Optimal Dynamic Taxes*

Mikhail Golosov
Yale
and NES

Maxim Troshkin
University of Minnesota
and FRB Minneapolis

Aleh Tsyvinski
Yale
and NES

October 2010

Preliminary and incomplete. Do not circulate without permission.

Abstract

We develop a methodology to derive formulas that facilitate interpretation of the forces determining optimal labor and savings distortions and taxes in dynamic settings. The formulas for the labor wedges extend the static optimal taxation analysis of Diamond (1998) and Saez (2001) to dynamic settings. Compared to the static analysis, the dynamic nature of the problem offers three novel insights. First, the opportunity to provide incentives dynamically adds a force lowering labor distortions. Second, labor distortions in dynamic settings may differ significantly from those in static settings because a key determinant of the former is the conditional rather than the unconditional distribution of skill shocks. The conditional distribution of shocks differs significantly from the unconditional one. Third, the persistence of shocks manifests itself as an increase in the redistributionary motive of the government. We also derive a novel formula to analyze the determinants of the savings distortions. In the i.i.d. case and under certain conditions in the case of persistent shocks, we show that the labor wedge tends to zero for sufficiently high skills. This is in sharp contrast to the static case with Pareto tail of the skill distribution of Diamond (1998) and Saez (2001), who show that taxes on the high skill agents are increasing and tend to potentially high levels depending on the parameters of the tail.

Our second set of results is to numerically simulate the optimal labor and savings distortions. The analysis is conducted for a realistically calibrated economy based on empirical income distributions. The computed optimal dynamic distortions differ significantly from the optimal static distortions, highlighting the importance of the forces in the theoretical analysis. The welfare gains compared to optimal linear taxes are non-trivial in the case of the utilitarian social planner and are significant (close to 5% of consumption) for a more redistributive Rawlsian criterion.

Our third contribution is a novel implementation of the optimal allocations. We show that a tax system based on consolidated income accounts (CIA) implements the optimum. The labor income tax depends on the current labor income and on the balance on the CIA. The savings tax depends only on the amount of savings. The CIA balance is updated as a function of the labor income and the previous balance.

*We thank Fernando Alvarez, V.V. Chari, Dirk Krueger, Larry Jones, Stephen Morris, James Poterba, Emmanuel Saez, Nancy Qian, Hongda Xiao, and Pierre Yared for helpful comments. We thank the audiences at ASU, Boston University, Chicago Booth, Cowles Foundation, EIEF, Minnesota Macro, NBER PF, NY Fed, NYU, Gerzensee, SED, Texas, Rochester, UCSD and University of Chicago. Golosov and Tsyvinski thank EIEF for hospitality and NSF for support.

1 Introduction

A sizeable New Dynamic Public Finance (NDPF) literature studies optimal taxation in dynamic settings[1]. The models in this literature extend the classic Mirrlees equity-efficiency trade-offs to dynamic settings in which agents' skills change stochastically over time.

This paper provides a methodology to derive simple formulas that facilitate interpretation of the forces behind the optimal taxation results in dynamic settings. The formulas easily connect to empirically observable data. Diamond (1998) and Saez (2001) significantly expanded the understanding and policy relevance of *static* Mirrlees models by deriving an easily interpretable formula in terms of elasticities and the shape of income distribution. Our paper extends their analysis to *dynamic* settings.

Our first contribution is to derive easily interpretable formulas for the first-order conditions for the dynamic labor and savings distortions. As in the static case, the shape of the income distribution, the redistributionary objectives of the government, and labor elasticity play important roles in the determination of labor distortions. However, the dynamic model adds three significant differences to the analysis of optimal distortions: (*i*) the use of dynamic incentives adds a force that tends to lower labor income wedges; (*ii*) conditional rather than unconditional distributions of skills is a key determinant of wedges; (*iii*) persistence of shocks acts as a larger redistributionary motive for the government. We also derive a novel representation of the savings wedge that allows the analysis of the forces determining it.

Specifically, we study T-period dynamic optimal taxation economies with i.i.d. and persistent shocks based on Golosov, Kocherlakota, and Tsyvinski (2003) with preferences represented by utility with no income effects. Consider first an illustrative case of i.i.d. shocks in two periods. There are two key insights in this part of the analysis for the nature of labor distortions in the first period (early in life): (*i*) the dynamic nature of the incentives represents itself as an additional term in the formula for the optimal distortions changing the weights assigned to agents by the social planner, and (*ii*) this reweighing represents the fact that using dynamic incentives allows to lower marginal taxes. The derivation of the easily interpretable formulas for the labor distortions is novel to the New Dynamic Public Finance literature as theoretical analysis primarily focused on the intertemporal (savings) distortion. Next, we derive a for-

[1] See, for example, Golosov, Kocherlakota, and Tsyvinski (2003) or reviews in Golosov, Tsyvinski, and Werning (2006) and Kocherlakota (2010).

mula representing the savings distortion. The derivation of the formula interpreting the savings distortion is new to the NDPF literature, as it provides a way to study the economic forces determining savings wedges. We show that there is a force driving savings distortion higher for the high income agents as a way to lower their labor distortion. The intuition is that the effort of the highly skilled agents is very valuable in production and deterring their deviations is particularly important. The use of the savings wedge allows provision of incentives while lowering the need for labor distortions which are overly costly for these agents.

We then study the case of persistent shocks. There are two additional key insights to the analysis of the static and the i.i.d. cases. The first difference is that the optimal labor distortions formulas now depend on conditional rather than on the unconditional distributions of skills. Empirical conditional and unconditional distributions differ significantly. Therefore, the optimal dynamic taxes may be very different from the static ones. The second insight is that persistence yields an additional force for redistribution to the optimal tax problem. The intuition is based on the optimal provision of dynamic incentives. An agent with a low skill early in life is likely to be low skill later in life; the same persistence is present for a high type. An agent who has a low income early in life and high income later in life is more likely to be a deviator, i.e., a high skilled agent pretending to be low skilled early in life. Changing the weights in the social welfare function by redistributing away from high income agents worsens benefits from such deviation and improves incentives.

In the i.i.d. case as well as under certain conditions in the case of persistent shocks, we show that the labor wedge tends to zero for sufficiently high skills. This is in sharp contrast to the static case with Pareto tail of the skills distribution of Diamond (1998) and Saez (2001), who show that the taxes on the high skill agents are increasing and tend to high levels (50-70%) depending on the parameters of the tail ratio of skills.

We note that our analysis of the case of the persistent shocks builds on the first-order approach developed in Kapicka (2008) and Pavan, Segal, and Toikka (2010). In numerical simulations, we verify its sufficiency.

The second contribution of the paper is to numerically simulate the optimal labor and savings wedges in a realistically calibrated economy based on the empirical income distributions. First, consider the case of the i.i.d. shocks. The results show that dynamic wedges are significantly different from the static taxes emphasizing the importance of the theoretical forces we study. We find that the labor distortion for the early periods are smaller than for the later periods. This result is also related to findings in Ales and Maziero (2007), who numerically solve a version of a life cycle economy with i.i.d. shocks drawn from a discrete, two-type distribution, and find that the labor distortions are lower early in life of the household. The second difference from the static model is that we provide calculations for the savings tax and find it numerically significant and increasing. The numerical simulations for the empirically calibrated persistent shocks add two important differences. The first is that the consideration of conditional rather than the unconditional empirical distributions of income and skills significantly alters the pattern of wedges compared to the static and the i.i.d. cases. The second difference is that agents face very different labor distortions conditional on the previous shocks. This is due to the differences among the conditional distributions and also due to the planner's increase in the redistributionary objectives to deter earlier deviations. Finally, we provide the calculations of the welfare gains of using the optimal policy. A natural benchmark to compare the constrained efficient optimum is an environment with the optimal linear taxes. First, consider the case of the utilitarian social planner. The optimal age-dependent linear labor wedges yield a welfare loss of 0.6% of consumption compared to the constrained optimum. The optimal age-*in*dependent labor distortion yields a welfare loss of 1.4%. While these magnitudes are non-trivial, linear taxes can still yield reasonably good policies. This is a well-known result in numerical simulations of the static Mirrlees models (e.g.,Mirrlees (1971), Atkinson and Stiglitz (1976), Tuomala (1990)) that illustrate that linear taxes with utilitarian social planner approximate the optimal policy rather well. This literature also points out that if the planner is more redistributive than utilitarian planner, the tax policy is substantially different from linear, and nonlinear taxes may yield large welfare gains. We also calculate welfare gains of using optimal policies when the social planner is more redistributive, in particular Rawlsian. The optimal age-dependent linear labor wedges yield a welfare loss of 4.5% compared to the constrained optimum. The optimal age-*in*dependent labor distortion yields a welfare loss of 5.1%. We conclude that the welfare gains of using optimal nonlinear policies are significant.

Our third contribution is a novel implementation of the optimal allocations – consolidated income accounts (CIA) tax system. In a given period, the labor income tax depends on labor income and on the balance of the CIA, the savings tax depends only on the amount of savings; the CIA balance is updated as a function of the labor income and the previous balance. The

CIA balance plays a role of the record-keeping device summarizing the previous labor choices of agents. The savings tax is constructed following Werning (2009) as an envelope of the best possible deviations. We then show that the CIA system takes a particularly simple form if the utility is exponential and the shocks are i.i.d. The tax system consists of a non-linear tax on savings income, a non-linear labor income tax, and a CIA account. In each period, a taxpayer can deduct the balance of the account from the total income tax bill. Thus, while all agents with the same labor income are facing the same marginal tax rate, the total tax bill is smaller for the agents with a higher CIA account. Similarly, updating the CIA balance follows a simple rule. In each period the increase on agent's CIA balance is determined solely by his labor income in that period. Moreover, if the distribution of the shocks does not change over time, taxes do not depend on age of the agent. We conclude with a numerical simulation of the optimal taxes. In our simulation, labor income taxes are lower than labor income distortions as they incorporate the effects of updating the CIA balance.

The recursive characterization of the problem, especially in the i.i.d. case, has similarities to the Mirrlees (1986) setup with two consumption goods. In Section 5, we further explore this connection and show the role that the nonseparability of preferences plays in the difference between static and dynamic models.

There are several papers related to our work. The first-order approach for persistent shocks is developed in Kapicka (2008) and Pavan, Segal, and Toikka (2010), who mainly focus on the risk-sharing properties of the taste shock model with exponential utility and Pareto shocks. There have been very limited theoretical analysis of the labor taxation in dynamic Mirrlees models. One important exception is Battaglini and Coate (2008) who provide a complete characterization of the optimal program with Markovian agents. While incorporating persistence in abilities, most of their analysis for tractability assumes only two ability types and risk neutral individuals.

An important contribution of Farhi and Werning (2010) is an analysis deriving a different way of characterizing the first order conditions of the optimal dynamic taxation model, provide numerical simulations, and also uses continuous time approach to derive additional insights. The analysis of Farhi and Werning (2010) and this paper are complementary. Our work focuses on a comprehensive study of cross-sectional properties of optimal wedges and on deriving elasticity based formulas following Diamond (1998) and Saez (2001). Farhi and Werning (2010) focus on the comprehensive study of the intertemporal properties of allocations and wedges.

Numerical simulations in our paper are also related to Weinzierl (2008). He derives theoretically and analyzes numerically an elasticity-based formula with which he studies optimal age-dependent taxation, in a dynamic Mirrlees setting. Albanesi and Sleet (2006) is a comprehensive numerical and theoretical study of optimal capital and labor taxes in a dynamic economy with i.i.d. shocks. Golosov and Tsyvinski (2006) study a disability insurance model with fully persistent shocks. Golosov, Tsyvinski, and Werning (2006) is a two-period numerical study of the determinants of dynamic optimal taxation in the spirit of Tuomala (1990). However, none of these papers, with the exception of Weinzierl (2008), base their analysis on an elasticity-based formula.

Our implementation is related to the work of Kocherlakota (2005), Albanesi and Sleet (2006), Grochulski and Kocherlakota (2007) and, importantly, Werning (2009). Both Kocherlakota (2005) and Werning (2009) discuss only capital taxation, and our construction of savings taxation builds directly on Werning (2009). The work by Albanesi and Sleet (2006) is another important predecessor to our decentralization. As in our paper, Albanesi and Sleet (2006) keep track of the summary of individuals' past histories summarized by one variable, which is individual's stock of wealth in their case. There are two main differences between our implementation and that in Albanesi and Sleet (2006). First, their implementation is applicable only for the i.i.d. shocks. Second, in Albanesi and Sleet (2006) savings of households play two roles: intertemporal smoothing and tracking history of previous labor incomes. This allows households to misrepresent their past histories by changing the amount of savings. To prevent households from doing this, the tax on capital should not only be non-linear, but the degree of non-linearity and the amount of capital taxes should depend on the realization of labor income. In our implementation, we keep track of the history of labor incomes in a separate CIA account and do not condition capital tax on income realization which substantially simplifies the tax system, as illustrated in the next section. The implementation we describe is relatively easy to implement in practice as it shares many of the existing elements of the current US tax code. It is also closely related to some long advocated ideas in public finance, such as income tax averaging, which go back at least to Vickrey (1939) and Vickrey (1947).

2 Environment

We consider an economy that lasts T periods, denoted by $t = 1, ..., T$ where T is a finite number. Each agent's preferences are described by a time separable utility function over consumption of a good c_t and labor l_t,

$$\mathbb{E}_1 \sum_{t=1}^{T} \beta^{t-1} U(c_t, l_t), \tag{1}$$

where $\beta \in (0,1)$ is a discount factor, $\mathbb{E}_1$ is an expectations operator and $U : \mathbb{R}_+^2 \to \mathbb{R}$. We assume that U is twice continuously differentiable, and partial derivatives with respect to c and l satisfy $U_c, U_{cc} > 0$, $U_l, U_{ll} < 0$.

In period $t = 1$, agents draw their initial type (skill), θ_1, from a distribution $F_1(\theta)$. For $t \geq 2$, skills follow a Markov process $F_t(\theta|\theta_{t-1})$, where θ_{t-1} is agent's skill realization in period $t-1$. We denote the probability density function by $f_t(\theta|\theta_{t-1})$ and assume that f_t is differentiable in both arguments. We assume that, in each period t, skills are non-negative: $\theta_t \in \Theta = \mathbb{R}_+$. The set of possible histories up to period t is denoted by Θ^t.

An agent of type θ_t who supplies l_t units of labor produces $y_t = \theta_t l_t$ units of output. The skill shocks and the history of shocks are privately observed by the agent. Output $y_t = \theta_t l_t$ and consumption c_t are observed by the planner. In period t, the agent knows only his skill realization for the first t periods $\theta^t = (\theta_1, ..., \theta_t)$. Denote by $c_t(\theta^t) : \Theta^t \to \mathbb{R}_+$ agent's allocation of consumption and by $y_t(\theta^t) : \Theta^t \to \mathbb{R}_+$ agent's allocation of output in period t. Denote by $\sigma_t(\theta^t) : \Theta^t \to \Theta^t$ agent's report in period t. We denote the set of all such reporting strategies in period t, $(\sigma_1(\theta^1), ..., \sigma_t(\theta^t))$ by Σ^t. Resources can be transferred between periods with a rate on savings $\delta > 0$. The observability of consumption implies that all savings are publicly observable. Hence, without loss of generality,we can assume that the social planner controls all the savings. We also assume that the social planner has a social welfare function $G : \mathbb{R} \to \mathbb{R}$, where G is increasing and concave. In particular since the lifetime utility of the agent is given by (1), the social welfare is given by $\int G\left(\mathbb{E}_1 \sum_{t=1}^{T} \beta^{t-1} U(c_t, l_t)\right) dF_1(\theta)$.

The optimal allocations solve the dynamic mechanism design problem (see, e.g., Golosov, Kocherlakota, and Tsyvinski (2003)):

$$\max_{\{c_t(\theta^t), y_t(\theta^t)\}_{\theta^t \in \Theta^t; t=1,..,T}} \int G\left(\mathbb{E}_1 \sum_{t=1}^{T} \beta^{t-1} U\left(c_t\left(\theta^t\right), y_t\left(\theta^t\right)/\theta_t\right)\right) dF_1(\theta) \tag{2}$$

subject to the incentive compatibility constraint:

$$\mathbb{E}_0 \left\{ \sum_{t=1}^{T} \beta^{t-1} U \left(c_t \left(\theta^t \right), y_t \left(\theta^t \right) / \theta_t \right) \right\} \geq \mathbb{E}_0 \left\{ \sum_{t=1}^{T} \beta^{t-1} U \left(c_t \left(\sigma_t \left(\theta^t \right) \right), y_t \left(\sigma_t \left(\theta^t \right) \right) / \theta_t \right) \right\}, \forall \sigma^T \in \Sigma^T, \tag{3}$$

and the feasibility constraint:

$$\mathbb{E}_0 \left\{ \sum_{t=1}^{T} \delta^{t-1} c_t \left(\theta^t \right) \right\} \leq \mathbb{E}_0 \left\{ \sum_{t=1}^{T} \delta^{t-1} y_t \left(\theta^t \right) \right\}. \tag{4}$$

Here the expectation $\mathbb{E}_0$ above is taken over all possible realizations of histories. The first constraint above is a dynamic incentive compatibility constraint that states that an agent prefers to truthfully report its history of shocks rather than to choose a different reporting strategy. The second constraint is the dynamic feasibility constraint.

We follow Fernandes and Phelan (2000) and Kapicka (2008) to re-write this problem recursively. Here, we briefly describe the recursive formulation and refer to these two paper for the technical details. Let $\omega \left(\theta | \tilde{\theta} \right) : \Theta \times \Theta \to \mathbb{R}$ denote *promised utility*. We use notation $\boldsymbol{\omega}(\theta)$ and $\boldsymbol{\omega}$ to denote functions $\omega(\theta|\cdot)$ and $\omega(\cdot|\cdot)$ respectively. Let $c : \Theta \to \mathbb{R}_+$ and $y : \Theta \to \mathbb{R}_+$.

The optimal allocations solve the cost minimization problem for period $t = 1$:

$$V_1 \left(\hat{\omega} \right) = \min_{c, y, \boldsymbol{\omega}} \int \left(c \left(\theta \right) - y \left(\theta \right) + \delta V_2 \left(\boldsymbol{\omega} \left(\theta \right), \theta \right) \right) f_1 \left(\theta \right) d\theta \tag{5}$$

subject to the incentive compatibility constraint:

$$U \left(c \left(\theta \right), y \left(\theta \right) / \theta \right) + \beta \omega \left(\theta | \theta \right) \geq U \left(c \left(\tilde{\theta} \right), y \left(\tilde{\theta} \right) / \theta \right) + \beta \omega \left(\tilde{\theta} | \theta \right), \ \forall \tilde{\theta} \in \Theta, \theta \in \Theta, \tag{6}$$

and to the promise keeping constraint:

$$\omega_0 \leq \int G \left(U \left(c \left(\theta \right), y \left(\theta \right) / \theta \right) + \beta \omega \left(\theta | \theta \right) \right) f_1 \left(\theta \right) d\theta. \tag{7}$$

The initial promised utility ω_0 is a solution to $V_1(\omega_0) = 0$.

For $t > 1$, the social planner takes the period $t-1$ realization of the shock which we denote by θ_-, and the chosen promised utility function $\hat{\boldsymbol{\omega}} \left(\theta_- \right)$ as given and solves:

$$V_t \left(\hat{\boldsymbol{\omega}} \left(\theta_- \right), \theta_- \right) = \min_{c, y, \boldsymbol{\omega}} \int \left(c \left(\theta \right) - y \left(\theta \right) + \delta V_{t+1} \left(\boldsymbol{\omega} \left(\theta \right), \theta \right) \right) f_t \left(\theta | \theta_- \right) d\theta \tag{8}$$

subject to the incentive compatibility constraint (6) and

$$\hat{\omega}(\theta_-|\tilde{\theta}) \geq \int \left(U\left(c\left(\theta\right), y\left(\theta\right)/\theta\right) + \beta\omega\left(\theta|\theta\right)\right) f_t\left(\theta|\tilde{\theta}\right) d\theta \text{ for all } \tilde{\theta} \in \Theta, \tag{9}$$

where constraint (9) must hold with equality for $\hat{\omega}(\theta_-|\theta_-)$.

