Optimal Dynamic Taxes*

Mikhail Golosov
Yale
and NES

Maxim Troshkin
University of Minnesota
and FRB Minneapolis

Aleh Tsyvinski
Yale
and NES

April 2011

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 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 wedge in the dynamic setting 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 redistributive motive of the planner. We also derive a novel formula to analyze the determinants of the savings distortions.

Our second set of results is to show that the labor wedge tends to zero for sufficiently high skills in both the i.i.d. case and, under certain conditions, in the case of persistent shocks. This is in sharp contrast to the static case with Pareto tail of the skill distribution in 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 third contribution 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.

*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. We gratefully acknowledge the use of software licences provided to one of us during the Institute on Computational Economics at the 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 income 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 formula

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

representing the savings distortion. The analysis of Diamond (1998) and Saez (2001) is static and thus does not consider intertemporal decisions – our formula for the savings wedge thus new to that literauture. The derivation of the formula interpreting the savings distortion is also 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.

Our second set of results is to show that the labor wedge tends to zero for sufficiently high skills in both the i.i.d. case and, under certain conditions, in the case of persistent shocks. 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 (2010) and Pavan, Segal, and Toikka (2010). In numerical simulations, we verify its sufficiency.

The third 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 0.9% of consumption compared to the constrained optimum. The optimal age-*in*dependent labor distortion yields a welfare loss of 1.6%. 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 4.3% compared to the constrained optimum. The optimal age-*in*dependent labor distortion yields a welfare loss of 5%. We conclude that the welfare gains of using optimal nonlinear policies are significant.

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 (2010) 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. Jacquet, Lehmann, and Van Der Linden (2010) study elasticity-based formulas in a static environment when individuals are allowed to respond along both the intensive and extensive labor margins and when income effects can prevail.

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.

2 Environment

We consider an economy that lasts T periods, denoted by $t = 1, ..., T$ ($T < \infty$)[2]. Each agent's preferences are described by a time separable utility function over consumption good $c_t \geq 0$ and labor $l_t \geq 0$,

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

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 $U_c > 0, U_{cc} < 0$, $U_l, U_{ll} < 0$.

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 his skill realization only 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 defined over lifetime utilities of the agents, $G : \mathbb{R} \to \mathbb{R}$, 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,

[2] The recursive formulation of the problem that follows makes it easy to extend the analysis to the case of infinitely lived agents. In fact, the calibration and numerical analysis is greatly simplified in the case of infinitely lived agents.

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:

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

The expectation $\mathbb{E}_0$ above is taken over all possible realizations of histories. Note that the expectation in the objective function is taken after the first period shocks are realized. The first constraint above is a dynamic incentive compatibility constraint. This constrain 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 (2010) to re-write this problem recursively. Here, we briefly describe the recursive formulation and refer to these two paper for the technical details. Let $\omega\left(\tilde{\theta}|\theta\right):\Theta\times\Theta\to\mathbb{R}$ denote *promised utility* to an agent of skill θ who reports skill $\tilde{\theta}$. 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 $t=1$:

$$V_1\left(\omega_0\right)=\min_{c,y,\boldsymbol{\omega}}\int\left(c\left(\theta\right)-y\left(\theta\right)+\delta V_2\left(\boldsymbol{\omega}\left(\theta\right),\theta\right)\right)f_1\left(\theta\right)d\theta \tag{5}$$

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 $\hat{\boldsymbol{\omega}}(\theta_-)$ 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 $\omega\left(\tilde{\theta}|\theta\right)$. Promise keeping constraints (9) ensure that next period allocations indeed deliver the expected utility $\omega\left(\tilde{\theta}|\theta\right)$ to any type $\tilde{\theta}$ who sends a report θ.

We proceed in this section by using the first order approach developed by Kapicka (2010) and Pavan, Segal, and Toikka (2010) to obtain a more manageable recursive formulation and 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" promised 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

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

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

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

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

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

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 reported type in period $t-1$.

