Accounting for Private Information*

Laurence Ales[†]
Tepper School of Business,
Carnegie Mellon University

Pricila Maziero
The Wharton School
University of Pennsylvania

First version: October 22, 2007
This version: June 01, 2009

Abstract

We study the quantitative properties of constrained efficient allocations in an environment where risk sharing is limited by the presence of private information. We consider a life cycle version of a standard Mirrlees economy where shocks to labor productivity have a component that is public information and one that is private information. The presence of private shocks has important implications for the age profiles of consumption and hours. 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 productivity must be private information.

JEL codes: D82, D91, D11, D58, D86, H21

Keywords: Private Information, Risk Sharing, Consumption Inequality.

*We are grateful to Larry Jones and Patrick Kehoe for their continuous help and support. We thank Fabrizio Perri and Dirk Krueger for providing the CEX data. We thank Rajesh Aggarwal, Francesca Carapella, V.V. Chari, Simona Cociuba, Roozbeh Hosseini, Mike Golosov, Katya Kartashova, Narayana Kocherlakota, Erzo Luttmer, Ellen R. McGrattan, Fabrizio Perri, Chris Phelan, Anderson Schneider, Martin Schneider, Chris Telmer, Pierre Yared, Warren Weber, participants at 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. Remaining mistakes are ours. The views expressed in this paper are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

†Contact Ales: ales@cmu.edu, Maziero: pricila@umn.edu.

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.[1] Given that the efficient level of insurance prescribed by the complete markets model 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 a component 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 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.[2] Looking at the cross section, we observe that inequality in consumption is increasing over age.[3] At the same time, the profile for the cross-sectional variance in hours worked is slightly decreasing over the working life. As shown in Cochrane (1991) and Storesletten, Telmer, and Yaron (2001), these facts suggest that workers are partially insured against idiosyncratic shocks.

The study of contractual arrangements that can explain the lack of full insurance is the underlying motivation for Green (1987), Thomas and Worrall (1990), Atkeson and Lucas (1992). These papers show that a repeated moral hazard environment with privately ob-

[1] See, for example, Cochrane (1991), Townsend (1994), Storesletten, Telmer, and Yaron (2001), and Attanasio and Davis (1996).

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

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

served 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 jointly the behavior of consumption and hours; for this reason we focus on an environment where the source of asymmetric information is the worker's labor productivity, as in Mirrlees (1971) and Golosov, Kocherlakota, and Tsyvinski (2003).

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

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

We solve the model numerically and use the simulated method of moments to determine parameter values. Our targets are the variances of consumption and hours along the life cycle. Our baseline estimated model can account for the increase in consumption inequality over the working life and the slight decrease in the inequality in hours that we observe in the data.

In the calibrated model, 99% of the labor productivity shock is private information. The result is robust to different specifications of the utility function, different target moments, heterogeneity in initial promised utility levels, and persistence of the publicly observable component of labor productivity.

This paper is related to a growing literature that studies the distortions implied by the optimal contract in dynamic versions of the original Mirrlees environment. The focus of most of this literature is normative, looking at decentralization through taxes in environments where the government is the sole provider of insurance. Few papers have looked at the empirical implications of the allocations of such constrained efficient problems.[4] 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.[5] The key difference of our paper is the focus on a Mirrlees environment, which generates jointly the behavior of consumption and hours worked and using both series allows us to identify the amount of private information. The second focuses on a moral hazard problem with hidden savings and shows, analyzing equilibrium restrictions, how private information can explain the excess smoothness in consumption in data from the United Kingdom.

This paper is also related to a recent literature that studies an environment where workers have access to insurance that is in addition to what is available through precautionary savings. Blundell, Pistaferri, and Preston (2008) and Heathcote, Storesletten, and Violante (2007) study environments where workers are subject to two types of shocks – some that are completely insured, some are entirely uninsured. With respect to these papers, the assets available in our environment, and hence the level of insurance provided at different ages, are determined endogenously.

The paper is structured as follows: section 2 describes the environment, section 3 studies

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

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

the qualitative implications of the environment, section 4 presents the data, section 5 presents our estimation strategy and results, and section 6 concludes.