Function $V_{T+1}\left(\boldsymbol{\omega}\left(\theta\right), \theta\right) = 0$ if $\boldsymbol{\omega}\left(\theta\right) = \mathbf{0}$ and $V_{T+1}\left(\boldsymbol{\omega}\left(\theta\right), \theta\right) = \infty$ otherwise. All other functions V_t are defined by backward induction. The function V_t is the resource cost of delivering promised utilities $\boldsymbol{\omega}\left(\theta\right)$.

The incentive compatibility constraint states that an agent prefers to reveal his true type θ, receive utility $U\left(c\left(\theta\right), y\left(\theta\right)/\theta\right)$ and a continuation utility $\omega\left(\theta|\theta\right)$ rather than claim a different type $\tilde{\theta}$, receive utility $U\left(c\left(\tilde{\theta}\right), y\left(\tilde{\theta}\right)/\theta\right)$ and a continuation utility $\omega\left(\theta|\tilde{\theta}\right)$. Promise keeping constraints (9) ensure that next period allocations indeed deliver the expected utility $\omega\left(\theta|\tilde{\theta}\right)$ to any type $\tilde{\theta}$ who sends a report θ.

We proceed in this section by using the first order approach developed by Kapicka (2008) and Pavan, Segal, and Toikka (2010) to obtain characterization of distortions. Since this approach only provides the necessary conditions, we verify numerically its sufficiency in the simulations in Section 4.2. Under the assumption that only local incentive constraints bind, the number of state variables reduces dramatically. One needs to keep track only of the "on the path" promise utility $\omega\left(\theta|\theta\right)$ and the utility from a local deviation $\omega_2\left(\theta|\theta\right)$ where $\omega_2\left(\theta|\theta\right)$ is the derivative of ω with respect to its second argument evaluated at $\left(\theta|\theta\right)$. Then defining functions $w : \Theta \to \mathbb{R}$ and $w_2 : \Theta \to \mathbb{R}$ the maximization problem (8) can be re-written as

$$V_t(\hat{w}, \hat{w}_2, \theta_-) = \min_{\{c(\theta), y(\theta), u(\theta), w(\theta), w_2(\theta)\}_{\theta\in\Theta}} \int \left(c\left(\theta\right) - y\left(\theta\right) + \delta V_{t+1}\left(w\left(\theta\right), w_2(\theta), \theta\right)\right) f_t\left(\theta|\theta_-\right) d\theta \tag{10}$$

$$u'\left(\theta\right) = U_l(c(\theta), y(\theta)/\theta)\left(-\frac{y\left(\theta\right)}{\theta^2}\right) + \beta w_2\left(\theta\right), \tag{11}$$

$$\hat{w} = \int u\left(\theta\right) f\left(\theta|\theta_-\right) d\theta, \tag{12}$$

$$\hat{w}_2 \geq \int u\left(\theta\right) f_2(\theta|\theta_-) d\theta, \tag{13}$$

$$u(\theta) = U(c(\theta), y(\theta)/\theta) + \beta w(\theta). \tag{14}$$

There are three state variables in this recursive formulation. The first, $\hat{w}$, is the promised utility associated with the promise-keeping constraint (12). The second, $\hat{w}_2$, is the state variable associated with the threat-keeping constraint (13). Finally, θ_- is the last period reported type.

Before characterizing the problem, we re-write the optimal problem in a different form that allows to highlight the effects of persistence.

Lemma 1. *Let* $\{c(\theta)^*, y(\theta)^*, u(\theta)^*, w(\theta)^*, w_2(\theta)^*\}_{\theta\in\Theta}$ *be a solution to (10). Then*

$$\{c(\theta)^*, y(\theta)^*, u(\theta)^*, w(\theta)^*, w_2(\theta)^*\}_{\theta\in\Theta}$$

is a solution to

$$\min_{\{c(\theta),y(\theta),u(\theta),w(\theta),w_2(\theta)\}_{\theta\in\Theta}} \int (c(\theta) - y(\theta) + \delta V_{t+1}(w(\theta), w_2(\theta), \theta)) f_t(\theta|\theta_-) d\theta \tag{15}$$

s.t. (11), (14) and

$$\hat{w} = \int \left(\zeta - \frac{f_2(\theta|\theta_-)}{f(\theta|\theta_-)}\right) u(\theta) f(\theta|\theta_-) d\theta$$

for some constant ζ*.*

Proof. In the Appendix. □

In problem (15), utility $u(\theta)$ is multiplied by the term $\left(\zeta - \frac{f_2(\theta|\theta_-)}{f(\theta|\theta_-)}\right)$. This pseudo-objective is equivalent to the objective function of a social planner that has (non-normalized) weights $\left(\zeta - \frac{f_2(\theta|\theta_-)}{f(\theta|\theta_-)}\right)$ instead of the utilitarian weights equal to 1 for all types θ in period t. To see the implications of these new weights, consider first an example of the function f that has a property that if $\theta_H > \theta_L$ then $f(\theta|\theta_L)/f(\theta|\theta_H)$ is decreasing in θ. For such function, the ratio $f_2(\theta|\theta_-)/f(\theta|\theta_-)$ is monotonically increasing in θ. The term $\left(\zeta - \frac{f_2(\theta|\theta_-)}{f(\theta|\theta_-)}\right)$ assigns the highest weight to the lowest type and monotonically decreases for the higher types. In other words, the planner's objective is more redistributionary towards the lower types in period t. The intuition for this change in weights is as follows. Consider a marginal deviation in period $t-1$. Suppose type $\theta_- + \varepsilon$ claims to be θ_- for some small ε. Under the above assumption on $f(\theta|\theta_-)$, this type is relatively more likely to receive high shocks θ and relatively less likely to receive low shocks θ in period t. The social planner who is more redistributive in period t and puts higher (pseudo) weights on the low types allocates relatively low utility to this agent. The type θ_- is not significantly affected, since his probability of having high shocks θ is relatively low. This agent benefits from more redistribution as for him the high shocks θ in period t are less likely. The same intuition generalizes for other stochastic processes. The general insight is that the social planner allocates relatively higher pseudo weights on those realizations of shocks θ for which there is large difference in the probability of occurrence between types $\theta\_$ and types close to θ_{-}.

Now we characterize optimal distortions. For an agent with the history of shocks θ^t at time t, we define a labor distortion:

$$1-T'_{D,t}\left(\theta^t\right)\equiv\frac{-U_l\left(c_t\left(\theta^t\right),y_t\left(\theta^t\right)/\theta_t\right)}{\theta_t U_c\left(c_t\left(\theta^t\right),y_t\left(\theta^t\right)/\theta_t\right)} \tag{16}$$

and a savings distortion

$$\tau_{S,t}\left(\theta\right)=1-\frac{\delta U_c\left(c_t\left(\theta^t\right),y_t\left(\theta^t\right)/\theta_t\right)}{\beta\mathbb{E}_t\left\{U_c\left(c_{t+1}\left(\theta^{t+1}\right),y_{t+1}\left(\theta^{t+1}\right)/\theta_{t+1}\right)\right\}}. \tag{17}$$

For the rest of the section we focus on the quasi-linear preferences of the form

$$U\left(c,l\right)=\bar{U}\left(c-\frac{1}{\gamma}l^{\gamma}\right), \tag{18}$$

where we denote derivatives of $\bar{U}$ by $\bar{U}_c$ and $\bar{U}_{cc}$.

When utility function satisfies (18), equation (11) becomes:

$$u'\left(\theta\right)=\bar{U}_c\left(c\left(\theta\right)-\frac{l\left(\theta\right)^{\gamma}}{\gamma}\right)\frac{l\left(\theta\right)^{\gamma}}{\theta}+w_2(\theta). \tag{19}$$

We can re-write the optimal problem (8) such that $u\left(\theta\right)$ is a state variable of optimization. Define function m implicitly by $\bar{U}_c\left(x\right)=m\left(\bar{U}\left(x\right)\right)$. For an agent with the skill θ, period utility of reporting the true type is given by $U\left(\theta\right)=U\left(c\left(\theta\right),y\left(\theta\right)/\theta\right)$. Since $u\left(\theta\right)=U\left(\theta\right)+\beta w\left(\theta\right)$, then

$$u'\left(\theta\right)=m\left(u\left(\theta\right)-\beta w\left(\theta\right)\right)\frac{l\left(\theta\right)^{\gamma}}{\theta}+\beta w_2(\theta). \tag{20}$$

From (18) we can express:

$$c\left(\theta\right)=\frac{1}{\gamma}l\left(\theta\right)^{\gamma}+\bar{U}^{-1}\left(u\left(\theta\right)-\beta w\left(\theta\right)\right). \tag{21}$$

Substituting (20) and (21) in the optimal program (8), we obtain:

$$V_t\left(\hat{w},\hat{w}_2,\theta_{-}\right)=$$
$$\min_{\{l(\theta),u(\theta),w(\theta),w_2(\theta)\}_{\theta\in\Theta}}\int\left(\frac{1}{\gamma}l\left(\theta\right)^{\gamma}+\bar{U}^{-1}\left(\left[u\left(\theta\right)-\beta w\left(\theta|\theta\right)\right]\right)-\theta l\left(\theta\right)+\delta V_{t+1}\left(w\left(\theta\right),w_2\left(\theta|\theta\right),\theta\right)\right)dF\left(\theta|\theta_{-}\right) \tag{22}$$

subject to (12), (13) and (20).

Note that we changed the variables of minimization. Variables $l(\theta)$ and $w(\theta|\theta)$, $w_2(\theta|\theta)$ are now control variables, and $u(\theta)$ is a state variable. We can apply optimal control techniques to characterize (22).

We characterize the solution for the general model in Section 4. However, we first focus in Section 3 on the illustrative example that highlights the key determinants of the dynamic distortions.

3 Illustrative example

In this section, we consider a two-period economy with i.i.d. shocks which are drawn from the same distribution $F(\theta)$. We assume that utility function satisfies

$$U(c,l) = -\frac{1}{\psi}\exp\left(-\psi\left(c - \frac{1}{\gamma}l^{\gamma}\right)\right) \tag{23}$$

where $\psi > 0$. We also assume that the social welfare function is linear, $G(x) = x$.

Most of the derivations in this section is a special case of the general model considered in Section 4, in particularly Propositions 1 and 2. We then compute optimal capital and labor distortions. The analysis of this section provides insights into the nature of the optimal capital and labor distortions that hold in a general model.

3.1 Characterizing optimal wedges

When shocks are i.i.d., analysis of (22) significantly simplifies. The i.i.d. shocks imply that $f_2(\theta|\theta_-) = 0$ for all θ_-. The value function in period 2 has as a state variable only the promised utility, w, instead of the three state variables, w, w_2, θ_-. For a given w, let $\bar{U}_{c,t}(\theta)$ be the marginal utility of consumption of the agent whose period t shock is θ. With exponential preferences (23) $\bar{U}_c = \exp\left(-\psi\left(c - \frac{1}{\gamma}l^{\gamma}\right)\right)$, but we keep a slightly more general notation for the ease of comparison with the results in section 4. After setting up a Hamiltonian to (22) and taking the first order conditions, it can be shown that the optimal labor distortions in period 2 are:

$$\frac{T'_{D,2}(\theta)}{1 - T'_{D,2}(\theta)} = \gamma\frac{1 - F(\theta)}{\theta f(\theta)}\int_{\theta}^{\infty}\left(1 - \frac{\bar{U}_{c,2}(x)}{\lambda_2}\right)\frac{f(x)\,dx}{1 - F(\theta)} \tag{24}$$

where

$$\lambda_2 = \int_0^{\infty}\bar{U}_{c,2}(x)\,f(x)\,dx.$$

In this expression $\bar{U}_{c,2}(x)$ is a function of w and in general $T'_{D,2}(\theta)$ would also be an implicit function of w and, indirectly, of the first period realization of the skill θ_1. With exponential preferences it can be shown that in the solution to (22), $\bar{U}_{c,2}(x)/\lambda_2$ is independent of w, so that $T'_{D,2}(\theta)$ is independent of w and depends only on the realization of θ in period 2.

The expression (24) for the optimal labor distortion in period 2 is identical to that obtained in the static model with quasi-linear preferences as in Diamond (1998). His analysis of the static Mirrlees problem applies in this setting. In particular, it can be shown that $\bar{U}_{c,2}(x) \to 0$ as $x \to \infty$, and the term

$$\int_{\theta}^{\infty} \left(1 - \frac{\bar{U}_{c,2}(x)}{\lambda_2}\right) \frac{f(x)\,dx}{1 - F(\theta)},$$

converges to 1 from below. This expression simplifies further if F has a Pareto tail with the coefficient a. Pareto distribution implies that the term $(1 - F(\theta))/(\theta f(\theta))$ is constant and equal to a^{-1}. For sufficiently large θ the term $T'_{D,2}/(1 - T'_{D,2})$ is increasing and converges to γ/a. Since $T'_{D,2}/(1 - T'_{D,2})$ is increasing in $T'_{D,2}$, it also implies that $T'_{D,2}$ increases for high θ and converges to a positive limit.

The labor distortion in period 1 is:

$$\frac{T'_{D,1}(\theta)}{1 - T'_{D,1}(\theta)} = \gamma \frac{1 - \tilde{F}(\theta)}{\theta \tilde{f}(\theta)} \int_{\theta}^{\infty} \left(1 - \frac{\bar{U}_{c,1}(x)}{\lambda_1}\right) \frac{\tilde{f}(x)\,dx}{1 - \tilde{F}(\theta)}, \tag{25}$$

where

$$\lambda_1 = \int_0^{\infty} \bar{U}_{c,1}(x)\,\tilde{f}(x)\,dx$$

and $\tilde{f}(x) = \Psi(x) f(x)$, where $\Psi(x)$ is

$$\Psi(\theta) = \frac{\exp\left(\beta \int_0^{\theta} -\psi \frac{w'(x)}{\bar{U}_c(x)} dx\right)}{\int_0^{\infty} \exp\left(\beta \int_0^{\tilde{\theta}} -\psi \frac{w'(x)}{\bar{U}_c(x)} dx\right) f(\tilde{\theta}) d\tilde{\theta}}.$$

We now compare the optimal labor distortion in period 1 given by (25) to the labor distortion 1 given by (24). The key difference is in an additional term $\Psi(\theta)$ that re-scales the density $f(\theta)$. This term depends on the promised utility $w(\theta)$ as a function of the current realization of the skill shock θ. Equation (25) shows that the optimal labor distortion in period 1 has the same form as the optimal labor distortion in a static economy in which types are drawn from a distribution $\tilde{F}(\theta)$, with the density $\tilde{f}(\theta)$. We immediately see that for the lowest type in the distribution $\Psi(0) = 1$, and that $\Psi'(\theta) = -\beta\psi \frac{w'(x)}{\bar{U}_c(x)} \Psi(\theta)$.

In general, it is difficult to determine the sign of marginal promised utility, $w'(\theta)$. It is instructive, however, to consider a case when w is increasing for all θ, and $w'(\theta) > 0$ which holds in all our numerical simulations in Section 3.2. In this case, $\Psi'(\theta) \leq 0$ for all θ, and the distribution F has a property:

$$\frac{1-\tilde{F}(\theta)}{\theta \tilde{f}(\theta)} \leq \frac{1-F(\theta)}{\theta f(\theta)},$$

with a strict inequality for interior θ.

It can be shown that similarly to period 2, the marginal utility of consumption in period 1 must be decreasing and converging to zero as the, which implies that $T'_{D,1}$ asymptotically increases to $\gamma \lim_{\theta\to\infty} \frac{1-\tilde{F}(\theta)}{\theta\tilde{f}(\theta)}$. This argument implies that labor distortions for all θ above some threshold $\hat{\theta}$ in period 1 are lower than in period 2.

There is also another force that affects the taxes. Note that, in period 1, the planner generally provides more redistribution than in the second period. The intuition for the additional redistribution is as follows. Let w^S be the value that a Utilitarian social planner can achieve in a static model. When there is an additional period, it is feasible for the planner to set $w(\theta) = w^S$ for all θ. The optimal allocations of labor and consumption in both periods coincides with those in the static economy and achieve the same welfare w^S. However, by varying $w(\theta)$ the planner generally is able to provide higher welfare in period 1. Higher welfare implies more redistribution which flattens $U_{c,1}(\theta)$ relative to the static model (and relative to period 2). The flatter profile of $U_{c,1}(\theta)$ brings the term $U_{c,1}(\theta)/\lambda_{1,t}$ closer to 1 and lowers the value of the integral term in (25). This effect generates a force for lower taxes in period 1.

We now consider the savings distortion. The first order conditions with respect to w in period 1 imply that

$$1-\tau_{S,1}(\theta) = z\left(1-\frac{\psi}{\gamma}T'_{D,1}(\theta)y_1(\theta)\right), \tag{26}$$

where z is a positive constant.

The first important term in the expression (26), $T'_{D,1}(\theta)$, shows that there is a force that increases the savings wedge for the agents with the highest incentive problem. That is for those with a large labor wedge. The intuition behind this force is that the savings distortions is an additional instrument used to alleviate the incentive problem. It is optimal to have higher distortions on the agents who face the most severe incentive constraints. The second term, $y_1(\theta)$, shows that the there is a force that increases the savings wedge for the higher skilled agents who produce a high level of output. The savings wedge is an additional instrument of providing incentives to these agents valuable to the planner. The reason is that the same *rate* of labor distortions leads to a larger output loss when applied to high types than to the low types. Therefore it is optimal to substitute from labor distortions to savings distortions when the social planner provides incentives to the more productive types.

To see the implication of (26) for the asymptotic behavior of $T'_{D,1}(\theta)$, note that when preferences are quasi-linear, equation (16) becomes

$$1 - T'_{D,t}(\theta^t) = \frac{y(\theta^t)^{\gamma-1}}{\theta^\gamma}.$$

Therefore, if $T'_{D,1}(\theta)$ does not converge to 1, $y_1(\theta)$ diverges to infinity. Since the definition of τ_S in (17) implies that $\tau_{S,1}(\theta) \leq 1$, equation (26) can hold only if $T'_{D,1}(\theta)$ converges to either zero or one. A simple perturbation argument can be used to rule out the latter case (see the proof of Proposition 2) which implies that $T'_{D,1}(\theta) \to 0$.

3.2 Numerical illustration

In this subsection, we compute optimal labor and savings wedges for the illustrative two-period example considered above. We study the differences between the labor distortions in a static and this dynamic model. This intuition is helpful to understand the results of the computations in the general economy with the shock processes that are estimated from the US data.

We extend the analysis of Saez (2001) to determine the cross-sectional distribution of skills and the unconditional distribution. We use the 1996 wave of the Panel Study of Income Dynamics (PSID) dataset. We treat heads of households and their spouses as separate observations. Our sample is restricted to include only the observations with the total labor income of at least $\$1,000$ and total hours worked at least 250. We estimate the effective marginal tax rates faced by the individuals using National Bureau of Economic Research's (NBER) program, TAXSIM. We compute individual liabilities under U.S. federal and state income tax laws, by supplying TAXSIM with the individual labor income as well as with other individual data from the PSID such as marital status, dependent exemptions, dividend income, other property income.[2] As in Saez (2001), given the actual effective marginal tax rates, we determine the skill distribution generating the labor income of the agents in the sample. We assume that the elasticity of labor

[2] See Section 4.2 and Appendix 9 for more details on the sample selection and the use of TAXSIM.

supply of 0.5 which implies that $\gamma = 3$. This allows us to compute the implied skill for each type from the individual first-order conditions as follows:

$$\theta_i = \frac{Y_i}{\left(Y_i \left(1 - T' \left(Y_i\right)\right)\right)^{1/\gamma}}. \tag{27}$$

In this equation, Y_i is the labor income of individual i and $T' \left(Y_i\right)$ is the effective marginal tax rate for that individual. Note that with the preferences of the form (23) there are no income effects. Hence, the individual labor supply decision is unaffected by the individual savings choice. The implied skills thus can be determined from the static consumption-labor margin.

