UNIVERSITÀ DEGLI STUDI DI ROMA "TOR VERGATA"

FACOLTA' DI ECONOMIA

DOTTORATO DI RICERCA IN

ECONOMETRIA ED ECONOMIA EMPIRICA

CICLO DEL CORSO DI DOTTORATO

XVIII

Titolo della tesi

Condivisione del rischio con asimmetrie informative

Nome e Cognome del dottorando

LAURENCE ALES

A.A. 2008/2009

Docente Guida/Tutor: Prof. Franco Peracchi

Coordinatore: Prof. Franco Peracchi

Acknowledgments

I'd like to thank Franco Peracchi and Roberto Monte for giving me the opportunity to start my career in economics.

I'd like to thank my advisor Larry Jones for his continuous help and support.

I also thank Rajesh Aggarwal, Francesca Carapella, V.V. Chari, Simona Cociuba, Roozbeh Hosseini, Katya Kartashova, Patrick Kehoe, Narayana Kocherlakota, Erzo Luttmer, Pricila Maziero, Ellen R. McGrattan, Fabrizio Perri, Chris Phelan, Anderson Schneider, Martin Schneider, Pierre Yared, Warren Weber, participants at the Midwest Macro conference in Cleveland, the SED meetings in Prague, LAEF conference on dynamic political economy and optimal taxation and participants at the Federal Reserve Bank of Minneapolis bag lunch for comments and suggestions.

Dedication

This dissertation is dedicated to my family.

Contents

List of Tables

List of Figures

Part I

Introduction

This thesis studies the optimal provision of insurance in environments with asymmetric information. Individuals are subject to idiosyncratic uncertainty and wish to enter a contract with other individuals that can provide them some insurance. The basic friction present in all of the environments studied here, is that the cost of observing the realization of the individual uncertainty varies across people: while each individual might incur in little or no cost to observe shocks that affect his own utility, it might be arbitrarily costly for him to observe shocks that affect any other individual in society. To overcome this friction the optimal contract will have to provide the right incentives to individuals to report their private information. The papers in this thesis study these incentives under particular specification of the nature of the asymmetric information. The emphasis will be both positive (trying to use the nature of the optimal contract to rationalize aspects we observe in the data) and normative (characterizing the infinite horizon nature of the optimal contract).

Consider and environment where the agent is subject to an idiosyncratic multidimensional random shock $\theta \in \Theta = \prod_{i=1}^{N} \Theta_i$ that affects his utility, where $\Theta_i = \{\theta_H^i, \theta_L^i\}$.[1] The agent can observe the shock without incurring any cost, the planner on the other hand has to pay a cost proportional on how precise his observation

[1] For more details on the environment refer to sections 2 and 9.

throughout a signal $\tilde{\theta}$ will be. The precision of the signal will be the conditional probability of the signal of returning the true value of the private shock. Denote the cost associated with a given level of precision with $T : [0,1]^N \to \mathbb{R}$, denote also with T_i the cost associated with observing the i-th component of θ. A precision equal to zero will characterize a signal independent of the true realization of the private shock, while a precision of 1 will denote a perfectly correlated signal.

By varying the property of the cost function we can nest numerous environments commonly studied in the optimal contract literature. For example if $T_i(x) = 0$ for all values of $x \in [0,1]$ then the i-th component of θ will be perfectly observed (as in Cochrane (1991) and Townsend (1994)); on the other hand if $T_i(x) = \infty$ for $x \in [0,1]$, then no additional information on the i-th component of the shock will be acquired by the planner and any insurance will have to be provided uniquely relying on the reports of the agent (in this last category fall environments such as the one in Mirrlees (1971), Thomas and Worrall (1990a), Atkeson and Lucas (1992a) and Golosov, Kocherlakota, and Tsyvinski (2003)). Finally we can also consider the case where the principal can perfectly observe the state paying a finite cost: that is when the cost function is such that $T_i(0) = 0$ and $T_i(x) = k$ for all $x > 0$ for some finite $k > 0$, this case is the standard costly state verification environment as in Townsend (1979) and Aiyagari and Alvarez (1995a).

In this thesis we will study two particular cases of this information structure. In part 2 a model with $N = 2$, $T_1 = 0$ and $T_2 = \infty$ will be studied; this will imply that one shock will be perfectly observable and the other entirely private information of the agent. In part 3 an intermediate cases is studied with $N = 1$ but with $T(x)$ being an incresing function with $T(0) = T'(0) = 0$ so that some but not perfect additional information will always be acquired by the planner.

The focus of part 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.[2] 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. The presence of private shocks has important implications for the age profiles of consumption and income. 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 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 productivities must be private information.

Part 3 studies how the optimal allocation of such problems is affected by changing

[2] This part is coauthored with Pricila Maziero

the amount of information available to the principal. In every period, the agent receives a privately observed taste shock, the principal provides insurance to the agent based on two inputs: the report of the agent and on a signal which is correlated with the realized taste shock. Different amounts of information are modeled by looking at different correlations between the signal and the taste shock. Two cases are studied, with the correlation level set exogenously, and endogenously. In both environments the optimal allocation is not qualitatively different from the case in which the principal only relies on the report of the agent, in the long run the agent is driven to misery. In addition, if the correlation of the signal is determined endogenously, the principal in the long run will invests no resources in gathering additional information and the signal will become uninformative.

Part II

Accounting For Private Information[3]

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.[4] 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

[3]This chapter is coauthored with **Pricila Maziero**

[4]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.[5] Looking at the cross section, we observe that inequality in consumption is increasing over age.[6] At the same time, the profile for the cross-sectional variance in hours worked is slightly decreasing over the working life. These facts suggest that workers are partially insured against shocks.

The study of contractual arrangements that can explain the lack of full insurance is the underlying motivation for Green (1987), Thomas and Worrall (1990a), Atkeson and Lucas (1992a). 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 in studying the joint behavior of consumption and hours; for this reason we focus on

[5]See, for example, Cochrane (1991), Dynarski and Gruber (1997), and Gervais and Klein (2006). See section 4 for details on the data.

[6]See, for example, Deaton and Paxson (1994) and Heathcote, Storesletten, and Violante (2005).

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 productivities, 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, hours and income 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 recent 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.[7] 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.[8] The key difference of our paper is the focus on a Mirrlees environment, which allows us to consider jointly the behavior of consumption and hours worked. The second focuses on a moral hazard problem with hidden savings and shows how private information can explain the excess smoothness in consumption in data from the United Kingdom.

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 (2006) and Heath-

[7]Some exceptions are Farhi and Werning (2006), Golosov, Tsyvinski, and Werning (2006), Golosov and Tsyvinski (2006) and Huggett and Parra (2006).

[8]Also, Ai and Yang (2007) study an environment with private information and limited commitment that can account for the elasticity of consumption growt to income growth found in U.S data.

cote, 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, in our environment the assets available 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 result, and section 11 concludes.

2 Environment

In this section we describe the main features of the environment and write 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 Albanesi 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.[9] A large

[9] In our environment, there is no moral hazard problem after retirement; hence, retirement can

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

$$W \{c_t, l_t\}_{t=1}^T = \mathbb{E}_0 \sum_{t=1}^{T} \beta^{t-1} [u(c_t, l_t)], \tag{2}$$

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.

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 $\theta^t \equiv (\theta_1, .., \theta_t)$ and $\eta^t \equiv (\eta_1, .., \eta_t)$ 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:

$$y_t = f(\theta_t, \eta_t) \cdot l_t, \tag{3}$$

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 θ_t and η_t. The shock θ is publicly observed by all workers (from now on, we will call it the public shock,), while the shock η is observed privately by the worker (the private shock). Let $\pi(\theta^T, \eta^T)$ denote the probability of drawing a particular sequence of productivity shocks θ^T and η^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.[10] The contribution of private information to labor productivity uncertainty is summarized by Ω, the fraction of the variance of labor productivity due to private information,

$$\Omega_t = \frac{\sigma_t^2(\eta)}{\sigma_t^2(\eta) + \sigma_t^2(\theta)} \in [0,1]. \quad (4)$$

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

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 θ^t and the reported history of private shock η^t, a level of required effective output and a level for consumption. We denote the contract with $\{c, y\} = \{c_t(\theta^t, \eta^t), y_t(\theta^t, \eta^t)\}_{t=1}^T$. Before any uncertainty is realized, each worker is associated with a number w_0, which

[10] In section 5.3 we relax this assumption by looking at the effects of a persistent public shock.

denotes his entitlement of discounted lifetime utility. As in Atkeson and Lucas (1992a), we solve the correspondent planner's problem for each level of promised utility w_0 for each worker of generation 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:

$$\sum_{t=1}^{T}\sum_{\theta^t,\eta^t}\pi_\theta(\theta^t)\pi_\eta(\eta^t)\beta^{t-1}u\left(c_t\left(\theta^t,\eta^t\right),\frac{y_t\left(\theta^t,\eta^t\right)}{f(\theta_t,\eta_t)}\right)\geq \tag{5}$$
$$\sum_{t=1}^{T}\sum_{\theta^t,\eta^t}\pi_\theta(\theta^t)\pi_\eta(\eta^t)\beta^{t-1}u\left(c_t\left(\theta^t,\tilde{\eta}^t\right),\frac{y_t\left(\theta^t,\tilde{\eta}^t\right)}{f(\theta_t,\eta_t)}\right),\quad \forall\tilde{\eta}^t\in H^t.$$

Note that in our environment, full insurance against productivity shocks is not incentive compatible. The intuition for this is clear 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 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 said to be feasible if it satisfies the following:

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

In this environment, the planner offers a contract for a worker with initial lifetime utility w_0 that solves the following problem:

$$S_1(w_0) = \max_{\{c,y\}_{t=0}^{T}} \sum_{t=1}^{T} \sum_{\theta^t,\eta^t} \pi_\theta(\theta^t)\pi_\eta(\eta^t)\beta^{t-1} u\left(c_t\left(w_0,\theta^t,\eta^t\right), \frac{y_t\left(w_0,\theta^t,\eta^t\right)}{f(\theta_t,\eta_t)}\right) \quad (7)$$

$$s.t. \quad (5) \quad and \quad (6)$$

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 maximization problem, we solve its dual cost minimization problem.

In the recursive formulation we need to distinguish between the problem faced 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 $\eta_t \in \{\eta_{H,t}, \eta_{L,t}\}$ with $\eta_{H,t} > \eta_{L,t}$ for all t. We also consider the relaxed problem, only considering incentive compatibility constraints for the agent that draws η_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.[11] The period T problem is

$$S_T(w) = \min_{c,y} \sum_{\theta_T} \pi(\theta_T) \sum_{\eta_T} \pi(\eta_T) \left[c_T(\theta_T, \eta_T) - y_T(\theta_T, \eta_T) \right], \tag{8}$$

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

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

The time $T-1$ problem is

$$S_{T-1}(w) = \min_{c,y,w'} \sum_{\theta} \pi(\theta) \sum_{\eta} \pi(\eta) \left[c_{T-1}(\theta, \eta) - y_{T-1}(\theta, \eta) + q S_T(w'_{T-1}(\theta, \eta)) \right], \tag{11}$$

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

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

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

[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.

few analytical results that can be derived relies on applying variational methods to the planner problem. This approach has been used by Rogerson (1985a) and in an environment similar to ours by Golosov, Kocherlakota, and Tsyvinski (2003). The key result is that the planner equates 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 (with a node we denote a particular history of labor productivity 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*

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

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

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:

$$\frac{1}{u_c\left(c\left(\theta^t,\eta^t\right)\right)} = \frac{q}{\beta} \sum_{\theta^{t+1},\eta^{t+1}} \pi\left(\theta^{t+1}|\theta^t\right)\pi\left(\eta^{t+1}|\eta^t\right)\frac{1}{u_c\left(c\left(\theta^{t+1},\eta^{t+1}\right)\right)}, \quad \forall\theta^t,\eta^t. \tag{16}$$

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 a 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, in the presence of private information, consumption depends on θ if effective labor is given by the function $y(\theta, \eta) = (\theta \cdot \eta) \cdot l$.

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 $c(\theta, \eta) \neq c(\hat{\theta}, \eta)$ for all $\theta, \hat{\theta}, \eta$.*

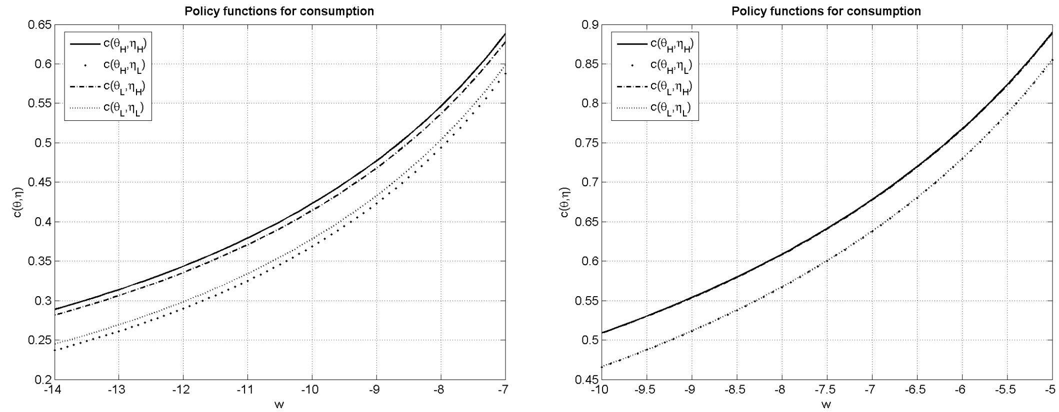
Proof. In appendix 7.4. ∎

The key intuition for this result is how different realizations of θ 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 θ, $\Delta(\theta)$ denotes the effective amount of labor productivity that the worker with realization η_H can misreport. This implies that if $\Delta(\theta)$ varies with θ, 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 θ. In figure 1-(a) we illustrate the results of the proposition. The plots displays typical policy function for the case with $f(\theta, \eta) = \theta \cdot \eta$.

Consider now a specification of the total labor productivity given by $f(\theta, \eta) = \theta + \eta$. This specification makes $\Delta(\theta)$ independent of θ. The policy functions for



(a)

(b)

Figure 1: Policy functions for consumption with $\Omega = 0.5$. In panel (a) $f(\theta, \eta) = \theta\eta$; in panel (b) $f(\theta, \eta) = \theta + \eta$.

consumption under this specification are displayed in figure 1-(b). In this case, we observe that 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 form for the utility function is Cobb-Douglas,

$$u(c, l) = \frac{\left[c^{\alpha}(1-l)^{1-\alpha}\right]^{1-\sigma}}{1-\sigma}, \tag{18}$$

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 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 allocation, setting $T = 1$. This is the static known Mirrleesian benchmark.[12] With only one period, the planner can provide incentives only 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:

$$\tau_{cl}(\theta,\eta) = 1 + \frac{1}{\theta\eta}\frac{u_l(c(\theta,\eta), l(\theta,\eta))}{u_c(c(\theta,\eta), l(\theta,\eta))}. \tag{19}$$

In the full information case $\tau_{cl} = 0$, hours are set efficiently according to current labor productivity, and consumption is determined equating marginal utility of con-

[12]For a detailed review of the literature on the static and dynamic Mirrlesian environment, we refer the reader to Tuomala (1990) and Golosov, Tsyvinski, and Werning (2006).

sumption across workers.[13] This induces a volatility of hours directly related to the volatility of labor productivity. The volatility of consumption depends on the cross partial between consumption and leisure.

Table 1: Population statistics for the static environment

Ω	$var(c)$	$var(l)$	$E[\tau_{cl}]$	$cov(c,\theta\eta)$	$cov(l,\theta\eta)$
1	0.067	0.017	0.069	0.060	0.030
0.75	0.052	0.030	0.054	0.050	0.036
0.25	0.023	0.059	0.021	0.032	0.052
0	0.011	0.080	0	0.024	0.065

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 Ω 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

[13]Note that 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.

consumption increases as Ω increases, while the response of hours decreases.

In this example, different distortions are achieved by varying Ω. 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 $T = 6$.[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 or 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:

$$\lambda_t + \frac{\mu_t(\theta)}{\pi(\eta_H)\pi(\theta)} = \frac{q}{\beta} S'_{t+1}(w'_t(\theta, \eta_H)), \quad \forall \theta, \tag{20}$$

$$\lambda_t - \frac{\mu_t(\theta)}{\pi(\eta_L)\pi(\theta)} = \frac{q}{\beta} S'_{t+1}(w'_t(\theta, \eta_L)), \quad \forall \theta, \tag{21}$$

with λ_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

[14] From here onwards, we assume that a period represents a five-year interval, with the initial period set at age 25.

of promised utility. A positive multiplier on the incentive constraint, and 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.[15] The evolution of promised utility for an exante 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.