2 Environment

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

Our environment is a standard dynamic Mirrlees economy similar to Golosov, Kocherlakota, and Tsyvinski (2003) and 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.[6] A large literature on dynamic optimal contracts considers contracts with infinite length. In our environment, solving a contract with finite length has important implications for the allocations of consumption, hours, and income, which will be explained in the next section.

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

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

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

Assume that u is twice continuously differentiable, increasing, and concave in both arguments. Agents discount future utility at the constant rate $\beta < 1$. Given a sequence of consumption and leisure $\{c_t, l_t\}_{t=1}^T$, the expected discounted utility over the working life is

[6] In our environment, there is no moral hazard problem after retirement; hence, retirement can be fully characterized by the continuation utility assigned at time T denoted by w_T. Our approach is to assume that the planner assigns to each worker the same level of w_T. There might be welfare gains from allowing the planner to choose w_T optimally as an additional instrument for providing incentives to agents at time T.

given by

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

where $\mathbb{E}_0$ denotes the expectation with respect to the information available at age $t = 0$.

Uncertainty in ages $1, \ldots, T$ is in the form of labor productivity shocks. At every age, a worker is subject to two idiosyncratic labor productivity shocks, $\theta_t \in \Theta_t$ and $\eta_t \in H_t$. Let $\theta^t \equiv (\theta_1, .., \theta_t) \in \Theta^t$ and $\eta^t \equiv (\eta_1, .., \eta_t) \in H^t$ denote the histories of the shocks up to age t . For a given realization of the labor productivity shocks, a worker can produce y units of effective output according to the following relation:

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

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 privately observed by the worker (the private shock). Let $\pi(\theta^T, \eta^T)$ denote the probability of drawing a particular sequence of productivity shocks 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.[7] The contribution of private information to

[7] In section 5.3 we relax this assumption by looking at the effects of a persistent public shock. Adding serial correlation to the privately observed shock is left for future work. Extending the model along this direction introduces several obstacles. From a computational point of view, the difficulty is in characterizing

labor productivity uncertainty is summarized by Ω, the fraction of the variance of labor productivity due to private information,

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

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. Denote the contract by $\{c, y\} = \{c_t(\theta^t, \eta^t), y_t(\theta^t, \eta^t)\}_{t=1}^{T}$. Note that the planner's problem is not subject to any aggregate uncertainty.

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

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

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

the optimal contract given the history dependent reporting strategies that must be considered by the planner and the lack of common prior on the type of the agent (some of these issues have been addressed in Fernandes and Phelan (2000) and Doepke and Townsend (2006)). Also, the presence of persistent private information introduces an additional age varying component in the provision of incentives (besides the one emphasized in this paper) making the identification of the amount of private information less transparent.

appendix A we extend this argument to the case with a nonseparable (Cobb-Douglas) utility function.

The planner has access to a technology that allows transferring resources linearly over time at the constant rate $1/q$. A contract $\{c, y\}$ is feasible if it satisfies the following:[8]

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

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

$$\begin{aligned} \max_{\{c,y\}_{t=0}^{T}} \quad & \sum_{t=1}^{T} \sum_{\theta^t, \eta^t} \pi_\theta(\theta^t) \pi_\eta(\eta^t) \beta^{t-1} u \left(c_t \left(\theta^t, \eta^t \right), \frac{y_t \left(\theta^t, \eta^t \right)}{f(\theta_t, \eta_t)} \right) \\ s.t. \quad & (5) \quad and \quad (6) \end{aligned} \tag{7}$$

2.1 Recursive formulation

To compute the solution to the planner's problem, it is convenient to rewrite the above problem recursively. We write the problem using as a state variable the continuation lifetime utility, as in Spear and Srivastava (1987) and Green (1987). In addition, instead of solving the above utility maximization problem, we solve its dual cost minimization problem. To allow for ex-ante heterogeneity, before any uncertainty is realized, each worker is associated with a number w_0, which denotes his entitlement of discounted lifetime utility. As in Atkeson and Lucas (1992), we solve the correspondent planner's problem for each level of promised utility w_0 for each worker of generation t.