We non-parametrically estimate the implied unconditional distribution of skills using a kernel density estimation method. There are two considerations to address. First, the PSID is "top coded". That is, there is an income cutoff level above which no observations are collected. Second, high income individuals are undersampled in the PSID. The analysis of Diamond (1998), Saez (2001) in the static settings, and our results above imply that the upper tail of the distribution is an important determinant of the shape of the optimal tax code. We follow Heathcote, Perri, and Violante (2009) and fit a Pareto tail in our skill distribution above the income level of $150,000. We find a Pareto parameter of 2.68 to be statistically significant at the 1% level.[3]

The coefficient of *absolute* risk aversion, ψ, is set equal to 10. We set the discount factor $\beta = 0.9852$ and chose the marginal rate of transformation across periods $\delta = 1.015$ so that the social planner at the solution of the optimal program chooses not to transfer resources between the two periods.

The results of the numerical simulation are presented in Figure 1. First, consider the optimal marginal labor distortions in period 2 (higher dashed line in Figure 1). The distortions coincide with those in a static economy and are similar to those in Diamond (1998) and Saez (2001). The optimal distortions exhibit a pronounced U-shaped pattern for lower incomes, increase above income of $75,000, reach 52% at income of $700,000 and tend to the analytic limit of 52.81% given the Pareto tail.

Next, consider the optimal marginal labor distortions in period 1 (lower dashed line in

[3] When the tail of the distribution of skills $F(\theta)$ converges to a, the tail of the distribution of income converges to $a/(1+\zeta)$ where ζ is the elasticity of labor supply (see Chapter 2 of Saez (1999)). This is consistent with the estimated coefficient for income distribution in the US at the values of $1.6 - 1.8$ found in Atkinson, Piketty, and Saez (2009).

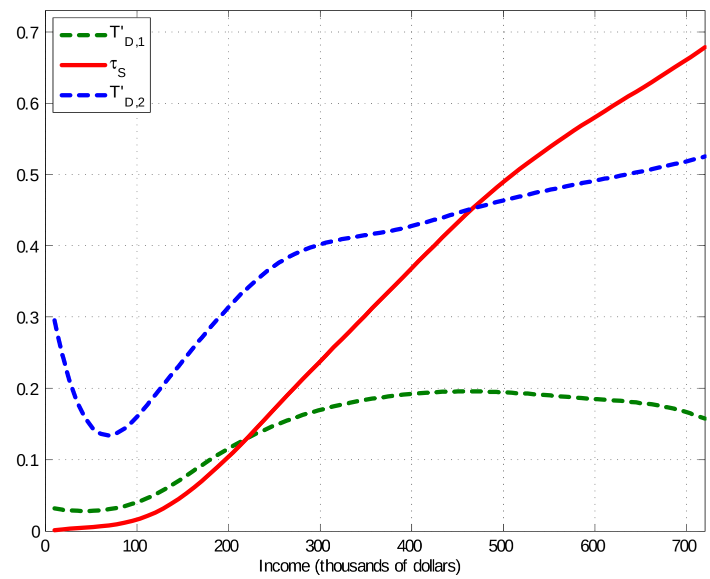

Figure 1: Optimal labor and capital distortions with iid shocks and $T = 2$

Figure 1). These distortions are lower for all individuals compared to those in period 2. This is consistent with the intuition of formulas (24) and (25). The labor distortion start to decline around annual income of $\$430,000$. Consistent with the theoretical results discussed above, they tend to zero at incomes above \$2 million.

Finally, the savings distortion is represented by the solid line in Figure 1. The savings wedge increases for all income levels and ranges from close to zero for low income individuals, increases with income and approaches 68% for income of $\$700,000$. This pattern is consistent with the discussion of the equation (26) whereas the optimal savings distortion is used by the planner to substitute away from the labor distortion for the most valuable high skilled agents.

4 General case with persistent shocks

We now return to the general problem stated in (22).

4.1 Characterizing optimal wedges

Before we state the proposition characterizing optimal wedges we define a coefficient of absolute risk aversion $\psi_t(x) = -\bar{U}_{cc,t}(x)/\bar{U}_{c,t}(x)$.

Proposition 1. *Suppose that $U(c,l)$ satisfies (18).*

Part 1. The optimal labor distortion in period t satisfies

$$\frac{T'_{D,t}(\theta)}{1-T'_{D,t}(\theta)} = \gamma \frac{1-\tilde{F}_t(\theta|\theta_-)}{\theta \tilde{f}_t(\theta|\theta_-)} \int_\theta^\infty \left(1 - \frac{\alpha_t(x)\bar{U}_{c,t}(x)}{\lambda_t}\right) \frac{\tilde{f}_t(x|\theta_-)\,dx}{1-\tilde{F}_t(\theta|\theta_-)} \tag{28}$$

where

$$\tilde{f}_t(\theta|\theta_-) = \frac{\Psi(\theta) f_t(\theta|\theta_-)}{\int_0^\infty \Psi(x') f_t(x'|\theta_-)\,dx'},$$

$$\Psi(\theta) = \exp\left(\beta \int_0^\theta -\psi_t(x) \frac{\omega_{1,t}(x|x)}{\bar{U}_{c,t}(x)} dx\right),$$

$$\tilde{F}_t(\theta) = \int_0^\theta \tilde{f}_t(x)\,dx,$$

$$\lambda_t = \int_0^\infty \alpha_t(x) \bar{U}_{c,t}(x)\,d\tilde{F}_t(x|\theta_-) > 0,$$

and

$$\alpha_t(x) = \begin{cases} G'(\bar{U}_{c,1}(x)) \text{ for } t=1 \\ \left(\zeta_t - \frac{f_{2,t}(x|\theta_-)}{f_t(x|\theta_-)}\right) \text{ for } t>1 \end{cases}$$

Part 2. The savings distortion in period $t < T$ satisfies

$$1 - \tau_{S,t}(\theta) = z_t(\theta)\left(1 - \frac{\psi_t(\theta)}{\gamma} T'_{D,t}(\theta) y_t(\theta)\right) \tag{29}$$

where

$$z_t(\theta) = \frac{\int_0^\infty \left(1 - \tilde{\zeta}_{t+1} \frac{f_{2,t+1}(x|\theta_-)}{f_{t+1}(x|\theta_-)}\right) \bar{U}_{c,t+1}(x)\, \tilde{f}_{t+1}(x|\theta) dx}{\int_0^\infty \bar{U}_{c,t+1}(x)\, f_{t+1}(x|\theta) dx}. \tag{30}$$

and $\tilde{\zeta}_t$ and ζ_t are constants.

The expressions for the optimal labor and savings distortions are similar to those obtained in the two period example in Section 3. Comparing (28) to (25) there are several differences. First, marginal utility $\bar{U}_c(x)$ in (28) is multiplied by the modified social welfare weights $\alpha_1(x) = G'\left(\bar{U}_c(x)\right)$ for $t = 1$, and $\alpha_t(x) = \left(\zeta - \frac{f_2(x|\theta_-)}{f(x|\theta_-)}\right)$ for $t > 1$. The term $G'\left(\bar{U}_c(x)\right)$ captures redistributive objective of the social planner. This terms was not present in (25) as we assumed that planner was Utilitarian in Section 3. For all periods $t > 1$, the additional term $\left(\zeta - \frac{f_2(x|\theta_-)}{f(x|\theta_-)}\right)$ changing the social welfare weights appears because of the Lemma 1. The second difference is that the conditional distribution of types $F_t(\theta|\theta_-)$ rather than the unconditional distribution $F(\theta)$ determines the shape of the skill distribution. Similarly to the i.i.d.

case, this distribution is adjusted by a dynamic term $\Psi(\theta)$. The term $\Psi(\theta)$ depends on the dynamic incentive provision term $\omega_1(\theta|\theta)$ which is a generalization of the static term $w'(\theta)$.

Similarly to the illustrative example in Section 3, we immediately notice that for the lowest type in the distribution $\Psi(0) = 1$ and that $\Psi'(\theta) = -\beta\psi(x)\frac{\omega_{1,t}(x|x)}{\bar{U}_{c,t}(x)}\Psi(\theta)$. Moreover, if $\omega_1(x|x) > 0$, then $\Psi'(\theta) \leq 0$ for all θ and the distribution F has a property:

$$\frac{1-\tilde{F}_t(\theta|\theta_-)}{\theta\tilde{f}_t(\theta|\theta_-)} \leq \frac{1-F_t(\theta|\theta_-)}{\theta f_t(\theta|\theta_-)}, \tag{31}$$

with a strict inequality for interior θ. The assumption that $\omega_{1,t}(x|x) > 0$ is stronger than in the case of i.i.d. shocks. In the i.i.d. case, one generally expects that the report of a higher skill in period t leads to higher rewards by the planner in period $t+1$ because of the dynamic incentive provision. In the case of persistent shocks an additional force is present. Conditional on the high realization of a shock in period t the planner may learn that the agent is likely to be very productive in the future. This may lead to more redistribution away from that type in the future. Still in vast majority of our simulations we found that $\omega_1(x|x) > 0$.

Now consider savings distortions (29). Similar forces determine these distortions as in the example of Section 3. The main difference is that generally terms $z(\theta)$ and $\psi(\theta)$ are endogenous and depend on the state (w, w_2, θ_-).

Next, we characterize asymptotic labor distortions. In section 3 we argued that, in a two period model, labor distortions behave significantly differently from static model and in particular are decreasing and small for high skill types. We now extend this result for the general model of this section.

Proposition 2. *Assume that $U(c,l)$ satisfies (18). Moreover, assume that there exists some $\underline{\psi} > 0$ s.t. $-\bar{U}_{cc}(x)/\bar{U}_c(x) \geq \underline{\psi}$ for all x.*

Part 1. Suppose that $f_t(\theta|\theta_-)$ is independent of θ_- for all t. Then $T'_{D,t}(\theta) \to 0$ as $\theta \to \infty$ for all $t < T$.

Part 2. Consider a family of distributions $f^{\varepsilon}_{t+1}(\theta|\theta_-)$ with a property that $\lim_{\varepsilon\to 0} f^{\varepsilon}_{2,t+1}(\theta|\theta_-) \to 0$ uniformly. Suppose the optimal $\left(w^{\varepsilon}, w^{\varepsilon}_2, T^{\varepsilon\prime}_{D,t}\right)$ are bounded. Then there exists $\xi'' > 0$ s.t. for all $|\varepsilon| < \xi''$ $T^{\varepsilon\prime}_{D,t}(\theta) \to 0$ as $\theta \to \infty$.

Part 3. Suppose that $U(c,l)$ satisfies (23). Suppose that there exists $\varpi > -\infty$ s.t. $f_{t+1,2}(\theta|\theta_-)/f_t(\theta|\theta_-) \geq \varpi$ for all θ, t. Then if there exists $\bar{x}$ s.t. $\alpha_t(x) > 0$ for all $x \geq \bar{x}$ then $T'_{D,t}(\theta) \to 0$ as $\theta \to \infty$.

Proof. In the Appendix. □

The first two parts of this Proposition show that if shocks are either i.i.d. or are sufficiently close to i.i.d. shocks the optimal labor distortions must converge to zero asymptotically in all periods except for the last period. The proof proceeds similarly to the arguments sketched in Section 3. We can show the that first order conditions implies that $T'_{D,t}(\theta)$ must converge to either 0 or 1. Then we consider an argument by contradiction to rule out $T'_{D,t}(\theta) \to 1$. If $T'_{D,t}(\theta) \to 1$, we construct an allocation that does not distort labor supply of any type above some $\bar{\theta}$ and collects as much resources $\int_{\bar{\theta}}^{\infty}(y(\theta) - c(\theta))dF_t(\theta)$ from those types as the original allocation. This perturbation is incentive compatible, leaves utility of all types below $\bar{\theta}$ unchanged, and makes the types above $\bar{\theta}$ strictly better off. With shocks that are i.i.d. or close to i.i.d., this argument implies that the ex-ante welfare must be higher.

Under conditions of Part 3, we can show that $T'_{D,t}(\theta)$ must converge to either 0 or 1 for much more general set of stochastic processes. The same perturbation argument as in i.i.d. case can be used to show that if $T'_{D,t}(\theta) \to 1$, there is an allocation which is incentive compatible in period t and improves expected utility of all types. This perturbation, however, may violate constraint (13). Intuitively, by making high types better off in period t, the planner also makes the deviation in period $t-1$ by high types more desirable, since with persistent shocks they are more likely to be high in period t. Therefore, this perturbation is welfare improving only as long as the Pareto weights $\alpha_t(x)$ are positive. Although it is difficult to find the conditions on the primitives in general, from the definition of α_t it can be easily seen that $\alpha_1(x) > 0$ for all x. Part 3 shows that with exponential preferences result holds for a much large class of Markov processes provided that $f_{2,t}(\cdot|\theta_-)/f_t(\cdot|\theta_-)$ is bounded and $\omega_{1,t}(\theta|\theta) > 0$.

4.2 Numerical simulations with persistent shocks

In this subsection, we provide numerical calculations for the case of the persistent shocks. The calculations are conducted for the same parameters as in the i.i.d. section. The number of time periods is $T = 40$. The details of our calibration and computational techniques can be found in the Appendix. Here, we highlight the main steps and discuss the results.

First, we obtain an entire unconditional (cross-sectional) distribution of skills, F. As in the illustrative example of Section 3.2, the dataset we use is the PSID. Since our fourty-period model has individuals from age 25 to age 65, we start with an empirical distribution of labor income for 25 year old individuals. We then follow the procedure described in the Appendix to impute the distribution of skills. To proceed to compute our main numerical problem with persistence of shocks, we also need to estimate the transition probabilities for the skills. We estimate two different transition probabilities: an earlier age transition for 25-45 year old individuals, and a later age transition for 45-65 year old individuals. That is, we allow age-dependent transition probabilities: younger individuals experience different transitions than older individuals. Within each age group, we assume age-independent transitions.

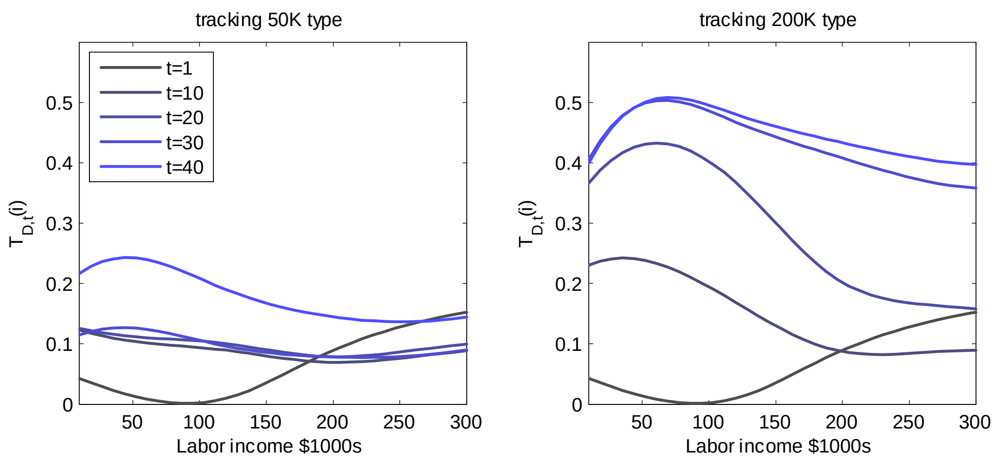

Figure 2: Labor distortions with persistent shocks and $T = 40$.

Figure 2 and Figure 3 present the results of our numerical simulations. Consider first Figure 2. The left panel in this figure presents the labor wedges in period t ($t = 1, 10, 20, 30, 40$ are displayed, with darker, generally lower lines representing earlier periods) for the agent with a history of shocks up to that period such that in each previous period he had income of \$50,000. The right panel displays the labor wedges in period t (once again, $t = 1, 10, 20, 30, 40$ are displayed, with darker, generally lower lines representing earlier periods) for the agent with a history of shocks up to that period such that in each previous period he had income of \$200,000. The lowest, and darkest, lines in each panel are the unconditional labor wedge in period $t = 1$, which is identical in both panels. There are three key features of interest with the labor wedges results.

First, both for the agent with the history of \$50,000 incomes and for the agent with the history of \$200,000 incomes, the average conditional labor wedges are increasing with age. This is consistent with our theoretical findings where the provision of incentives dynamically allows to lower labor wedges early in life.

Second, the conditional labor wedges for the agent with a history of \$50,000 incomes are lower than those for the agent with a history of incomes of \$200,000. There are two forces driving the differences in taxes for these two agents that follow from the discussion of Proposition 1: (i) the differences between conditional and unconditional distributions of skills as well as the differences in conditional distributions among agents, specifically between those who were relatively low income and relatively high income; (ii) the additional redistribution implied by the term $\left(\zeta - \frac{f_2(x|\theta_-)}{f(x|\theta_-)}\right)$ in equation (28). The intuition behind the first force is as follows. From (28), conditional probabilities are one of the key determinants of the optimal marginal labor distortions. Let us consider the estimated probabilities of transition. Consider first an individual with $\$50,000$ income. His income and skills are likely to remain low in the next period. The conditional probability of having the same income is relatively high as the conditional distribution for this agent is very concentrated just above $\$50,000$. For those income levels, the ratio $(1 - F(\theta))/\theta f(\theta)$ for conditional probabilities is lower than for the unconditional probabilities in the i.i.d. case, which is a force driving taxes lower on those types. For higher types the term $1 - F(\theta)$ is close to zero, which drives the labor wedge lower than in the case of unconditional distribution even if such individuals make incomes significantly above $\$50,000$ in the next period. Similarly, an agent who had a relatively high income of $\$200,000$ is likely to receive high income in the next period. Therefore, for such individual the term $f(\theta)$ is small for low θ while the term $1 - F(\theta)$ is large. The ratio $(1 - F(\theta))/\theta f(\theta)$ works towards higher marginal taxes. Once such agents reach sufficiently high income levels where the conditional probability is substantial, labor distortions fall. The second force, is represented by the higher welfare weights that the planner assigns to the agents with relatively low incomes to provide appropriate threat-keeping constraints.

Third, consistent with Proposition 1, the labor wedge decreases for the high incomes. This is one of the important differences in the pattern of labor wedges in the dynamic economy that is in contrast to the static analysis. Recall, that in the static case with the Pareto tail of the skills the labor wedges are increasing and can reach rather high levels.

Next, consider the savings distortions. Just like in the case of labor wedges, Figure 3

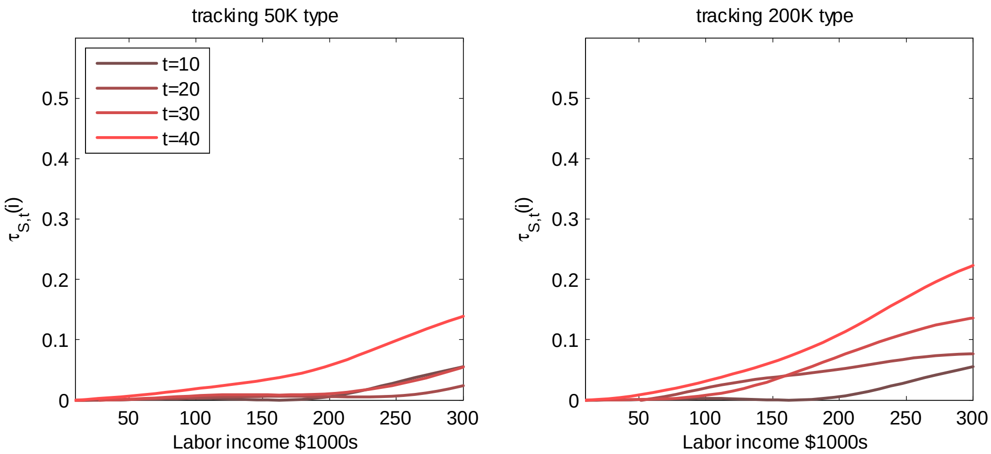

Figure 3: Savings distortions with persistent shocks and $T = 40$.

displays two panels. The left panel in this figure presents the savings wedges in period t ($t = 10, 20, 30, 40$ are displayed, with darker, generally lower lines representing earlier periods) for the agent with a history of shocks up to that period such that in each previous period he had income of \$50,000. The right panel displays the savings wedges in period t (once again, $t = 10, 20, 30, 40$ are displayed, with darker, generally lower lines representing earlier periods) for the agent with a history of shocks up to that period such that in each previous period he had income of \$200, 000. The lowest, and darkest, lines in each panel are the unconditional labor wedge in period $t = 1$, which is identical in both panels. In both cases, the conditional savings distortions are positive and increasing in current period realization of income: they are close to zero for current incomes below \$50, 000 and increase up to 14% and 22% at \$300, 000 income for the agents with a history of \$50, 000 and \$200, 000 incomes respectively.[4]

Finally, we perform the calculations of the welfare gains of using the optimal dynamic non-linear policy. We conduct these in a version of the model with the empirically calibrated persistent shocks that lasts $T = 5$ periods (each period represents 10 years of agent's life). A natural benchmark for comparison is optimal linear taxes.

