuperscript{78}

\textsuperscript{76}This chapter is coauthored with Laurence Ales.

\textsuperscript{77}For a review refer to Duffie (1999).

\textsuperscript{78}On March 10, 2009 ICE Trust\textsuperscript{TM} began operating as a central clearing house for credit default swaps in North America.
This observation is the starting point for this paper, in particular we ask what are the restrictions on the equilibrium insurance contract that arises once we dispense with the exclusivity assumption. To do this we extend the standard Rothschild and Stiglitz (1976) (RS henceforth) environment. We allow agents to engage in multiple insurance contract simultaneously and the terms of these contracts are not observed by other insurance providers. Insurance providers behave non-cooperatively and compete offering menus of insurance contracts from an unrestricted contract space. We show that the Rothschild and Stiglitz equilibrium allocation is not an equilibrium in the presence of non-exclusive contracting, since insurance providers will offer latent contracts to prevent deviation by other firms that prevent separation of the agents. This possibility also implies that latent menus can prevent cream-skimming strategies, however pooling equilibrium still fails to exists. We derive conditions under which a separating equilibrium exists and fully characterize it. The equilibrium allocation consists of agents with a lower probability of accident purchasing no insurance and agents with higher accident probability buying the actuarially fair competitive level of insurance. To sustain the equilibrium allocation firms must offer latent contracts. The equilibrium allocation also constitute a linear price schedule for insurance.

**Related literature**

This paper is related to the work of Wilson (1976). He extends the equilibrium
concept used in RS beyond static Nash equilibrium by allowing insurance providers to take into account how a change in their policy offers might effect the set of policies offered by other insurance providers. Latent contracts introduced in this paper play a similar role by enabling a reaction of insurance providers to deviations by other insurance providers.

This paper is related to a series of paper that analyze the effect of non-exclusive contracting in the purchase of goods and insurance. In chapter 3 we study a dynamic environment with private information (but differently from this paper the realization of private information is realized after the agents sign the contract) where agents can engage in multiple non-exclusive contracts for both labor and credit relationships. They show that a unique equilibrium always exists requires latent contracts and involves linear wages and bond prices. In a recent paper, Attar, Mariotti, and Salanie (2008) extend the environment of Akerlof (1970) to include non-exclusive contracting. They focus on goods market (using a linear utility) and show, contrary to this paper, that a unique equilibrium always exists, it involves linear prices and is sustained by latent contracts. Also, Arnott and Stiglitz (1991) and Bisin and Guaitoli (2004) study static moral hazard environments. In particular the last paper shows that latent contracts are necessary to sustain the equilibrium and lead to positive profit for the insurance providers.

This paper is organized as follows: section 15 introduces the environment. Characterization and implementation are studied respectively in section 16 and 16.1. We
conclude in section 16.2 by looking at the relation between the current paper and Rothschild and Stiglitz (1976).

15 Environment

Consider an economy populated by a continuum of measure one of agents and $I$ firms (insurance providers), where $I$ is a natural number. Agents are ex ante heterogeneous: there is a fraction $p_g$ of type $G$ consumers (the good type) and a fraction $p_b$ of type $B$ (the bad type). The economy lasts for 1 period. Agents utility $u$ is defined over consumption $c$. Assume $u : \mathbb{R}_+ \to \mathbb{R}$ is twice continuously differentiable, increasing and a strictly concave function. At time 1 agent of type $j = b, g$ receives endowment $\omega_H$ with probability $\pi_j$ and $\omega_L$ with probability $1 - \pi_j$, with $\omega_H > \omega_L$, and realization of the endowment occurs at the end of the period. Assume $\pi_g > \pi_b$ and that these probabilities are private information of the agent. The realization of the endowment is publicly observed. Each firm $i \in \{1, ..., I\}$ offers contracts to agents to insure against the endowment shock. A contract prescribes consumption transfers conditional on the realization of the endowment: formally $(\tau_L, \tau_H)$ are the transfers pair conditional on a realization of $\omega_L$ or $\omega_H$.