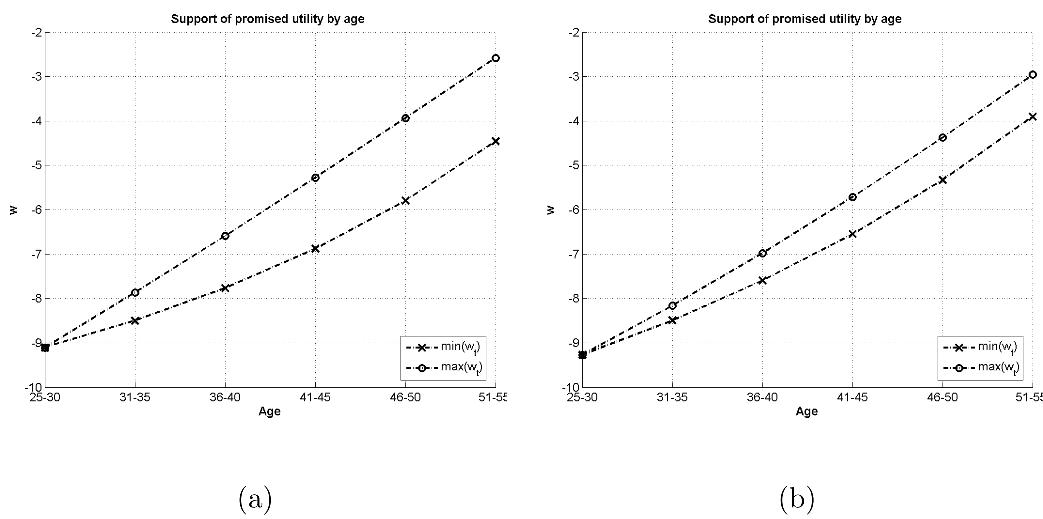


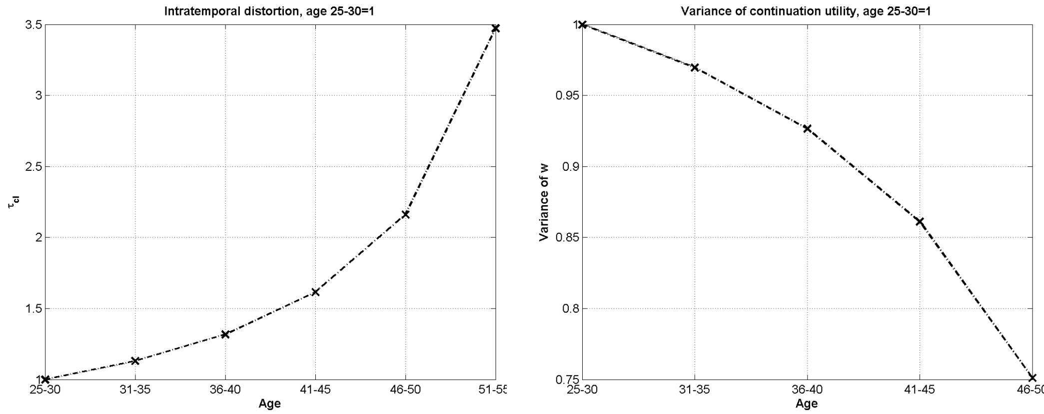
Figure 2: Support of promised utility by age. In panel (a) $\Omega = 1$; in panel (b) $\Omega = \frac{1}{2}$.

The mechanism in play is the following: as workers age, the planner provides insurance substituting progressively from incentives provided on the intertemporal

[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 (1990a), Atkeson and Lucas (1992a), Aiyagari and Alvarez (1995b), and Phelan (1998).

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 look at 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 intra-temporal 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, we observe that 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 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 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

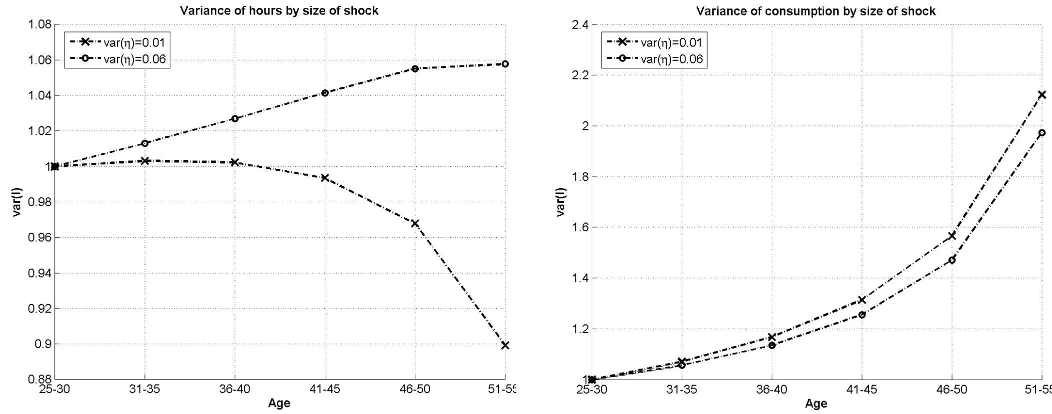


(a) (b)

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.

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. The variance 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



(a) (b)

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.

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.[16]

Finally, we look at the relationship between consumption and output. From equation (3) when $f(\theta,\eta)=\theta\eta$, we have that

$$\log y = \log\theta + \log\eta + \log l. \tag{22}$$

[16] 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 display a roughly linear increase of variance of consumption, Deaton and Paxson (1994) show that this increase is convex for the United Kingdom and Taiwan.

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

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

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; overall the first term will dominate, 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 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 σ, through its effect on the cross-partial, increases the level of the covariance while increasing Ω increase its growth rate.

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. Describing how this optimal allocation is implemented is beyond the scope of this paper. Several papers have proposed decentralizations for environments similar to ours. Prescott and Townsend (1984)

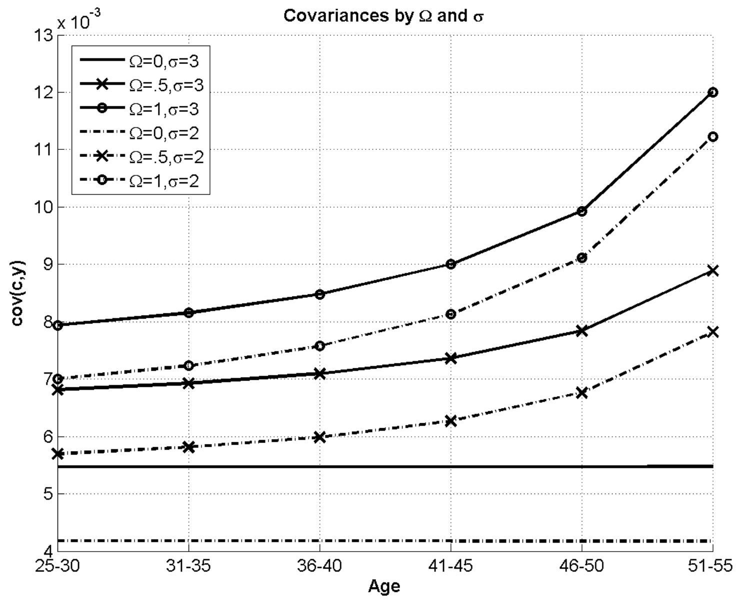


Figure 5: Covariances for different Ω and σ.

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 allocation 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 depend 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. 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.[17] All the earnings variables and hours refer to the household head, while the expenditure variables are total household expenditure per adult equivalent.[18] 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 1980 to 2003.[19] 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.

[17]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.

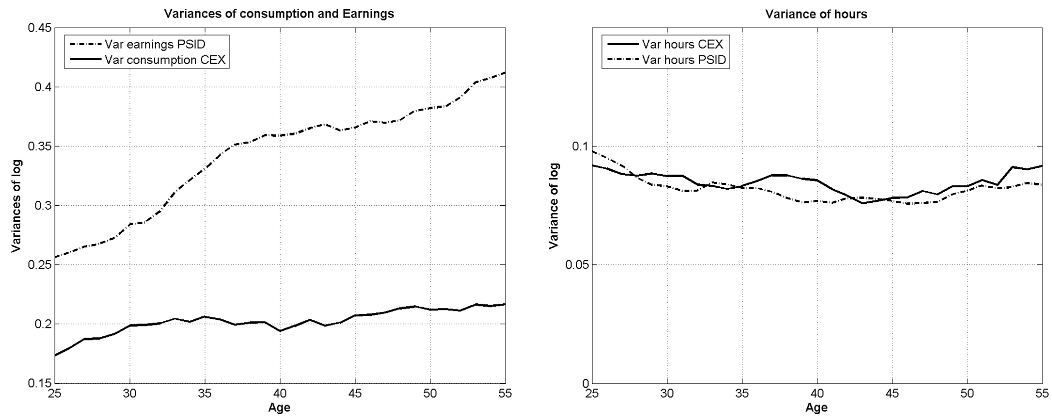
[18]We use the Census definition of adult equivalence.

[19]By stopping at age 55 we also minimize the disconnect 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).

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) (b)

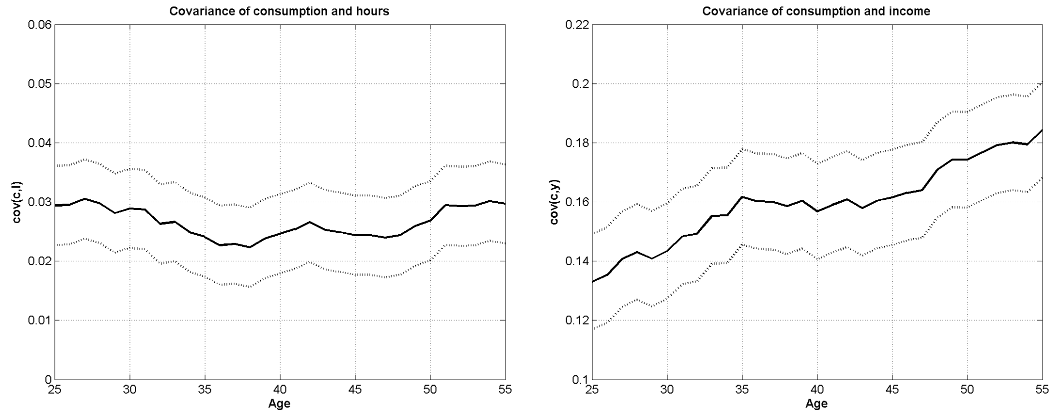
Figure 6: Life-cycle profiles, source: CEX and PSID. Panel (a) displays the variances of consumption expenditure and earnings, panel (b) the variances of hours.

a significant increase in the covariance of consumption and earnings.

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 pin down 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



(a) (b)

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.

and simulate a population. This enables us to determine preference parameters and Ω using a minimum distance estimator, which consists of 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 and covariances of consumption hours and labor income by age. 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, $g_x(\Gamma) = \sigma^2_x(\Gamma) - \hat{\sigma}^2_x$. 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 Γ will be given by

$$\Gamma^* \equiv \arg\min_{\Gamma} g(\Gamma) \cdot W \cdot g(\Gamma)', \tag{24}$$

where W is a $J \times J$ positive semi-definite weighting matrix. Once we obtain the value of Γ^* we compute the properties of the gradient at the minimum determining 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 will estimate the curvature parameter (σ) and the share of consumption (α). For each shock we restrict to two realizations: $\theta_H > \theta_L$ for the public shock and $\eta_H > \eta_L$ for the private shock. The average value of labor productivity is held constant along the life cycle, but we do allow for an increasing variance for both private and public shocks. The two shocks are parametrized by the following, for every age t

$$
\begin{aligned}
\theta_{t,H} - \theta_{t,L} &= 2h_\theta[1 + g_v(t-1)], && (25)\\
\eta_{t,H} - \eta_{t,L} &= 2h_\eta[1 + g_v(t-1)]. && (26)
\end{aligned}
$$

With this formulation we can determine the evolution of the two shocks with only three parameters: h_θ, h_η 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 two shocks enables us also to maintain a constant Ω across the working life, from equation (4) we have that

$$
\Omega_t = \Omega = \frac{h_\eta^2}{h_\eta^2 + h_\theta^2}. \quad (27)
$$

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 δ as follows

$$
\begin{aligned}
\log(w_H) &= \log(w_0) + \log(\delta)\\
\log(w_L) &= \log(w_0) - \log(\delta)
\end{aligned}
$$

where for a given δ the value w_0 is determined so that the zero profit condition of the planner holds

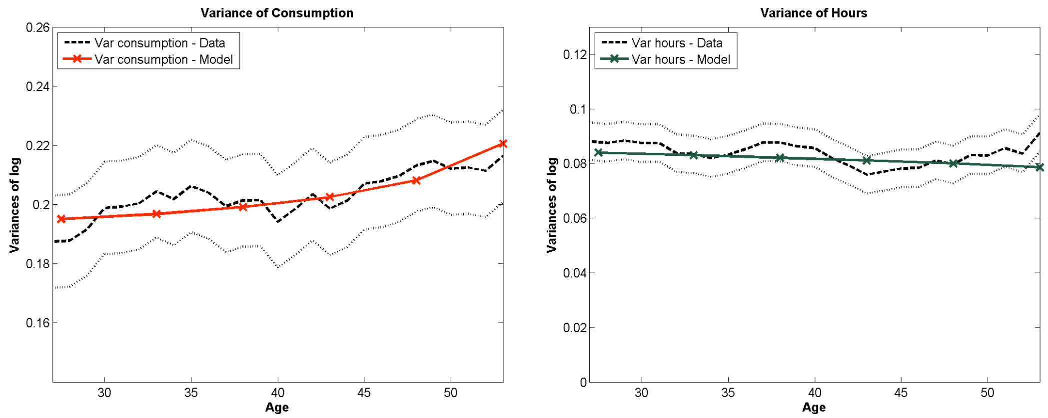
$$
\sum_{i=H,L} S_1(w_i) = 0. \quad (28)
$$

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

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 age 25 to 55. In choosing these moments, we are motivated by the following. 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. Looking at these moments then directly relates to 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 at the minimum has full rank). The value of Ω 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. The curvature parameter (σ) and the share of consumption (α) 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 (note 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 ϕ we can approximate the Frisch elasticity of labor supply by multiplying ϕ by $\frac{1-\alpha}{\alpha}$, the implied 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 over age. Figure 8 displays the fit of the model with respect to the targeted moments.



(a) (b)

Figure 8: Benchmark environment fit on matched moments.

The benchmark environment successfully accounts for 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 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). Of the two quantities consumption and hours, the second is the most likely to be 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

$$\tau_{cl} = 1 - \frac{1}{\theta\eta}\frac{1-\alpha}{\alpha}\frac{c}{L-l}, \tag{29}$$

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). The value of τ_{cl} in the original sample does not display any clear pattern over the ages 25-55. However, we observe that two factors are important in determining this measure: family composition and housing. Once we control for both (by focusing only on single households and removing services imputed from housing in the definition of consumption) τ_{cl} clearly displays an increasing trend over age as shown in figure 9.

The growth rate of τ_{cl} is decreasing in α. For $\alpha = 0.69$, as estimated in the benchmark environment, τ_{cl} in the model increases as τ_{cl} in the data up to ages 40-45, then the model overestimates the increase as shown in figure 9-(b). Moving to a lower value of α increases the growth rate of τ_{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 disutility of leisure, displays a flat profile for τ_{cl}. The only way to induce an increase in this quantity is by introducing individual taste shocks in the value of α that increase in variance as the worker age (as for example in Badel and Huggett (2006)).

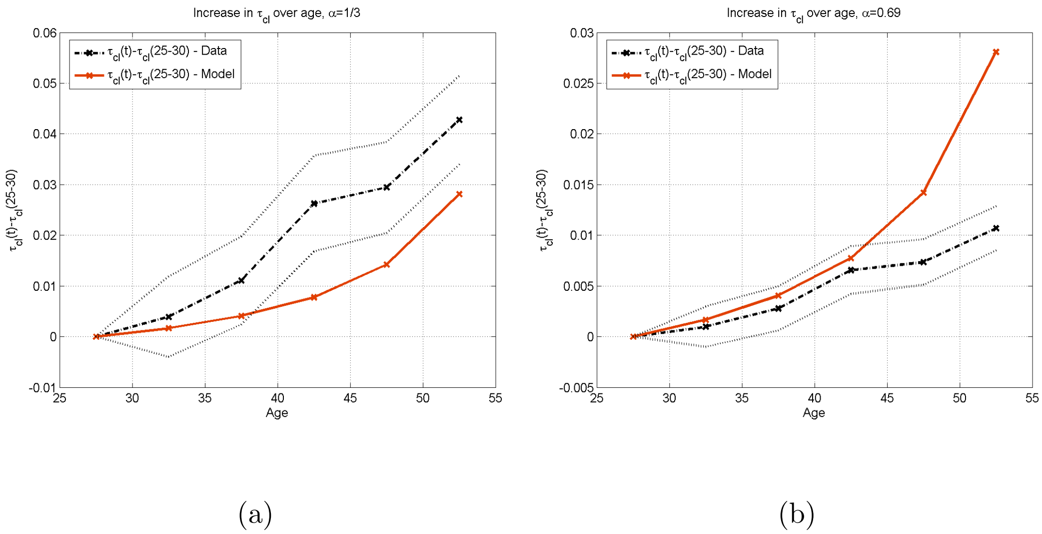


Figure 9: Evolution of the average distortion on the marginal of substitution between consumption and leisure over age.