First, consider the case of the utilitarian social planner. Using the optimal age-dependent linear labor wedges instead of the constrained optimal wedges results in a welfare loss of 0.6%

[4] Recall the classical result of Judd (1985) and Chamley (1986) obtained in representative-agent macroeconomic Ramsey settings: the Chamley-Judd result states that in the long-run capital should go untaxed. Judd (1999) extends that analysis to environments whith no steady state.

of consumption. The optimal age-*in*dependent labor distortions increase the welfare loss to 1.4%. While these magnitudes are non-trivial, linear taxes can still yield reasonably good policies. This is a well-known result in numerical simulations of the the static Mirrlees models (e.g.,Mirrlees (1971), Atkinson and Stiglitz (1976), Tuomala (1990)) who find that linear taxes with utilitarian social planner approximate the optimal policy rather well. Additionally, we find that age-dependence cuts the welfare loss by more than half.

The static literature also points out that if the planner is more redistributive than utilitarian, the tax policy is substantially different from linear, and non-linear taxes may yield large welfare gains. In particular, we calculate the welfare gains of using optimal policies when the social planner is Rawlsian. The optimal age-dependent linear labor wedges yield a welfare loss of 4.5% of consumption compared to the constrained optimum. The optimal age-*in*dependent labor distortion yields a welfare loss of 5.1%. We conclude that the welfare gains of using optimal non-linear policies are significant.

5 Extensions and generalizations

In this subsection, we further explore the relationship between dynamic and static models. First we focus on the case when shocks are i.i.d..

An important point to note that the recursive formulation of the dynamic model is equivalent to a static model with two goods: consumption, c, and promised utility, w. These two goods are perfect substitutes in production. The preferences over these two goods are given by

$$\bar{U}\left(c - \frac{l^\gamma}{\gamma}\right) + h(w)^5. \tag{32}$$

An equivalent way to think about the dynamic economy (in the recursive formulation) is as of a static economy in which a planner imposes a non-linear tax on l and w. We now compare the asymptotic properties of labor taxes in a two good model versus a one good model.

First, consider a static model with preferences given by (18). Suppose that the marginal labor distortion converges to a linear tax $T' < 1$ for sufficiently high types. As we argued above this tax implies that $\bar{U}_c(\theta) \to 0$. Asymptotically, the planner does not value utility allocated to the highest types. For this reason, the optimal tax extracts the maximal revenues from the

[5] Recall that with iid shocks the value function $V_t(w, w_2, \theta_-)$ simplifies to $V_t(w)$. In this case $h(w) = V_t^{-1}(w)$.

high types to allocate to the lower types.

Consider a perturbation in which for types between $[\theta^*, \theta^* + d\theta]$ the planner raises the marginal taxes from T' to $T' + d\tau$. Note that since preferences are quasi-linear such a perturbation does not affect labor supply of any types above $\theta^* + d\theta$, and it raises their tax liability by $dy(\theta)\, d\tau$. The increase in tax revenues is given by $M = d\tau dy(\theta) \int_{\theta^*}^{\infty} f(\theta) d\theta$. It can be shown that $dy(\theta) = (1+\zeta)\frac{y(\theta)}{\theta} d\theta$, so that

$$M = d\tau d\theta\, (1+\zeta)\, \frac{y(\theta^*)}{\theta}(1 - F(\theta^*)),$$

where ζ is the elasticity of labor supply.[6]

All the types in the interval $[\theta^*, \theta^* + d\theta]$ reduce their labor supply by $dy(\theta^*) = -\zeta \frac{y(\theta^*)}{1-T'} d\tau$. The total revenue loss from the changes in the labor supply is given by

$$B = \frac{T'}{1-T'} y(\theta^*) f(\theta^*) d\tau d\theta.$$

If T' is chosen optimally, then this perturbation should leave the tax revenues unchanged, so that $M + B = 0$ or

$$\frac{T'(\theta^*)}{1 - T'(\theta^*)} = \left(1 + \frac{1}{\zeta}\right) \frac{1 - F(\theta^*)}{\theta^* f(\theta^*)} = \gamma \frac{1 - F(\theta^*)}{\theta^* f(\theta^*)}.$$

For general preferences $U(c,l)$ the analysis is similar except that now there is an additional income effect that affects the labor supply of agents above $\theta^* + d\theta$ (see Saez (2001)).

Consider a two-good economy with preferences (32). In this case, there is an additional tax on good w, $P(w|y)$, which we assume to converge to P' for high types. The same perturbation of taxes T' generally decreases consumption of good w through income effects by amount $dw(\theta)/d\tau$. There is an additional revenue effect $P' \int_{\theta}^{\infty} \frac{dw(\theta)}{d\tau} f(\theta) d\theta$. If $P' > 0$, this effect decreases tax revenues and results in lower optimal T'. From the static multi-good analysis (see Mirrlees (1986)), $P' > 0$ is optimal if $U_{cl} < 0$, which is satisfied in the case of quasi-linear preferences.

This reasoning suggests that the cross partial elasticity U_{cl} is important to understand the distinction between the static and dynamic economies. We investigate this further and consider general preferences $U(c,l)$, where U is increasing, twice differentiable and jointly concave in c

[6] In particular, $\zeta = 1/(\gamma - 1)$.

and $-l$. To make comparisons with the previous result more straightforward, we assume that $U(c,l)$ satisfies

$$U_{cl}\frac{U_l}{U_c} = U_{cl}\frac{U_{cl}}{U_{cc}}. \tag{33}$$

We also define a Frisch elasticity of labor supply η^{Fr}:

$$\eta^{Fr} = \frac{U_l}{l\left(U_{ll} - U_{cl}^2/U_{cc}\right)}.$$

In general, η^{Fr} depends on the allocations $c(\theta)$ and $y(\theta)$ as well as the type θ. We denote η^{Fr} evaluated at the optimal values of $c(\theta)$ and $y(\theta)$ for type θ by $\eta^{Fr}(\theta)$. For many standard preferences, η^{Fr} is constant. For example, $1 + 1/\eta^{Fr} = \gamma$ in the quasi-linear utility case and in a separable utility case $U(c,l) = U(c) - \frac{l^\gamma}{\gamma}$.

Proposition 3. *Suppose that $U\left(c,l\right)$ satisfies (33). Then*

$$\frac{T'_{D,t}\left(\theta\right)}{1-T'_{D,t}\left(\theta\right)} = \left(1+\frac{1}{\eta^{Fr}(\theta)}\right)\frac{1-\tilde{F}_t\left(\theta|\theta_-\right)}{\theta\tilde{f}_t\left(\theta|\theta_-\right)}\int_\theta^\infty \frac{U_c(\theta)}{U_c(x)}\left(1-\frac{\alpha_t(x)\bar{U}_{c,t}\left(x\right)}{\lambda_t}\right)\frac{\tilde{f}_t\left(x|\theta_-\right)dx}{1-\tilde{F}_t\left(\theta|\theta_-\right)}$$

where $\tilde{f}_t, \alpha_t, \lambda_t$ are defined as in Proposition 1, and

$$\Psi(\theta) = \exp\left(\int_0^\theta -\frac{U_{cl}(x)}{U_c(x)U_l(x)}\left(U'_t(x)+\beta\omega_{1,t}\left(x|x\right)\right)dx\right).$$

When $U(c,l)$ is quasi-linear, as in the analysis above, it is easy to show that the same result as in the previous analysis. When $U(c,l)$ does not satisfy (33), the optimal labor distortions satisfy the same equation but $\eta^{Fr}(\theta)$ cannot longer be interpreted as a Frisch elasticity.

This proposition shows the influence of the cross-partial U_{cl} on the optimal labor wedges. Note that the only place in which a dynamic term appears in the expressions is the term $\omega_1(x|x)$ in the expression for $\Psi(\theta)$. When $U_{cl} < 0$, it creates a force that makes the tail ratio $\frac{1-\tilde{F}_t(\theta|\theta_-)}{\theta\tilde{f}_t(\theta|\theta_-)}$ thinner and lowers the optimal labor wedges. When $U_{cl} > 0$, there is a force that makes the tail ratio fatter and increases the optimal labor wedges.

When preferences are separable, $U_{cl} = 0$, the dynamic labor wedges are similar to those in static model. In this case $\Psi(\theta) = 1$ and $\tilde{f} = f$. With i.i.d. shocks (or as long as $\alpha_t \geq 0$ with persistent shocks) the integral $\int_\theta^\infty \frac{U_c(\theta)}{U_c(x)}\left(1-\frac{\alpha_t(x)\bar{U}_{c,t}(x)}{\lambda_t}\right)\frac{\tilde{f}_t(x|\theta_-)dx}{1-\tilde{F}_t(\theta|\theta_-)}$ goes to 1 for sufficiently high types, and labor wedges converge as in the static model to

$$\left(1+\frac{1}{\eta^{Fr}(\theta)}\right)\frac{1-F_t\left(\theta|\theta_-\right)}{\theta f_t\left(\theta|\theta_-\right)},$$

and with $U(c,l) = U(c) - \frac{l^\gamma}{\gamma}$ this expression becomes $\gamma\frac{1-F_t(\theta|\theta_-)}{\theta f_t(\theta|\theta_-)}$.

6 Consolidated Income Accounts - A Theory of Decentralization

Previous sections characterized the optimal labor and savings distortions (wedges). In dynamic Mirrleesian taxation models, optimal wedges do not necessarily coincide with the taxes implementing the optimal allocations (see, e.g., Kocherlakota (2005), Albanesi and Sleet (2006), Golosov and Tsyvinski (2006)). In this section, we provide a novel implementation of the optimal allocations – a consolidated income accounts (CIA) tax system. In a given period, the labor income tax depends on labor income and on the balance of the CIA, the savings tax depends only on the amount of savings; the CIA balance is updated as a function of the labor income and the previous balance. We then show that the CIA system takes a particularly simple form if the utility is exponential and the shocks are i.i.d. The tax system consists of a non-linear tax on capital income, non-linear labor income tax, and a CIA account. In each period a taxpayer can deduct the balance of the account from the total income tax bill. Thus, while all agents with the same labor income are facing the same marginal tax rate, the total tax bill is smaller for the agents with a higher CIA balance. Updating the CIA balance follows a simple rule. In each period a change in the CIA balance is determined solely by the individual's labor income in that period. Moreover, if the distribution of the shocks does not change over time, taxes do not depend on age of the agent. We conclude with a numerical simulation of the optimal taxes. In our simulation, labor income taxes are lower than labor income distortions as they incorporate the effects of updating the CIA balance.

6.1 Implementation in a general case

We consider a T period model of the economy described in the previous sections. The optimal allocations $\left\{c_t^*\left(\theta^t\right), y_t^*\left(\theta^t\right)\right\}_{\theta^t \in \Theta^t}$ solve a generalized version of (2):

$$\max_{\left\{c_t\left(\theta^t\right), y_t\left(\theta^t\right)\right\}_{\theta^t \in \Theta^t}} \mathbb{E}_0 \left\{\sum_{t=1}^{T} \beta^t U\left(c_t\left(\theta^t\right), y_t\left(\theta^t\right)/\theta_t\right)\right\} \tag{34}$$

subject to the incentive compatibility constraint

$$\begin{aligned} & \mathbb{E}_0 \left\{\sum_{t=1}^{T} \beta^t U\left(c_t\left(\theta^t\right), y_t\left(\theta^t\right)/\theta_t\right)\right\} \\ \geq\ & \mathbb{E}_0 \left\{\sum_{t=1}^{T} \beta^t U\left(c_t\left(\sigma\left(\theta^t\right)\right), y_t\left(\sigma_t\left(\theta^t\right)\right)/\theta_t\right)\right\} \text{ for all } \sigma_t\left(\theta^t\right), \end{aligned} \tag{35}$$

and the feasibility constraint

$$\mathbb{E}_0 \sum_{t=1}^{T} \delta^{t-1} c_t\left(\theta^t\right) \leq \mathbb{E}_0 \sum_{t=1}^{T} \delta^{t-1} y_t\left(\theta^t\right).$$

As in Kocherlakota (2005), we impose the following assumption on the optimal allocations.

Assumption 1. *For any θ^{t-1}, if $\theta \neq \theta'$ then $y^*\left(\theta^{t-1}, \theta\right) \neq y^*\left(\theta^{t-1}, \theta'\right)$.*

The tax system that we propose consists of three elements. First, taxes depend on a consolidated labor income account (CIA) which keeps track of agents' past earning. We denote by $\omega_t \in \mathbb{R}$ the balance on that account, which is updated according to a rule

$$\omega_t = g_t\left(y_t | \omega_{t-1}\right), \tag{36}$$

where $g_t : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$. That is, given a previous balance ω_{t-1} and the current amount of labor income y_t, the agent is assigned a new CIA balance of ω_t according to the function g_t. The second element of the tax system is a nonlinear tax on labor income y_t in period t, $T_t\left(y_t | \omega_t\right)$, where $T_t : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$. The third element of the tax system is a savings tax $\tau_t\left(k_t | \omega_t\right)$, where $\tau_t : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$.

Next, we formally define the CIA tax system.

Definition 1. *A CIA tax system, $\left\{\left(\omega_t, g_t\left(y_t|\omega_{t-1}\right)\right), T_t\left(y_t|\omega_t\right), \tau_t\left(k_t\right)\right\}_t$, consists of, $\forall t$: (1) the updating rule for the CIA balance $g_t\left(y_t|\omega_{t-1}\right)$ with the associated balance $\omega_t = g_t\left(y_t|\omega_{t-1}\right)$ and the initial balance $\omega_0 = 0$, (2) the labor income tax $T_t\left(y_t|\omega_t\right)$, and (3) the savings tax $\tau_t\left(k_t\right)$.*

We now formally define the notion of implementation.

Definition 2. *A CIA tax system $\left\{\left(\omega_t, g_t\left(y_t|\omega_{t-1}\right)\right), T_t\left(y_t|\omega_t\right), \tau_t\left(k_t\right)\right\}_t$* implements an allocation *$\left\{\tilde{c}_t\left(\theta^t\right), \tilde{y}_t\left(\theta^t\right)\right\}_{\theta^t \in \Theta^t}$ if this allocation solves*

$$\max_{\left\{c_t\left(\theta^t\right), y_t\left(\theta^t\right), k_{t+1}\left(\theta^t\right)\right\}_{\theta^t \in \Theta^t}} \mathbb{E}_0 \left\{\sum_{t=1}^{T} \beta^t U\left(c_t\left(\theta^t\right), y_t\left(\theta^t\right)/\theta_t\right)\right\} \tag{37}$$

subject to the budget constraint

$$c_t\left(\theta^t\right) + k_{t+1}\left(\theta^t\right) \leq y_t\left(\theta^t\right) + \delta^{-1} k_t\left(\theta^{t-1}\right) - T_t\left(y_t\left(\theta^t\right)|\omega_t\right) - \tau_t\left(k_t\left(\theta^{t-1}\right)\right), \forall t,\ \theta^t \in \Theta^t \text{ and } \theta^t \geq \theta^{t-1}, \tag{38}$$

the updated balances given by (36), and $k_1, \omega_0 = 0$.

Note that the agents' problem above assumes that the agent can borrow or lend at the interest rate δ^{-1}.[7] We are now ready to prove the main result of this section that a CIA system implements the optimum. The proof of the theorem shows how to construct this system.

Theorem 1. *Suppose that solution to the optimal program (34), $\left\{c_t^*\left(\theta^t\right), y_t^*\left(\theta^t\right)\right\}_{\theta^t\in\Theta^t}$, satisfies Assumption 1. Then there exists a CIA system that implements $\left\{c_t^*\left(\theta^t\right), y_t^*\left(\theta^t\right)\right\}_{\theta^t\in\Theta^t}$.*

Proof. We prove this theorem in two steps. The first step is to construct a tax system $\tilde{T}_t\left(y|y_1,...,y_{t-1}\right), \tau_t\left(k\right)$ in which labor income taxes in period t depend on the whole history of labor earnings. We show that such a system implements the solution to (34),$\left\{c_t^*\left(\theta^t\right), y_t^*\left(\theta^t\right)\right\}_{\theta^t\in\Theta^t}$. In the second step, we invoke a standard set theory result to show that the constructed tax system can be equivalently written as $g_t\left(y|\omega_{t-1}\right), T_t\left(y|\omega_{t-1}\right), \tau_t\left(k\right)$. That is, it is sufficient to condition the tax system on the consolidated income account.

Step 1.

We start by recursively constructing the amounts of savings in the optimal allocation:

$$\begin{aligned} &k_1^* = 0, \\ &k_t^* = \delta^{-1}k_{t-1}^* + \mathbb{E}_0\left\{y_t^*\left(\theta^t\right) - c_t^*\left(\theta^t\right)\right\}, \ \forall t > 1. \end{aligned} \tag{39}$$

For any history $\theta^t \in \Theta^t$, construct the labor tax

$$\tilde{T}_t\left(y_t^*\left(\theta^t\right)|y_1^*\left(\theta^1\right),...,y_{t-1}^*\left(\theta^{t-1}\right)\right) = y_t^*\left(\theta^t\right) - c_t^*\left(\theta^t\right) + \delta^{-1}k_t^* - k_{t+1}^*$$

and set

$$\tilde{T}_t\left(y|y_1^*\left(\theta^1\right),...,y_{t-1}^*\left(\theta^{t-1}\right)\right) = \infty$$

if there exists no $\theta \in \Theta_t$ such that $y = y_t^*\left(\theta^{t-1},\theta\right)$. Note that this tax is well defined because of Assumption 1 since we can invert history of incomes $y_1^*,...,y_{t-1}^*$ into the sequence of shocks θ^{t-1}.

[7] To derive the implications for taxes implementing the optimum one needs to take a stand on what private insurance contract are available. As it is well known in the literature (see, e.g., Prescott and Townsend (1984) or Golosov and Tsyvinski (2007)), if unrestricted multi-period contracts are allowed, the competitive equilibrium allocations are efficient. At the same time, in practice private markets do not provide efficient insurance (see, e.g., Kocherlakota (2005) for some of the arguments). To better understand the implications of the idiosyncratic uncertainty on the optimal non-linear taxes, we consider an important benchmark in which without government intervention no insurance is available to the households beyond borrowing and lending with a risk free rate. This is the most commonly used benchmark in the literature on implementation.