An important feature of our environment is that agents can sign contracts with more than one firm simultaneously, and the terms of the contract between an agent and a firm $i$ are not observed by other firms. We do not impose any restriction on the contracts offered by each firm. As described in chapter 3, due to the delegation
principle, we can restrict the analysis to menu games. In a menu game, each firm offers a menu consisting of collection of transfers pairs, with each element referring to the transfer conditional on the realization of the endowment. A menu is a set $C^i$ in $\mathcal{P}(\mathbb{R}^2)$ (the power set of $\mathbb{R}^2$). The problem of firm $i$ is to choose its menu, taking as given the menus offered by other firms $C^{-i}$:

$$V(C^{-i}) = \max_{C^i \in \mathcal{P}(\mathbb{R}^2)} - \sum_{j=B,G} p_j [\pi_j \tau^i_{H,j*} + (1 - \pi_j) \tau^i_{L,j*}]$$

s.t. $$(\tau^i_{L,j*}, \tau^i_{H,j*}) \in C^i$$

where $(\tau^i_{L,j*}, \tau^i_{H,j*})$ are the transfers pair chosen by an agent of type $j$. Let $\Pi(C^i, C^{-i}) = - \sum_{j=B,G} p_j [\pi_j \tau^i_{H,j*} + (1 - \pi_j) \tau^i_{L,j*}]$.\(^{79}\)

Let $U^j(C)$ be the expected utility of the agent of type $j = B, G$ if he accepts menus $C = C^1 \times \ldots \times C^I$:

$$U^j(C) = \max_{(\tau^i_L, \tau^i_H) \in C} \left[ \pi_j u \left( \omega_H + \sum_{i=1}^I \tau^i_H \right) + (1 - \pi_j) u \left( \omega_L + \sum_{i=1}^I \tau^i_L \right) \right].$$

where $(\tau^i_L, \tau^i_H)$ are the transfers chosen from firm $i$.

**Definition 4 (Equilibrium of Menu Games).** A pure strategy equilibrium of a menu game is a collection of menus $C$ and agent’s choices $\tau^i_j = (\tau^i_L, \tau^i_H) \in C^i \ \forall i \in \{1,\ldots,I\}$, for each $j = B, G$:

1. Agents’ choices at time $(\tau^i_L, \tau^i_H) \in C^i \ \forall i \in \{1,\ldots,I\}$ solve the agent problem (134).

\(^{79}\)Note that we do not allow for the use of random contracts.
2. For each \( i \in \{1, ..., I\} \), \( C^i \) solves (133) taking as given \( C^{-i} \) chosen by firms \( -i \) and the agents’ choice \( (\tau^i_L, \tau^i_H) \in C^i \) \( \forall i \in \{1, ..., I\} \).

Note that a menu might contain more alternatives than the number of types, implying that some alternatives are not chosen in equilibrium. We denote a contract as \textit{latent} if it is offered in equilibrium by a firm but is not chosen in equilibrium by any agent.

In this environment, two distinct sources of uncertainty can be insured: the private realization of the type and the public realization of the endowment shock. If the private type is perfectly insured, both agents will receive identical contracts. Following the adverse selection literature, we call this allocation a \textbf{pooling} equilibrium. If different contracts are chosen by different types we call the resulting allocation a \textbf{separating} equilibrium. In the characterization we show that no pooling equilibrium allocation exists. The unique equilibrium in this environment is the following separating equilibrium: the bad type receives full insurance (at his actuarially fair price) against the public realization of the endowment and the good type receives no insurance.

16 Characterization of Equilibrium

The following lemma provides the necessary conditions for a pooling equilibrium.\(^{80}\)

\(^{80}\)All the proofs for this section are provided in Appendix 17.1.
Lemma 6. In any pooling equilibrium allocation \( c = (c_L, c_H) \), the following conditions must be satisfied:

\[
\frac{1 - \pi_b}{\pi_b} u'(c_L) \leq \frac{1 - \pi_b}{\pi_b}, \tag{135}
\]

\[
\frac{1 - \pi_g}{\pi_g} u'(c_L) \geq \frac{1 - \hat{p}_H}{\hat{p}_H}, \tag{136}
\]

where \( \hat{p}_H = p_g \pi_g + p_b \pi_b \).

The first condition implies that the marginal rate of substitution between the two states for the \( B \) agent is less than or equal to the actuarially fair price for the insurance only if \( B \) agents accept. If not, entrants can provide some additional insurance at slightly less than that price. The second condition requires that the marginal rate of substitution between states for the \( G \) agent is greater than the price for insurance when all agents accept the contract (the actuarially fair pooling price). Otherwise, entrants can profitably provide alternative insurance contracts at a slightly lower price. A direct implication of this lemma is that there is no pooling equilibrium, since there is no allocation that satisfies these two conditions at the same time.

Proposition 10. There is no pooling equilibrium.

Proof. Suppose there exists a pooling equilibrium \( c = (c_L, c_H) \). This equilibrium has to satisfy condition (135) and (136). This implies:

\[
\frac{1 - \pi_g}{\pi_g} \frac{\hat{p}_H}{1 - \hat{p}_H} \geq 1 \implies \frac{1 - \pi_g}{\pi_g} \geq \frac{1 - \hat{p}_H}{\hat{p}_H},
\]

\(^{81}\)Note that equation 135 implies that \( c_L \geq c_H \).
which is a contraction since \( \pi_b < \pi_g \).