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 from the PSID.[20] Figure 10 shows the relation between these two variables. Overall,

[20]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 per year observation received a transfer,

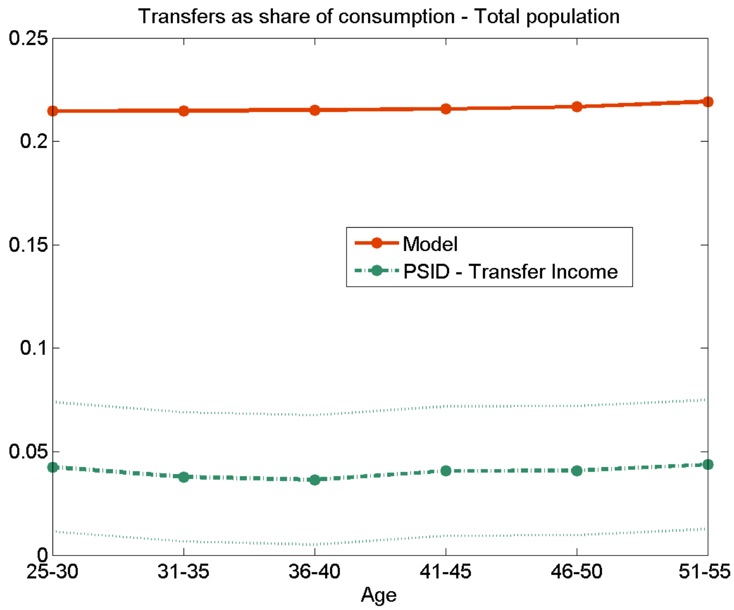


Figure 10: Transfers over age.

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 for these runs we set the initial heterogeneity in w equal to zero. 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.[21] We first look at the effect of setting $\delta = 0$ (column (2)). Then we look at the effect

and in total 67% of the households received a transfer at some stage. These transfers are significant, averaging at $1930 (1983 dollars) and account for 70% to 90% of total food expenditures.

[21] A similar limitation is also discussed in Phelan (1994).

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 α 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 looks at additional robustness checks.

5.2 Discussion

To understand why the minimum distance estimator returns a high value for Ω, we first consider the implications for the profiles of the variances of consumption and hours when Ω is equal to zero. In this case, we can solve directly for these moments. The problem is characterized by the following first-order conditions:

$$u_c\left(c\left(\theta\right), l\left(\theta\right)\right) = -\frac{1}{\theta} u_l\left(c\left(\theta\right), l\left(\theta\right)\right), \quad \forall \theta, \tag{30}$$

$$u_c\left(c\left(\theta\right), l\left(\theta\right)\right) = \frac{1}{\lambda}, \quad \forall \theta, \tag{31}$$

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

$$\ln c_\theta = \ln \theta + \ln \frac{\alpha}{1-\alpha} + \ln\left(1 - \frac{y_\theta}{\theta}\right), \tag{32}$$

similarly, from (31) we have

$$\ln c_\theta = \ln\lambda + \ln\alpha + \alpha\,(1-\sigma)\ln c_\theta + (1-\alpha)\,(1-\sigma)\ln\left(1-\frac{y_\theta}{\theta}\right). \qquad (33)$$

Combining the previous two equations, we get

$$Var\,[\ln c_\theta] = (\phi-1)^2\,Var\,[\ln\theta]\,, \qquad (34)$$

$$\Delta Var\,[\ln c_\theta] = (\phi-1)^2\,\Delta Var\,[\ln\theta]\,, \qquad (35)$$

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

$$Var\,[\ln(1-l_\theta)] = \phi^2 Var\,[\ln\theta]\,, \qquad (36)$$

$$\Delta Var\,[\ln(1-l_\theta)] = \phi^2 \Delta Var\,[\ln\theta]\,. \qquad (37)$$

If the amount of time devoted to leisure is greater than labor hours, we have that $Var\,[\ln(l)] > Var\,[\ln(1-l)]$. 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. This feature highlights the difficulty of a full information insurance environment in describing the profile of consumption and hours.[22] Private information is necessary to provide an increasing variance in consumption while at the same time keeping the variance of hours constant.

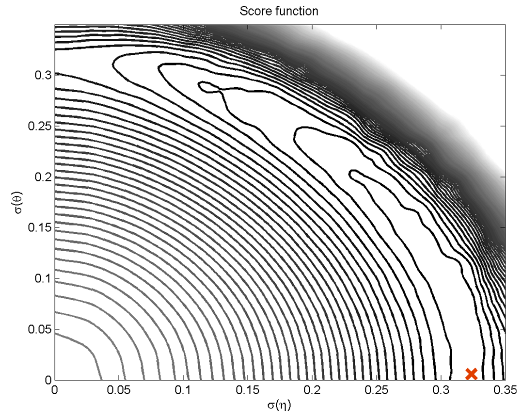
We now try to provide some intuition on the values obtained for the preference parameters. During the minimization procedure starting for example from an initial

[22]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).

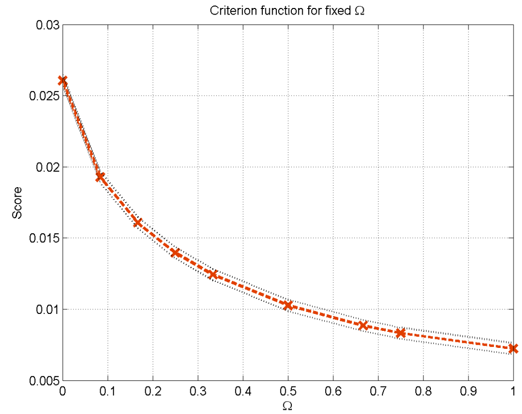
guess of ($\sigma = 3,\ \alpha = \frac{1}{3}$), we observe that the minimization path ultimately progresses, decreasing risk aversion and increasing the elasticity of leisure (decreasing σ and increasing α), for the following reasons. For a given level of uncertainty, a high elasticity of leisure makes the spread in hours at the optimum larger. This translates, in the presence of private information, to a more severe moral hazard problem; this implies larger distortions on both the intra-temporal 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 σ 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 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.

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 increase. For each point in the graph, the



(a)



(b)

Figure 11: Surface plot of score function for benchmark estimation (a); Criterion function for benchmark estimation with respect to Ω (b).

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 Ω: the score function displays a semi circular ridge at a constant distance from the origin, this pins downs 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 Ω close to 1.

In figure 11-(b) we plot the criterion function with respect to Ω. Each point in this curve is generated by keeping the value of Ω 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 denotes that the moments chosen are sensitive to increases of Ω as we move away from the full information case ($\Omega = 1$) but become less responsive as we move to higher values of Ω.

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 an identity weighting matrix in the minimization criterion. We also 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

$$\begin{aligned} \Gamma^* \quad &\equiv \quad \arg\min_{\Gamma} g(\Gamma) \cdot W(\Gamma)^{-1} \cdot g(\Gamma)' \\ s.t. \quad & \qquad W(\Gamma) = E\left[g(\Gamma) \cdot g(\Gamma)'\right]. \end{aligned}$$

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 (ϵ) to 1. We relax this implicit constraint by looking at the more general CES utility function,

$$u(c,l) = \frac{\left[\alpha c^{\nu} + (1-\alpha)(1-l)^{\nu}\right]^{\frac{1-\sigma}{\nu}}}{1-\sigma}, \tag{38}$$

where now $\epsilon = \frac{1}{1-\nu}$. 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. 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%.[23] The results are in column (8) of tables 4 and 5.

Allowing persistence of the public shock

[23] 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.

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 ρ. 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 Ω is now equal to .79, the estimated value for ρ is 0.99, public shocks to labor productivity are permanent.

Alternative utility function

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:

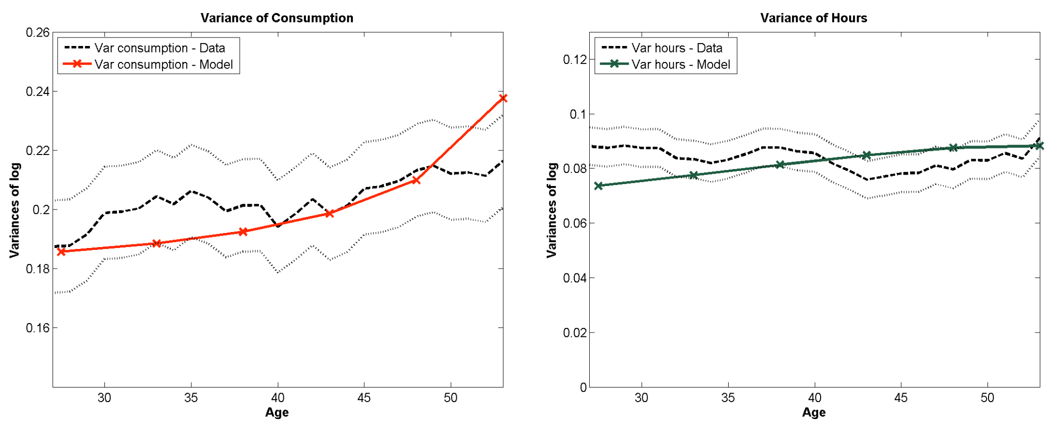
$$u(c, l) = \alpha \frac{c^{1-\sigma}}{1-\sigma} - \frac{l^{1+\alpha_l}}{1+\alpha_l}. \tag{39}$$

This specification is commonly used in the labor literature.[24] The coefficient of risk aversion is given by σ and $1/\alpha_l$ 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

[24] See Browning, Hansen, and Heckman (1999).

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 now slightly increasing and underestimated early in the working life.



(a) (b)

Figure 12: Estimation results for utility given in equation (39): panel (a) variance of consumption; panel (b) variance of hours.

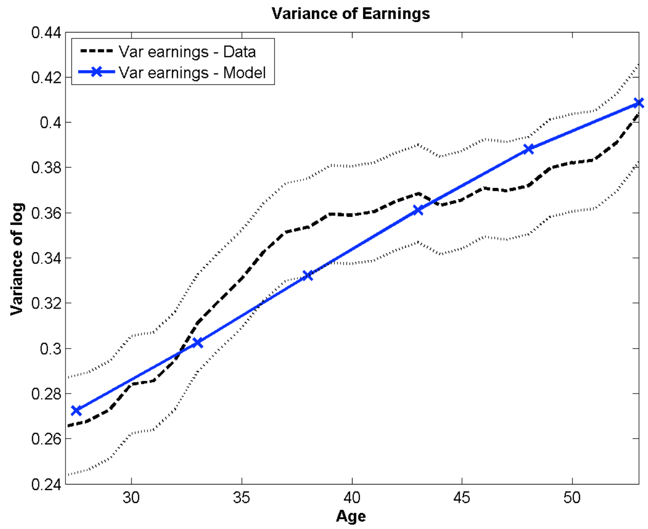


Figure 13: Estimation results for utility given in equation (39): variance of earnings.

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 is of interest since it reflects the insurance possibilities of workers against income fluctuationsas. We compute the response of consumption growth to consumption growth (α_2) from the following regression:[25]

$$\Delta \log c_t^i = \alpha_1 + \alpha_2 \Delta \log y_t^i + controls. \tag{40}$$

[25] 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.

We compute the value of α_2 from the CEX data using an OLS and instrumental variable approach as in Dynarski and Gruber (1997). Results are in table 6. We perform the same estimation on the panel generated by the model. For our baseline environment with non separable utility, we find a value of α_2 equal to 0.107 which falls within our estimates using OLS and IV on CEX data.[26] A more stark interpretation of the link between α_2 and the level of insurance available to workers can be derived in an environment with separable utility. In this case, if workers are fully insured against income shocks, the value of α_2 is 0 (marginal utility of consumption is held constant). In our environment with private information and separable preferences α_2 is equal to 0.067, which directly implies that the insurance possibilities available to workers are reduced.

[26] Gervais and Klein (2006) show that the standard IV estimates overstates the true value α_2. Using a projection method they estimate in the CEX a value of $\alpha_2 = 0.1$. Also note, Ai and Yang (2007), in an environment with private information and limited commitment, find a value of $\alpha_2 = 0.269$.

Table 6: Consumption response to income shocks

Source	α_2
Data - OLS	0.028 (.004)
Data - IV	0.177 (.021)
Data - IV-20%	0.134 (.018)
Model - separable	0.067
Model - non separable	0.107

Estimation of α_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 this environment. The model predicts that the value of α_2 should be increasing in age due to progressive more importance given to within period incentives. We calculate this statistic with the same restriction imposed to generate picture 9 (restricting to single household and removing services from housing from consumption). Figure 14 displays the result. The value of α_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.

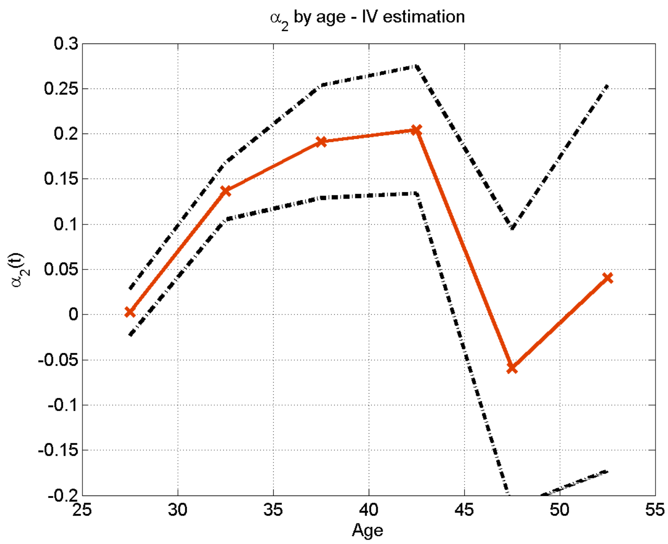


Figure 14: Estimation of α_2 by age. Source: CEX.

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 provision of incentives differs along the lifecycle: 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. 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.

An interesting question is what determines worker's heterogeneity early in life. In this paper we do not model how initial conditions, in the form of promised utility, are determined. Huggett, Ventura, and Yaron (2007) show that differences in human capital account for a large fraction of inequality early in life. Motivated by this, in current work (Ales and Maziero (2007)) we study how much of the inequality observed at age 25 can be explained as the result of a constrained efficient environment in which workers are privately informed about their cognitive learning abilities. In this environment early in life inequality is necessary to provide incentives for an efficient level of investment in human capital.