In the recursive formulation we need to distinguish between the problem faced in period T, when the planner chooses current consumption and output, and all other periods $t < T$ when the planner chooses current consumption, output, and continuation utility. We refer to the problem for any $t < T$ as the $T - 1$ problem. From here onward we make the additional assumption that the private information labor productivity shock can take only two values per period $\eta_t \in \{\eta_{H,t}, \eta_{L,t}\}$ with $\eta_{H,t} > \eta_{L,t}$ for all t.[9] We also consider the relaxed problem, only considering incentive compatibility constraints for the agent that draws η_H. In appendix

[8]This feasibility constraint abstracts from inter-generational transfers.

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

B we show that the relaxed problem is equivalent to the original if the utility function is separable over consumption and leisure.[10] The period T problem is

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

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

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

The time $T-1$ problem is

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

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

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

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 few analytical results that can be derived relies on applying variational methods to the planner problem. This approach has been used by Rogerson (1985) and in an environment similar to ours by Golosov, Kocherlakota, and Tsyvinski (2003). The key result is that it is optimal for the planner to equate expected marginal cost whenever possible. Equating marginal costs requires the planner to be able to transfer resources between different nodes of the contract

[10] 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.

in an incentive feasible way (by a node we refer a particular history of labor productivity shocks at a given age). For example, the Euler equation for marginal cost derived by Golosov, Kocherlakota, and Tsyvinski (2003) requires the planner at every period t to be able to transfer resources between all of the states at time $t+1$ and the current period. Since time is observable, this transfer can be performed in an incentive feasible way.

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

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

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

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

Proof. In appendix C. ■

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 when $f(\theta, \eta) = \theta \cdot \eta$, in the presence of private information, consumption depends on θ.

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 D. ■

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 plot displays typical policy function for the case with $f(\theta, \eta) = \theta \cdot \eta$.

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

(a) (b)

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

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 values for the elasticity of substitution, and with nonconstant elasticity of substitution.

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

3.1 The static allocation

We start with the static Mirrleesian benchmark, setting $T = 1$.[11] With only one period, the planner can provide incentives only by distorting consumption and hours with respect

[11]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).

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:.[12]

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

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

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 2. 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 2. Looking at the covariance between consumption and hours with labor productivity, we observe that the response of 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.

[12]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.

3.2 The multi-period allocation

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

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

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

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 of promised utility. A positive multiplier on the incentive constraint, together with the cost function of the planner being increasing and convex imply a spreading out of continuation utilities. In addition, the convexity of the cost function implies that this spreading out is asymmetric.[14] 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.

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

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

[14] The spreading out of continuation utility has been shown numerically by Phelan and Townsend (1991). Asymptotic limit results have also been studied by Thomas and Worrall (1990), Atkeson and Lucas (1992), Aiyagari and Alvarez (1995), and Phelan (1998).

(a) (b)

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

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

$$\begin{aligned} var(w_{t+1}) &= E[var(w_{t+1}|w_t)] + var(E(w_{t+1}|w_t)) \\ &= E[var(w_t')] + var\left(E(w_t')\right). \end{aligned}$$

The first term, as stated, is decreasing while the second one is increasing over age.

The behavior of promised utility affects directly the allocations of consumption and hours. The increasing variance of promised utility contributes to an increase in the variance of both over age. However the way incentives are provided has different implications for the variance of consumption and hours. As noted in the static environment, a high distortion on

(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.

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 of hours a quantitative question. This relation between private information and the evolution of the inequality consumption and hours over age is at the basis of our identification strategy described in section 5. 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.[15] 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 is also robust once we introduce persistence in the publicly observable component of labor productivity shocks) is a specific prediction of this

[15]The increase in the variance of consumption and the effect on the variance of hours depend on the assumption that consumption and leisure are complements ($\sigma > 1$). Under this assumption, the binding incentive constraints are high productivity agents misreporting as low productivity, which are the constraints in the relaxed problem we solve.

(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.