We now construct savings taxes following the contribution by Werning (2009). Let $\tau_t(k_t^*) = 0$. First, consider a savings decision in period $t = 1$ that is different from the optimal amount k_2^*. We denote such saving by $(k_1^* + \varepsilon)$ for some $\varepsilon \neq 0$. Let $W_1(\varepsilon, \tau)$ denote the highest ex-ante utility that any agent can achieve who saves ε additional units in period $t = 1$ on which he pays tax τ in period $t = 2$:

$$\begin{aligned} &W_1(\varepsilon,\tau) = \\ &\qquad \sup_{\sigma(\theta^T)\in\Sigma} \mathbb{E}_0 \left\{ U\left(c_1^*\left(\sigma_1\left(\theta^1\right)\right) - \varepsilon, \frac{y_1^*\left(\sigma_1\left(\theta^1\right)\right)}{\theta_1}\right) \right. \\ &\qquad + \beta U\left(c_2^*\left(\sigma_2\left(\theta^2\right)\right) + \delta^{-1}\varepsilon - \tau, \frac{y_2^*\left(\sigma_2\left(\theta^2\right)\right)}{\theta_2}\right) \\ &\qquad \left. + \sum_{t=3}^{T} \beta^{t-1} U\left(c_t^*\left(\sigma_t\left(\theta^t\right)\right), \frac{y_t^*\left(\sigma\left(\theta^t\right)\right)}{\theta_t}\right) \right\} \end{aligned}$$

Note from (35) that

$$W_1(0,0) = \mathbb{E}_0 \left\{ \sum_{t=1}^{T} \beta^{t-1} U\left(c^*\left(\theta^t\right), y^*\left(\theta^t\right)/\theta_t\right) \right\}.$$

Since $W_1(\varepsilon, \tau)$ is monotonically decreasing in τ (for a given amount of savings in period $t = 1$, ε), there exists τ_ε^* such that the agent is indifferent between the best deviation and the optimal allocation:[8]

$$W_1(\varepsilon, \tau_\varepsilon^*) = W_1(0,0).$$

For all ε, set $\tau_2(k_1^* + \varepsilon) = \tau_\varepsilon^*$

Similarly, define by $W_2(\varepsilon, \tau)$ the utility under best deviation (for a given amount of savings ε) for period $t = 2$:

$$\begin{aligned} W_2(\varepsilon,\tau) = &\sup_{\sigma(\theta^T)\in\Sigma,\varepsilon_1} \mathbb{E}_0 \left\{ U\left(c_1^*\left(\sigma_1\left(\theta^1\right)\right) - \varepsilon_1, \frac{y_1^*\left(\sigma_1\left(\theta^1\right)\right)}{\theta_1}\right) \right. \\ &+ \beta U\left(c_2^*\left(\sigma_2\left(\theta^2\right)\right) + \delta^{-1}\varepsilon_1 - \tau_2\left(k_1^* + \varepsilon_1\right) - \varepsilon, \frac{y_2^*\left(\sigma_2\left(\theta^2\right)\right)}{\theta_2}\right) \\ &+ \beta^2 U\left(c_3^*\left(\sigma_3\left(\theta^3\right)\right) + \delta^{-1}\varepsilon - \tau, \frac{y_3^*\left(\sigma_3\left(\theta^3\right)\right)}{\theta_3}\right) \\ &\left. + \sum_{t=4}^{T} \beta^{t-1} U\left(c_t^*\left(\sigma_t\left(\theta^t\right)\right), \frac{y_t^*\left(\sigma\left(\theta^t\right)\right)}{\theta_t}\right) \right\} \end{aligned}$$

[8] To be absolutely precise, this also requires $W_1(\varepsilon, 0) \geq W_1(0,0)$, which is clearly true here.

Given the amount of savings in period $t = 2$, ε, and the tax in period $t = 3$, τ, the agent chooses the reporting strategy $\sigma\left(\theta^T\right) \in \Sigma$ and the savings in period $t = 1$, ε_1. The agent pays the savings tax $\tau_2\left(k_1^* + \varepsilon_1\right)$ on savings ε_1.

As before, we find τ_ε^* (for a given amount of savings in period $t = 2$, ε) such that an agent is indifferent between the best deviation and the optimal allocation:

$$W_2\left(\varepsilon, \tau_\varepsilon^*\right) = W_2\left(0, 0\right)$$

We then set the savings tax in period $t = 3$, $\tau_3\left(k_2^* + \varepsilon\right) = \tau_\varepsilon^*$. We proceed by induction to define $\tau_t\left(k_t\right)$ for all $t > 1$.

Now, we verify that constructed tax system $\left\{\tilde{T}_t, \tau_t\right\}_t$ implements the optimum. Consider any sequence of labor incomes $y_t\left(\theta^t\right)$, $\forall \theta^t \in \Theta^t$. We want to show that for any such sequence, and the equilibrium choice of consumption $c\left(\theta^t\right)$ and savings k_t, agent's utility is weakly less than utility from choosing the optimal allocation $\left\{c_t^*\left(\theta^t\right), y_t^*\left(\theta^t\right)\right\}_{\theta^t \in \Theta^t}$ and the corresponding sequence $\left\{k_t^*\right\}_t$. First, we can restrict our attention to the sequences $y_t\left(\theta^t\right)$ that satisfy $y_t\left(\theta^t\right) = y_t^*\left(\hat{\theta}^t\right)$ for some $\hat{\theta}^t$. From the definition of the labor taxes, all other sequences involve labor taxes sufficiently large that an agent does not choose them. Construction of capital taxes implies that the agent's choice of savings satisfies $k_t = k_t^*$, for all t. This implies that the agent's choice of consumption $c_t\left(\theta^t\right) = c_t^*\left(\hat{\theta}^t\right)$. Let $\hat{\sigma}_t\left(\theta^t\right) = \hat{\theta}^t$ for all t. Then the utility of the agent from such a labor choice is

$$\mathbb{E}_0\left\{\sum_{t=1}^{T} \beta^t U\left(c\left(\hat{\sigma}\left(\theta^t\right)\right), y\left(\hat{\sigma}\left(\theta^t\right)\right)/\theta_t\right)\right\}$$

which, by (35), is less then his utility of optimal allocation.

$$\mathbb{E}_0\left\{\sum_{t=1}^{T} \beta^t U\left(c_t^*\left(\theta^t\right), y_t^*\left(\theta^t\right)/\theta_t\right)\right\}.$$

In other words, the above reasoning is as follows. Labor taxes are constructed such that agent chooses income from the menu of the optimal allocations under different strategies. Capital taxes are constructed such that the agent's savings and consumption are among the menu of the optimal allocations. Finally, the incentive compatibility constraint insures that the best choice of such consumption and labor menu is indeed the optimal allocation.

Step 2.

We now construct a CIA tax system. Recall that a t-dimensional space of real numbers is equivalent to $\mathbb{R}$, i.e., there is a bijection $q_t : \mathbb{R}^t \to \mathbb{R}$ (see, e.g. Problem 9 on page 20 in Kolmogorov and Fomin (1975)). Let $\omega_t = q^t(y_1, ..., y_t)$ and denote by q_t^{-1} the inverse of q_t. Define taxes as

$$T_t(y|\omega_{t-1}) = \tilde{T}_t\left(y|q_{t-1}^{-1}(\omega_{t-1})\right)$$

and

$$\tau_t(k|\omega_{t-1}) = \tau_t\left(k|q_{t-1}^{-1}(\omega_{t-1})\right).$$

Finally define a consolidated labor income account as

$$g_t(y|\omega_{t-1}) = q_t\left(q_{t-1}^{-1}(\omega_{t-1}), y\right).$$

It is clear from the construction that this system implements the same allocations as the tax system constructed in Step 1. $\square$

Theorem 1 is quite general. In particularly, as long as Assumption 1 is satisfied, the theorem does not depend on the details of the stochastic process for skills or the form of preferences and holds irrespective of whether only local incentive constraints bind. However, despite the simplicity of construction, the tax functions would not be necessarily well behaved with respect to ω_t. For example, the proof does not say anything about whether these functions are continuous, monotonic, or differentiable.

Theorem 1 establishes that the planner needs relatively limited amount of information to design taxes that implement the optimal allocation. It is possible to set up a rule for the consolidated income account that keeps track of a one dimensional summary statistics of the past labor incomes, together with nonlinear taxes on labor (which is a function of that summary statistics) and savings taxes to achieve the constrained-efficient allocations.

6.2 Implementation in a simple case

At the level of generality in the previous section, it is difficult to characterize the CIA rule g_t or the dependence of labor taxes on the CIA balance $T_t(y_t|\omega_t)$ in more details. We turn to a special case of our model which allows to understand the properties of a CIA system in simple cases.

Proposition 4. *Suppose θ_t are i.i.d. with the distribution $F(\theta_t)$, and the utility satisfies (23). Then there exists a CIA tax system that implements the optimal allocation and satisfies*

$$T_t(y_t|\omega_{t-1}) = T_t(y_t) - \omega_{t-1},$$

and

$$\omega_t = g_t(y_t) + \omega_{t-1}.$$

Proof. In the Appendix □

This Proposition shows that when shocks and utility satisfy assumptions of Proposition 4, the tax system takes a particularly simple form. It consists of a non-linear tax on capital income, a non-linear labor income tax, and a CIA account. In each period a taxpayer can deduct the balance of the account from the total income tax bill. Thus, while all agents with the same labor income are facing the same marginal tax rate, the total tax bill is smaller for the agents with a higher CIA account. Similarly, rules for the updating the CIA balance follow a simple rule. In each period the increase on your account is determined solely by the individual's labor income in that period.

To gain further understanding of the CIA tax system implementing the optimal allocation in this case, we provide a numerical simulations in a two period version of our economy with i.i.d. shocks. We compute T and g functions for an economy with the parameters described in Section 3.2 and present them in Figure 4.

The marginal CIA function g' is represented by a dashed line in Figure 4. The function g' is at a minimum of 3% at the level of income of \$10,000. It then reaches 10% at \$61,000 income, then increases reaching a maximum of 11% at \$112,000 income, and then declines for higher incomes reaching 8% at \$300,000 income. We note that for higher income levels - at least above \$50,000 - the marginal CIA function g' stays close to 10%. The value of g' equal to 10% has the following interpretation: an agent's \$1 of additional income in period $t = 1$ reduces the agent's tax liability in period $t = 2$ by 10 cents in real terms.

The solid lines in Figure 4 plot the marginal labor tax functions T_1' and T_2'. We note that the marginal labor tax functions T_1' and T_2' strongly resemble the shapes of the respective optimal labor distortions we obtained in Section 3.2.

To facilitate the analysis, we reproduce optimal marginal labor distortions from subsection 3.2 together with the marginal labor tax functions T_1' and T_2' in Figure 5 for comparison. The

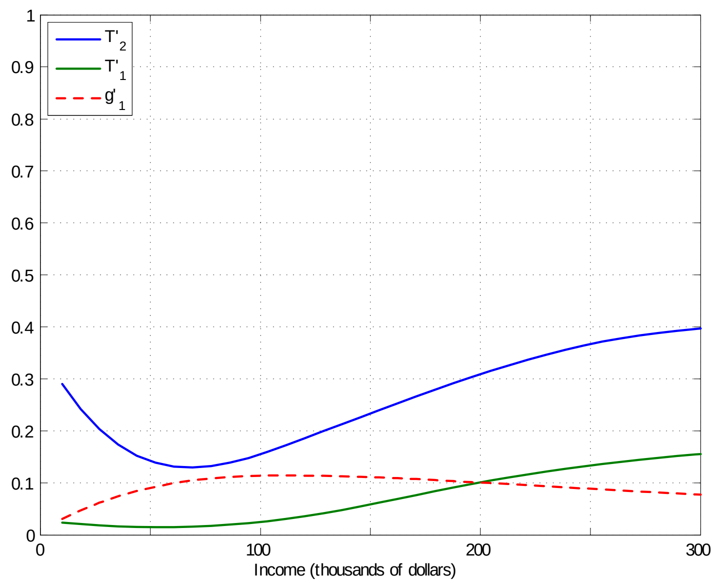

Figure 4: Implementation with a CIA tax system.

labor distortion in period $t = 1$ is represented by a dashed line in Figure 5. A dot-dashed line in Figure 5 - which coincides with T_2' - represents the labor distortion in period $t = 2$.

First, notice is that the marginal labor taxes in period $t = 1$, $T_1'(y)$, are lower than the marginal labor distortions, $T_{D,1}'(y)$. To understand this result, consider the first order conditions for problem (37):

$$-U_l \frac{1}{\theta} = \left(1 - T_1'(y_1)\right) U_c + \zeta g_1'(y_1),$$

where ζ is a Lagrange multiplier on the balance of the CIA (36). As we saw from Figure 1, in the optimal allocation agents with higher y_1 receive higher period $t = 2$ utility, which, from construction of g_1, implies that $g_1'(y_1) > 0$. Therefore

$$T_1'(y_1) < T_{D,1}'(y_1).$$

The labor taxes are lower than then the distortions characterized in equation (25). This relationship is not surprising. The labor choice of individuals is distorted by both the marginal tax rate on labor income, and by the distortions arising from the CIA. The CIA is used to provide incentives dynamically – agents with higher income are assigned a higher CIA balance. The sum of the two distortions is equal to the optimal labor distortion characterized in Section 3. Therefore, the marginal tax rate are lower than the optimal labor distortions.

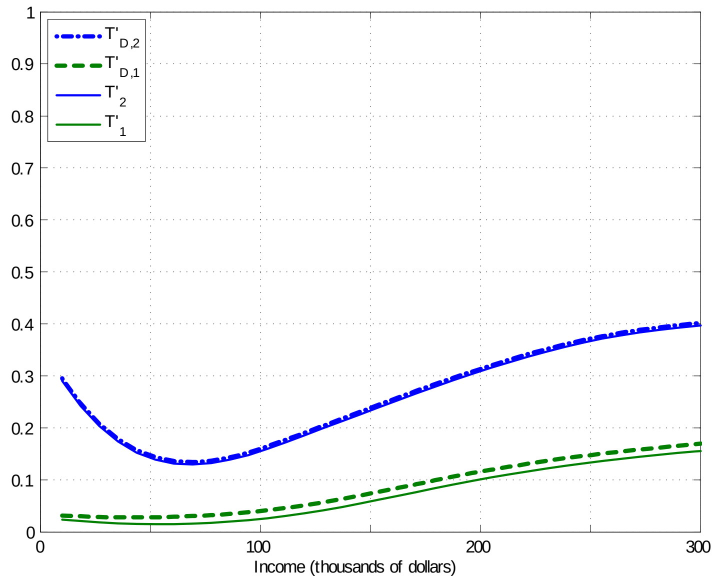

Figure 5: Implementation with a CIA tax system vs. optimal labor distortions.

We now compare our decentralization to the ideas of cumulative income tax averaging proposed by Vickrey (1939) and Vickrey (1947). One of the main motivations behind Vickrey's proposal was to avoid *"inequality of burden as between taxpayers of fluctuating and of steady incomes"*. Although the implementation details of our proposal differ from those of Vickrey's, conceptually they are closely related. Table 1 compares two individuals, with a similar present value realization of labor income, and the total labor taxes they pay. The present value of taxes for these individuals are similar, which is in line with Vickrey's motivation for the cumulative average income tax. The details differ. Under Vickrey's proposal, an individual in period $t = 2$ should pay tax based on the lifetime present value of the labor income. In our decentralization, the labor tax in period $t = 2$ is on labor income in that period, but the first individual gets a bigger CIA credit for the higher taxes he paid in period $t = 1$. Our system is a natural extension of Vickrey since it allows to provide better insurance against idiosyncratic shocks by incorporating savings distortions.

Table 1. Two realizations of labor income, labor taxes, and CIA balances (optimal allocation is computed in Section 3.2, units are normalized so that $y_1(\theta_1 = 10) = 1$)

	θ_1	θ_2	$y_1(\theta_1) + \delta y_2(\theta_2)$	$T_1(y_1) + \delta T_2(y_2)$	$g_1(y_1)$
Realization 1:	10	9	1.66	0.64	0.14
Realization 2:	9	10	1.63	0.58	0.09

It is also relatively easy to extend the implementation to the case of the persistent shocks in the environment with exponential utility. There, the tax functions would also need to keep track of the CIA depending on the previous realization of income.

7 Conclusion

This paper provides a methodology to study the determinants of optimal distortions and taxes using the first-order conditions of the optimal mechanism design problem. The dynamic optimal taxes differ significantly from the static ones. Our formulas for the labor and the savings wedges show the forces determining these wedges. We then provide numerical simulations for a realistically calibrated economy. Finally, we derive a novel implementation using consolidated income accounts.

8 Appendix

Preliminary and incomplete

8.1 Proof of Lemma 1

Consider a Hamiltonian to (10) and use (14) to substitute for $w(\theta)$

$$
\begin{aligned}
H &= -\left(c(\theta) - y(\theta) + \delta V_{t+1}\left(\beta^{-1}\left(u(\theta) - U(c(\theta), y(\theta)/\theta)\right), w_2(\theta), \theta\right)\right) f_t(\theta|\theta_-) \\
&\quad + \mu(\theta) U_l(c(\theta), y(\theta)/\theta)\left(-\frac{y(\theta)}{\theta^2}\right) + \beta w_2(\theta) \\
&\quad - p u(\theta) f(\theta|\theta_-) - p_2 u(\theta) f_2(\theta|\theta_-) \\
&= -\left(c(\theta) - y(\theta) + \delta V_{t+1}\left(\beta^{-1}\left(u(\theta) - U(c(\theta), y(\theta)/\theta)\right), w_2(\theta), \theta\right)\right) f_t(\theta|\theta_-) \\
&\quad + \mu(\theta) U_l(c(\theta), y(\theta)/\theta)\left(-\frac{y(\theta)}{\theta^2}\right) + \beta w_2(\theta) \\
&\quad \left(-\frac{p}{p_2} - \frac{f_2(\theta|\theta_-)}{f(\theta|\theta_-)}\right) p_2 u(\theta) f(\theta|\theta_-)
\end{aligned}
$$

and let $(c^*, y^*, w_2^*, \mu^*, p^*, p_2^*)$ be a solution. Let $\zeta = -p^*/p_2^*$. Using direct substitution it is straightforward to verify that $(c^*, y^*, w_2^*, \mu^*, p_2^*)$ is a solution to a Hamiltonian for (15).

8.2 Proof of Proposition 1

Form the Hamiltonian to (22):

$$
\begin{aligned}
\mathcal{H} &= -\left(\frac{1}{\gamma} l(\theta)^\gamma + \bar{U}^{-1}\left([u(\theta) - \beta w(\theta)]\right) - \theta l(\theta) + \delta V_{t+1}(w(\theta), w_2(\theta), \theta)\right) f(\theta|\theta_-) \\
&\quad + p_2 u(\theta) f(\theta|\theta_-)\left(p - \frac{f_2(\theta|\theta_-)}{f(\theta|\theta_-)}\right) + \mu(\theta)\left(m(u(\theta) - \beta w(\theta))\frac{l(\theta)^\gamma}{\theta} - \beta w_2(\theta)\right)
\end{aligned}
$$

where $p_2 > 0$ and $-pp_2$ are the adjoint functions associated with (13) and (12) respectively.