We now characterize the necessary conditions a separating equilibrium satisfies. In this equilibrium each type of agents receives different allocations: \( c^B = (c^B_L, c^B_H) \), \( c^G = (c^G_L, c^G_H) \), which is denoted by \( C = \{c^B, c^G\} \). In this case, each agent must prefer his own allocation:

\[
U^B(c^B) \geq U^B(c^G), \quad U^G(c^G) \geq U^G(c^B).
\]

(137)

The equilibrium must also deliver non-negative profits: \( \Pi(C) \geq 0 \). The following lemma characterizes three necessary conditions an equilibrium must satisfy. The first condition is equivalent to (135) in the pooling equilibrium, requiring the equilibrium allocation for the \( B \) type to be on the region of over-insurance. The second is that the indifference curve of the \( G \) agent, at the equilibrium allocation, must be less steep than the average zero profit line. The last condition requires that the allocation for the \( G \) type must be on the under-insurance region.

**Lemma 7.** Any separating equilibrium allocation must satisfy:

1. For the \( B \) agent:

\[
\frac{1 - \pi_b}{\pi_b} \frac{u'(c^B_L)}{u'(c^B_H)} \leq \frac{1 - \pi_b}{\pi_b},
\]

(138)

2. For the \( G \) agent:

\[
\frac{1 - \pi_g}{\pi_g} \frac{u'(c^G_L)}{u'(c^G_H)} \leq \frac{1 - \hat{p}_H}{\hat{p}_H},
\]

(139)

\[
\frac{1 - \pi_g}{\pi_g} \frac{u'(c^G_L)}{u'(c^G_H)} \geq \frac{1 - \pi_g}{\pi_g}.
\]

(140)
Using the above restrictions, we now characterize the equilibrium and show that it is unique. Let a candidate equilibrium be \( c^B = (\omega^B, \omega^B), \ c^G = (\omega_L, \omega_H) \), where \( \omega_B = \pi_b \omega_H + (1 - \pi_b) \omega_L \). Figure 16 illustrates this equilibrium. This equilibrium provides no insurance to the \( G \) agent and the actuarially fair insurance to the \( B \) type. Note that this allocation satisfies conditions (137), (138), (140) and delivers zero profits. To satisfy (139), the following parameter restriction is needed:

\[
\frac{1 - \pi_g}{\pi_g} \frac{u'(\omega_L)}{u'(\omega_H)} \leq \frac{1 - \hat{p}_H}{\hat{p}_H}.
\] (141)

This condition is satisfied if, for example, \( \pi_g \) is large relative to \( \pi_b \) or if the spread between \( \omega_L \) and \( \omega_H \) is sufficiently small. Note that there exists a non-empty set of parameter values for which it holds.\(^82\) Also note that the \( G \) agent prefers autarky

\(^82\)Note that, as in Rothschild and Stiglitz (1976), equilibrium might fail to exist for this envi-
to the allocation designed for the $B$ agent: $\pi_g u(\omega_H) + (1 - \pi_g)u(\omega_L) \geq u(\omega_B)$. \footnote{Suppose that $\pi_g u(\omega_H) + (1 - \pi_g)u(\omega_L) < u(\omega_B)$. This implies that in the consumption space $(c_l, c_h)$, the indifference curve for the high type passing through the endowment point passes below the point $\omega_B$. In this space, we represent indifference curves as a one-dimensional function, denoted by $U_{\text{aut}}(c_l)$. Let $c_{lb}$ be the level of consumption so that $U'_{\text{aut}}(c_{lb}) = -\frac{1 - \pi_b}{\hat{p}_b}$. By the contradicting assumption, this point must lie on the right of $\omega_L$, which follows from the fact that at $c_{lb}$, the indifference curve and the zero profit line for the bad type are at the maximum distance. Denote by $c_{tm}$ the value of consumption in the low state so that $U''_{\text{aut}}(c_{tm}) = -\frac{\hat{p}_H}{\hat{p}_H}$. Since the slope of the indifference curve is decreasing in the consumption of the low state (keeping the level of utility constant), we have that $\omega_L > c_{lb} > c_{tm}$, since condition (141) requires $U''_{\text{aut}}(\omega_L) \leq -\frac{\hat{p}_H}{\hat{p}_H}$, we reach a contradiction.}

**Proposition 11.** Let $\{\pi_g, \pi_b, \omega_h, \omega_l, u\}$ satisfy condition (141); then an equilibrium allocation of the menu game satisfies:

1. $c^B = (\omega_B^H, \omega_B^L)$, where $\omega_B = \pi_b \omega_H + (1 - \pi_b)\omega_L$;

2. $c^G = (\omega_L, \omega_H)$.