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

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

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 with age; overall the first term dominates, increasing the covariance between consumption and output over the working life. The covariance between c and y over age is plotted in figure 5 for different values of the curvature parameter and different amounts of private information.

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

FIGURE 5: Covariances for different Ω and σ.

From figure 5 we observe how without private information the covariance between c and y remains flat over age (when $\Omega = 0$ the level of the covariance is set by the total variance of the uncertainty and by the cross-partial between consumption and leisure in the utility function). Increasing σ, through its effect on the cross-partial, increases the level of the covariance while increasing Ω increases 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. By analyzing the data through the lens of our model, our goal is to verify if the allocation in the data is compatible with the predictions of the model without taking a stand on how this allocation is actually implemented. Several papers have proposed decentralizations for environments similar to ours. Prescott and Townsend (1984) show, for a general class of economies, that a competitive equilibrium in which firms are allowed to offer history-dependent contracts is Pareto optimal. Following the seminal work of Mirrlees (1971), the public finance literature has focused on implementing the constrained optimal 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 depends on the entire history of productivity shocks (the former) or on the current productivity shock and wealth level (the latter). In a similar environment in which the worker's disability is unobservable and permanent, Golosov and Tsyvinski (2006) show that the constrained efficient allocation can be decentralized as a competitive equilibrium with an asset-tested disability policy. Grochulski (2007) shows that the informational constrained allocation can be implemented using an institutional arrangement that resembles the US personal bankruptcy code. Following Kocherlakota (1998) and Prescott and Townsend (1984), Kapicka (2007) shows that the optimal allocation can be decentralized with workers sequentially trading one-period income-contingent assets.

4 The Data

We use two different data sources, the Michigan Panel Study of Income Dynamics (PSID) and the Consumer Expenditure Survey (CEX). Our main source for consumption expenditures and hours is the CEX; labor income is taken from the PSID. In order to make the data from both surveys as comparable as possible, we apply the same sample selection to both. We consider household heads (reference person in the CEX) as those between ages 25 and 55 who worked more than 520 hours and less than 5096 hours per year and with positive labor income.[17] We exclude households with wage less than half of the minimum wage in any given year. Table 3 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 13 (in appendix F) we present some descriptive statistics from both surveys.[18] All the earnings variables and hours refer to the household head, while the expenditure variables are total household expenditure per adult equivalent.[19] 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

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

[18] From table 13 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.

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

Table 3: 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.

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

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

(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.

In figure 6 we report the cross-sectional variances for consumption, hours, and earnings over the working life. The first fact to be noted is the large increase in the variance of income over the working life (14 log points), consumption increases less (3 log points), while the variance of hours is roughly constant with a slight decrease over the working life. In order to compare the two data sets used we plot the cross-sectional variance of hours from both; we observe that this moment is very similar in both datasets over the ages considered. In figure 7 we observe that the covariance of hours and consumption does not display any particular trend, remaining essentially flat across the life cycle. We also observe a significant increase in the covariance of consumption and earnings.

5 Estimation Strategy and Results

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

Due to the nonlinearity of the optimal contract, it is not possible to separately determine

(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.

from equilibrium conditions the size of the private information and the preference parameters. On the other hand, for any combination of parameters we can solve for the optimal contract and simulate a population. This enables us to determine preference parameters and Ω using a minimum distance estimator, minimizing the distance between moments generated by the model and moments observed in the data. We follow the procedure described in Gourinchas and Parker (2002) and estimate our model using the method of simulated moments.

Denote by Γ the vector of parameters to be estimated. From our model we determine the individual values of consumption, hours, and income as functions of the parameters and promised utility, denoted respectively by $c_{it}(w_{it}, \Gamma)$, $l_{it}(w_{it}, \Gamma)$, and $y_{it}(w_{it}, \Gamma)$. Our target moments are the cross-sectional variances of consumption hours and labor income by age.[20] 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

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

[20]Targeting effective hours worked is consistent with our informational assumption. Although in our model hours are assumed to be non-observable by the planner, since we consider direct truth-telling mechanisms, the number of hours worked is actually known along the equilibrium path.