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:

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

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 $u\left(c(\theta_H),\frac{y(\theta_H)}{\theta_H}\right)<u\left(c(\theta_L),\frac{y(\theta_L)}{\theta_H}\right)$. Recall that for the Cobb-Douglas utility we have

$$u_c\left(c,l\right) = \frac{\alpha}{c}\left(1-\sigma\right)u\left(c,l\right), \quad (41)$$

$$u_l\left(c,l\right) = -\frac{1-\alpha}{1-l}\left(1-\sigma\right)u\left(c,l\right). \quad (42)$$

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

$$(1-l(\theta))=\frac{(1-\alpha)}{\alpha}\frac{c(\theta)}{\theta}, \quad \forall\theta, \quad (43)$$

$$c(\theta_L)=c(\theta_H)\left(\frac{\theta_L}{\theta_H}\right)^{\frac{(1-\alpha)(1-\sigma)}{-\sigma}}. \quad (44)$$

Using (43) we can rewrite the utility function as

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

Substituting (44) in the above for $\theta=\theta_L$,

$$\begin{aligned}
u\left(c(\theta_L),l(\theta_L)\right) &= \frac{c(\theta_H)^{1-\sigma}\left(\frac{(1-\alpha)}{\alpha\theta_L}\right)^{(1-\alpha)(1-\sigma)}\left(\frac{\theta_L}{\theta_H}\right)^{\frac{(1-\alpha)(1-\sigma)^2}{-\sigma}}}{1-\sigma} = \\
&= \left[c(\theta_H)^{1-\sigma}\left(\frac{(1-\alpha)}{\alpha}\right)^{(1-\alpha)(1-\sigma)}\left(\frac{1}{\theta_L}\right)^{\frac{(1-\alpha)(1-\sigma)}{\sigma}}\left(\frac{1}{\theta_H}\right)^{\frac{(1-\alpha)(1-\sigma)^2}{-\sigma}}\right]\frac{1}{1-\sigma} \\
&= \left[c(\theta_H)^{1-\sigma}\left(\frac{(1-\alpha)}{\alpha}\right)^{(1-\alpha)(1-\sigma)}\left(\frac{\theta_H}{\theta_L}\right)^{\frac{(1-\alpha)(1-\sigma)}{\sigma}}\left(\frac{1}{\theta_H}\right)^{(1-\alpha)(1-\sigma)}\right]\frac{1}{1-\sigma} \\
&= u\left(c(\theta_H),l(\theta_H)\right)\left(\frac{\theta_H}{\theta_L}\right)^{\frac{(1-\alpha)(1-\sigma)}{\sigma}}. \qquad (45)
\end{aligned}$$

By assumption $0<\left(\frac{\theta_H}{\theta_L}\right)^{\frac{(1-\alpha)(1-\sigma)}{\sigma}}<1$. This implies that $u\left(c(\theta_H),\frac{y(\theta_H)}{\theta_H}\right)<u\left(c(\theta_L),\frac{y(\theta_L)}{\theta_L}\right)$. Given that $\sigma>1$, $u\left(c(\theta_H),l(\theta_H)\right)$ and $u\left(c(\theta_L),l(\theta_L)\right)$ are both negative. From this result it follows that

$$u\left(c(\theta_H),\frac{y(\theta_H)}{\theta_H}\right)<u\left(c(\theta_L),\frac{y(\theta_L)}{\theta_L}\right)<u\left(c(\theta_L),\frac{y(\theta_L)}{\theta_H}\right). \qquad (46)$$

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 $(P1)$.

$$
\begin{aligned}
S_T(w) &= \min_{c,y} \sum_{\theta_T} \pi(\theta_T) \sum_{\eta_T} \pi(\eta_T) \left[c_T(\theta_T, \eta_T) - y_T(\theta_T, \eta_T) \right], \quad (47) \\
s.t. \quad & \sum_{\theta_T} \pi(\theta_T) \sum_{\eta_T} \pi(\eta_T) u \left(c_T(\theta_T, \eta_T), \frac{y_T(\theta_T, \eta_T)}{f(\theta_T, \eta_T)} \right) = w, \quad (48) \\
& u \left(c_T(\theta_T, \eta_H), \frac{y_T(\theta_T, \eta_H)}{f(\theta_T, \eta_H)} \right) \geq u \left(c_T(\theta_T, \eta_L), \frac{y_T(\theta_T, \eta_L)}{f(\theta_T, \eta_H)} \right), \quad \forall \theta_T, \quad (49) \\
& u \left(c_T(\theta_T, \eta_L), \frac{y_T(\theta_T, \eta_L)}{f(\theta_T, \eta_L)} \right) \geq u \left(c_T(\theta_T, \eta_H), \frac{y_T(\theta_T, \eta_H)}{f(\theta_T, \eta_L)} \right), \quad \forall \theta_T, \quad (50)
\end{aligned}
$$

The time $T-1$ problem is

$$
\begin{aligned}
S_{T-1}(w) &= \min_{c,y,w'} \sum_{\theta} \pi(\theta) \sum_{\eta} \pi(\eta) \left[c_{T-1}(\theta, \eta) - y_{T-1}(\theta, \eta) + q S_T(w'_{T-1}(\theta, \eta)) \right], \quad (51) \\
s.t. \quad & \sum_{\theta} \pi(\theta) \sum_{\eta} \pi(\eta) \left[u \left(c_{T-1}(\theta, \eta), \frac{y_{T-1}(\theta, \eta)}{f(\theta, \eta)} \right) + \beta w'_{T-1}(\theta, \eta) \right] = w, \quad (52) \\
& u \left(c_{T-1}(\theta, \eta_H), \frac{y_{T-1}(\theta, \eta_H)}{f(\theta, \eta_H)} \right) + \beta w'_{T-1}(\theta, \eta_H) \geq \\
& u \left(c_{T-1}(\theta, \eta_L), \frac{y_{T-1}(\theta, \eta_L)}{f(\theta, \eta_H)} \right) + \beta w'_{T-1}(\theta, \eta_L), \quad \forall \theta_{T-1}, \quad (53) \\
& u \left(c_{T-1}(\theta, \eta_L), \frac{y_{T-1}(\theta, \eta_L)}{f(\theta, \eta_L)} \right) + \beta w'_{T-1}(\theta, \eta_L) \geq \\
& u \left(c_{T-1}(\theta, \eta_H), \frac{y_{T-1}(\theta, \eta_H)}{f(\theta, \eta_L)} \right) + \beta w'_{T-1}(\theta, \eta_H), \quad \forall \theta_{T-1}. \quad (54)
\end{aligned}
$$

Let the relaxed maximization problem be the original problem without constraints (50) and (54). Denote it by $(P2)$.

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

(54) for some θ_{T-1}. Then

$$u\left(c_{T-1}(\theta,\eta_H)\right) - v\left(\frac{y_{T-1}(\theta,\eta_H)}{f(\theta,\eta_L)}\right) + \beta w'_{T-1}(\theta,\eta_H) >$$
$$u\left(c_{T-1}(\theta,\eta_L)\right) - v\left(\frac{y_{T-1}(\theta,\eta_L)}{f(\theta,\eta_L)}\right) + \beta w'_{T-1}(\theta,\eta_L). \tag{55}$$

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

$$v\left(\frac{y_{T-1}(\theta,\eta_H)}{f(\theta,\eta_H)}\right) - v\left(\frac{y_{T-1}(\theta,\eta_H)}{f(\theta,\eta_L)}\right) > v\left(\frac{y_{T-1}(\theta,\eta_L)}{f(\theta,\eta_H)}\right) - v\left(\frac{y_{T-1}(\theta,\eta_L)}{f(\theta,\eta_L)}\right),$$
$$v\left(\frac{y_{T-1}(\theta,\eta_H)}{f(\theta,\eta_H)}\right) - v\left(\frac{y_{T-1}(\theta,\eta_L)}{f(\theta,\eta_H)}\right) > \left(\frac{y_{T-1}(\theta,\eta_H)}{f(\theta,\eta_L)}\right) - v\left(\frac{y_{T-1}(\theta,\eta_L)}{f(\theta,\eta_L)}\right) \tag{56}$$

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

$$v(x) - v\left(x-\varepsilon\right) > v(\hat{x}) - v\left(\hat{x}-\varepsilon\right),$$
$$v(x) - v\left(x-\varepsilon\right) > v(\hat{x}) - v\left(\hat{x}-\varepsilon\frac{f(\theta,\eta_L)}{f(\theta,\eta_H)}\right).$$

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:

$$v\left(\frac{y_{T-1}(\theta,\eta_H)}{f(\theta,\eta_L)}\right) - v\left(\frac{y_{T-1}(\theta,\eta_L)}{f(\theta,\eta_L)}\right) > v\left(\frac{y_{T-1}(\theta,\eta_H)}{f(\theta,\eta_H)}\right) - v\left(\frac{y_{T-1}(\theta,\eta_L)}{f(\theta,\eta_H)}\right).$$

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

7.3 Proof of Proposition 1

Proof. The proof follows Rogerson (1985a) 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 (θ,η). Let $C\left(\bar{u}\right)$ 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 $\bar{u}(\theta^t, \left[\eta^{t-1}, \eta_t\right]) = \bar{u}^*(\theta^t, \left[\eta^{t-1}, \eta_t\right]) - \Delta$ and $\bar{u}(\left[\theta^t, \theta_{t+1}\right], \eta^{t+1}) = u^*(\left[\theta^t, \theta_{t+1}\right], \eta^{t+1}) + \Delta/\beta$ for all θ^t and $u(\theta^t, \left[\eta^{t-1}, \tilde{\eta}_t\right]) = u^*(\theta^t, \left[\eta^{t-1}, \tilde{\eta}_t\right])$ 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 θ is publicly observed. This contract minimizes the cost for the planner if the following holds:

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

which implies

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

Equation (14) then follows given that $C'\left(u\left(\theta^t, \eta^t\right)\right) = \frac{1}{u_c(c(\theta^t, \eta^t))}$. 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 η_t and θ^{t-1} let $u(\left[\theta^{t=1}, \theta_t\right], \eta^t) = u^*(\left[\theta^{t=1}, \theta_t\right], \eta^t) - \Delta$ 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 Δ are contingent on observables and the total utility of the worker is unchanged. Optimality of this contract requires

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

Equation (15) then follows from

$$\sum_{\eta^t} \pi\left(\eta^t|\eta^{t-1}\right) C'(u([\theta^{t-1}, \tilde{\theta}_t], \eta^t)) = \sum_{\eta^t} \pi\left(\eta^t|\eta^{t-1}\right) C'(u\left(\left[\theta^{t-1}, \theta_t\right], \eta^t\right)), \quad \forall \tilde{\theta}_t, \theta_t. \tag{60}$$

∎

7.4 Proof of Proposition 2

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

$$u\prime(c(\theta, \eta_H))\left[\lambda\pi(\theta)\pi(\eta_H) + \mu\left(\theta\right)\right] = \pi(\theta)\pi(\eta_H), \quad \forall \theta.$$

This implies that $\mu\left(\theta\right) = \mu(\hat{\theta}) = \mu$ 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 $\mu\left(\theta\right) = \mu(\hat{\theta}) = \mu$ and $c(\theta, \eta_H) = c(\hat{\theta}, \eta_H)$.) From the first-order conditions (FOCs) for $w'(\theta, \eta)$

$$\begin{aligned} \pi(\theta)\pi(\eta_H)\beta\lambda + \beta\mu &= \pi(\theta)\pi(\eta_H)qS_T'(w'(\theta, \eta_H)), \quad \forall \theta, \\ \pi(\theta)\pi(\eta_L)\beta\lambda - \beta\mu &= \pi(\theta)\pi(\eta_L)qS_T'(w'(\theta, \eta_L)), \quad \forall \theta. \end{aligned}$$

This implies

$$w'(\theta, \eta_H) = w'(\hat{\theta}, \eta_H), \quad w'(\theta, \eta_L) = w'(\hat{\theta}, \eta_L). \tag{61}$$

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

$$\begin{aligned} &\frac{1}{\theta\eta_H} v\prime\left(\frac{y(\theta, \eta_H)}{\theta\eta_H}\right)\left[\lambda\pi(\theta)\pi(\eta_H) + \mu\right] = -\pi(\theta)\pi(\eta_H), \quad \forall \theta, \\ &\frac{1}{\theta\eta_H} v\prime\left(\frac{y(\theta, \eta_H)}{\theta\eta_H}\right) = \frac{1}{\hat{\theta}\eta_H} v\prime\left(\frac{y(\hat{\theta}, \eta_H)}{\hat{\theta}\eta_H}\right). \end{aligned}$$

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 $y(\theta,\eta_L)$

$$\begin{aligned}
&\frac{\lambda\pi(\theta)\pi(\eta_L)}{\theta\eta_L}v'\left(\frac{y(\theta,\eta_L)}{\theta\eta_L}\right) - \frac{\mu}{\theta\eta_H}v'\left(\frac{y(\theta,\eta_L)}{\theta\eta_H}\right) = -\pi(\theta)\pi(\eta_L), \quad \forall\theta,\\
&\frac{\lambda\pi(\theta)\pi(\eta_L)}{\theta\eta_L}v'\left(\frac{y(\theta,\eta_L)}{\theta\eta_L}\right) - \frac{\mu}{\theta\eta_H}v'\left(\frac{y(\theta,\eta_L)}{\theta\eta_H}\right) =\\
&\frac{\lambda\pi(\hat{\theta})\pi(\eta_L)}{\hat{\theta}\eta_L}v'\left(\frac{y(\hat{\theta},\eta_L)}{\hat{\theta}\eta_L}\right) - \frac{\mu}{\hat{\theta}\eta_H}v'\left(\frac{y(\hat{\theta},\eta_L)}{\hat{\theta}\eta_H}\right),\\
&\frac{y(\theta,\eta_L)^\gamma}{\theta^{1+\gamma}}\left(\frac{\lambda\pi(\theta)\pi(\eta_L)}{\eta_L^{1+\gamma}} - \frac{\mu}{\eta_H^{1+\gamma}}\right) = \frac{y(\hat{\theta},\eta_L)^\gamma}{\hat{\theta}^{1+\gamma}}\left(\frac{\lambda\pi(\hat{\theta})\pi(\eta_L)}{\eta_L^{1+\gamma}} - \frac{\mu}{\eta_H^{1+\gamma}}\right).
\end{aligned}$$

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 θ. Summing the incentive-compatibility constraint for both θ, using (61) and the fact that the consumption is independent of θ we have

$$\begin{aligned}
v\left(\frac{y(\hat{\theta},\eta_H)}{\hat{\theta}\eta_H}\right) - v\left(\frac{y(\theta,\eta_H)}{\theta\eta_H}\right) &= v\left(\frac{y(\hat{\theta},\eta_L)}{\hat{\theta}\eta_H}\right) - v\left(\frac{y(\theta,\eta_L)}{\theta\eta_H}\right),\\
\frac{y(\hat{\theta},\eta_H,w)^{\gamma+1}}{(\hat{\theta}\eta_H)^{1+\gamma}} - \frac{y(\theta,\eta_H,w)^{\gamma+1}}{(\theta\eta_H)^{1+\gamma}} &= \frac{y(\hat{\theta},\eta_L,w)^{\gamma+1}}{(\hat{\theta}\eta_H)^{1+\gamma}} - \frac{y(\theta,\eta_L,w)^{\gamma+1}}{(\theta\eta_H)^{1+\gamma}}.
\end{aligned}$$

Substituting in this expression equations (62) and (63)

$$\frac{y(\hat{\theta},\eta_H)^{\gamma+1}}{(\hat{\theta}\eta_H)^{1+\gamma}} - \frac{y(\hat{\theta},\eta_H)^{\gamma+1}}{(\hat{\theta}\eta_H/\theta\eta_H)^{\frac{(1+\gamma)^2}{\gamma}}}\frac{1}{(\theta\eta_H)^{1+\gamma}} =$$
$$\frac{y(\hat{\theta},\eta_L)^{\gamma+1}}{(\hat{\theta}\eta_H)^{1+\gamma}} - \frac{y(\hat{\theta},\eta_L)^{\gamma+1}}{(\hat{\theta}\eta_H/\theta\eta_H)^{\frac{(1+\gamma)^2}{\gamma}}}\frac{1}{(\theta\eta_H)^{1+\gamma}},$$
$$y(\hat{\theta},\eta_H)^{\gamma+1}\left(\frac{1}{(\hat{\theta}\eta_H)^{1+\gamma}} - \frac{1}{(\hat{\theta}\eta_H/\theta\eta_H)^{\frac{(1+\gamma)^2}{\gamma}}}\frac{1}{(\theta\eta_H)^{1+\gamma}}\right) =$$
$$y(\hat{\theta},\eta_L)^{\gamma+1}\left(\frac{1}{(\hat{\theta}\eta_H)^{1+\gamma}} - \frac{1}{(\hat{\theta}\eta_H/\theta\eta_H)^{\frac{(1+\gamma)^2}{\gamma}}}\frac{1}{(\theta\eta_H)^{1+\gamma}}\right).$$

So that

$$y(\hat{\theta},\eta_H) = y(\hat{\theta},\eta_L). \tag{64}$$

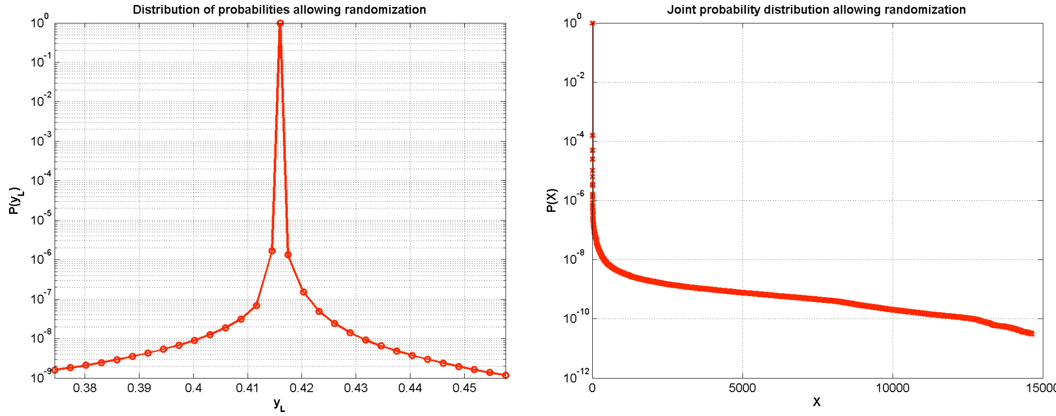
Note that the above, together with $c(\eta_H) > c(\eta_L)$, $w'(\eta_H) > w'(\eta_L)$ (these relations come from the FOCs and $\mu > 0$), implies:

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

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.