### 16.1 Implementation of Equilibrium

The following proposition shows that $(c^B, c^G)$ can be implemented in equilibrium and is necessary to have more than one firm active in equilibrium with each of these firms offering latent contracts.

**Proposition 12.** Let $\{\pi_g, \pi_b, \omega_h, \omega_l, u\}$ satisfy condition (141); then there exists an equilibrium of the menu game.
Proof. The following strategies for the firms implement the equilibrium allocation.

Let firms $i = 1, 2$ offer the menu: $C^i = \{(\frac{\tau_L B}{2}, \frac{\tau_H B}{2}), (\tau_L B, \tau_H B), (0, 0)\}$, where $\tau_L B = \pi_b(\omega_H - \omega_L)$ and $\tau_H B = (1 - \pi_b)(\omega_L - \omega_H)$. Let all remaining firms $i \neq 1, 2$ offer the null menu: $C^i = \{(0, 0)\}$. The agents’ strategies, given these menus, are:

- Type $B$ chooses $(\tau_L B, \tau_H B)$ from firms 1 and 2;
- Type $G$ chooses $(0, 0)$ from all firms.

In this equilibrium, all firms make zero profits and agents $B$ and $G$ get allocations $c^B$ and $c^G$, respectively.

We now show that there are no profitable deviations by any firm (either entrant or incumbent). Given that there is always at least one firm offering a collection of menus that contains $(\tau_L B, \tau_H B)$, there is no profitable deviation that an entrant or incumbent can make that makes the $B$ agent better off since the allocation $c^B$ is the unique solution of maximizing the $B$ agent’s utility subject to non-negative profits.

The only potentially profitable deviation that can attract $G$ agents is if a firm offers additional insurance as the contract $\tilde{C} = \{(\varepsilon, -\alpha \varepsilon)\}$ for $\varepsilon$ small enough and with $\alpha$ such that $\frac{1 - \pi_a}{\pi_g} < \alpha < \frac{1 - \pi_a}{\pi_g} \frac{u'(\omega_l)}{u'_{\omega_h}}$. This deviation increases the utility of $G$ agents and is profitable as long as only $G$ agents accept it. However, given that the incumbent firms offer $(\tau_L B, \tau_H B)$, if a firm (either one of the incumbents or an entrant) offers contract $\tilde{C}$, the $B$ agent will choose the contract $(\tau_L B, \tau_H B)$ together with $\tilde{C}$, since the slope of his indifference curve at $c_B$ is $\frac{1 - \pi_a}{\pi_b}$ and condition (141) holds. In this case, this firm will make negative profits; hence, this deviation is not offered.

Using a similar argument, there is no profitable deviation that can increase agents’
utility. Note that if there is only one active firm, this firm would deviate and offer additional insurance to the G agent as $\tilde{C}$. 

The proof shows that the allocation $(c^B, c^G)$ is an equilibrium and to be implemented it is necessary more than one active firm and latent contracts to prevent deviations by potential entrants and incumbents.

**Corollary 2.** The allocation $(c^B, c^G)$ can be implemented as an equilibrium only if latent contracts are offered and if there is more than one active firm.

### 16.2 The Rothschild-Stiglitz Equilibrium

A natural benchmark to compare our results is the standard case with exclusive contracts characterized in RS. We show that in the presence of non-exclusive contracting, opposite to RS, in a pooling equilibrium, exist latent contracts can be used to rule out deviations that only attract good types (cream skimming). We also show that the separating equilibrium described in RS is not an equilibrium in our environment.

RS show that there is no pooling equilibrium when contracts are exclusive because there is always an alternative contract that can be offered by an entrant that is profitable and attracts only $G$ types. When agents can sign non-exclusive contracts, these cream skimming deviations can be prevented by latent contracts. We illustrate this argument in Figure 17. The solid lines represent the indifference curves.
of B and G agents at the best pooling equilibrium. Any contract in the green area in Figure 17 is profitable to a firm as long as only G agents accept it and it is preferred to the pooling equilibrium by G agents but not by B agents. When contracts are exclusive, this argument is enough to show that there are no pooling equilibria. Under non-exclusivity, a firm can offer a latent contract, as point A in Figure 18, that makes any deviation in the green area unprofitable. If an incumbent firm offers such contract together with the pooling allocation, no other firm finds it profitable to offer a contract in the green area. Since in this case, that contract would be chosen by all agents: good types prefer it to the pooling allocation and bad types prefer it to the pooling allocation once it is combined with the latent contract A. This implies that this deviation is no longer profitable hence it would not be offered.