First we characterize the labor wedge (28). Define $r(\theta) = c(\theta) - \frac{1}{\gamma} l(\theta)^\gamma$. The first order condition with respect to $u(\theta)$ is

$$
\left(-\frac{1}{\bar{U}_c(r(\theta))} + p_2\left(p - \frac{f_2(\theta|\theta_-)}{f(\theta|\theta_-)}\right)\right) f(\theta|\theta_-) + \mu(\theta) m'(u(\theta) - \beta w(\theta))\frac{l(\theta)^\gamma}{\theta} = -\mu'(\theta).
$$

This is a second order differential equation, that has a solution

$\mu(\theta) =$

$$\int_{\theta}^{\infty} -\frac{1}{\bar{U}_c(r(x))} \exp\left(\int_{\theta}^{x} m'\left(u\left(\tilde{\theta}\right) - \beta w\left(\tilde{\theta}\right)\right) \frac{l\left(\tilde{\theta}\right)^{\gamma}}{\tilde{\theta}} d\tilde{\theta}\right) \left(1 - p_2\left(p - \frac{f_2(\theta|\theta_-)}{f(\theta|\theta_-)}\right) \bar{U}_c(r(x))\right) dF(x|\theta_-). \tag{40}$$

We proceed to take the first order condition of the Hamiltonian with respect to labor $l(\theta)$:

$$\left[l(\theta)^{\gamma-1} - \theta\right] f(\theta) = \gamma\mu(\theta) m\left(u(\theta) - \beta w(\theta|\theta)\right) \frac{l(\theta)^{\gamma-1}}{\theta}. \tag{41}$$

>From the definition of labor distortion (16),

$$T_D'(\theta) = 1 - \frac{l(\theta)^{\gamma-1}}{\theta}, \tag{42}$$

which implies together with (41) that

$$\begin{aligned} \frac{T_D'(\theta)}{1 - T_D'(\theta)} &= -\frac{\gamma\mu(\theta) m\left(u(\theta) - \beta w(\theta)\right)}{\theta f(\theta|\theta_-)} \\ &= -\mu(\theta)\bar{U}_c(r(\theta)) \frac{\gamma}{\theta f(\theta|\theta_-)}. \end{aligned} \tag{43}$$

We now proceed to characterize $\mu(\theta)\bar{U}_c(r(\theta))$ by substituting the expression for $\mu(\theta)$ from (40) and using the fact that $m'(\theta) = \frac{\bar{U}_{cc}(r(\theta))}{\bar{U}_c(r(\theta))}$.

$$\mu(\theta)\bar{U}_c(r(\theta)) = \int_{\theta}^{\infty} -\frac{\bar{U}_c(r(\theta))}{\bar{U}_c(r(x))} \exp\left(\int_{\theta}^{x} \frac{\bar{U}_{cc}(r(\tilde{\theta}))}{\bar{U}_c(r(\tilde{\theta}))} \frac{l\left(\tilde{\theta}\right)^{\gamma}}{\tilde{\theta}} d\tilde{\theta}\right) \left(1 - p_2\left(p - \frac{f_2(\theta|\theta_-)}{f(\theta|\theta_-)}\right) \bar{U}_c(r(x))\right) f(x|\theta_-)\, dx. \tag{44}$$

Observe that

$$\frac{\bar{U}_c(r(\theta))}{\bar{U}_c(r(x))} = \exp\left(\ln\left(\frac{\bar{U}_c(r(\theta))}{\bar{U}_c(r(x))}\right)\right) = \exp\left(-\int_{\theta}^{x} \frac{\bar{U}_{cc}(r(\tilde{\theta}))}{\bar{U}_c(r(\tilde{\theta}))} dr(\tilde{\theta})\right).$$

To find $dr(\tilde{\theta})$ observe that $\bar{U}(r(\tilde{\theta})) = U(\tilde{\theta})$ and therefore $\bar{U}_c(r(\tilde{\theta}))dr(\tilde{\theta}) = U'(\tilde{\theta})$. This implies that

$$dr(\tilde{\theta}) = \frac{U'(\tilde{\theta})}{\bar{U}_c(r(\tilde{\theta}))} d\tilde{\theta} = \frac{u'(\tilde{\theta}) - \beta w'(\tilde{\theta})}{\bar{U}_c(r(\tilde{\theta}))} d\tilde{\theta}.$$

Therefore,

$$\frac{\bar{U}_c(r(\theta))}{\bar{U}_c(r(x))} = \exp\left(-\int_{\theta}^{x} \frac{\bar{U}_{cc}(r(\tilde{\theta}))}{\bar{U}_c(r(\tilde{\theta}))} \frac{u'(\tilde{\theta}) - \beta w'(\tilde{\theta})}{\bar{U}_c(r(\tilde{\theta}))} d\tilde{\theta}\right). \tag{45}$$

Substitute expression for $u'(\tilde{\theta})$ from (20) to (45) and observe that since $w(\theta) = \omega(\theta|\theta)$ and $w_2(\theta) = \omega_2(\theta|\theta)$, then

$$\begin{aligned} w'(\theta) &= \omega_1(\theta|\theta) + \omega_2(\theta|\theta) \\ &= \omega_1(\theta|\theta) + w_2(\theta). \end{aligned} \tag{46}$$

Then expression (45) implies that

$$\frac{\bar{U}_c\left(r(\theta)\right)}{\bar{U}_c(r(x))} = \exp\left(-\int_\theta^x \frac{\bar{U}_{cc}(r(\tilde{\theta}))}{\bar{U}_c(r(\tilde{\theta}))}\left(\frac{l\left(\tilde{\theta}\right)^\gamma}{\tilde{\theta}} - \frac{\beta\omega_1(\tilde{\theta}|\tilde{\theta})}{\bar{U}_c(r(\tilde{\theta}))}\right)d\tilde{\theta}\right).$$

Substitute (45) into (44)

$$\mu(\theta)\bar{U}_c(r(\theta)) = \int_\theta^\infty -\exp\left(\int_\theta^x \frac{\bar{U}_{cc}(r(\tilde{\theta}))}{\bar{U}_c(r(\tilde{\theta}))}\frac{\beta\omega_1\left(\tilde{\theta}|\tilde{\theta}\right)}{\bar{U}_c(r(\tilde{\theta}))}d\tilde{\theta}\right)\left(1 - p_2\left(p - \frac{f_2(\theta|\theta_-)}{f\left(\theta|\theta_-\right)}\right)\bar{U}_c(r(x))\right)f\left(x|\theta_-\right)dx. \tag{47}$$

Substitute (47) into (43)

$$\begin{aligned} \frac{T_D'\left(\theta\right)}{1 - T_D'\left(\theta\right)} &= \frac{\gamma}{\theta f\left(\theta|\theta_-\right)} \qquad\qquad (48)\\ &\quad\times\int_\theta^\infty \exp\left(\int_\theta^x \frac{\bar{U}_{cc}(r(\tilde{\theta}))}{\bar{U}_c(r(\tilde{\theta}))}\frac{\beta\omega_1\left(\tilde{\theta}|\tilde{\theta}\right)}{\bar{U}_c(r(\tilde{\theta}))}d\tilde{\theta}\right)\left(1 - p_2\left(p - \frac{f_2(\theta|\theta_-)}{f\left(\theta|\theta_-\right)}\right)\bar{U}_c(r(x))\right)f\left(x|\theta_-\right)dx \\ &= \frac{\gamma}{\theta f\left(\theta|\theta_-\right)\exp\left(\int_0^\theta \frac{\bar{U}_{cc}(r(\tilde{\theta}))}{\bar{U}_c(r(\tilde{\theta}))}\frac{\beta\omega_1\left(\tilde{\theta}|\tilde{\theta}\right)}{\bar{U}_c(r(\tilde{\theta}))}d\tilde{\theta}\right)} \\ &\quad\times\int_\theta^\infty \exp\left(\int_0^x \frac{\bar{U}_{cc}(r(\tilde{\theta}))}{\bar{U}_c(r(\tilde{\theta}))}\frac{\beta\omega_1\left(\tilde{\theta}|\tilde{\theta}\right)}{\bar{U}_c(r(\tilde{\theta}))}d\tilde{\theta}\right)\left(1 - p_2\left(p - \frac{f_2(\theta|\theta_-)}{f\left(\theta|\theta_-\right)}\right)\bar{U}_c(r(x))\right)f\left(x|\theta_-\right)dx \end{aligned}$$

Define

$$\Psi(\theta) = \exp\left(\int_0^x \frac{\bar{U}_{cc}(r(\tilde{\theta}))}{\bar{U}_c(r(\tilde{\theta}))}\frac{\beta\omega_1\left(\tilde{\theta}|\tilde{\theta}\right)}{\bar{U}_c(r(\tilde{\theta}))}d\tilde{\theta}\right)$$

and

$$\tilde{f}(\theta|\theta_-) = \frac{\Psi(\theta)f(\theta|\theta_-)}{\int_0^\infty \Psi(x)f(x|\theta_-)dx}.$$

Then (48) can be re-written as

$$\frac{T_D'\left(\theta\right)}{1 - T_D'\left(\theta\right)} = \gamma\frac{1 - \tilde{F}(\theta|\theta_-)}{\theta\tilde{f}\left(\theta|\theta_-\right)}\int_\theta^\infty \left(1 - p_2\left(p - \frac{f_2(\theta|\theta_-)}{f\left(\theta|\theta_-\right)}\right)\bar{U}_c\left(x\right)\right)\frac{f\left(x|\theta_-\right)dx}{1 - \tilde{F}(\theta|\theta_-)} \tag{49}$$

Using $\mu(0) = 0$, we obtain

$$0 = \int_0^\infty \left(1 - p_2 \left(p - \frac{f_2(\theta|\theta_-)}{f(\theta|\theta_-)}\right) \bar{U}_c(x)\right) \tilde{f}(x|\theta_-)\, dx. \tag{50}$$

Let $\lambda \equiv 1/p_2$. Then (50) implies that

$$\lambda = \frac{1}{p_2} = \int_0^\infty \left(p - \frac{f_2(\theta|\theta_-)}{f(\theta|\theta_-)}\right) \bar{U}_c(x)\, \tilde{f}(x|\theta_-)\, dx. \tag{51}$$

Substitute (51) into (49) to obtain (28).

We proceed to derive the marginal capital wedge (29). Take the first order condition of the Hamiltonian with respect to $w(\theta)$

$$\left(-\frac{\beta}{\bar{U}_c(r_t(\theta))} - \frac{\partial V_{t+1}(w(\theta), w_2(\theta), \theta)}{\partial w(\theta)}\right) f(\theta|\theta_-) = -\beta\mu_t(\theta) \frac{\bar{U}_{cc}(r_t(\theta))}{\bar{U}_c(r_t(\theta))} \frac{l_t(\theta)^\gamma}{\theta}. \tag{52}$$

Rearrange (52)

$$-\frac{\delta}{\beta} \frac{\bar{U}_c(r_t(\theta))\, \partial V_{t+1}(w(\theta), w_2(\theta), \theta)}{\partial w(\theta)} = 1 - \frac{1}{f(\theta|\theta_-)} \beta\mu_t(\theta) \bar{U}_c(r_t(\theta)) \frac{\bar{U}_{cc}(r_t(\theta))}{\bar{U}_c(r_t(\theta))} \frac{l_t(\theta)^\gamma}{\theta}.$$

Now use the expression (42) and (43)

$$-\frac{\delta}{\beta} \frac{\bar{U}_c(r_t(\theta))\, \partial V_{t+1}(w(\theta), w_2(\theta), \theta)}{\partial w(\theta)} = 1 - \psi(\theta) \frac{1}{\gamma} T'_{D,t}(\theta)\, y_t(\theta). \tag{53}$$

>From the envelope theorem

$$\frac{\partial V_{t+1}(w(\theta), w_2(\theta), \theta)}{\partial w(\theta)} = -p_{t+1} p_{2,t+1},$$

where $p_{t+1} p_{2,t+1}$ can be determined from the equivalent of equation (50) for period $t+1$

$$p_{t+1} p_{2,t+1} = \int_0^\infty \left(1 - \tilde{\zeta} \frac{f_{2,t+1}(x|\theta)}{f_{t+1}(x|\theta)}\right) \bar{U}_c(r_{t+1}(x))\, \tilde{f}_{t+1}(x|\theta) dx$$

where $\tilde{\zeta}$ is a constant. Then the wedge on capital is given by

$$1 - \tau_{S,t}(\theta) = z_t(\theta) \left(1 - \frac{\psi(\theta)}{\gamma} T'_{D,t}(\theta) y_t(\theta)\right),$$

where

$$z_t(\theta) = \frac{\int_0^\infty \left(1 - \tilde{\zeta} \frac{f_{2,t+1}(x|\theta)}{f_{t+1}(x|\theta)}\right) \bar{U}_c(r_{t+1}(x))\, \tilde{f}_{t+1}(x|\theta) dx}{\int_0^\infty \bar{U}_c(r_{t+1}(x))\, f_{t+1}(x|\theta) dx}.$$

For future references it also will be useful to have the following first order condition with respect to $w_2(\theta)$

$$\begin{aligned} \frac{\partial V_{t+1}(w(\theta), w_2(\theta), \theta)}{\partial w_2(\theta)} &= -\frac{\beta}{\delta} \frac{\mu(\theta)}{f(\theta|\theta_-)} \\ &= \frac{\beta}{\delta\gamma} \frac{T'_{D,t}(\theta)}{1 - T'_{D,t}(\theta)} \frac{\theta}{\bar{U}_c(\theta)}. \end{aligned}$$

8.3 Proof of Proposition 2

Proof of part 1.

Suppose that $f_{t+1}(\theta|\theta_-)$ is independent of θ_- so that $f_{2,t+1}(\theta|\theta_-) = 0$. In this case the optimal allocation in period $t+1$ minimizes (22) subject to (20) and (12). This problem is independent of $w_{2,t+1}$ and it is decreasing in w_{t+1}, therefore $\partial V_{t+1}(w_{t+1}, w_{2,t+1}, \theta)/\partial w_{t+1} < 0$. Let $p_1^0 = -\partial V_{t+1}(w_{t+1}, w_{2,t+1}, \theta)/\partial w_{t+1}$.

Let (l^*, u^*, w^*, w_2^*) denote the solution to (22). Then they must satisfy equation (53). The left hand side of that expression is positive, since $\partial V_{t+1}(w_{t+1}, w_{2,t+1}, \theta)/\partial w_{t+1}$ is negative. Since $\psi(\theta)$ is bounded away from zero, this implies that $T'_{D,t}(\theta)y(\theta)$ is bounded from above. Since $y(\theta) = \theta^{\gamma/(\gamma-1)}\left(1 - T'_{D,t}(\theta)\right)^{1/(\gamma-1)}$, it must be true that $\theta^{\gamma/(\gamma-1)}\left(1 - T'_{D,t}(\theta)\right)^{1/(\gamma-1)} T'_{D,t}(\theta)$ is bounded from above, which is possible only if $T'_{D,t}(\theta)$ converges to either 0 or 1.

Next we show that $T'_{D,t}(\theta) \to 0$.

Suppose $T'_{D,t}(\theta) \to 1$. Since $\theta^{\gamma/(\gamma-1)}\left(1 - T'_{D,t}(\theta)\right)^{1/(\gamma-1)} T'_{D,t}(\theta)$ is bounded, it must be true that $\theta^{\gamma}\left(1 - T'_{D,t}(\theta)\right)$ is bounded, and therefore $\theta\left(1 - T'_{D,t}(\theta)\right) \to 0$.

Since $\theta\left(1 - T'_{D,t}(\theta)\right) = l^*(\theta)^{\gamma-1}$ and $\gamma > 1$, this implies that $l^*(\theta) \to 0$. Moreover,

$$\begin{aligned} \theta^{\gamma}\left(1 - T'_{D,t}(\theta)\right) &= \theta^{\gamma-1} l^*(\theta)^{\gamma-1} \\ &= y^*(\theta)^{\gamma-1}, \end{aligned}$$

so that $y^*(\theta)$ is bounded. Let $\bar{y}$ be the least upper bound of y.

Pick $\varepsilon > 0$ and choose $\bar{\theta}$ so that for all $\theta \geq \bar{\theta}$

$$\frac{1}{\gamma}\left(\frac{\bar{y}}{\theta}\right)^{\gamma} \leq \varepsilon \tag{54}$$

and

$$\bar{y} + \varepsilon \leq \frac{\gamma - 1}{\gamma}\theta^{\gamma/(\gamma-1)} \tag{55}$$

Such $\bar{\theta}$ exists because the left hand side of (55) is increasing in θ, while the left is decreasing in θ.

For our purposes it will be convenient to consider a dual to (22), which can be written as

$$\max_{c,y,w} \int \left(U(c(\theta), y(\theta)) + \beta w(\theta)\right) dF_t(\theta)$$

s.t. (20) and

$$\int \left(c(\theta) - y(\theta) + \delta V_{t+1}(w(\theta))\right) dF(\theta) \leq V_t(\theta).$$

The proof proceed in two steps. First we establish the bound for the utility at the tail of the distribution $[\bar{\theta}, \infty)$ that the allocation may achieve if $T'_{D,t}(\theta) \to 1$. Second we show that an unconstrained allocation is incentive compatible and achieves higher utility.

Step 1.

Let $K_{\bar{\theta}} = \int_{\bar{\theta}}^{\infty} \left(y^*(\theta) - c^*(\theta)\right) dF(\theta) / (1 - F(\bar{\theta}))$ and $\Omega_{\bar{\theta}} = \int V_{t+1}(w^*(\theta)) dF(\theta) / (1 - F(\theta))$.

Define allocation $\left(c^{fb}, w^{fb}\right)$ as a solution to

$$W_{\bar{\theta}}^{fb} = \max_{c,w} \int_{\bar{\theta}}^{\infty} \left(U(c(\theta), y^*(\theta)) + \beta w(\theta)\right) \frac{dF_t(\theta)}{1 - F_t(\bar{\theta})} \tag{56}$$

s.t.

$$\int_{\bar{\theta}}^{\infty} c(\theta) \frac{dF_t(\theta)}{\left(1 - F_t(\bar{\theta})\right)} = \int_{\bar{\theta}}^{\infty} y^*(\theta) \frac{dF_t(\theta)}{\left(1 - F_t(\bar{\theta})\right)} - K_{\bar{\theta}}$$

and

$$\int_{\bar{\theta}}^{\infty} V_{t+1}(w(\theta)) \frac{dF_t(\theta)}{\left(1 - F_t(\bar{\theta})\right)} = \Omega_{\bar{\theta}}$$

It can be shown that V is concave, so $w^{fb}(\theta) = \Omega_{\bar{\theta}}$ for all θ is a solution. Since this is an unconstrained maximization problem that consumes the same amount of resources as (c^*, y^*, w^*), it must be true that

$$W_{\bar{\theta}}^{fb} \geq \int_{\bar{\theta}}^{\infty} u^*(\theta) \frac{dF_t(\theta)}{1 - F_t(\bar{\theta})}. \tag{57}$$

Since $u^*(\theta)$ is increasing in θ because of incentive compatibility, (57) implies

$$W_{\bar{\theta}}^{fb} \geq u^*\left(\bar{\theta}\right). \tag{58}$$

The first order conditions to (56) also imply that

$$c^{fb}(\theta) - \frac{1}{\gamma}\left(\frac{y^*(\theta)}{\theta}\right)^{\gamma} = c^{fb}\left(\theta'\right) - \frac{1}{\gamma}\left(\frac{y^*\left(\theta'\right)}{\theta'}\right)^{\gamma} \tag{59}$$

for all $\theta, \theta' \geq \bar{\theta}$. Therefore from (54) for all $\theta, \theta' \geq \bar{\theta}$

$$c^{fb}(\theta) - c^{fb}\left(\theta'\right) \leq \varepsilon.$$

Since $\int_{\bar{\theta}}^{\infty} c^{fb}(\theta) \frac{dF(\theta)}{(1-F(\bar{\theta}))} = \int_{\bar{\theta}}^{\infty} y^*(\theta) \frac{dF(\theta)}{(1-F(\bar{\theta}))} - K_{\bar{\theta}}$, this implies that for all $\theta \geq \bar{\theta}$

$$\begin{aligned} c^{fb}(\bar{\theta}) &\leq \int_{\bar{\theta}}^{\infty} y^*(\theta) \frac{dF_t(\theta)}{(1-F_t(\bar{\theta}))} - K_{\bar{\theta}} + \varepsilon \\ &\leq \bar{y} - K_{\bar{\theta}} + \varepsilon. \end{aligned}$$

Therefore

$$\begin{aligned} U\left(c^{fb}(\theta), l^{fb}(\theta)\right) &< U\left(c^{fb}(\theta), 0\right) \\ &\leq \bar{U}(\bar{y} - K_{\bar{\theta}} + \varepsilon) \end{aligned}$$

and

$$W_{\bar{\theta}}^{fb} \leq \int_{\bar{\theta}}^{\infty} \left(\bar{U}(\bar{y} - K_{\bar{\theta}} + \varepsilon) + w^{fb}\right) \frac{dF_t(\theta)}{1-F_t(\bar{\theta})}.$$

Now, define an autarkic allocation. Let $l^{aut}(\theta) = \theta^{1/(\gamma-1)}$ and $y^{aut}(\theta) = \theta^{\gamma/(\gamma-1)}$, $c^{aut}(\theta) = y^{aut}(\theta) - K_{\bar{\theta}}$ and $w^{aut}(\theta) = w^{fb}$ for all $\theta \geq \bar{\theta}$, and $y^{aut}(\theta) = y^*(\theta)$, $c^{aut}(\theta) = c^*(\theta)$, $w^{aut}(\theta) = w^*(\theta)$ for all $\theta < \bar{\theta}$ Also define

$$U^{aut}(\theta) = \bar{U}\left(\frac{\gamma-1}{\gamma}\theta^{\gamma/(\gamma-1)} - K_{\bar{\theta}}\right)$$

and

$$W_{\bar{\theta}}^{aut} = \int_{\bar{\theta}}^{\infty} U^{aut}(\theta) \frac{dF(\theta)}{1-F(\bar{\theta})} + \beta w^{fb}.$$

Note that $W_{\bar{\theta}}^{aut} > W_{\bar{\theta}}^{fb}$ since

$$\begin{aligned} W_{\bar{\theta}}^{aut} - W_{\bar{\theta}}^{fb} &= \int_{\bar{\theta}}^{\infty} \left(U^{aut}(\theta) - U\left(c^{fb}(\theta), l^{fb}(\theta)\right)\right) \frac{dF(\theta)}{1-F(\bar{\theta})} \\ &> \int_{\bar{\theta}}^{\infty} \left(\bar{U}\left(\frac{\gamma-1}{\gamma}\theta^{\gamma/(\gamma-1)} - K_{\bar{\theta}}\right) - \bar{U}(\bar{y} + \varepsilon - K_{\bar{\theta}})\right) \frac{dF(\theta)}{1-F(\bar{\theta})} \\ &\geq 0 \end{aligned}$$

where the last inequality follows from (55).