(a) (b)

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 η_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 2: Sample selection for PSID and CEX

	PSID	CEX
Baseline sample	192,897	69,816
Exclude SEO sample	109,342	NA
Hours restriction	85,811	46,559
Earnings<=0	NA	46,002
Labor income<=0	76,633	45,745
Minimum wage restriction	67,023	43,802
Age >21 and <=55	56,628	36,871
Food<=0	47,757	NA
Final sample	47,757	36,871

Numbers indicate total observations remaining at each stage of the sample selection.

Table 3: Benchmark estimation

Parameter	(1)	(2)	(3)	(4)	(5)
β, q	0.9	0.9	0.9	0.9	0.9
σ	1.46	1.27	3.88	1.59	1.42
α	0.69	0.67	$\frac{1}{3}$	0.46	0.49
g_v	-0.0073	-0.014	-0.001	-0.006	0
δ	1.22	–	–	–	–
Ω	0.99	0.99	1	1	1

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 β and q are held fixed. For columns (2) to (5) only the increase in the variance of consumption is targeted

Table 4: Summary of moments

	(1)	(2)	(3)	(4)	Data
$\Delta var(c)$	0.0287	0.017	0.0214	0.0238	0.0285
$\Delta var(l)$	-0.0062	0.0058	-0.0036	0.0022	-0.004
$\Delta cov(c,l)$	0.0045	0.002	-0.0031	-0.001	-0.0004
$E[l]$	2960	1540	2124	2244	2123
	(5)	(6)	(7)	(8)	Data
$\Delta var(c)$	0.0284	0.0281	0.025	0.0258	0.0285
$\Delta var(l)$	-0.01	-0.0013	-0.003	-0.006	-0.004
$\Delta cov(c,l)$	0.0049	0.0048	0.0021	0.0035	-0.0004
$E[l]$	2132	2831	2564	3119	2123

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

Parameter	(5)	(6)	(7)	(8)	(9)
α	0.47	0.64	0.56	1.27	0.83
σ	2.84	1.46	1.77	0.681	0.82
α_l	–	–	–	–	1.66
g_v	0.007	-0.005	-0.007	-0.008	0.12
δ	–	–	–	–	1.13
ν	–	–	0.0725	–	–
ρ	0.99	–	–	–	–
Ω	0.79	1	0.98	1	0.99

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.

	PSID (68-91)	CEX (80-04)	PSID (80-91)	CEX (80-91)
Age	35.71 (9.47)	39.17 (8.74)	35.52 (8.92)	38.14 (8.79)
Education				
High school dropout	11.96	6.99	10.34	8.27
High school graduate	36.61	29.26	35.52	31.92
College	44.99	60.46	51.11	55.93
Race				
White	90.42	86.95	91.13	88.18
Black	7.42	9	7.06	8.55
Family composition	3.19 (1.56)	3.07 (1.58)	3.02 (1.42)	3.15 (1.62)
Average earnings ($)	26,594 (18,168)	30,340 (20,406)	26,519 (20,474)	28,491 (16,908)
Average consumption ($)	NA	13,542 (6,842)	NA	13,166 (6,541)
Food ($)	4,493 (2,344)	3,791 (1,965)	4,218 (2,340)	3,998 (2,036)
Rent ($)	935 (1,890)	262 (487)	1,007 (2,028)	258 (469)
Hours	2,203 (588)	2.123 (567)	2,191 (580)	2,178 (559)

Note - All dollar amounts in 1983 dollars.

Part III

Noisy Information In a Dynamic Principal Agent Problem

8 Introduction

A risk neutral principal has to provide insurance to a risk averse agent against idiosyncratic taste shocks; the principal is constrained by a limited amount of resources and by lack of direct information on the realization of the taste shocks. This type of optimal contract problem has been extensively studied among others, by Green (1987), Thomas and Worrall (1990b) , Atkeson and Lucas (1992b). In this environment full insurance is not incentive compatible, the standard result for this class of environments is that that incentive compatible contract requires intertemporal distortions between current and continuation utilities. This feature creates a drift for continuation utilities, in the case in which principal and agent discount utility at the same rate, this drift eventually leads to the agent's continuation utility to reach it's lower bound (the so called immiseration result). This feature removes the existence of a steady state.[27]

[27] Phelan (2006) and Farhi and Werning (2007) establish that when the principal discounts future utilities at a lower rate than the agent there exist a steady state with continuation utilities bounded away from the lower bound.

Previous literature on optimal contracts with asymmetric information has focused on the case in which the cost of acquiring additional information on the realization of the shock is infinite. This paper studies the effect of relaxing this assumption. The principal receives a signal correlated with the realized taste shock, and in the first part of the paper this correlation is set exogenously. If the signal is independent of the true realization of the shock the model reverts to the standard asymmetric information environment; in the other extreme, if the signal is perfectly correlated with the taste shock, the model is equivalent to an environment with complete information. In this case complete risk sharing is incentive compatible and continuation utilities of the agent remain constant over time. The case we will focus on, has the signal not being perfectly correlated with the shock, the main result is that the optimal contract shares the same qualitative implications as the one studied by Thomas and Worrall (1990b): the agent is driven to misery.

In the second part of the paper the principal can costly choose the correlation of the signal (as these cost become infinite this environment reduces to the standard one with complete asymmetric information) with the cost being increasing in the precision of the signal. In this environment the main result is the principal will always invest resources in acquiring additional information throughout the signal; but asymptotically the planner will set the signal to be uncorrelated: in the long run no resources will be used in acquiring additional information, the additional signal becomes noise and the agent is driven to misery. A similar result was shown in Aiyagari and Alvarez (1995a) studying optimal unemployment insurance in environments with stochastic monitoring.

This paper is related to the literature on costly state verification (CSV). Static version of CSV have been studied by Townsend (1979), while more recently a dynamic version was analyzed by Aiyagari and Alvarez (1995a) and by Wang (2005). In a CSV model the principal pays a cost proportional to the probability of perfectly observing the realization of the shock. A CSV model cannot easily handle the case when the utility of the agent is unbounded below, this would allow the principal to inflict an expected unlimited punishment on the agent found lying even with an infinitesimal small monitoring probability, thus full insurance is sustainable paying an infinitesimal cost. To prevent this Aiyagari and Alvarez (1995a) study the case with bounded utility, and Wang (2005) restricts the principal to monitor with probability either 0 or 1. In this paper the planner can continuously change the correlation level of the signal from being perfectly correlated to random noise, this allows us to study the optimal allocation in all the intermediate cases between full information and complete asymmetric information. Also, given that perfect correlation is infinitely costly for the principal, we can look at a wide range of utility functions, not necessarily bounded below.

The paper is organized as follows: in section 9 studies the environment with an exogenously fixed amount of correlation. The long and short run effects are studied. In section 10 the correlation is treated as an endogenous variable. Section 11 concludes.

9 Exogenous Quality of Information

There is a single agent. The agent is characterized by a taste parameter $\theta \in \{\theta_1 \dots \theta_H\} = \Theta$ with H finite, Θ is ordered so that $\theta_i > \theta_j$ iff $i > j$. In each period the agent receives a taste shock with probability π, the probability distribution on Θ. Taste shocks are i.i.d. over time.

Preferences of the agent are defined over consumption and taste shocks according to $u : \mathbb{R}_+ \times \Theta \to \mathbb{R}$, future utility is discounted at rate β. Assume that $u(\cdot, \theta)$ is strictly concave and strictly increasing for each $\theta \in \Theta$. Also for each $b \in \mathbb{R}_+$, $u(b, \cdot)$ and $u'(b, \cdot)$ are strictly decreasing functions.[28] Also define $\underline{u} = \inf_x\{\min \theta[u(x, \theta)]\}$, note that $\underline{u}$ is not restricted to be finite.

At time zero the agent chooses a reporting strategy $\sigma = \{\sigma_t\}_{t=0}^{\infty}$ such that for all t, $\sigma_t : \Theta \to \Theta$. Denote with σ_t^* the truth telling strategy, that is $\sigma_t^*(\theta) = \theta$ for all $\theta \in \Theta$ and for all t.

Given a realized θ the principal also observes a signal $\tilde{\theta} \in \Theta$, distributed with conditional probability $\tilde{\pi}$. The signal is informative, that is $\tilde{\pi}(\tilde{\theta} = \theta_i | \theta = \theta_i) > \tilde{\pi}(\tilde{\theta} = \theta_j | \theta = \theta_i)$ for all $i \neq j$ and $i, j = 1 \dots N$. The signal is not observed by the agent. Signals are i.i.d. over time. A history up to time t for the principal is denoted

[28] Under this assumption, in an environment with incomplete information first best is not sustainable, the agent will always have incentive to misreport to a low θ.

as h^t, for all $t > 0$ is constructed as follows $h^t = (h^{t-1}, h_t)$, with $h_t = (\theta_t, \tilde{\theta}_t)$ note that h_t combines both signal and report also $h_0 = \emptyset$. Denote $\theta^t = (\theta_1, \ldots, \theta_t)$ and define similarly $\tilde{\theta}^t$ respectively the history of states and of reports up to time t.

At time t, contingent on h^t, the principal assigns consumption b (the transfer of the principal is the agent's only source of income.) The principal is required to provide a lifetime utility level equal to v to the agent. Consider $v > \underline{u}$ as a parameter that exogenously characterizes the agent. The planner problem is to minimize resources to provide level v of promised utility in an incentive compatible way:[29]

$$\min_{\{b_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \sum_{\theta_t \in \Theta} \pi(\theta_t) \sum_{\tilde{\theta}_t \in \Theta} \tilde{\pi}(\tilde{\theta}_t|\theta_t) b_t(\sigma_t^*(\theta^t), \tilde{\theta}^t), \tag{65}$$

subject to promise keeping

$$\sum_{t=0}^{\infty} \beta^t \sum_{\theta_t \in \Theta} \pi(\theta_t) \sum_{\tilde{\theta}_t \in \Theta} \tilde{\pi}(\tilde{\theta}_t|\theta_t) u(b(\sigma_t^*(\theta^t), \tilde{\theta}^t), \theta_t) = v, \tag{66}$$

and incentive compatibility

$$\sum_{t=0}^{\infty} \beta^t \sum_{\theta_t \in \Theta} \pi(\theta_t) \sum_{\tilde{\theta}_t \in \Theta} \tilde{\pi}(\tilde{\theta}_t|\theta_t) u(b(\sigma_t^*(\theta^t), \tilde{\theta}^t), \theta_t) \geq$$

$$\sum_{t=0}^{\infty} \beta^t \sum_{\theta_t \in \Theta} \pi(\theta_t) \sum_{\tilde{\theta}_t \in \Theta} \tilde{\pi}(\tilde{\theta}_t|\theta_t) u(b(\sigma_t(\theta^t), \tilde{\theta}^t), \theta_t), \quad \forall \sigma \in \Sigma. \tag{67}$$

The next step is to derive a recursive formulation. The state and the signal received by the principal are i.i.d. over time, also all actions are deterministic. So in every period t the principal's action will be contingent only on h_t the value of continuation

[29] In appendix 12.1 it is shown that this environment does not require the use of lotteries to overcome the non convexity introduced with the incentive constraint.

utility w, this value will summarize previous outcomes and will be used as the state variable in the planner problem. The principal problem in recursive form is

$$P(v) = \min_{\{b,w\}} \sum_{\theta \in \Theta} \pi(\theta) \sum_{\tilde{\theta} \in \Theta} \tilde{\pi}(\tilde{\theta}|\theta)[b(\theta,\tilde{\theta}) + \beta P(w(\theta,\tilde{\theta}))] \tag{68}$$

$$s.t. \quad \sum_{\theta \in \Theta} \pi(\theta) \sum_{\tilde{\theta} \in \Theta} \tilde{\pi}(\tilde{\theta}|\theta)[u(b(\theta,\tilde{\theta}),\theta) + \beta w(\theta,\tilde{\theta})] = v \tag{69}$$

$$\begin{aligned} &\sum_{\tilde{\theta} \in \Theta} \tilde{\pi}(\tilde{\theta}|\theta)[u(b(\theta,\tilde{\theta}),\theta) + \beta w(\theta,\tilde{\theta})] \geq \\ &\sum_{\tilde{\theta} \in \Theta} \tilde{\pi}(\tilde{\theta}|\theta)[u(b(\hat{\theta},\tilde{\theta}),\theta) + \beta w(\hat{\theta},\tilde{\theta})], \quad \forall \theta, \hat{\theta} \in \Theta, \end{aligned} \tag{70}$$

and non negativity on b. Let λ be the lagrange multiplier on the promise keeping constraint (69), and $\mu(\theta,\hat{\theta})$ the multipliers on the incentive constraints (70) for all $\theta, \hat{\theta} \in \Theta$. With the non negativity constraint on b, the value function is bounded below by zero. If signals are not perfectly correlated with the shock we have

Theorem 1. *1. P is differentiable.*

2. $P' > 0$ and $P'' > 0$.

3. $\lim_{w \to \underline{u}} P'(w) = 0$ and $\lim_{w \to \infty} P'(w) = \infty$.

Proof. Thomas and Worrall (1990b). ∎

The full information solution is given by solving (68) subject only to (69). The solution is given by

$$w(\theta) = v; \quad u'(b(\theta),\theta) = u'(b(\hat{\theta}),\hat{\theta}), \qquad \forall\, \theta, \hat{\theta} \in \Theta$$

If the taste shock is private information of the agent, the first best solution can still be sustained if the principal receives a perfectly correlated signal, in this case for all $\theta \in \Theta$, $\tilde{\pi}(\tilde{\theta}|\theta) = 1$ if $\tilde{\theta} = \theta$. To see this is note that the principal can make payments contingent only on the signal and not on the report. Also the first best is sustainable in this case since now the principal can administer the maximum punishment to an agent found lying without harming him in other realization of the state, that is:

Lemma 1. *With perfectly correlated signal, without loss of generality for all v and for all $\theta \in \Theta$, $b(\hat{\theta}, \theta) = 0$ and $w(\hat{\theta}, \theta) = \underline{u}$ if $\hat{\theta} \neq \theta$.*

When the signal is not perfectly correlated with the shock, is never optimal to inflict the maximum punishment, since now with positive probability the agent that is reporting truthfully is also punished. In general there is always some punishment, for all $\theta \in \Theta$ $w(\theta, \theta) \geq w(\hat{\theta}, \theta)$ and $b(\theta, \theta) \geq b(\hat{\theta}, \theta)$ as will be shown in Lemma 4.

Lemma 2. $\mu(\theta, \theta_H) = 0$ *for all* $\theta \in \Theta$.

Given the above results we can characterize the optimal contract.