To show directly that the separating equilibrium allocation defined in RS is not an equilibrium in our environment, let this allocation be defined as follows.

**Definition 5.** The RS separating allocation is \((\tilde{c}^B, \tilde{c}^G)\) which satisfies \(\tilde{c}^B = (\omega^B, \omega^B)\) and \(\tilde{c}^G = (c^G_L, c^G_H)\); where \(\omega^B = \pi_b \omega_h + (1 - \pi_b) \omega_l\) and such that \(U^B(c^B) = U^B(c^G)\) and \(\pi_g (\omega_H - c^G_H) + (1 - \pi_g) (\omega_L - c^G_L) = 0\).

If \((\hat{c}^B, \hat{c}^G)\) were an equilibrium, an entrant can offer a small additional contract \(\hat{c}\) that attracts B agents and that delivers strictly positive profits. This contract

---

84 The pooling equilibrium that delivers the highest expected utility when agents are weighted equally. The same argument holds for other pooling equilibria that deliver non-negative profits.
is such that $B$ agents accept the $G$ contract from the incumbent together with $\hat{c}$, delivering negative profits to the incumbent. Let the contract $\hat{c}$ be $\hat{c} = (\varepsilon, -\alpha \varepsilon)$ for some small $\varepsilon$ and where $\alpha$ satisfies:

$$\frac{1 - \pi b}{\pi b} \frac{u'(c^G_H)}{u'(c^G_L)} > \alpha > \frac{1 - \pi b}{\pi b}$$

(142)

This $\alpha$ exists since $\tilde{c}^G_H > \tilde{c}^G_L$. This deviation is preferred by $B$ agents since

$$U^B (c^G + \hat{c}) = \pi_b u'(c^G_H - \alpha \varepsilon) + (1 - \pi_b) u'(c^G_L + \varepsilon)$$

$$= \pi_b u'(c^G_H) + (1 - \pi_b) u'(c^G_L) - \pi_b u'(c_H) \alpha + (1 - \pi_b) u'(c_L)$$

$$> U^B (c^G) - \pi_b u'(c_H) \frac{1 - \pi_b}{\pi_b} \frac{u'(c_L)}{u'(c_H)} + (1 - \pi_b) u'(c_L)$$

$$> U^B (c^G) = U^B (c^B)$$

The deviation delivers strictly positive profits to the entrant even if only $B$ agents
accept it since:

$$\Pi^B(\hat{c}) = \pi_b \alpha \varepsilon - (1 - \pi_b) \varepsilon \Rightarrow \frac{\Pi^B(\hat{c})}{\pi_b \varepsilon} = \alpha - \frac{1 - \pi_b}{\pi_b} > 0,$$

with the last inequality follows from (142).

Note that in this case no latent contract can be used to prevent the deviation described above since this deviation is profitable also in the worst case for the firm when only $B$ agents accept it. Hence a latent contract can prevent this deviation only if it makes all the agents to strictly prefer the latent to accepting the deviation. But this contradicts the contract being latent because it would be preferred to the equilibrium allocation.

17 Appendix

17.1 Proofs of Section 16

Proof of Lemma 6

Proof. 1. Suppose (135) does not hold; this implies:

$$\frac{1 - \pi_b}{\pi_b \varepsilon} u'(c_L) > \frac{1 - \pi_b}{\pi_b}.$$  \hspace{1cm} (143)

Consider the following menu offered by an entrant: $\hat{c} = (\varepsilon, -\alpha \varepsilon)$ for some small $\varepsilon$ and where $\alpha$ satisfies:

$$\frac{1 - \pi_b}{\pi_b} \frac{u'(c_L)}{u'(c_H)} > \alpha > \frac{1 - \pi_b}{\pi_b}. \hspace{1cm} (144)$$

Parameter $\alpha$ can be interpreted as the slope of a line passing between the zero profit line of the bad type and the slope of his indifference curve through $c$ (the original pooling equilibrium). This deviation is chosen by the $B$ agents.
since
\[ U^B(c + \hat{c}) = \pi_b u(c_H - \alpha \varepsilon) + (1 - \pi_b) u(c_L + \varepsilon) = \pi_b u(c_H) + (1 - \pi_b) u(c_L) - \pi_b u'(c_H) \alpha + (1 - \pi_b) u'(c_L) \]
\[ > U^B(c) - \pi_b u'(c_H) \frac{1 - \pi_b}{\pi_b} u'(c_L) + (1 - \pi_b) u'(c_L) \quad (145) \]

using the first inequality in (144) we have
\[ U^B(c + \hat{c}) > U^B(c) \]

The minimum profits for the entrant occur when only the bad types accept this contract, since this increases the probability of a positive transfer from the insurance provider to the agent.\(^{85}\) This deviation delivers strictly positive profits to the entrant even if only \(B\) agents accept it:
\[ \Pi^B(\hat{c}) = \pi_b \alpha \varepsilon - (1 - \pi_b) \varepsilon \Rightarrow \frac{1}{\pi_b \varepsilon} \Pi^B(\hat{c}) = \alpha - \frac{(1 - \pi_b)}{\pi_b}. \]
The first inequality of (144) then implies \(\Pi^B(\hat{c}) > 0\). Hence, no equilibrium contract can prevent this deviation.