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 estimate the curvature parameter (σ) and the share of consumption (α). We restrict shock to two realizations per period: $\theta_{t,H} > \theta_{t,L}$ for the public shock and $\eta_H > \eta_L$ for the private shock. The average value of labor productivity is held constant (normalized to 1) along the life cycle, but we do allow for an increasing variance for both private and public shocks. The two shocks are parametrized by the following, for every age t

$$\theta_{t,H} - \theta_{t,L} = 2h_\theta[1 + g_v(t-1)], \tag{25}$$

$$\eta_{t,H} - \eta_{t,L} = 2h_\eta[1 + g_v(t-1)]. \tag{26}$$

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 shocks enables us also to maintain a constant Ω across the working life, from equation (4) we have that

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

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

$$\log(w_H) = \log(w_0) + \log(\delta)$$

$$\log(w_L) = \log(w_0) - \log(\delta)$$

where for a given δ the value w_0 is determined so that the feasibility condition holds

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

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 ages 25 to 55. In the previous section it was shown how introducing private information enables us to have an increase in the variance of consumption without increasing the variance of hours. Hence, looking at these moments, directly exploits the mechanism induced by private information. Column (1) of table 5 displays the results.

TABLE 5: 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

The model is locally identified (the gradient of the score function (24) 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.[21] Estimation results are further discussed in section 5.2. 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

[21] The difference in the value of the score function for values of Ω=0.99 and Ω=1 is within numerical rounding.

leisure equal to 0.90.[22] 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 resulting value is then equal to 0.40. This value is well within the common estimates in the labor literature (refer to Browning, Hansen, and Heckman (1999) table 3.3). In addition, the value of g_v is close to zero and negative; thus, the total variance of labor productivity decreases (slightly) over age. Figure 8 displays the fit of the model with respect to the targeted moments.

(a) (b)

FIGURE 8: Benchmark environment fit on matched moments.

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

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

[22]The implied coefficient of risk aversion is $\rho = 1 - \alpha + \alpha\sigma$ and the Frisch elasticity of leisure $\phi = \rho/\sigma$.

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).[23] When calculating how the value of τ_{cl} evolves over the ages 25-55 we observe that two factors are important: family composition and the definition of consumption (including or not housing). Our benchmark calculation of τ_{cl}, plotted in figure 9, considers only single households and does not include services imputed from housing in the definition of consumption.[24] We observe that τ_{cl} clearly displays an increasing trend over age in the data.

(a) (b)

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

The growth rate of τ_{cl} is decreasing in α. For the benchmark estimation ($\alpha = 0.69$), for ages up to 40-45, τ_{cl} in the model increases at the same rate as τ_{cl} in the data. After that, 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

[23] By doing so, we are not taking into account incentives provided through wages.

[24] By restricting the sample to single households we do not have to consider the joint decision of hours worked within a household. This joint decision might have significant effect when calculating τ_{cl} since it includes ratios of marginal utility.

Figure 10: Transfers over age.

of consumption to the marginal disutility of leisure as an exogenously incomplete markets model with endogenous labor, 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)).

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

In table 5 we report some initial robustness checks. We first look at the effect of setting

[25] The PSID for the years 1969 to 1985 continuously reports any additional transfer income the household received during the previous year. This variable includes transfers from publicly funded programs (food stamps, child nutrition programs, supplemental feeding programs, supplemental social security income, AFDC, earned income tax credit) and transfers received by family and nonfamily members. In our sample, 24% of the household-year observation received a transfer, and in total 67% of the households received a transfer at some time. These transfers are significant, averaging \$1930 (1983 dollars) and account for 70% to 90% of total food expenditures.

the initial heterogeneity in w equal to zero ($\delta = 0$) (column (2)). The environment without initial heterogeneity cannot account for the entire level in the heterogeneity in consumption, so in this test we only look at the increase in the variance of consumption over age.[26] For the remaining cases, we keep $\delta = 0$. We next look at the effect of fixing a lower share of consumption in the utility function (column (3)). We target average hours worked (column (4)) and restrict to a stationary process for labor productivity, fixing $g_v = 0$ (column (5)). All the cases confirm that shocks to labor productivity are entirely private information. In column (3) of table 7, 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 5 and 7). Section 5.3 considers 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