Not also that since the expression inside of the integral is positive for all $\theta \geq \bar{\theta}$, this implies

$$U\left(c^{aut}(\bar{\theta}), y^{aut}(\bar{\theta})/\bar{\theta}\right) + \beta w^{aut}(\bar{\theta}) > W_{\bar{\theta}}^{fb} \geq u^*(\bar{\theta}) \tag{60}$$

By construction

$$\begin{aligned} &\int_{\bar{\theta}}^{\infty} \left(c^*(\theta) - y^*(\theta) + \delta V(w^*(\theta))\right) dF_t(\theta) \\ = &\int_{\bar{\theta}}^{\infty} \left(c^{fb}(\theta) - y^*(\theta) + \delta V(w^{fb})\right) dF_t(\theta) \\ = &\int_{\bar{\theta}}^{\infty} \left(c^{aut}(\theta) - y^{aut}(\theta) + \delta V(w^{aut})\right) dF_t(\theta) \end{aligned}$$

Also

$$\begin{aligned} & \int_{\bar{\theta}}^{\infty} \left(U(c^{aut}(\theta), y^{aut}(\theta)/\theta) + \beta w^{aut}\right) dF_t(\theta) \\ \geq & \int_{\bar{\theta}}^{\infty} \left(U(c^{fb}(\theta), y^{*}(\theta)/\theta) + \beta w^{fb}\right) dF_t(\theta) \\ \geq & \int_{\bar{\theta}}^{\infty} \left(U(c^{*}(\theta), y^{*}(\theta)/\theta) + \beta w^{*}(\theta)\right) dF_t(\theta) \end{aligned}$$

Thus, if $\left(c^{aut}, y^{aut}, w^{aut}\right)$ it is incentive compatible, it is both feasible and give higher welfare than (c^*, y^*, w^*), which implies that (c^*, y^*, w^*) cannot be optimal.

To show that $\left(c^{aut}, y^{aut}, w^{aut}\right)$ is incentive compatible, we need to verify that

$$U\left(c^{aut}(\theta), \frac{y^{aut}(\theta)}{\theta}\right) + \beta w^{aut}(\theta) \geq U\left(c^{aut}(\theta'), \frac{y^{aut}(\theta')}{\theta}\right) + \beta w^{aut}(\theta') \text{ for all } \theta, \theta' \geq \bar{\theta} \tag{61}$$

and

$$U\left(c^{aut}(\theta), \frac{y^{aut}(\theta)}{\theta}\right) + \beta w^{aut}(\theta) \geq U\left(c^{*}(\theta'), \frac{y^{*}(\theta')}{\theta}\right) + \beta w^{*}(\theta') \text{ for all } \theta \geq \bar{\theta}, \theta' < \bar{\theta} \tag{62}$$

Equation (61) follows from construction of allocation $\left(c^{aut}, y^{aut}, w^{aut}\right)$.

Now consider equation (62). The single crossing property of U imply that if for $\theta > \bar{\theta} > \theta'$

$$U\left(c^{aut}(\theta), \frac{y^{aut}(\theta)}{\theta}\right) + \beta w^{aut}(\theta) \geq U\left(c^{aut}(\bar{\theta}), \frac{y^{aut}(\bar{\theta})}{\theta}\right) + \beta w^{aut}(\bar{\theta})$$

and

$$U\left(c^{aut}(\bar{\theta}), \frac{y^{aut}(\bar{\theta})}{\bar{\theta}}\right) + \beta w^{aut}(\bar{\theta}) \geq U\left(c^{*}(\theta'), \frac{y^{*}(\theta')}{\bar{\theta}}\right) + \beta w^{*}(\theta') \text{ for } \theta' < \bar{\theta} \tag{63}$$

then

$$U\left(c^{aut}(\theta), \frac{y^{aut}(\theta)}{\theta}\right) + \beta w^{aut}(\theta) \geq U\left(c^{*}(\theta'), \frac{y^{*}(\theta')}{\bar{\theta}}\right) + \beta w^{*}(\theta').$$

Equation (63) is true because (60) and incentive compatibility of (c^*, y^*, w^*) imply

$$\begin{aligned} u^{aut}(\bar{\theta}) &> u^{*}(\bar{\theta}) \\ &\geq U\left(c^{*}(\theta'), \frac{y^{*}(\theta')}{\bar{\theta}}\right) + \beta w^{*}(\theta') \text{ for all } \theta' < \bar{\theta}. \end{aligned}$$

Therefore (62) holds.

Proof of part 2

Since $f_2^{\varepsilon}(\theta|\theta_-)$ converges uniformly to 0, then $f^{\varepsilon}(\theta|\theta_-)$ converges uniformly to $f(\theta)$ (Theorem 7.17 in Rudin (1976)). By Berge's Theorem of Maximum then $T_D^{\varepsilon\prime}(\theta) \to T_D^{0\prime}(\theta)$ in sup norm. Since from part 1 $T_D^{0\prime}(\theta) \to 0$ as $\theta \to \infty$, that implies that $T_D^{\varepsilon\prime}(\theta) \to 0$ as $\theta \to \infty$ for all ε sufficiently small.

8.4 Proof of Proposition ??

First, we consider the first order conditions to the Hamiltonian (??) for which we explicitly substitute (23) and (??).

The first order condition with respect to w is

$$\left(-\frac{\beta}{\psi U(\theta)} - \delta\frac{\partial V_{t+1}(w(\theta), w_2(\theta), \theta)}{\partial w(\theta)}\right) f(\theta|\theta_-) = -\mu(\theta)\beta\psi\frac{l(\theta)^{\gamma}}{\theta}$$

and the first order condition with respect to w_2 is

$$\delta\frac{\partial V_{t+1}(w(\theta), w_2(\theta), \theta)}{\partial w_2(\theta)} f(\theta|\theta_-) = \beta\mu(\theta)$$

Substitute (??) to obtain

$$\frac{\partial V_{t+1}(w(\theta), w_2(\theta), \theta)}{\partial w(\theta)} = a_t'\left(\frac{w_2}{w}|\theta_-\right)\left(\frac{w_2}{w}\right)\left(-\frac{1}{w}\right) - \frac{b_t}{\psi}\frac{1}{w}$$

and

$$\frac{\partial V_{t+1}(w(\theta), w_2(\theta), \theta)}{\partial w_2(\theta)} = a_t'\left(\frac{w_2}{w}|\theta_-\right)\left(\frac{1}{w}\right).$$

Substitute these expressions into the FOCs for w_2 and w :

$$a_t'\left(\frac{w_2}{w}|\theta_-\right)\left(\frac{1}{w}\right) = \frac{\beta}{\delta}\frac{\mu(\theta)}{f(\theta|\theta_-)} \tag{64}$$

and

$$\begin{aligned}
\left(-\frac{\beta}{\psi U(\theta)} - \delta a_t'\left(\frac{w_2}{w}|\theta_-\right)\left(\frac{w_2}{w}\right)\left(-\frac{1}{w}\right) + \delta\frac{b_t}{\psi}\frac{1}{w}\right) f(\theta|\theta_-) &= -\mu(\theta)\beta\psi\frac{l(\theta)^{\gamma}}{\theta} \\
\left(\frac{1}{U(\theta)} + \psi\delta a_t'\left(\frac{w_2}{w}|\theta_-\right)\left(\frac{w_2}{w}\right)\left(-\frac{1}{w}\right) - \delta b_t\frac{1}{w}\right) &= \frac{\mu(\theta)}{f(\theta|\theta_-)}\psi^2\frac{l(\theta)^{\gamma}}{\theta} \\
\left(1 + \frac{\delta\psi}{\beta}a_t'\left(\frac{w_2}{w}|\theta_-\right)\left(\frac{w_2}{w}\right)\left(-\frac{U(\theta)}{w}\right) - \frac{\delta}{\beta}b_t\frac{U(\theta)}{w}\right) &= \frac{\mu(\theta)U(\theta)}{f(\theta|\theta_-)}\psi^2\frac{l(\theta)^{\gamma}}{\theta}
\end{aligned}$$

>From the previous analysis we showed that

$$\frac{T'(\theta)}{1-T'(\theta)} = \frac{\mu(\theta)\psi U(\theta)\gamma}{\theta f(\theta|\theta_-)} \tag{65}$$

and

$$\frac{\mu(\theta)\,U(\theta)}{f(\theta|\theta_-)}\psi^2\frac{l(\theta)^\gamma}{\theta}=\frac{\psi}{\gamma}T'_{D,t}(\theta)y(\theta).$$

Note that last expression implies that $\mu(\theta)\leq 0$. Therefore (64) implies that $a'_t\left(\frac{w_2}{w}|\theta_-\right)\geq 0$.

After substitutions,

$$\left(1+\frac{\delta\psi}{\beta}a'_t\left(\frac{w_2}{w}|\theta_-\right)\left(\frac{w_2}{w}\right)\left(-\frac{U(\theta)}{w}\right)-\frac{\delta}{\beta}b_t\frac{U(\theta)}{w}\right)=\frac{\psi}{\gamma}T'_{D,t}(\theta)y(\theta)$$

or

$$1-\frac{\delta}{\beta}b_t\frac{U(\theta)}{w}=\frac{\psi}{\gamma}T'_{D,t}(\theta)y(\theta)+\frac{\delta\psi}{\beta}a'_t\left(\frac{w_2}{w}|\theta_-\right)\left(\frac{w_2}{w}\right)\left(\frac{U(\theta)}{w}\right). \tag{66}$$

Use (64) and (65) to get

$$\begin{aligned}\psi a'_t\left(\frac{w_2}{w}|\theta_-\right)\left(\frac{U(\theta)}{w}\right) &= \frac{\beta}{\delta}\frac{\mu(\theta)U(\theta)\psi\gamma}{\theta f(\theta|\theta_-)}\frac{\theta}{\gamma}\\ &= \frac{\beta}{\delta}\frac{\theta}{\gamma}\frac{T'(\theta)}{1-T'(\theta)}\end{aligned}$$

Substitute it into (66)

$$1-\frac{\delta}{\beta}b_t\frac{U(\theta)}{w}=\frac{\psi}{\gamma}T'_{D,t}(\theta)y(\theta)+\left(\frac{w_2}{w}\right)\frac{\theta}{\gamma}\frac{T'_{D,t}(\theta)}{1-T'_{D,t}(\theta)}$$

or

$$1-\frac{\delta}{\beta}b_t\frac{U(\theta)}{w(\theta)}=\frac{T'_{D,t}(\theta)\theta}{\gamma}\left(\psi l(\theta)+\left(\frac{w_2(\theta)}{w(\theta)}\right)\frac{1}{1-T'_{D,t}(\theta)}\right) \tag{67}$$

We prove the result of the Proposition three steps.

Step 1: If f satisfies (**??**), then $w_2(\theta)/w(\theta)$ is bounded from below

Consider $w_2(\theta)$:

$$\begin{aligned}w_2(\theta) &= \int_0^\infty u(\theta)\frac{f_2(\theta|\theta)}{f(\theta|\theta)}f(\theta|\theta)d\theta\\ &= \int_{f_2\geq 0}u(\theta)\frac{f_2(\theta|\theta)}{f(\theta|\theta)}f(\theta|\theta)d\theta+\int_{f_2<0}u(\theta)\frac{f_2(\theta|\theta)}{f(\theta|\theta)}f(\theta|\theta)d\theta\end{aligned}$$

Since the first integral is negative and the second is positive

$$\begin{aligned}w_2(\theta|\theta) &\leq \int_{f_2<0}u(\theta)\frac{f_2(\theta|\theta)}{f(\theta|\theta)}f(\theta|\theta)d\theta\\ &\leq -\varpi\int_{f_2<0}u(\theta)f(\theta|\theta)d\theta\end{aligned}$$

This implies (since $w(\theta) = \int u(\theta) f(\theta|\theta) d\theta < 0$) that

$$\frac{w_2(\theta|\theta)}{w(\theta|\theta)} \geq -\varpi \frac{\int_{f_2<0} u(\theta) f(\theta|\theta) d\theta}{\int u(\theta) f(\theta|\theta) d\theta} \geq -\varpi$$

Step 2. If $w_2(\theta)/w(\theta)$ is bounded from below, then either $T'_{D,t}(\theta) \to 0$ or $T'_{D,t}(\theta) \to 1$.

Suppose $T'_{D,t}(\theta)$ does not converge to either zero or 1. Then there must exist some $\zeta > 0$ and an infinite subsequence θ_i s.t. $T'_{D,t}(\theta_i) \in [\zeta, 1-\zeta]$. This implies that

$$\frac{T'_{D,t}(\theta_i)\theta_i}{\gamma\psi} \to \infty.$$

>From the first order condition on labor

$$l(\theta)^{\gamma-1} = \theta(1 - T'_{D,t}(\theta)) \tag{68}$$

we get $l(\theta_i) \to \infty$, while

$$\left(\frac{w_2(\theta_i)}{w(\theta_i)}\right) \frac{1}{1 - T'_{D,t}(\theta_i)}$$

is bounded from below. Therefore the right hand side of (67) goes to infinity. Since $U(\theta), w(\theta) < 0$, the left hand side of (67) is bounded by 1, which leads to a contradiction.

Step 3. $T'_{D,t}(\theta)$ cannot converge to 1.

This steps is analogous to Step 3 of the proof of Proposition 2.

8.5 Value functions in exponential case

In this section we will prove the form of the value function when utility is exponential that were used in the paper. We prove it for the case when shocks are persistent. With i.i.d. shocks the proof is analogous.

Following the same steps as the proof of Proposition **??**, we can show that $V_T(w, w_2, \theta_-) = a_T\left(\frac{w_2}{w}|\theta_-\right) - \frac{1}{\psi}\ln(-w)$. The rest of the proof is by induction. Re-write expression (10) for $t < T$ as

$$V_t(w_t, w_{2t}, \theta_-) = \max_{\{y(\theta), c(\theta), w'(\theta), w'_2(\theta)\}} -\int \left(c(\theta) - y(\theta) + \delta V_{t+1}\left(w'(\theta), w'_2(\theta), \theta\right)\right) f(\theta|\theta_-) d\theta \tag{69}$$

$$U(c(\theta), y(\theta)/\theta) + \beta w'(\theta) \geq U\left(c(\theta'), y(\theta')/\theta\right) + \beta w'_2(\theta') \tag{70}$$

$$w_t = \int \left(U(c(\theta), y(\theta)/\theta) + \beta w'(\theta)\right) f\left(\theta|\theta_-\right) d\theta, \tag{71}$$

$$w_{2t} = \int \left(U(c(\theta), y(\theta)/\theta) + \beta w'(\theta)\right) f_2(\theta|\theta_-) d\theta. \tag{72}$$

Suppose that $V_{t+1}(w, w_2, \theta)$ satisfies $a_{t+1}\left(\frac{w_2}{w}|\theta\right) - \frac{1+\delta+...+\delta^{T-t-1}}{\psi}\ln(-w)$.

Let $c(\theta) = \tilde{c}(\theta) - \frac{1}{\psi}\ln(-w)$ and $w'(\theta) = -\tilde{w}'(\theta)\, w$ and $w_2' = -\tilde{w}_2'(\theta)\, w$ Substitute the newly defined variables to (70):

$$U\left(\tilde{c}(\theta), y(\theta)/\theta\right) + \beta\tilde{w}'(\theta) \geq U\left(\tilde{c}\left(\theta'\right), y\left(\theta'\right)/\theta\right) + \beta\tilde{w}_2'\left(\theta'\right) \tag{73}$$

and (71) becomes and (72) become

$$-1 = \int \left(U(\tilde{c}(\theta), y(\theta)/\theta) + \beta\tilde{w}'(\theta)\right) f\left(\theta|\theta_-\right) d\theta \tag{74}$$

and

$$-\frac{w_2}{w} = \int \left(U(\tilde{c}(\theta), y(\theta)/\theta) + \beta\tilde{w}'(\theta)\right) f_2\left(\theta|\theta_-\right) d\theta \tag{75}$$

The objective function in (69) satisfies

$$\begin{aligned}
&\int \left(c(\theta) - y(\theta) + \delta\left\{a_{t+1}\left(\frac{w_2}{w}|\theta\right) - \frac{1+\delta+...+\delta^{T-t-1}}{\psi}\ln\left(-w'(\theta)\right)\right\}\right) dF\left(\theta|\theta_-\right) \\
&\quad = \int \left(\tilde{c}(\theta) - y(\theta) + \delta\left\{a_{t+1}a_{t+1}\left(\frac{\tilde{w}_2}{\tilde{w}}|\theta\right) - \frac{1+\delta+...+\delta^{T-t-1}}{\psi}\ln\left(-\tilde{w}'(\theta)\right)\right\}\right) \\
&\qquad\qquad\qquad\qquad dF\left(\theta|\theta_-\right) - \frac{1+\delta+...+\delta^{T-t}}{\psi}\ln(-w)
\end{aligned}$$

Therefore for any w, (69) can be re-written as

$$V_t(w) = \min_{\{\tilde{c}(\theta), y(\theta), \tilde{w}'(\theta)\}} \int \left(c(\theta) - y(\theta) + \delta V_{t+1}\left(\tilde{w}'(\theta)\right)\right) dF\left(\theta|\theta_-\right) - \frac{1+\delta+...+\delta^{T-t}}{\psi}\ln(-w)$$

s.t. (73), (74) and (75). Solution $\{\tilde{c}(\theta), y(\theta), \tilde{w}'(\theta)\}$ depends only on θ_- and w_2/w, therefore, $V_t(w, w_2, \theta_-) = a_t(\frac{w_2}{w}|\theta_-) - \frac{1+\delta+...+\delta^{T-t}}{\psi}\ln(-w)$

8.6 Proof of Proposition 4

First, consider the recursive formulation of the optimal problem

$$V_t(w) = \min_{\{c(\theta), y(\theta), w'(\theta)\}} \int \left(c(\theta) - y(\theta) + \delta V_{t+1}\left(w'(\theta)\right)\right) dF_t(\theta) \tag{76}$$

s.t. incentive constraints (6) and

$$w=\int\left(U(c(\theta),y(\theta)/\theta)+\beta w'(\theta)\right)dF_t(\theta). \tag{77}$$

By Proposition **??**, $V_T(w)=a_T-\frac{1}{\psi}\ln(-w)$ where a_T is a constant. First we want to show that $V_t(w)=a_t-\frac{1+\delta+...+\delta^{T-t}}{\psi}\ln(-w)$ for all t where a_t are constants. The proof of this fact is by induction. Suppose that $V_{t+1}(w)=a_{t+1}-\frac{1+\delta+...+\delta^{T-t-1}}{\psi}\ln(-w)$ and let $a_t=V_t(-1)$.