2. Suppose (136) does not hold; this implies:
\[ \frac{1 - \pi_g}{\pi_g} \frac{u'(c_L)}{u'(c_H)} < \frac{1 - \hat{p}_H}{\hat{p}_H}. \quad (146) \]
Consider the following deviating menu \(\hat{c} = (c_L - \varepsilon, c_H + \alpha \varepsilon)\) where \(\alpha\) satisfies:
\[ \frac{1 - \pi_g}{\pi_g} \frac{u'(c_L)}{u'(c_H)} < \alpha < \frac{1 - \hat{p}_H}{\hat{p}_H}. \quad (147) \]
Differently from the previous case, this deviation will be accepted as a substitute allocation by \(G\) agents rather than an additional allocation as in the previous case, to see this:
\[ U^G(\hat{c}) = \pi_g u(c_H + \alpha \varepsilon) + (1 - \pi_g) u(c_L - \varepsilon) = \pi_g u(c_H) + (1 - \pi_g) u(c_L) + \pi_g u'(c_H) \alpha - (1 - \pi_g) u'(c_L) \]
\[ > U^G(c) + \pi_g u'(c_H) \frac{1 - \pi_g}{\pi_g} u'(c_L) - (1 - \pi_g) u'(c_L) > U^G(c) \quad (148) \]

\(^{85}\)In general, given a contract of the type \(\Gamma = (a\varepsilon, b\varepsilon)\) let \(\pi^* = \arg\min_{a, \pi_a} \{ \pi_a \cdot a\varepsilon + (1 - \pi_a) \cdot b\varepsilon \}. \) The minimum profits are given when the type of agent has an accident probability equal to \(\pi^*\); that is, \(\pi^* = \pi_b\) if \(a > b\) and \(\pi^* = \pi_g\) if \(a < b\).
This deviation delivers strictly positive profits to the entrant if all agents accept it, since:

\[ \Pi(\hat{c}) = \hat{p}_H (\omega_H - c_H - \alpha \epsilon) + (1 - \hat{p}_H) (\omega_L - c_L + \epsilon) \]
\[ = \Pi(C) - \hat{p}_H \alpha \epsilon + (1 - \hat{p}_H) \epsilon > 0. \]

The deviation also delivers strictly positive profits to the entrant also if only \( G \) agents accept it, since:

\[ \Pi^G(\hat{c}) = \pi_g (\omega_H - c_H - \alpha \epsilon) + (1 - \pi_g) (\omega_L - c_L + \epsilon) \]
\[ = \pi_g (\omega_H - c_H) + (1 - \pi_g) (\omega_L - c_L) - \epsilon (\pi_g (\omega_H - c_H) + (1 - \pi_g) (\omega_L - c_L)) \]
\[ \implies \Pi(\hat{c}) > (\omega_H - c_H) + (1 - \pi_g) (\omega_L - c_L) - \epsilon \left( \frac{1 - \pi_g}{\hat{p}_H} \right), \]

so that

\[ \Pi(\hat{c}) > (\omega_H - c_H) + (1 - \pi_g) (\omega_L - c_L) - \epsilon \left( \frac{1 - \pi_g}{\hat{p}_H} \right). \]

Since \( \frac{1 - \pi_g}{\hat{p}_H} > 0 \) if \( (\omega_H - c_H) + (1 - \pi_g) (\omega_L - c_L) > 0 \), then there exists \( \epsilon \) small enough so that \( \Pi(\hat{c}) > 0 \), we show this next. Suppose for contradiction that \( \pi_g (\omega_H - c_H) + (1 - \pi_g) (\omega_L - c_L) \leq 0 \). Equation (135) implies that \( c_L > c_H \) so:

\[ (\omega_H - c_H) > (\omega_L - c_L) \]
\[ \pi_g (\omega_H - c_H) + (1 - \pi_g) (\omega_L - c_L) > \pi_b (\omega_H - c_H) + (1 - \pi_b) (\omega_L - c_L) \geq 0. \]