Lemma 3. *Necessary conditions for an entirior optimal contract are*

$$w(\theta_H, \tilde{\theta}) = w_H, \quad b(\theta_H, \tilde{\theta}) = b_H, \quad \forall\ \tilde{\theta} \in \Theta$$

Proof. The envelope condition gives $P'(v) = \lambda$, and the first order condition are

$$-\pi(\theta)\tilde{\pi}(\tilde{\theta}|\theta) + \pi(\theta)\tilde{\pi}(\tilde{\theta}|\theta)u'(b(\theta, \tilde{\theta}), \theta)\lambda +$$

$$+\sum_{\hat{\theta}\in\Theta} \mu(\theta, \hat{\theta})\tilde{\pi}(\tilde{\theta}|\theta)u'(b(\theta, \tilde{\theta}), \theta) - \sum_{\hat{\theta}\in\Theta} \mu(\hat{\theta}, \theta)\tilde{\pi}(\tilde{\theta}|\hat{\theta})u'(b(\theta, \tilde{\theta}), \hat{\theta}) = 0 \quad \forall\ \theta, \tilde{\theta} \qquad (71)$$

$$-\pi(\theta)\tilde{\pi}(\tilde{\theta}|\theta)P'(w(\theta,\tilde{\theta}))+\lambda\pi(\theta)\tilde{\pi}(\tilde{\theta}|\theta)+\sum_{\hat{\theta}\in\Theta}\mu(\theta,\hat{\theta})\tilde{\pi}(\tilde{\theta}|\theta)-\sum_{\hat{\theta}\in\Theta}\mu(\hat{\theta},\theta)\tilde{\pi}(\tilde{\theta}|\hat{\theta}) = 0 \,\forall\, \theta,\tilde{\theta} \tag{72}$$

rearranging (71)

$$u'(b(\theta,\tilde{\theta}),\theta)\left[\pi(\theta)\tilde{\pi}(\tilde{\theta}|\theta)\lambda+\sum_{\hat{\theta}\in\Theta}\mu(\theta,\hat{\theta})\tilde{\pi}(\tilde{\theta}|\theta)\right] = \pi(\theta)\tilde{\pi}(\tilde{\theta}|\theta)+\sum_{\hat{\theta}\in\Theta}\mu(\hat{\theta},\theta)\tilde{\pi}(\tilde{\theta}|\hat{\theta})u'(b(\theta,\tilde{\theta}),\hat{\theta}) \tag{73}$$

If $\theta = \theta_H$ then $\mu(\theta,\theta_H) = 0$ for all θ so that

$$u'(b(\theta_H,\tilde{\theta}),\theta_H) = \frac{\pi(\theta_H)}{\pi(\theta_H)\lambda+\sum_{\hat{\theta}\in\Theta}\mu(\theta,\hat{\theta})}$$

The highest level of consumption is independent of the signal $\tilde{\theta}$, similarly from (72)

$$P'(w(\theta,\tilde{\theta})) = \lambda+\frac{1}{\pi(\theta)\tilde{\pi}(\tilde{\theta}|\theta)}\left[\tilde{\pi}(\tilde{\theta}|\theta)\sum_{\hat{\theta}\in\Theta}\mu(\theta,\hat{\theta})-\sum_{\hat{\theta}\in\Theta}\mu(\hat{\theta},\theta)\tilde{\pi}(\tilde{\theta}|\hat{\theta})\right] \tag{74}$$

If $\theta = \theta_H$ similarly as before

$$P'(w(\theta_H,\tilde{\theta})) = \lambda+\frac{1}{\pi(\theta_H)}\sum_{\hat{\theta}\in\Theta}\mu(\theta_H,\hat{\theta})$$

here the continuation utility is independent on the realization of $\tilde{\theta}$. ∎

This is an expected result, since if nobody has incentive to lie to the highest state, an agent that reports high will be telling the truth. The signal has no direct effect on the allocation of the highest type. On the other hand if $\theta = \theta_{H-1}$(the second highest type) given that $\mu(\theta_H, \theta_{H-1})$ is greater than zero, we have

$$
\begin{aligned}
&P'(w(\theta_{H-1}, \tilde{\theta})) = \\
&= \lambda + \frac{1}{\pi(\theta_{H-1})\tilde{\pi}(\tilde{\theta}|\theta_{H-1})} \left[\tilde{\pi}(\tilde{\theta}|\theta_{H-1}) \sum_{\hat{\theta}<\theta_{H-1}} \mu(\theta_{H-1}, \hat{\theta}) - \mu(\theta_H, \theta_{H-1})\tilde{\pi}(\tilde{\theta}|\theta_H) \right] \\
&= \lambda + \frac{1}{\pi(\theta_{H-1})} \sum_{\hat{\theta}<\theta_{H-1}} \mu(\theta_{H-1}, \hat{\theta}) - \frac{\mu(\theta_H, \theta_{H-1})}{\pi(\theta_{H-1})} \frac{\tilde{\pi}(\tilde{\theta}|\theta_H)}{\tilde{\pi}(\tilde{\theta}|\theta_{H-1})}
\end{aligned}
$$

The continuation utility now depends on the realization of the signal trough the likelihood ration of $\tilde{\theta}$.

In the optimal contract some amount of punishment is always optimal, for all $\theta, \hat{\theta} \in \Theta$ $w(\theta, \theta) \geq w(\hat{\theta}, \theta)$ and $b(\theta, \theta) \geq b(\hat{\theta}, \theta)$. To see this I look at the simple case where $\Theta = \{\theta_L, \theta_H\}$ with $\theta_H > \theta_L$. Let $\pi(\theta_H) = \pi(\theta_L) = \pi$. For the signal $\tilde{\theta}$, define the following conditional probabilities: $\tilde{\pi}(\theta_H|\theta_H) = \tilde{\pi}(\theta_L|\theta_L) = \alpha$ and $\tilde{\pi}(\theta_H|\theta_L) = \tilde{\pi}(\theta_L|\theta_H) = 1 - \alpha$ (I am assuming that the conditional probabilities are symmetric). The signal being informative requires $\alpha \neq 1/2$. Also, without loss of generality we can restrict ourself to $\alpha \in (\frac{1}{2}, 1]$; since if $\alpha \in [0, \frac{1}{2})$ (the signal is negatively correlated with the state) then relabel the H signal as L and viceversa, and all the result will then follow. I now show the following:

Lemma 4. *Necessary conditions for the optimal contract are*

1. $w_H > v > w(\theta_L, \theta_L) > w(\theta_L, \theta_H)$;

2. $b(\theta_L, \theta_L) > b(\theta_L, \theta_H)$;

3. $E_{\tilde{\theta}}[b(\theta_L, \tilde{\theta})] > \theta_H$.

Proof. In appendix. ∎

The spreading out of promised utility will be one of the crucial ingredients in determining the lung run behavior of the allocation. What remains to establish immiseration, is an intertemporal relation on the evolution of promised utility, this relation is the Euler equation for marginal cost.

9.1 The martingale condition and misery

Combining the envelope condition with the first order condition for continuation utility (72) and summing over $\theta, \tilde{\theta}$, since

$$\sum_{\theta,\tilde{\theta},\hat{\theta}\in\Theta} \mu(\theta,\hat{\theta})\tilde{\pi}(\tilde{\theta}|\theta) = \sum_{\theta,\tilde{\theta},\hat{\theta}\in\Theta} \mu(\hat{\theta},\theta)\tilde{\pi}(\tilde{\theta}|\hat{\theta}), \tag{75}$$

we get

$$\sum_{\theta\in\Theta} \pi(\theta) \sum_{\tilde{\theta}\in\Theta} \tilde{\pi}(\tilde{\theta}|\theta) P'(w(\theta,\tilde{\theta})) = P'(v), \tag{76}$$

an Euler equation for the marginal cost.

Proposition 4. $\lim_{t\to\infty} w_t = \underline{u}$ *almost surely.*

Proof. The result follows directly from Thomas and Worrall (1990b). Let ν be a probability distribution on $\Theta \times \Theta$, from (76) we have

$$\sum_{(\theta,\tilde{\theta})} \nu(\theta,\tilde{\theta}) P'(w(\theta,\tilde{\theta})) = P'(v), \tag{77}$$

P' is a non negative martingale and from Doob convergence theorem it must converge almost surely. As shown in Thomas and Worrall (1990b) such convergence cannot be in the interior of the space of the feasible promised utilities since it would contradict the spreading out of continuation utilities (Lemma 4 in this paper). ∎

9.2 Short run behavior of contracts

The focus of the previous section was on the long run behavior of the agent continuation utility. This section characterizes the behavior of promised utility further, in particular I now look at the short run behavior of the continuation utility. It will be shown that the behavior of promised utility will be qualitatively unaffected by the addition of a correlated signal.

This section extends the analysis of Rogerson (1985b), in that case the behavior of expected wage was studied in a two period model with asymmetric information. The first step is to introduce a quantity that summarizes the expected drift of promised utility

Definition 1. *The expected drift (Δ) is defined as:* $\Delta(v) = E[w(\theta, \tilde{\theta})] - v$

Complete insurance requires that the expected drift is $\Delta^{FI}(v) = E_\theta[w(\theta)] - v = 0$ for all values of v.

To characterize the drift in this framework for tractability as before I look at the case where $\Theta = \{\theta_L, \theta_H\}$. From the first order equation derived in the previous section, the expected continuation utility for a given shock realization evolves according to the following equations

$$E_{\tilde{\theta}}[P'(w(\theta_H, \tilde{\theta}))] = P'(v) + \frac{1}{\pi(\theta_H)}[\mu(\theta_H, \theta_L) - \mu(\theta_L, \theta_H)] = P'(v) + \frac{1}{\pi(\theta_H)}A, \quad (78)$$

$$E_{\tilde{\theta}}[P'(w(\theta_L, \tilde{\theta}))] = P'(v) + \frac{1}{\pi(\theta_L)}[\mu(\theta_L, \theta_H) - \mu(\theta_H, \theta_L)] = P'(v) - \frac{1}{\pi(\theta_L)}A, \quad (79)$$

where $A(v) = \mu(\theta_H, \theta_L) - \mu(\theta_L, \theta_H)$. I will now prove under which conditions there is a negative drift; first I show that $A \neq 0$, then show how this result together with a convex P' delivers the result.

Lemma 5. *For all $\alpha \in (0,1)$, $A \neq 0$*

Proof. A can be zero only if both constraint are binding and the multipliers are equal, or if both constraint are not binding. Neither of these cases can occur.

Step 1: *The upward constraint is not binding*

Proven in Lemma 2.

Step 2: *At least one constraint has to be binding*

Suppose there exist a value of α^* such that both multipliers on the incentive constraint are zero. In the case of the H incentive constraint not binding we have $w(\theta_L, \theta_L) = w(\theta_H, \theta_L) = w(\theta_H, \theta_L) = w(\theta_H, \theta_H)$, for the H-IC to hold we then must have

$$\alpha\ u(b(\theta_H, \theta_H), \theta_H) + (1-\alpha)u(b(\theta_H, \theta_L), \theta_H) \geq$$

$$\alpha\ u(b(\theta_L, \theta_H), \theta_H) + (1-\alpha)u(b(\theta_L, \theta_L), \theta_H), \quad (80)$$

substituting the first best allocation into (80) we get $u(b(\theta_L, \theta_L), \theta_L) \geq u(b(\theta_L, \theta_L), \theta_H)$ which (given the assumptions on u) is a contradiction for all values of α. ∎

An intuition for the previous result is the following: the only way to have the IC not binding is to have high enough punishment for agent thought to be lying. But in order for this punishment to be optimal, the multipliers have to be greater than zero else consumption will be smoothed as in the first best case. The shape of value function can now be characterized, define

$$\Gamma(x,\theta) = \frac{u'''(x,\theta)u'(x,\theta)}{u''^2}. \tag{81}$$

Assumption 2. *Let u such that $\Gamma(x,\theta) < 3$ for all values of x and $\theta \in \Theta$.*

If $u(x,\theta) = f(\theta)u_c(x)$ then for a large class of utility functions u_c, $\Gamma(x,\theta)$ is independent of (x,θ), in particular if $u_c(x) = \log(x)$ then $\Gamma = 2$, for CARA utility $(u_c(x) = -\frac{1}{a}e^{-ax})$ $\Gamma = 1$, for CRRA utility $(u_c(x) = \frac{x^{1-\sigma}}{1-\sigma})$ then $\Gamma = \frac{1+\sigma}{\sigma}$.

Lemma 6. *If u satisfies Assumption 1, then P' is strictly convex.*

Proof. Let $c(v) = u^{-1}(v)$. The value function is given by $P(v) = Kc(v)$, for some constant $K > 0$ (a proof of this statement can be found in Atkeson and Lucas (1992b). We then have the following

$$c'(x) = \frac{1}{u'(c(x))}, \qquad c''(x) = -\frac{u''(c(x))}{u'(c(x))^3} \tag{82}$$

and

$$c'''(x) = \frac{c'(x)}{[u'(c(x))]^4}[-u'''(c(x))u'(c(x)) + 3u''(c(x))^2] \tag{83}$$

Under Assumption 1 and our guess on P the result follows. Note that $c'' > 0$ and thus P' is increasing. ∎

Theorem 2. *For all* $0 < \alpha < 1$ *and for all* v, $\Delta(v) < 0$

Proof. As before I only show the proof in the case $\alpha \in [\frac{1}{2}, 1)$. I first show this result in the case $\alpha = 1/2$ and then extend it for all remaining values of α. In the case $\alpha = 1/2$ the principal's allocation rules are only conditional on the report since the signal is just random noise. We then have

$$P'(w(\theta_H)) = P'(v) + \frac{1}{\pi(\theta_H)}A, \quad (84)$$

$$P'(w(\theta_L)) = P'(v) - \frac{1}{1-\pi(\theta_H)}A. \quad (85)$$

If $A = 0$ then $w(\theta_H) = w(\theta_L) = v$ so $\Delta(v) = 0$. If $A \neq 0$ summing the above two equation and using strict convexity of P', we have

$$\pi(\theta_H)P'(w(\theta_H)) + (1-\pi(\theta_H))P'(w(\theta_L)) = \quad (86)$$

$$P'(v) > P'(\pi(\theta_H)w(\theta_H) + (1-\pi(\theta_H))w(\theta_L)) = P'(E\theta[w]),$$

since P' is increasing, $E[w] < v$ and $\Delta < 0$.

For the case $\alpha \in (\frac{1}{2}, 1)$, equation (78) and (79) give an equivalent version of (76) for the two state case. The result then follows by strict convexity of P' (technically what happens is that a N state case can be thought as a $N * N$ case the moment signals are introduced.) ∎

The previous theorem holds for all values of $\alpha \in [\frac{1}{2}, 1)$; the result is unaffected if the principal receives a correlated signal or none at all.

Assumption 1, for example, is violated for a CRRA function when $\sigma < \frac{1}{2}$, with this utility function if follows that $\Delta(v) > 0$; this is not a contradiction with the immiseration result derived in the previous section: a positive expected drift is not enough to prevent promised utility reaching $\underline{u}$ almost surely, since in this case the lower bound of utility is finite (in a similar fashion as the gambler's ruin problem.[30]

10 Endogenous Quality of Information

In the previous section it was shown that an accurate but imprecise signal does not qualitatively change the long and short run behavior of promised utility. In this section we continue looking at the long run behavior of promised utility allowing the decision on the precision signal to be determined endogenously.

Through this section we assume that the taste shock can take only two values. The timing in this section is as follows, before the realization of the signal, conditional on the current value of promised utility and the report of the agent, the principal sets $\alpha(\theta) \in \left[\frac{1}{2}, 1\right]$ paying a cost $T(\alpha(\theta))$.

Assumption 3. *Let $T : \left[\frac{1}{2}, 1\right] \to \mathbb{R}_+$, a continuous, twice differentiable function with: $T\left(\frac{1}{2}\right) = 0$, $T'(\frac{1}{2}) = 0$, $T' > 0$, $T'' > 0$ and $\lim_{\alpha \to 1^-} T'(\alpha) = \infty$.*

The above assumption on T, as it will be shown, implies that is always optimal to choose an interior value of α. In this environment allowing the additional choice of α introduces a non convexity in the planner problem. The constraint set now contains terms that feature the product of two choice variables (for example the planner chooses at the same time the value of continuation utility and the probability of

[30]Thanks to Chris Phelan for suggesting this example.

delivering it, for example: $\alpha_H w(\theta_H, \theta_H)$). Given that choosing lotteries on the allocation might be Pareto improving, we allow the planner to randomize on the allocation and choice of α. The principal maximizing profits chooses the following lottery

$$\Pi(\alpha_H, \alpha_L, b(\theta, \tilde{\theta}), w(\theta, \tilde{\theta})) : \mathbb{R}^8 \times \left[\frac{1}{2}, 1\right]^2 \to [0,1]^{10}$$

To write the problem more compactly let $V(\theta_i, \theta_j, \theta_k) = u(b(\theta_i, \theta_j), \theta_k) + \beta w(\theta_i, \theta_j)$ (where θ_i denotes the the report of the agent, θ_j the realization of the public signal and θ_k the true state) also define $K(\theta_i, \theta_j) = b(\theta_i, \theta_j) + \beta P(w(\theta_i, \theta_j))$ for $i, j, k \in \{H, L\}$. The principal problem is given by

$$P(w) = \min_{\Pi} \quad \textstyle\sum_{\{\cdot\}} \Pi(\alpha_H, \alpha_L, b(\theta, \tilde{\theta}), w(\theta, \tilde{\theta}))\{ \tag{87}$$

$$\pi_H \left[\alpha_H K(\theta_H, \theta_H) + (1 - \alpha_H) K(\theta_H, \theta_L) + T(\alpha_H)\right] +$$

$$+\pi_L \left[\alpha_L K(\theta_L, \theta_L) + (1 - \alpha_L) K(\theta_L, \theta_H) + T(\alpha_L)\right]\},$$

subject to promise keeping

$$\sum_{\{\cdot\}} \Pi(\cdot)\{\pi_H \left[\alpha_H V(\theta_H, \theta_H, \theta_H) + (1 - \alpha_H) V(\theta_H, \theta_L, \theta_H)\right] +$$
$$+\pi_L \left[\alpha_L V(\theta_L, \theta_L, \theta_L) + (1 - \alpha_L) V(\theta_L, \theta_H, \theta_L)\right]\} = w, \tag{88}$$

incentive constraints

$$\sum_{\{\cdot\}} \Pi(\cdot)\{\alpha_H V(\theta_H,\theta_H,\theta_H) + (1-\alpha_H)V(\theta_H,\theta_L,\theta_H)\} \geq$$
$$\sum_{\{\cdot\}} \Pi(\cdot)\{\alpha_L V(\theta_L,\theta_H,\theta_H) + (1-\alpha_L)V(\theta_L,\theta_L,\theta_H)\}, \qquad IC_H, \tag{89}$$
$$\sum_{\{\cdot\}} \Pi(\cdot)\{\alpha_L V(\theta_L,\theta_L,\theta_L) + (1-\alpha_L)V(\theta_L,\theta_H,\theta_L)\} \geq$$
$$\sum_{\{\cdot\}} \Pi(\cdot)\{\alpha_H V(\theta_H,\theta_L,\theta_L) + (1-\alpha_H)V(\theta_H,\theta_H,\theta_L)\}, \qquad IC_L, \tag{90}$$

and additional constraints on the probability distribution

$$\sum_{\{\cdot\}} \Pi(\cdot) = 1, \qquad \Pi(\cdot) \geq 0. \tag{91}$$

The cost function of the planner satisfies the following properties

Proposition 5.