Let $c(\theta)=\tilde{c}(\theta)-\frac{1}{\psi}\ln(-w)$ and $w'(\theta)=-\tilde{w}'(\theta)w$. Substitute the newly defined variables to (6):

$$U(\tilde{c}(\theta),y(\theta)/\theta)+\beta\tilde{w}'(\theta)\geq U\left(\tilde{c}\left(\theta'\right),y\left(\theta'\right)/\theta\right)+\beta\tilde{w}'\left(\theta'\right) \tag{78}$$

and (77) becomes

$$-1=\int\left(U(\tilde{c}(\theta),y(\theta)/\theta)+\beta\tilde{w}'(\theta)\right)dF_t(\theta). \tag{79}$$

The objective function in (76) satisfies

$$\begin{aligned}&\int\left(c(\theta)-y(\theta)+\delta\left\{a_{t+1}-\frac{1+\delta+...+\delta^{T-t-1}}{\psi}\ln\left(-w'(\theta)\right)\right\}\right)dF_t(\theta)\\ =&\int\left(\tilde{c}(\theta)-y(\theta)+\delta\left\{a_{t+1}-\frac{1+\delta+...+\delta^{T-t-1}}{\psi}\ln\left(-\tilde{w}'(\theta)\right)\right\}\right)dF_t(\theta)-\frac{1+\delta+...+\delta^{T-t}}{\psi}\ln(-w)\end{aligned}$$

Therefore for any w, (76) can be re-written as

$$V_t(w)=\min_{\{\tilde{c}(\theta),y(\theta),\tilde{w}'(\theta)\}}\int\left(c(\theta)-y(\theta)+\delta V_{t+1}\left(\tilde{w}'(\theta)\right)\right)dF_t(\theta)-\frac{1+\delta+...+\delta^{T-t}}{\psi}\ln(-w)$$

s.t. (78) and (79). Solution $\{\tilde{c}(\theta),y(\theta),\tilde{w}'(\theta)\}$ is independent of w, therefore, $V_t(w)=a_t$ $-\frac{1+\delta+...+\delta^{T-t}}{\psi}\ln(-w)$

For the rest of the proof it is useful to note that this implies that for any two $\check{w},\hat{w}$,

$$\frac{\hat{w}}{\check{w}}=\frac{w'(\theta;\hat{w})}{w'(\theta;\check{w})} \tag{80}$$

Now we define the CIA tax system. First we define all variables as a function of shocks, and then we invert them. Let

$$\tilde{\omega}_t\left(\theta^t\right)=\frac{1}{\psi}\ln\left(-w'\left(\theta^t\right)\right)$$

and have $\tilde{g}_t(\tilde{\omega}_{t-1},\theta_t)$ satisfy

$$\tilde{\omega}_t\left(\theta^t\right)=\tilde{g}_t(\tilde{\omega}_{t-1},\theta_t). \tag{81}$$

Consider two histories, $\check{\theta}^t$ and $\hat{\theta}^t$ with a property that $\check{\theta}_t = \hat{\theta}_t$. Then

$$\begin{aligned}\tilde{\omega}_t\left(\check{\theta}^t\right) - \tilde{\omega}_t\left(\hat{\theta}^t\right) &= \frac{1}{\psi}\ln\left(\frac{w'\left(\check{\theta}^t\right)}{w'\left(\hat{\theta}^t\right)}\right) \\ &= \frac{1}{\psi}\ln\left(\frac{w'\left(\check{\theta}^{t-1}\right)}{w'\left(\hat{\theta}^{t-1}\right)}\right) \\ &= \tilde{\omega}_t\left(\check{\theta}^{t-1}\right) - \tilde{\omega}_t\left(\hat{\theta}^{t-1}\right)\end{aligned}$$

where we used (80) on the second line. This implies that (81) can be re-written

$$\tilde{\omega}_t\left(\theta^t\right) = \tilde{g}_t\left(\theta_t\right) + \tilde{\omega}_{t-1}. \tag{82}$$

Finally, we are ready to define the CIA system. Let

$$\omega_t\left(y\left(\theta^t\right)\right) = \tilde{\omega}_t\left(\theta^t\right)$$

and

$$g_t\left(\omega_{t-1}\left(y\left(\theta^{t-1}\right)\right), y\left(\theta_t\right)\right) = \tilde{g}_t\left(\tilde{\omega}_{t-1}\left(\theta^{t-1}\right), \theta_t\right).$$

>From (82), g_t and ω_t satisfy[9]

$$\omega_t\left(y\right) = g_t\left(y\right) + \omega_{t-1}.$$

Define k_t^* as in (39). Finally, we define taxes on labor as

$$\tilde{T}_t\left(\theta^{t-1}, \theta_t\right) = y_t\left(\theta^{t-1}, \theta_t\right) - c_t\left(\theta^{t-1}, \theta_t\right) + \delta^{-1}k_t^* - k_{t+1}^* - \tilde{\omega}_{t-1}\left(\theta^{t-1}\right) + A_t$$

where A_t is a constant. Since $y_t\left(\theta^{t-1}, \theta_t\right)$ does not depend on θ^{t-1} since $y\left(\theta^{t-1}, \theta_t\right)$ is a solution to (**??**), and $c_t\left(\theta^{t-1}, \theta_t\right) + \tilde{\omega}_{t-1}\left(\theta^{t-1}\right)$ do not depend on θ^{t-1} by construction, $\tilde{T}_t$ depends only on the current realization of θ_t. Therefore, we can set

$$T_t\left(y\left(\theta_t\right)\right) = \tilde{T}_t\left(\theta^{t-1}, \theta_t\right).$$

Finally we choose A_t so that

$$A_t = \int \tilde{\omega}_{t-1}\left(\theta^{t-1}\right) dF\left(\theta^{t-1}\right).$$

[9] If y_t is chosen such that there is no history θ for which $y\left(\theta\right)$ is a solution, we can set $g_t(y) = \underline{\omega}$ where $\underline{\omega}$ is a sufficiently low number

9 Numerical Appendix: Calibration and Computation

Here, we provide the details of our calibrations and computations that enable us to produce the numerical results in Sections 3 and 4. First, we discuss the calibration including the estimation of unconditional and conditional distributions. Then, we discuss our computational procedures.

9.1 Calibration

First, we need to obtain an entire unconditional (cross-sectional) distribution of types, F. We start with an empirical distribution of labor income. The dataset we use is the Panel Study of Income Dynamics (PSID).[10] We treat heads of households and their spouses as separate observations and restrict our sample to include only the observations with the total labor income of at least $\$1,000$ and total hours of at least 250. We later relax these restrictions on our sample to check the robustness of our numerical results. To obtain the initial unconditional distribution we consider all 25-30 year old individuals. We later vary this subsample to consider 20-25 and 20-30 year old individuals to ensure the robustness of our results.

Next, we estimate the actual effective marginal tax rates faced by the individuals in our sample. We use the National Bureau of Economic Research's (NBER) program TAXSIM.[11] To compute individual liabilities under U.S. federal and state income tax laws, we supply TAXSIM with individual labor income as well as with other individual data from the PSID such as marital status, dependent exemptions, dividend income, other property income, etc.

Then, as in Saez (2001), given the actual effective marginal tax rates, we determine the skill distribution generating the labor income of the agents in the sample. We compute the implied skill for each type using the individual first-order conditions as follows:

$$\theta_i = \frac{Y_i}{\left(Y_i\left(1 - T'\left(Y_i\right)\right)\right)^{1/\gamma}}. \tag{83}$$

In this formula, Y_i is the labor income of individual i and $T'\left(Y_i\right)$ is the effective marginal tax rate for that individual. Note that with the preferences of the form (23) there are no income effects. Hence, the individual labor supply decision is unaffected by the individual savings

[10] See http://psidonline.isr.umich.edu/

[11] It is freely available for use at http://www.nber.org/~taxsim/.

choice. The implied skills thus can be determined from the static consumption-labor margin as in (83).

We then estimate the implied unconditional distribution of skills non-parametrically using a kernel density estimation method. There are two considerations that we have to address. First, the PSID is "top coded", i.e., it has an income cutoff level above which no observations are collected. Second, high income individuals are undersampled in the PSID. At the same time, the analysis of Diamond (1998), Saez (2001) in the static settings, and our results above imply that the upper tail of the distribution is an important determinant of the shape of the optimal tax code. We follow Heathcote, Perri, and Violante (2009) and fit a Pareto tail in our skill distribution above the income level of $\$150,000$. We combine the fitted Pareto upper tail with the non-parametrically estimated lower part of the distribution.

We choose $\gamma = 3$, which corresponds to the Frisch elasticity of labor supply equal to 0.5. The coefficient of absolute risk aversion, ψ, is set equal to 10. We set the discount factor $\beta = 0.9852$. We chose the marginal rate of transformation across periods $\delta = 1.015$ so that the social planner at the solution of the optimal program chooses not to transfer resources between the two periods.[12]

To proceed to compute our main numerical problem of Section 4, we also need to estimate the transition probabilities for the skills. We estimate two different transition probabilities: an earlier age transition for 25-45 year old individuals, and a later age transition for 45-65 year old individuals. That is, we allow age-dependent transition probabilities: younger individuals experience different transitions than older individuals. Within each age group, however, we assume age-independent transitions.

We expand the dataset constructed above with observations for individuals of each age group and keep track of each individual income and personal characteristics over time. Since the PSID is a panel data set, with this extended sample we are now able to compute the distribution of income (in real terms) conditional on income realization two years prior (since PSID data comes in two-year waves). Next, we follow the procedure described above to impute the distribution of skills. We estimate the conditional distribution of skills non-parametrically applying a kernel density estimation method. Then, we extrapolate the tails to match the data

[12] To fascilitate comparison with the case of persistent shocks in Section 4, in Section 3 we take one period to be 10 years, so that the discount factor is β^{10} and the marginal rate of tranformation between periods is δ^{10}.

on the transitions of top earners from CBO (2007), CBO (2008), and Treasury (2008). Finally, assuming age independent transitions within each age group, we compute one-year transitions from the constructed two-year transitions so that we obtain transitions for each period as we consider $T = 40$.

9.2 Computation

It is well known that incentive problems of mechanism design pose a significant challenge for numerical analysis - classical expositions of the computational approaches and challenges are Wilson (1996) and Judd (1998). Significantly, we have restricted ourselves here (as it is generally done in the optimal taxation literature) to agents who are heterogeneous along a single dimension - their skills.[13] Nevertheless, it is difficult to find global numerical solutions to these types of problems with more than a small number of agent types living for a few periods. To make progress with our numerical analysis, we rely on our theoretical analysis as well as on recent contributions of Su and Judd (2007) to inform our choices of computational techniques.

First, we discuss how we compute the optimal labor and savings distortions in the illustrative example of Section 3 with i.i.d. shocks and two periods ($T = 2$). The small number of periods and uncorrelated shocks make the problem size tractable to be solved in sequential form by direct optimization. Then, we discuss how we solve our main numerical problem in Section 4 with persistent shocks and forty periods ($T = 40$) by exploiting the recursive structure of the dual formulation of the planner's problem.

Because in the illustrative example of Section 3 we only have two periods and the distribution in the second period does not depend on the shock realization in the first period, we can attack the computational problem directly as a large nonlinearly constrained optimization problem. We start by writing the primal formulation of the planner's problem in the form of problem (2). The mechanism design problem in its general form is a bi-level maximization problem (alternatively, a mathematical programming problem with equilibrium constraints). The outer-level maximization of the planner has to take into account the best response of the

[13] One notable exception is Judd and Su (2006). They show that when types are heterogeneous along more than one dimension, the planner's problem is a difficult nonlinear optimization problem since the linear independence constraint qualification does not need to hold at the solution. Judd and Su (2006) find that the results with multidimensional types may differ substantially.

agents, which is the outcome of the inner-level maximization of each agent type with respect to the type reported. In other words, incentive constraints are individual agent type maximization problems with type report as a choice variable. We follow the usual convention of computationally approaching these types of problems (see e.g. Judd (1998)) by writing the incentive constraints as inequalities (without relying on simplifying the incentive compatibility constraints with the envelope theorem) as in problem (2).

The next step is to numerically find the allocation that is a global solution to the resulting large nonlinearly constrained maximization problem. We use KNITRO with a crossover option that has been shown to be able to accurately and robustly solve this type of problems. We use the interior-point local method with conjugate gradient iteration. The acceptance criterion is a penalty function. Our globalization strategy is to explore multiple (1000) feasible starting points. Once the problem is correctly scaled, we observe quadratic convergence to the solutions reported in Section 3.[14]

More specifically, we solve the planner's maximization problem using KNITRO's implementation of the interior-point algorithm with the conjugate gradient iteration to compute the optimization step. Conjugate gradient iteration offers a way of dealing with possible Jacobian and Hessian singularities. The interior-point approach is one of the most efficient and stable methods that are currently available for solving large nonlinear optimization problems. The interior-point algorithm uses a trust-region Newton method to solve the barrier problem and an l_1 penalty barrier function. We find that the interior-point algorithm provides a good approximate estimate of the solution and the optimal set of active constraints. To compute accurate estimates of the solution, including Lagrange multipliers, we proceed to use KNITRO's crossover option. That is, we switch to an active-set iteration that uses the output of the interior-point algorithm as its input. The implementation of the active-set algorithm is based on the sequential linear quadratic programming.

Once we obtain the constrained optimal allocation, the final step is to compute the optimal labor and savings distortions from their definitions in equations (16) and (17) respectively.

The illustrative numerical example of Section 3 provides us with a way to build intuition in the dynamic model as well as with a benchmark test case for the computational algorithm

[14] To streamline communication with KNITRO we use AMPL. We gratefully acknowledge the use of software licences provided to one of us during participation in the Institute on Computational Economics 2009 organized by Ken Judd at the University of Chicago.

for our main numerical problem of Section 4 that we discuss next.

Our main numerical problem in Section 4 has persistent shocks and forty periods ($T = 40$). To be able to solve the problem of this size and complexity we exploit the recursive structure of the dual formulation of the planner's problem as well as the quasi-linear nature of preferences (18).

We proceed in three stages. The first stage is a value function iteration. That is, we start from period T and proceed by backward induction. First, we must solve period $t = T$ problem for a fixed set of values of the state vector and compute V_T for each one of then. Then we can approximate V_T and proceed to period $t = T - 1$ where we use the approximation as the basis for the interpolation of V_T to any value of the state vector to solve for V_{T-1}. And so on until we compute V_1. However, we know that with the exponential preferences that we use for our numerical simulations we have

$$V_t\left(w, w_2, \theta_-\right) = a_t(\frac{w_2}{w}|\theta_-) - \frac{1 + \delta + ... + \delta^{T-t}}{\psi} \ln\left(-w\right)$$

and in particular

$$V_T(w, w_2, \theta_-) = a_T\left(\frac{w_2}{w}|\theta_-\right) - \frac{1}{\psi}\ln(-w).$$

This means two things for our computations. First, if we discretize the type space Θ, then we only need to consider w and $\frac{w_2}{w}$ as the state variables. Second, we do not need to approximate V_t as a whole, rather we only need to approximate a_t, which tremendously improves the quality of approximation of V_t. We approximate a_t's using a shape-preserving LAD method with Chebyshev polynomials.[15]

To be able to compute the full constrained optimal allocation, we find w_0 such that $V_1\left(w_0\right) = 0$. That is the second stage. Given V_1 computed in the first stage, we search for an interval containing zero by quadratically increasing steps. Then we converge to w_0 by bisection.

The third and final stage is to compute the optimal allocations and then the optimal labor and savings distortions using their respective definitions. Given V_t's and w_0 from the first two stages, we can now compute the optimal allocations by forward induction. We start with w_0 computed in the second stage and roll out the solution from period $t = 1$ all the way to period $t = T$.

[15] For more on this, see e.g. Judd (1996) and Judd (1998).

We note that for this three-stage computational procedure to be feasible it is absolutely essential to have an exceptionally fast, efficient, and robust optimization algorithm to solve all of the separate period t problems of each stage. We use KNITRO implementation of the interior-point method with conjugate gradient iteration and a crossover to the active-set algorithm as described above. Note also that we can use our direct optimization approach to the primal planner's problem in the example above as a first check.

Finally, we verify that the first order approach is valid by solving for a competitive equilibrium given the optimal distortions and verifying that no individual agent is better off at any time by deviating.

References

Albanesi, S., and C. Sleet (2006): "Dynamic optimal taxation with private information," *Review of Economic Studies*, 73(1), 1–30.

Ales, L., and P. Maziero (2007): "Accounting for private information," *working paper*.

Atkinson, A., T. Piketty, and E. Saez (2009): "Top incomes in the long run of history," *NBER Working Paper*.

Atkinson, A., and J. Stiglitz (1976): "The design of tax structure: direct versus indirect taxation," *Journal of Public Economics*, 6(1-2), 55–75.

Battaglini, M., and S. Coate (2008): "Pareto efficient income taxation with stochastic abilities," *Journal of Public Economics*, 92(3-4), 844–868.

CBO (2007): "Historical Effective Federal Tax Rates, 1979 to 2005," *Congressional Budget Office*.

——— (2008): "Historical Effective Federal Tax Rates, 1979 to 2005: Supplement with Additional Data on Sources of Income and High-Income Households," *Congressional Budget Office*.

Chamley, C. (1986): "Optimal taxation of capital income in general equilibrium with infinite lives," *Econometrica*, 54(3), 607–622.

Diamond, P. (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88(1), 83–95.

Farhi, E., and I. Werning (2010): "Insurance and Taxation over the Life Cycle," *working paper*.

Fernandes, A., and C. Phelan (2000): "A recursive formulation for repeated agency with history dependence," *Journal of Economic Theory*, 91(2), 223–247.

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.

——— (2007): "Optimal Taxation With Endogenous Insurance Markets*," *The Quarterly Journal of Economics*, 122(2), 487–534.

Golosov, M., A. Tsyvinski, and I. Werning (2006): "New dynamic public finance: A user's guide," *NBER Macroeconomics Annual*, 21, 317–363.

Grochulski, B., and N. Kocherlakota (2007): "Nonseparable Preferences and Optimal Social Security Systems," *NBER Working Paper*.

Heathcote, J., F. Perri, and G. Violante (2009): "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006," *forthcoming Review of Economic Dynamics*.

Judd, K. (1985): "Redistributive taxation in a simple perfect foresight model," *Journal of Public Economics*, 28(1), 59–83.

——— (1996): "Approximation, perturbation, and projection methods in economic analysis," *Handbook of Computational Economics*, 1, 509–585.

——— (1998): *Numerical methods in economics*. The MIT Press.

——— (1999): "Optimal taxation and spending in general competitive growth models," *Journal of Public Economics*, 71(1), 1–26.

Judd, K., and C. Su (2006): "Optimal Income Taxation with Multidimensional Taxpayer Types," *working paper*.

Kapicka, M. (2008): "Efficient allocations in dynamic private information economies with persistent shocks: A first order approach," *working paper*.

Kocherlakota, N. (2005): "Zero expected wealth taxes: A Mirrlees approach to dynamic optimal taxation," *Econometrica*, 73(5), 1587–1621.

Kocherlakota, N. (2010): *The New Dynamic Public Finance*. Princeton University Press, USA, forthcomming.

Kolmogorov, A., and S. Fomin (1975): *Introductory real analysis*. Dover publications.

Mirrlees, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(2), 175–208.

——— (1986): "The theory of optimal taxation," *Handbook of mathematical economics*, 3, 1197–1249.

Pavan, A., I. Segal, and J. Toikka (2010): "Dynamic mechanism design: Incentive compatibility, profit maximization and information disclosure," .

Prescott, E., and R. Townsend (1984): "Pareto optima and competitive equilibria with adverse selection and moral hazard," *Econometrica*, 52(1), 21–45.

Saez, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–229.

Su, C., and K. Judd (2007): "Computation of Moral-Hazard Problems," *working paper*.

Treasury (2008): "Income Mobility in the U.S. from 1996 to 2005," *Report of the Department of the Treasury*.

Tuomala, M. (1990): *Optimal income tax and redistribution*. Oxford University Press, USA.

Vickrey, W. (1939): "Averaging of income for income-tax purposes," *The Journal of Political Economy*, pp. 379–397.

Vickrey, W. (1947): *Agenda for progressive taxation*. The Ronald Press Company.

Weinzierl, M. (2008): "The surprising power of age-dependent taxes," *working paper*.

Werning, I. (2009): "Nonlinear Capital Taxation," *working paper*.

Wilson, R. (1996): "Nonlinear pricing and mechanism design," *Handbook of Computational Economics*, 1, 253–293.