Before characterizing the problem, we re-write the optimal problem 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) for* $t > 1$*. 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 type 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 a relatively 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:

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

We can re-write the optimal problem (10) 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:

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

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

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

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

Note that we changed the variables of minimization. Variables $l(\theta)$ and $w(\theta)$, $w_2(\theta)$ are now control variables, and $u(\theta)$ is a state variable. We can apply optimal control techniques to characterize (10).

We characterize the solution for the general model in Section 4. However, we first focus in Section 3 on the illustrative example with two periods and i.i.d. shocks that highlights the key determinants of the dynamic distortions.

3 Illustrative example

In this section, we consider an illustrative example of a two-period economy with i.i.d. shocks which are drawn from the same distribution $F(\theta)$ in each period. 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 are a special case of the general model considered in Section 4, in particularly Propositions 1 and 2. Specifically, we proceed as follows. We first derive expressions that determine the forces behind the optimal capital and labor distortions. We then show how the dynamic nature of the problem modifies and changes the results of the static analysis of Diamond (1998) and Saez (2001). An important insight here is that the dynamic problem is similar to the static Mirrleesian problem with two goods – today's consumption and future promises. Importantly, we show that for the high incomes the labor distortion tends to zero. This is in contrast to the analysis of Diamond (1998) and Saez (2001) where the labor distortion is increasing and converges to a high level for a calibrated distribution of skills. We then compute the optimal wedges and show how they are affected by the key parameters of the calibration. The analysis of this section provides insights into the nature of the optimal capital and labor distortions that hold in the general model.

3.1 Characterizing optimal wedges

When shocks are i.i.d., the 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 $\tilde{f}(x) = \frac{\Psi(x)f(x)}{\int_0^{\theta}\Psi(x')f(x')dx'}$, where $\Psi(x)$ is

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

We now compare the optimal labor distortion in period 1 given by (25) to the labor distortion in period 2 (which coincides with the static distortion) 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 $\Psi'(\theta) = -\beta\psi\frac{w'(\theta)}{\bar{U}_c(\theta)}\Psi(\theta)$.

In general, it is difficult to determine the sign of marginal promised utility, $w'(\theta)$. It is instructive, however, to consider a case where 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 the following 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 similar to period 2, the marginal utility of consumption in period 1 must be decreasing and converging to zero, 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 distortions. 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 social welfare 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 high incentive problems. That is for those with a large labor wedge. Recall that in Mirrleesian models distortions are imposed because of the incentive problems. Higher wedge thus relates to the more severe incentive problems on an agent of a particular type. 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

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

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 $T'_{D,1}(\theta) \to 0$. This result is important as it contrasts the analysis of Diamond (1998) and Saez (2001) in which the labor distortion is increasing. In their calibrated examples, the labor income wedge can be as high as $50-70\%$ for high income types.

3.2 Empirical illustration

We now compute optimal labor and savings wedges for the illustrative two-period example considered above. We study the differences between labor distortions in a static and in this dynamic model and how the results depend on the main forces identified in the analysis above. The intuition developed here is later helpful in understanding the results of the simulations in the general economy with correlated shocks in Section 4.

We start by estimating the unconditional distribution of skills $f(\theta)$ implied by the U.S. micro level data. To do that we need to construct a dataset of implied individual skills θ. Given exponential preferences (23),[3] the implied skill θ for an individual i can be computed from the individual first order conditions as follows:

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

where Y_i is the labor income of individual i observed in the data and $T'(Y_i)$ is the effective marginal tax rate that the individual was facing when she earned her labor income.