1. *$P(\cdot)$ is continuous, increasing and convex*

2. *P is differentiable*

In equilibrium the planner will not use any randomization on the optimal allocation.

Definition 2. Π *is degenerate if there exist an* $x \in \mathbb{R}^8 \times [\frac{1}{2}, 1]^2$ *such that* $\Pi(x) = 1$.

Proposition 6. *Let* Π^* *be the solution to the planner problem (87),* Π^* *is degenerate.*

Proof. In the appendix 12.1, related to the section with exogenous α, it was argued that for any value of α randomization is not optimal for the remaining allocation of consumption. The same argument also holds in this case, so that after any outcome of the value of α will never be optimal to randomize on the remaining allocation.

What we need to show is that the planner will not find optimal to randomize on the choice of α. The first case we look at, rules out randomization of α if the planner does not condition the value of current consumption and continuation utility on the realization of α.

- *It is not optimal to randomize on the choice of α if the allocation of consumption and continuation utility is independent of α.* Suppose that in the solution of the planner problem α_1 and α_2 so that $\Pi^*(\alpha) > 0$ iff $\alpha \in \{\alpha_1, \alpha_2\}$.[31] Consider the following alternative $\tilde{\Pi}(\hat{\alpha}) = 1$ with $\hat{\alpha} = \Pi^*(\alpha_1)\alpha_1 + \Pi^*(\alpha_2)\alpha_2$. Such a choice of α leaves unchanged the constraint set; to see this consider, for example, the left hand side of IC_H (equation (89)):

$$\sum_{\alpha_H=\alpha_{H1},\alpha_{H2}} \Pi^*(\alpha_H)\{\alpha_H V(\theta_H,\theta_H,\theta_H) + (1-\alpha_H)V(\theta_H,\theta_L,\theta_H)\}, \tag{92}$$

substituting the definition of $\tilde{\Pi}$ and $\hat{\alpha}$ in the above

$$\begin{aligned}&\tilde{\Pi}(\hat{\alpha})\{\hat{\alpha}V(\theta_H,\theta_H,\theta_H) + (1-\hat{\alpha})V(\theta_H,\theta_L,\theta_H)\} = \\ &\Pi^*(\alpha_1)\alpha_1 V(\theta_H,\theta_H,\theta_H) + \Pi^*(\alpha_2)\alpha_2 V(\theta_H,\theta_H,\theta_H) + \\ &[1 - \Pi^*(\alpha_1)\alpha_1 - \Pi^*(\alpha_2)\alpha_2]\, V(\theta_H,\theta_L,\theta_H),\end{aligned}$$

given that $\Pi^*(\alpha_1) + \Pi^*(\alpha_2) = 1$ we get back (92). Note that the cost of implementing $\hat{\alpha}$ is now lower since the only change in the cost function of the planner is in the cost of choosing $\hat{\alpha}$, but from convexity of T we have

$$T\left(\Pi^*(\alpha_1)\alpha_1 + \Pi^*(\alpha_2)\alpha_2\right) < \sum_{\alpha_H=\alpha_{H1},\alpha_{H2}} \Pi^*(\alpha_H)T(\alpha_H) \tag{93}$$

[31] The same argument follows if there are more points in the support of Π^*.

thus Π^* cannot be the optimal lottery.

- Suppose now the planner randomizes on α and makes the remaining allocation $x \in \mathbb{R}^8 \times [\frac{1}{2}, 1]$ contingent on the realization of α. Suppose that for a report $\theta \in \{\theta_H, \theta_L\}$ there are two values $\alpha_1(\theta)$ and $\alpha_2(\theta)$ which are chosen with positive probability respectively (π_1, π_2). We now show how it is possible to choose an alternative value $\tilde{\alpha}$ chosen with probability 1. Which delivers the same on equilibrium value, lowers costs and preserves on equilibrium utility level. Without loss of generality let $\alpha_1 > \alpha_2$. As notation let $V^{\alpha_i}, V^{-\alpha_i}$ for $i = 1, 2$ be the utility value delivered to the agent under realization of α_i, respectively for the signal returning the true value or the opposite. Under this notation the expected utility for agent of type θ is

$$w(\theta) = \pi_1 \alpha_1 V^{\alpha_1} + \pi_1(1 - \alpha_1)V^{-\alpha_1} + \pi_2 \alpha_2 V^{\alpha_2} + \pi_2(1 - \alpha_2)V^{-\alpha_2}.$$

The first case we look at is the one in which $V^{\alpha_1} > V^{\alpha_2} > V^{-\alpha_2} > V^{-\alpha_1}$. In this case let $\tilde{\alpha}$ so that $\tilde{\alpha}V^{\alpha_1} + (1 - \tilde{\alpha})V^{-\alpha_2} = w(\theta)$, that is the planner will now randomize between the values V^{α_1} and $V^{-\alpha_2}$ using the signal as a randomization instrument; this implies

$$\tilde{\alpha}V^{\alpha_1} + (1-\tilde{\alpha})V^{-\alpha_2} = \pi_1\alpha_1 V^{\alpha_1} + \pi_1(1-\alpha_1)V^{-\alpha_1} + \pi_2\alpha_2 V^{\alpha_2} + \pi_2(1-\alpha_2)V^{-\alpha_2},$$

this implies

$$\tilde{\alpha}[V^{\alpha_1} - V^{-\alpha_2}] = \pi_1\alpha_1[V^{\alpha_1} - V^{-\alpha_1}] + \pi_2\alpha_2[V^{\alpha_2} - V{-\alpha_2}] + \pi_1[V^{-\alpha_1} - V^{-\alpha_2}],$$

so that

$$\tilde{\alpha} = \pi_1\alpha_1 + \pi_2\alpha_2 + $$
$$+ \frac{\{\pi_1\alpha_1(V^{-\alpha_2} - V^{-\alpha_1}) + \pi_2\alpha_2(V^{\alpha_2} - V^{\alpha_1}) + \pi_1(V^{-\alpha_1} - V^{-\alpha_2})\}}{V^{\alpha_1} - V^{-\alpha_2}}. \quad (94)$$

In order for the $\tilde{\alpha}$ to be less costly it is sufficient that $\tilde{\alpha} \leq \pi_1\alpha_1 + \pi_2\alpha_2$. This will hold if the term in brackets in (94) is negative. We have that

$$\pi_1\alpha_1(V^{-\alpha_2} - V^{-\alpha_1}) + \pi_2\alpha_2(V^{\alpha_2} - V^{\alpha_1}) + \pi_1(V^{-\alpha_1} - V^{-\alpha_2}) =$$
$$\pi_1(1-\alpha_1)(V^{-\alpha_1} - V^{-\alpha_2}) + \pi_2\alpha_2(V^{\alpha_2} - V^{\alpha_1}) < 0.$$

The other relevant case is when $V^{-\alpha_1} > V^{-\alpha_2}$. In this case $\tilde{\alpha}$ is constructed using the convex combination of utility levels of V^{α_1} and $V^{-\alpha_1}$, so that

$$\tilde{\alpha}V^{\alpha_1} + (1-\tilde{\alpha})V^{-\alpha_1} = \pi_1\alpha_1V^{\alpha_1} + \pi_1(1-\alpha_1)V^{-\alpha_1} + \pi_2\alpha_2V^{\alpha_2} + \pi_2(1-\alpha_2)V^{-\alpha_2}$$

this implies

$$\tilde{\alpha} = \pi_1\alpha_1 + \frac{\pi_2}{(V^{\alpha_1} - V^{-\alpha_1})}\left\{-V^{-\alpha_1} + \alpha_2V^{\alpha_2} + (1-\alpha_2)V^{-\alpha_2}\right\}$$

in order for $\tilde{\alpha}$ to use less resources we need

$$\frac{1}{(V^{\alpha_1} - V^{-\alpha_1})}\left\{-V^{-\alpha_1} + \alpha_2V^{\alpha_2} + (1-\alpha_2)V^{-\alpha_2}\right\} < \alpha_2$$
$$\alpha_2\frac{V^{\alpha_2} - V^{-\alpha_2}}{V^{\alpha_1} - V^{-\alpha_1}} + \frac{V^{-\alpha_2} - V^{-\alpha_1}}{V^{\alpha_1} - V^{-\alpha_1}} < \alpha_2$$

which is the case since $V^{\alpha_2} - V^{-\alpha_2} < V^{\alpha_1} - V^{-\alpha_1}$ and $V^{-\alpha_2} < V^{-\alpha_1}$.

Finally note that the incentive constraints holds after the planner deviates to $\tilde{\alpha}$ as long as the GDRA property for u holds.[32]

[32]See appendix 12.1.

$\blacksquare$

Given that lotteries will not be used in equilibrium, the following proposition allows to further simplify the planner problem

Proposition 7.

(a) *If* $\alpha_\theta = \frac{1}{2}$ *then for all* $\tilde{\theta}$, $b(\theta, \tilde{\theta}) = b(\theta)$ *and* $w(\theta, \tilde{\theta}) = w(\theta)$.

(b) *If* $\alpha\theta > \frac{1}{2}$ *then* $b(\theta, \theta) > b(\theta, \tilde{\theta})$ *and* $w(\theta, \theta) > w(\theta, \tilde{\theta})$

(c) *If* $\alpha_L > \frac{1}{2}$ *if and only if* IC_H *binds*

Proof. In appendix. $\blacksquare$

Given this proposition rewrite the planner problem

$$P(v) = \min_{\{\alpha_L, b, w\}} \quad \pi_H K(\theta_H, \theta_H) + \pi_L \left\{ \alpha_L K(\theta_L, \theta_L) + (1 - \alpha_L) K(\theta_L, \theta_H) + T(\alpha_L) \right\}$$

subject to promise keeping and the incentive compatibility constraint

$$V(\theta_H, \theta_H, \theta_H) \geq \alpha_L V(\theta_L, \theta_H, \theta_H) + (1 - \alpha_L) V(\theta_L, \theta_L, \theta_H) \tag{95}$$

necessary condition for an interior solution for α_L is given by are

$$\begin{aligned} T'(\alpha_L) &= \pi_L \left[K(\theta_L, \theta_H) - K(\theta_L, \theta_L) \right] + \lambda \left[V(\theta_L, \theta_L, \theta_L) - V(\theta_L, \theta_H, \theta_L) \right] + \\ &+ \mu(\theta_H, \theta_L) \left[V(\theta_L, \theta_L, \theta_H) - V(\theta_L, \theta_H, \theta_H) \right] \end{aligned} \tag{96}$$

From the above it follows that the amount of resources used to increase the precision of the signal is directly related with the magnitude of the multiplier on the incentive constraint: stronger the incentive problem, higher will be the precision of the signal. An immediate implication of the immiseration result shown in the next section will be that at the limit no resources will be spent on improving the precision of the signal.

The feature of progressively less information being acquired on the value of the private shock is consistent with dynamic costly state verification models, for example in Aiyagari and Alvarez (1995a), it is shown that there is no monitoring for low enough values of continuation utility. The key difference of this paper with Aiyagari and Alvarez (1995a) and related literature on dynamic costly state verification models, is what type of signals are available to the planner: in CSV models the planner has the ability acquiring perfect information paying a finite cost, this implies that there exist a non empty interval of continuation utilities starting from $\underline{u}$, where there is no monitoring. In this paper on the other hand, there is no monitoring only for $u = \underline{u}$. If $\alpha_L = \frac{1}{2}$ for any other $u < \underline{u}$, using point (c) of the previous lemma we would reach a contradiction since it would imply a non binding incentive constraint.

Lemma 7. $\lim_{w \to \underline{u}} \alpha(\theta_L) = \frac{1}{2}$

Proof. **Case 1:** *$\underline{u}$ is finite*

By inspection of the promise keeping constraint, to deliver this level of promised utility (for interior value of α) transfers from the principal have to be zero and the level of continuation utility for each state-signal realization have to be equal $\underline{u}$. In this case the IC is not binding. By the previous lemma the result follows.

Case 2: $\underline{u} = -\infty$

In the limit as $v \to \underline{u}$, this level of promised utility can be delivered either by setting any of the transfer to zero, setting any continuation utility to $\underline{u}$ or by doing any combination of the previous. Cost minimization then requires setting transfers to zero and continuation utilities to zero; which implies $\alpha(\theta_L) = \frac{1}{2}$ ∎

As a direct implication of the previous lemma we have that $\lim_{v \to \underline{u}} ||P(v) - P^{FB}(v)|| = 0$. The results of this section imply that setting α endogenously does not introduce any qualitative difference in the planners problem with respect to the case with exogenous α, thus an immiseration result will still hold and can be shown by extending the martingale condition shown in section 9.1.

10.1 The martingale condition and immiseration

The corresponding version of (76) is given by

$$\sum_{\theta \in \Theta} \pi(\theta) \sum_{\tilde{\theta} \in \Theta} \tilde{\pi}(\tilde{\theta}|\theta) P'(w(\theta, \tilde{\theta})) = P'(v). \tag{97}$$

With α chosen endogenously, the martingale condition has a particular difference with respect to section 9.1. The conditional probabilities are now chosen endogenously, so now we must redefine with respect to which filtration P' is a martingale. To do this I borrow from the notation of Golosov, Kocherlakota, and Tsyvinski (2003). An event in the probability space will now be an entire vector of draws for θ and $\tilde{\theta}$. Let the probability space be $\Omega = \Theta^T \times \Theta^T$ with $T = \infty$. In defining the probability distribution ν on Ω we must be careful that the probability used to draw a particular $\tilde{\theta}_t$ has to be consistent with the entire previous draws of the state θ^t, previous signals $\tilde{\theta}^{t-1}$ and the optimal behavior of the principal. Let $\tilde{\pi}^*(\tilde{\theta}_t|\theta^t, \tilde{\theta}^{t-1})$ be the optimal choice of the principal for the conditional probability (the choice of α) having observed reports θ^t and signals $\tilde{\theta}^{t-1}$. Then the probability of drawing $(\theta^T, \tilde{\theta}^T) \in \Omega$ is given by

$$\nu(\theta^T, \tilde{\theta^T}) = \pi(\theta_1)\tilde{\pi}^*(\tilde{\theta}_1|\theta_1) \prod_{t=2}^{T} \pi(\theta_t)\tilde{\pi}^*(\tilde{\theta}_t|\theta^t, \tilde{\theta}^{t-1}). \tag{98}$$

Rewrite (97) as

$$\sum_{\theta^T, \tilde{\theta}^T} \nu(\theta^T, \tilde{\theta^T}) P'(w(\theta, \tilde{\theta})) = P'(v). \tag{99}$$

Given this probability structure now define a filtration as $\mathcal{F}_t = \mathcal{B}(h^t)$. The Borel sets on the entire history of reports and signals up to time t. Now from (98) P' is a martingale with respect to $\mathcal{F}$ and we can then apply the procedure described in Thomas and Worrall (1990b) (as shown in the previous section) to conclude the long run immiseration of the agent.

11 Conclusion

The objective of this paper is to study the constraint efficient allocation when the principal has some additional, but imprecise information about the true realization of the private shock. The additional information is modeled by a signal correlated with the realized shock. In this environment, either with additional information determined endogenously or fixed exogenously, the continuation utility of the agent converges to the lower bound almost surely, as this happens the resources used by the principal to monitor the agent go to zero. This feature is shared by environments with complete asymmetric information. The technical reason for this result is that the inter-temporal Euler equation and the shape of the planners cost function in equilibrium are qualitatively unaffected by the presence of an additional noisy signal.

In this paper the quantitative implication of additional information available to the principal are not explored. Different correlation levels affect the speed of convergence to misery, together with the size of the intertemporal distortion on the optimal allocation. Current research is focused in quantitatively estimating the size of these distortion, together with determining the implication on the tax schedule that implements the optimal allocation.

12 Appendix

12.1 Optimality of degenerate lotteries

The planner problem in (65) although being a non convex program does not rely on the use of lotteries (or it uses only a degenerate one); this simplification, following Kehoe, Levine, and Prescott (2002) (KLP), is not restrictive. In KLP it is shown that if the utility satisfies a decreasing risk aversion property then lotteries are not used in equilibrium. Their result also apply to our environment.