The last inequality comes from the fact that total profits under \( c \) must be non-negative, so under the contradicting assumption, it must be that \( \pi_b (\omega_H - c_H) + (1 - \pi_b) (\omega_L - c_L) \geq 0 \), reaching a contradiction. So it must be true that \( \pi_g (\omega_H - c_H) + (1 - \pi_g) (\omega_L - c_L) > 0 \), and consequently condition (136) must hold.

\[ \boxed{\text{Proof of Lemma 7}} \]

**Proof.** The proof of condition (138) follows the proof of (135) in Lemma 6. Suppose that condition (139) is violated. If so, there exists an \( \alpha \) so that

\[ \frac{1 - \pi_g}{\pi_g} \frac{u'(c^*_H)}{u'(c^*_g)} > \alpha > \frac{1 - \hat{p}_H}{\hat{p}_H}. \]  \hspace{1cm} (149)

Consider an entrant firm offering the menu \( \hat{c} = (\epsilon, -\alpha \epsilon) \). This menu is profitable even if all agents accept it (as it will be shown next the \( G \) type always accept it), since

\[ \Pi(\hat{c}) = \hat{p}_H \alpha \epsilon - (1 - \hat{p}_H) \epsilon \Rightarrow \frac{\Pi(\hat{c})}{\epsilon \hat{p}_H} = \alpha - \frac{(1 - \hat{p}_H)}{\hat{p}_H} > 0, \]

157
where the last inequality follows from (149). In addition, \( \hat{C} \) is accepted by the good type, since

\[
U^G(C+\hat{C}) = \pi_g u(c^G_H - \alpha \varepsilon) + (1-\pi_g)u(c^G_L + \varepsilon) = U^G(C) - \alpha \pi_g u'(c^G_H) + (1-\pi_g)u'(c^G_L) > U^G(C),
\]

where the last inequality follows from (149): \( (1-\pi_g)u'(c^G_L) > \alpha \pi_g u'(c^G_H) \).

Suppose that condition (140) is violated. If so, there exists an \( \alpha \) so that

\[
\frac{1-\pi_g}{\pi_g} \frac{u'(c^G_L)}{u'(c^G_H)} < \frac{1-\pi_g}{\pi_g}. \tag{150}
\]

Consider an entrant firm offering the menu \( \hat{c} = (-\varepsilon, \alpha \varepsilon) \). This menu is profitable even if all types accept it:

\[
\Pi(\hat{c}) = -\hat{p}_H \alpha \varepsilon + (1-\hat{p}_H)\varepsilon \Rightarrow \Pi(\hat{c}) = -\alpha + \frac{(1-\hat{p}_H)}{\hat{p}_H} > 0,
\]

where the last inequality follows from (150): \( \alpha < \frac{1-\pi_g}{\pi_g} < \frac{(1-\hat{p}_H)}{\hat{p}_H} \). Also \( u^G(c^G + \hat{c}) > u^G(c^T) \), so that \( G \) agent is willing to accept it. \( \blacksquare \)

**Proof of Proposition 11**

**Proof.**

**Part 1.** We first show that the allocation for the \( B \) agent generates 0 profits. Consider the Cartesian plane \((c_L, c_H)\). For any allocation \( x = (x_L, x_H) \), let \( \sigma(x) = \frac{\omega_H - x_H}{\omega_L - x_L} \) be the slope of the line connecting the endowment point and allocation \( x \). The allocation for the bad type must satisfy \( \sigma(c^B) \geq \frac{1-\pi_b}{\pi_b} \) (that is, it cannot lie above the zero profit line for the bad type). Suppose not, in this case, the incumbent firm can deviate to the following allocation \( \tilde{c} = (c^B_H - \varepsilon, c^B_L - \varepsilon) \), this allocation increases profits by \( \varepsilon \). Since it delivers strictly lower utility than the original allocation, it will not be preferred by the good type if the original was not. \(^{86}\) Similar arguments apply if the allocation is the result of transfers originating from multiple insurance providers, that is, \( c^B = \left( \sum_{i=1}^{N} c^B_{i,L}, \sum_{i=1}^{N} c^B_{i,H} \right) \). In this case, if \( \sigma(c^B) < \frac{1-\pi_b}{\pi_b} \), then there exists at least one insurance provider for which \( \sigma(c^B) < \frac{1-\pi_b}{\pi_b} \). \(^{87}\) From here onward we consider \( \sigma(c^B) \geq \frac{1-\pi_b}{\pi_b} \). Suppose the previous

---

\(^{86}\) Even if the \( G \) agent would accept the allocation, it would only increase profits, since \( \varepsilon \) can be chosen large enough so that the allocation lies under the average zero profit line.