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

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

Combining the previous two equations, we get

$$\begin{aligned} Var\left[\ln c_\theta\right] &= (\phi-1)^2\, Var\left[\ln\theta\right], && (34)\\ \Delta Var\left[\ln c_\theta\right] &= (\phi-1)^2\, \Delta Var\left[\ln\theta\right], && (35)\end{aligned}$$

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

$$\begin{aligned} Var\left[\ln\left(1-l_\theta\right)\right] &= \phi^2 Var\left[\ln\theta\right], && (36)\\ \Delta Var\left[\ln\left(1-l_\theta\right)\right] &= \phi^2 \Delta Var\left[\ln\theta\right]. && (37)\end{aligned}$$

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.[27] This feature highlights the difficulty of a full information insurance environment in describing the profile of consumption and hours.[28] In our environment, private information is necessary to provide an increasing variance in consumption while at the same time keeping the variance of hours constant.

We now provide some intuition on the values obtained for the preference parameters. In the minimization procedure starting, for example, from an initial guess of ($\sigma = 3$, $\alpha = \frac{1}{3}$), we observe that the minimization path ultimately progresses, decreasing risk aversion and increasing the elasticity of leisure (decreasing 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; which implies larger distortions on both the intratemporal and inter-temporal margin causing a larger spreading out of consumption and continuation utility. Also, a low value of risk aversion, although reducing the need to provide insurance, increases the elasticity of intertemporal substitution, making it less costly to the planner to provide incentives intertemporally. The additional tension that determines the value of the risk aversion and elasticity is given by the cross-partial derivative between consumption and leisure. As 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.

[27] If the amount of time devoted to leisure is greater than hours worked, we have that $Var\left[\ln(l)\right] > Var\left[\ln(1-l)\right]$.

[28] 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).

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

(a) (b)

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

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

5.3 Robustness checks

In this section we look at additional robustness checks.

Optimal weighting matrix

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

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

The parameter results are reported in column (6) of table 9. A summary of the fit of the model is displayed in table 7. 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 7 and 9. 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

Table 7: 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).

In section 5 we used the level of the cross-sectional variance of hours as a target moment. However, the presence of measurement error can bias the level upward. To control for this effect, we reestimate the benchmark environment by cutting the cross-sectional variance of hours by 30%.[29] The results are in column (8) of tables 7 and 9.

Allowing persistence of the public shock

So far we have assumed that the labor productivity process is independent over age. We relax this assumption by introducing persistence in the public component of labor productivity. We model the public shock with a two state Markov chain. The transition matrix is bistochastic, and the probability of remaining in the same state is given by 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 9, 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, indicating that public shocks to labor productivity are permanent. We interpret this result as a supportive argument for the importance of private information shocks: even when the only shock allowed

[29]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.

to have serial correlation is the public one, we still require a significant fraction of the shock to be privately observed.

Separable utility function and income variance

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

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

This specification is commonly used in the labor literature.[30] 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 markets are perfectly competitive and workers are paid their marginal productivity. The parameters' estimates are given in table 9, column (9). The fit of the model is displayed in figures 12 and 13.

Table 9: 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).

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

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

variance of consumption at age 55. Also the variance of hours is 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.

5.4 The response of consumption to income shocks

Under the assumption that workers are paid at their marginal productivity, we can also study how consumption responds to income changes. This measure, which has been extensively studied, reflects the insurance possibilities of workers against income fluctuations and has important policy implications.[31] In particular, Kaplan and Violante (2008) show that a calibrated version of a Bewley economy is not compatible with the degree of consumption insurance observed in the US data. We compute the response of consumption growth to income growth (α_2) from the following regression:[32]

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

We compute the value of α_2 from the CEX data using an OLS and instrumental variable estimation as in Dynarski and Gruber (1997). Results are in table 12. We perform the

[31]See Kaplan and Violante (2008) and references therein.

[32]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.