Definition 3. *The utility function u satisfies generalized decreasing risk aversion*

(GDRA) if $\forall\ \theta, \theta' \in \Theta$ *and* $\forall\ \varepsilon > 0$

$$u(\hat{x}(\theta') + \varepsilon, \theta) \leq \int_{x \in X} u(x(\theta), \theta) d\mu_x(x(\theta)), \tag{100}$$

where $\hat{x}$ *is the certainty equivalent of the lottery* μ_x *on* X*.*[33]

To apply the results of KLP to our environment we need to make sure that the *expectation* of u with respect to the signal satisfies GDRA.[34] Let

$$V(x(\theta), \theta) = \sum_{\tilde{\theta} \in \Theta} \pi(\tilde{\theta}|\theta) u(x(\theta, \tilde{\theta}), \theta). \tag{101}$$

Lemma 8. *If the utility function* u *satisfies GDRA, then the expected utility* V *satisfies GDRA.*

Proof. Suppose that V does not satisfies GDRA while u does. This implies that for some θ and θ'

$$\sum_{\tilde{\theta} \in \Theta} \pi(\tilde{\theta}|\theta) u(\hat{x}(\theta', \tilde{\theta}) + \varepsilon, \theta) > \sum_{\tilde{\theta} \in \Theta} \pi(\tilde{\theta}|\theta) \int u(x(\theta, \tilde{\theta}), \theta) d\mu_x(x(\theta, \tilde{\theta})),$$

so that for at least one value of $\tilde{\theta}$ we violate GDRA for u. ∎

This lemma enables us to apply directly the KLP result

Proposition 8 (KLP)**.** *If* u *satisfies GDRA then the solution to the planner problem allowing for randomization involves degenerate lotteries.*

Proof. The results follows from Theorem 9.1 in KLP and lemma 8. ∎

[33] Note that this definition of GDRA is slightly different than the one in KLP which only applies to environments with endowment shocks.

[34] This property for example, in environments with taste shocks or with endowment shocks is satisfied as soon as the utility function satisfies non increasing absolute risk aversion.

12.2 Proof of Lemma 4

Proof. 1. The envelope and first order conditions from (68) simplify to

$$
\begin{aligned}
P'(w(\theta_H,\theta_H)) &= P'(v) + \frac{1}{\pi}\mu(\theta_H,\theta_L), && (102)\\
P'(w(\theta_L,\theta_H)) &= P'(v) - \frac{1}{\pi}\frac{\alpha}{1-\alpha}\mu(\theta_H,\theta_L), && (103)\\
P'(w(\theta_H,\theta_L)) &= P'(v) + \frac{1}{\pi}\mu(\theta_H,\theta_L), && (104)\\
P'(w(\theta_L,\theta_L)) &= P'(v) - \frac{1}{\pi}\frac{1-\alpha}{\alpha}\mu(\theta_H,\theta_L). && (105)
\end{aligned}
$$

As noted before $w(\theta_H,\theta_H) = w(\theta_H,\theta_L) = w_H$. For all $\alpha > \frac{1}{2}$. Also $\frac{1-\alpha}{\alpha} < 1 < \frac{\alpha}{1-\alpha}$, so that $w(\theta_L,\theta_H) < w(\theta_L,\theta_L)$; from convexity of P we have $w_H > v > w(\theta_L,\theta_L) > w(\theta_L,\theta_H)$.

2. From the necessary first order conditions for b

$$
\begin{aligned}
u'(b(\theta_H,\theta_H),\theta_H) &= u'(b(\theta_H,\theta_L),\theta_H) = \frac{\pi}{\pi\lambda + \mu(\theta_H,\theta_L)} && (106)\\
u'(b(\theta_L,\theta_H),\theta_L)\lambda &= 1 + \frac{\alpha}{1-\alpha}\frac{1}{\pi}\mu(\theta_H,\theta_L)u'(b(\theta_L,\theta_H),\theta_H) && (107)\\
u'(b(\theta_L,\theta_L),\theta_L)\lambda &= 1 + \frac{1-\alpha}{\alpha}\frac{1}{\pi}\mu(\theta_H,\theta_L)u'(b(\theta_L,\theta_L),\theta_H) && (108)
\end{aligned}
$$

From (106), $b(\theta_H,\theta_H) = b(\theta_H,\theta_L) = b_H$. We now show that $b(\theta_L,\theta_H) < b(\theta_L,\theta_L)$. The result is easy to show in the multiplicative taste shock case, in this case (107) and (108) become

$$
u'(b(\theta_L,\theta_H)) = \frac{1}{\frac{\lambda}{\theta_L} - K\frac{\alpha}{1-\alpha}}, \qquad u'(b(\theta_L,\theta_L)) = \frac{1}{\frac{\lambda}{\theta_L} - K\frac{1-\alpha}{\alpha}}, \tag{109}
$$

where $K = \frac{1}{\pi}\mu(\theta_H,\theta_L)$. Given $\frac{\alpha}{1-\alpha} > 1 > \frac{1-\alpha}{\alpha}$, we get $b(\theta_L,\theta_L) > b(\theta_L,\theta_H)$.

3. From the binding incentive constraint we have

$$u(b_H, \theta_H) + \beta w_H = \alpha u(b(\theta_L, \theta_H), \theta_H) +$$

$$+(1-\alpha)u(b(\theta_L, \theta_L), \theta_L) + \beta[\alpha w(\theta_L, \theta_H) + (1-\alpha)w(\theta_L, \theta_L)]$$

Given $w_H > \alpha w(\theta_L, \theta_H) + (1-\alpha)w(\theta_L, \theta_L)$

$$u(b(\theta_H, \theta_H), \theta_H) < \alpha u(b(\theta_L, \theta_H), \theta_H) + (1-\alpha)u(b(\theta_L, \theta_L), \theta_L)$$

The result then follows from concavity of u.

∎

12.3 Proof of Proposition 7

Part (a) and (b) follow the same steps as for the case with fixed α. We look at part (c). If IC_H does not bind then the first best level allocation is incentive compatible which implies that no additional information is required thus $\alpha_L = 0$. Suppose now that IC_H binds, that is $\mu(\theta_H, \theta_L) > 0$ suppose by contradiction that $\alpha_L = \frac{1}{2}$, we will show that the marginal benefit of increasing by ε amount α_H will be larger than the marginal cost of doing so. Consider the following perturbation on the optimal allocation: increase α_L by ε, raise $w(\theta_L, \theta_L)$ by δ/β and decrease $w(\theta_L, \theta_H)$ by $\Gamma = \frac{\frac{1}{2}+\varepsilon}{\frac{1}{2}-\varepsilon}\frac{\delta}{\beta}$, where both ε and δ are arbitrarily small. This perturbation leaves the promise keeping constraint unchanged. The original constraint is given by

$$\pi_H V(\theta_H, \theta_H, \theta_H) + \pi_L \left[\frac{1}{2}V(\theta_L, \theta_L, \theta_L) + \frac{1}{2}V(\theta_L, \theta_H, \theta_L)\right] = w$$

notice that given point (a), $V(\theta_L, \theta_L, \theta_L) = V(\theta_L, \theta_H, \theta_L)$, under the perturbation the previous relation becomes

$$\pi_H V(\theta_H, \theta_H, \theta_H) +$$
$$+\pi_L \left[\left(\frac{1}{2} + \varepsilon \right) [V(\theta_L, \theta_L, \theta_L) + \delta] + \left(\frac{1}{2} - \varepsilon \right) \left[V(\theta_L, \theta_H, \theta_L) - \frac{\Gamma}{\beta} \right] \right] = (\text{d}110)$$

on the other hand the incentive constraint will be relaxed following the deviation

$$V(\theta_H, \theta_H, \theta_H) \geq \left(\frac{1}{2} + \varepsilon \right) \left[V(\theta_L, \theta_L, \theta_H) - \frac{\Gamma}{\beta} \right] + \left(\frac{1}{2} - \varepsilon \right) [V(\theta_L, \theta_H, \theta_H) + \delta] \tag{111}$$

where the inequality holds since

$$-\frac{\left(\frac{1}{2} + \varepsilon\right)^2}{\left(\frac{1}{2} - \varepsilon\right)} \delta + \left(\frac{1}{2} - \varepsilon \right) \delta < 0 \tag{112}$$

this implies that the perturbation, by relaxing the incentive constraint will have a marginal benefit equal to $\mu(\theta_H, \theta_L)$. To look at the cost needed for this deviation we look at the cost function. Let $P(w, \varepsilon)$ the cost function after increasing α by a small amount ε, we have that

$$P(w, \varepsilon) = P(w) + \pi_L \left(\frac{1}{2} + \varepsilon \right) [P(w(\theta_L, \theta_L) + \delta) - P(w(\theta_L, \theta_L))] + \tag{113}$$
$$\pi_L \left(\frac{1}{2} - \varepsilon \right) \left[P(w(\theta_L, \theta_L) - \frac{\Gamma}{\beta}) - P(w(\theta_L, \theta_L)) \right] + T \left(\frac{1}{2} + \varepsilon \right) - T \left(\frac{1}{2} - \varepsilon \right). \tag{114}$$

for small δ and ε, $P(w, \varepsilon) = P(w)$ since the term in brackets both are equal to $P'(w(\theta_L, \theta_L))$ and $T'\left(\frac{1}{2}\right) = 0$ by assumption. This implies that for small increase for α from $1/2$ there is no marginal cost but only the marginal benefit from relaxing the IC, contradicting the original allocation with $\alpha = 1/2$ being optimal.

References

Aguiar, M., and E. Hurst (2005): "Consumption versus Expenditure," *Journal of Political Economy*, 113(5), 919–948.

Ai, H., and F. Yang (2007): "Private Information, Limited Commitment, and Risk Sharing," Manuscript, Duke University and SUNY-Albany.

Aiyagari, R., and F. Alvarez (1995a): "Efficient dynamic monitoring of unemployment insurance claims," *Manuscript*.

Aiyagari, S., and F. Alvarez (1995b): "Stationary Efficient Distributions with Private Information and Monitoring: A Tale of Kings and Slaves," Manuscript, Federal Reserve Bank of Minneapolis.

Albanesi, S., and C. Sleet (2006): "Dynamic Optimal Taxation with Private Information," *Review of Economic Studies*, 73(1), 1–30.

Ales, L., and P. Maziero (2007): "Skill, Luck and Inequality: Private Information as a Source of Lifetime Inequality," Manuscript, University of Minnesota.

Arnott, R., and J. Stiglitz (1988): "Randomization with Asymmetric Information," *RAND Journal of Economics*, 19(3), 344–362.

Atkeson, A., and R. Lucas (1992a): "On Efficient Distribution with Private Information," *Review of Economic Studies*, 59(3), 427–453.

——— (1992b): "On Efficient Distribution with Private Information," *The Review of Economic Studies*, 59(3), 427–453.

Attanasio, O., and S. Davis (1996): "Relative Wage Movements and the Distribution of Consumption," *Journal of Political Economy*, 104(6), 1227–1262.

Attanasio, O., and N. Pavoni (2007): "Risk Sharing in Private Information Models with Asset Accumulation: Explaining the Excess Smoothness of Consumption," *NBER Working Paper*, 12994.

Badel, A., and M. Huggett (2006): "Interpreting Life-Cycle Inequality Patterns as Efficient Allocation: Mission Impossible?," Manuscript, Georgetown University.

Blundell, R., L. Pistaferri, and I. Preston (2006): "Consumption Inequality and Partial Insurance," Manuscript, University College London.

Bound, J., C. Brown, G. Duncan, and W. Rodgers (1994): "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data," *Journal of Labor Economics*, 12(3), 345–368.

Browning, M., L. Hansen, and J. Heckman (1999): "Micro Data and General Equilibrium Models," *in Handbook of Macroeconomics, vol. 1A, ed. J. Taylor and M. Woodford, Amsterdam: Noth-Holland.*

Cochrane, J. (1991): "A Simple Test of Consumption Insurance," *Journal of Political Economy*, 99(5), 957–976.

Deaton, A., and C. Paxson (1994): "Intertemporal Choice and Inequality," *Journal of Political Economy*, 102(3), 437–467.

Dynarski, S., and J. Gruber (1997): "Can Families Smooth Variable Earnings?," *Brookings Papers on Economic Activity*, 1997(1), 229–303.

Farhi, E., and I. Werning (2006): "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation," *Working paper 06-15, MIT*.

——— (2007): "Inequality and Social Discounting," *Journal of Political Economy*, 115(3), 365–402.

Gervais, M., and P. Klein (2006): "Measuring Consumption Smoothing in CEX Data," Manuscript, University of Western Ontario.

Golosov, M., N. Kocherlakota, and A. Tsyvinski (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70(3), 569–587.

Golosov, M., and A. Tsyvinski (2006): "Designing Optimal Disability Insurance: A Case for Asset Testing," *Journal of Political Economy*, 114(2), 257–279.

Golosov, M., A. Tsyvinski, and I. Werning (2006): "New Dynamic Public Finance: A User's Guide," *in NBER Macroeconomics Annual 2006, ed. D. Acemoglu, K. Rogoff and M. Woodford, Cambridge: MIT Press, National Bureau of Economic Research*, 2006.

Gourinchas, P., and J. Parker (2002): "Consumption over the Life Cycle," *Econometrica*, 70(1), 47–89.

Green, E. (1987): "Lending and the Smoothing of Uninsurable Income ," *in Contractual Arrangements for Intertemporal Trade, ed. E. Prescott and N. Wallace, Minnesota Studies in Macroeconomics series*, vol. 1, 3–25.

Grochulski, B. (2007): "Optimal Personal Bankruptcy Design: A Mirrlees Approach," Manuscript, Federal Reserve Bank of Richmond.

Hansen, L., J. Heaton, and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14(3), 262–280.

Heathcote, J., K. Storesletten, and G. Violante (2005): "Two Views of Inequality over the Life Cycle," *Journal of the European Economic Association*, 3(2–3), 765–775.

——— (2007): "Consumption and Labor Supply with Partial Insurance: An Analytical Framework," Manuscript, NYU.

Huggett, M., and J. Parra (2006): "Quantifying the Inefficiency of the US Social Security System," Manuscript, Georgetown University.

Huggett, M., G. Ventura, and A. Yaron (2007): "Sources of Lifetime Inequality," *NBER Working Paper*, w13224.

Kapicka, M. (2007): "Decentralizing Dynamic Mirrleesian Economies with Income Dependent Assets," Manuscript, UCSB.

Kehoe, T., D. Levine, and E. Prescott (2002): "Lotteries, Sunspots, and Incentive Constraints," *Journal of Economic Theory*, 107(1), 39–69.

Kocherlakota, N. (1998): "The Effects of Moral Hazard on Asset Prices When Financial Markets Are Complete," *Journal of Monetary Economics*, 41(1), 39–56.

——— (2005): "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica*, 73(5), 1587–1621.

Lagarias, J. C., J. A. Reeds, M. H. Wright, and P. E. Wright (1998): "Convergence properties of the Nelder-Mead Simplex Algorithm in Low Dimensions," *SIAM Journal on Optimization*, 9(1), 112–147.

Mirrlees, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(114), 175–208.

Phelan, C. (1994): "Incentives, Insurance, and the Variability of Consumption and Leisure," *Journal of Economic Dynamics and Control*, 18(3–4), 581–599.

——— (1998): "On the Long Run Implications of Repeated Moral Hazard," *Journal of Economic Theory*, 79(2), 174–191.

——— (2006): "Opportunity and Social Mobility," *Review of Economic Studies*, 73(2), 487–505.

Phelan, C., and R. Townsend (1991): "Computing Multi-Period, Information-Constrained Optima," *Review of Economic Studies*, 58(5), 853–881.

Prescott, E., and R. Townsend (1984): "Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard," *Econometrica*, 52(1), 21–46.

Rogerson, W. (1985a): "Repeated Moral Hazard," *Econometrica*, 53(1), 69–76.

——— (1985b): "Repeated Moral Hazard," *Econometrica*, 53(1), 69–76.

Spear, S., and S. Srivastava (1987): "On Repeated Moral Hazard with Discounting," *Review of Economic Studies*, 54(4), 599–617.

Storesletten, K., C. Telmer, and A. Yaron (2001): "How Important Are Idiosyncratic Shocks? Evidence from Labor Supply," *American Economic Review*, 91(2), 413–417.

Thomas, J., and T. Worrall (1990a): "Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-Agent Problem," *Journal of Economic Theory*, 51(2), 367–390.

——— (1990b): "Income fluctuation and asymmetric information: An example of a repeated principal-agent problem.," *Journal of Economic Theory*, 51(2), 367–390.

Townsend, R. (1979): "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, 21(2), 265–293.

——— (1994): “Risk and Insurance in Village India,” *Econometrica*, 62(3), 539–591.

TUOMALA, M. (1990): *Optimal Income Tax and Redistribution*. Oxford University Press, Clarendon Press.

WANG, C. (2005): “Dynamic costly state verification,” *Economic Theory*, 25(4), 887–916.