\(^{87}\) To illustrate this, consider the case with \( N = 2 \) and denote with \( t \) the transfers from the insurance provider to the agent. Suppose that \( \sigma(c^B) = \frac{t^B_H + t^B_H}{t^B_L + t^B_L} < \frac{1-\pi_b}{\pi_b} \) and \( \frac{t^B_H}{t^B_L} > \frac{\pi_b}{1-\pi_b} \). If

\[
\frac{t^B_H}{t^B_L} > \frac{t^B_H + t^B_H}{t^B_L + t^B_L},
\]

which implies \( \frac{t^B_H}{t^B_L} > \frac{t^B_H}{t^B_L} \), thus reaching a contradiction.
relation holds with strict inequality: \( c^B \) is under the zero profit line for the \( B \) agent, 
suppose also the allocation delivers \( \varepsilon \) profits, then an entrant can offer the allocation 
\( \tilde{c} = (c^B_L - \omega_L + \varepsilon/2, c^B_H - \omega_H + \varepsilon/2) \). This allocation delivers strictly positive profits 
even if all agents accept it. Hence, the equilibrium \( \tilde{c} \) contract is preferred by \( B \) agents, which is a contradiction.

If \( c^B \) is on the zero profit line and \( c^B_L > c^B_H \) (the opposite inequality is excluded 
by (138)), there exists \( \alpha \) such that 
\( \frac{1-\pi_b}{\pi_b} \frac{u(c^B_L)}{u(c^B_H)} < \alpha < \frac{1-\pi_b}{\pi_b} \). An entrant can offer the 
contract \( \tilde{c} = (c^B_L - \varepsilon, c^B_H + \alpha \varepsilon) \). This contract is preferred by \( B \) agents and delivers 
positive profits even if all agents accept it. Hence, the equilibrium \( c^B \) must be on 
the intersection of the zero profit line for \( B \) agents and the 45° line: \( c^B = (\omega^B, \omega^B) \).

**Part 2.** In equilibrium the allocation for the \( G \) agent must satisfy \( U^G(c^G) \geq U^G(c^B) \) 
and \( U^G(c^G) \geq U^G(\omega^L, \omega^H) \). This implies that in the cartesian plane \((c_L, c_H)\), \( c^G \)
must lie in the triangle delimited by the endowment point, the intersection of the 
indifference curve of the \( B \) agent at \( c^B \) with the zero profit line for the \( G \) agent and 
the intersection of the indifference curve of the \( B \) agent at \( c^B \) with the intersection 
of indifference curve of the \( G \) agent at the endowment.

Suppose \( c^G \) is in this triangle and \( c^G \neq (\omega^L, \omega^H) \). An entrant can offer the 
following contract: \( \tilde{c} = (c^G_L - \varepsilon, c^G_H - \omega_L + \varepsilon, c^G_L - \omega_H + \varepsilon) \) for a small \( \varepsilon > 0 \). This allocation 
will be chosen by the \( B \) type together with \( c^G \) since \( \tilde{c} + c^G = c^B + (\varepsilon, 0) \). Also, this 
allocation is profitable for the entrant if any type chooses it. To see this, rewrite 
\[ \tilde{c} = (c^B_L - \omega_L + \varepsilon, c^B_H - \omega_H + \varepsilon, c^B_L - \omega_L + \varepsilon, c^B_H - \omega_H + \varepsilon) = (t^B_{L1} - t^G_L + \varepsilon, t^B_{H1} - t^G_H + \varepsilon), \]
where \( t^B_{L1}, t^B_{H1} > 0 \) and \( t^B_{H1}, t^B_{H1} < 0 \). Minimum profits occur when only the bad type 
accepts it. Suppose profits are negative: 
\( -\pi_b(t^B_{H1} - t^G_{H1}) - (1 - \pi_b)(t^B_{L1} - t^G_{L1} + \varepsilon) < 0 \). Since 
\( t^G_{L1} = -\frac{1-\pi_b}{\pi_b} t^B_{L1} \), \( \pi_b t^G_{H1} + (1 - \pi_b)(t^G_{L1} - \varepsilon) < 0 \). For sufficiently small \( \varepsilon \) this 
implies \( -\frac{t^G_{L1}}{t^L_{L1}} > \frac{1-\pi_b}{\pi_b} \), since \( t^G_{L1} > 0 \), \( \frac{t^G_{L1}}{t^L_{L1}} > \frac{1-\pi_b}{\pi_b} \). This is a contradiction with 
\( \frac{\omega^L - c^G_L}{\omega^L - c^G_L} \leq \frac{1-\pi_b}{\pi_b} \), since \( c^G \) lies above the average zero profit line given the parameter 
restriction. ■
References