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

same estimation on panel data 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.[33] 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 emphasizes the limited insurance possibilities available to workers.

[33] 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 $\alpha_2 = 0.269$.

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

Table 11: 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 our environment. The model predicts that the value of α_2 should be increasing in age due to the progressive 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.

6 Concluding Remarks

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

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

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, S., and F. Alvarez (1995): "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.

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

Atkeson, A., and R. Lucas (1992): "On Efficient Distribution with Private Information," *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 (2008): "Consumption inequality and partial insurance," *American Economic Review*, 98(5), 1887–1921.

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.

Doepke, M., and R. Townsend (2006): "Dynamic Mechanism Design with Hidden Income and Hidden Actions," *Journal of Economic Theory*, 126(1), 235–285.

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*.

Fernandes, A., and C. Phelan (2000): "A Recursive Formulation for Repeated Agency with History Dependence," *Journal of Economic Theory*, 91(2), 223–247.

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.

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

Kaplan, G., and G. Violante (2008): "How Much Insurance in Bewley Models?," Mimeo, NYU.

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.

Kocherlakota, N., and L. Pistaferri (2007): "Household heterogeneity and real exchange rates," *Economic Journal*, 117(519), C1–C25.

Kocherlakota, N., and L. Pistaferri (2009): "Asset Pricing Implications of Pareto Optimality with Private Information," *forthcoming Journal of Political Economy.*

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.

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. (1985): "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 (1990): "Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-Agent Problem," *Journal of Economic Theory*, 51(2), 367–390.

Townsend, R. (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.

Appendix

A 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

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

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

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

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

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

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.

B 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 $(P1)$.

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

The time $T-1$ problem is

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

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

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

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

$$\begin{aligned}
v\left(\frac{y_{T-1}(\theta,\eta_H)}{f(\theta,\eta_H)}\right)-v\left(\frac{y_{T-1}(\theta,\eta_H)}{f(\theta,\eta_L)}\right) &> v\left(\frac{y_{T-1}(\theta,\eta_L)}{f(\theta,\eta_H)}\right)-v\left(\frac{y_{T-1}(\theta,\eta_L)}{f(\theta,\eta_L)}\right),\\
v\left(\frac{y_{T-1}(\theta,\eta_H)}{f(\theta,\eta_H)}\right)-v\left(\frac{y_{T-1}(\theta,\eta_L)}{f(\theta,\eta_H)}\right) &> \left(\frac{y_{T-1}(\theta,\eta_H)}{f(\theta,\eta_L)}\right)-v\left(\frac{y_{T-1}(\theta,\eta_L)}{f(\theta,\eta_L)}\right). \qquad (56)
\end{aligned}$$

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

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

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. ■

C Proof of Proposition 1

Proof. The proof follows Rogerson (1985) closely. Considering the planner's problem as allocating utility levels to workers, let $\bar{u}(\theta,\eta)=u(c(\theta,\eta))$ be the utility derived from consumption in state (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 $\bar{u}(\theta^t,[\eta^{t-1},\eta_t])=\bar{u}^*(\theta^t,[\eta^{t-1},\eta_t])-\Delta$ 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 $u(\theta^t,[\eta^{t-1},\tilde{\eta}_t])=u^*(\theta^t,[\eta^{t-1},\tilde{\eta}_t])$ and $u([\theta^t,\tilde{\theta}_{t+1}],\eta^{t+1})=u^*([\theta^t,\tilde{\theta}_{t+1}],\eta^{t+1})$ for $\eta_t\neq\tilde{\eta}_t$ and for $\theta_{t+1}\neq\tilde{\theta}_{t+1}$. The labor allocations are left unchanged. This contract is still incentive compatible given that the time and the shock θ is publicly observed. This contract minimizes the cost for the planner if the following holds:

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

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 $C'\left(u\left(\theta^t,\eta^t\right)\right)=\frac{1}{u_c\left(c\left(\theta^t,\eta^t\right)\right)}$. 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

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

Equation (15) then follows from

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

■

D 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 $y(\theta, \eta_H)$

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

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)

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

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. ■

E 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.

F Data

TABLE 13: 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.