Thus our first step is to obtain labor income. Our main data source is the Panel Study of Income Dynamics (PSID). We use all of the PSID waves from 1990 to 2006 and combine them all into a single cross section for the illustrative example of this section. Later we exploit the longitudinal feature of the PSID when we turn to the life cycle model with persistent shocks in Section 4. In our sample, we treat heads of households and their spouses as separate observations. We restrict our sample to include only observations with the total labor income of at least $\$1,000$ in 1990 dollars and at least 250 total hours worked in a year. In total, we have more than 50 thousand observations. We also check our results for robustness with an alternative sample where individuals older than 65 are excluded. We obtain individual labor incomes Y_i together with the corresponding individual characteristics such as marital status,

[3] 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.

state of residence, number of children, and other types of income. These serve as inputs for the next step – the estimation of the effective marginal tax rates $T'(Y_i)$ faced by the individuals in our dataset. We do that by using National Bureau of Economic Research's program TAXSIM.[4] Supplied with the dataset we constructed in the previous step, TAXSIM allows us to estimate actual effective individual liabilities under the U.S. federal and state income tax laws. Next, we assume the Frisch elasticity of labor supply of 0.5, which implies $\gamma = 3$, and arrive at the dataset of individual skills θ_i.

We estimate the implied unconditional distribution of skills $f(\theta)$ non-parametrically using a kernel density estimation with the normal kernel function and $\mathbb{R}_+$ as the support. The analysis of Diamond (1998) and 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 schedule. One concern is that high income individuals are undersampled in the PSID or even that the PSID is "top coded", i.e., there is an income cutoff level above which no observations are collected. To address this concern we fit a Pareto distribution to the right tail of our skill distribution above the income level of $150,000. Our maximum likelihood estimate of the Pareto parameter is 3.11 with the expected standard error of 0.14.[5] We splice the lower income non-parametric part with the higher income Pareto tail to obtain the unconditional distribution of skills in Figure 1.

We set the coefficient of *absolute* risk aversion, ψ, equal to 10.[6] The discount factor is $\beta = 0.9852$ and the marginal rate of transformation across periods is $\delta = 1.015$ so that the social planner at the solution of the optimal program chooses not to transfer resources between the two periods.[7]

In this illustrative example we only have two periods, and the distribution in the second period does not depend on the shock realization in the first period. This is the reason we

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

[5] Note that f is the distribution of skills, not income. With exponential preferences (23), it can be shown that if the tail of the skill distribution is Pareto distributed with parameter a then the tail of income distribution is Pareto with parameter $b = a(\gamma - 1)/\gamma$. In this case, $a = 3.11$ corresponds to $b = 2.07$, which is consistent with the historical range of $1.38 - 2.63$ of estimated coefficients for the U.S income distribution in Atkinson, Piketty, and Saez (2009).

[6] In our numerical simulations consumption ranges from 0.1 to 1 implying that relative risk aversion ranges from 1 to 10 when $\psi = 10$.

[7] To fascilitate comparison with the case of persistent shocks over the life cycle in Section 4, we take one period to be 20 years, so that the discount factor is β^{20} and the marginal rate of tranformation between periods is δ^{20}.

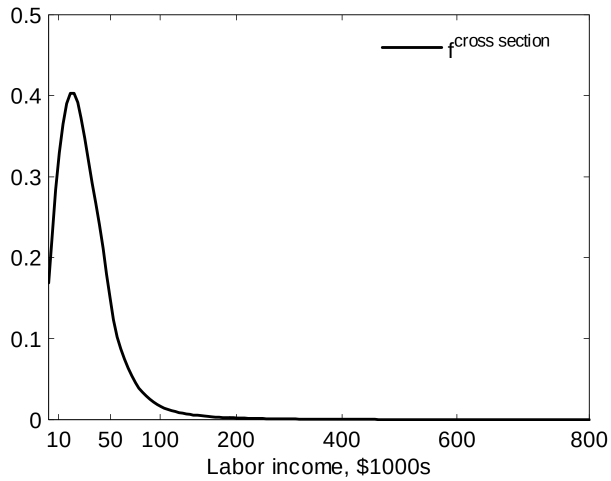

Figure 1: Estimated unconditional distribution of skills

can solve the problem directly as a large nonlinear constrained optimization problem, i.e., the primal formulation of the planner's problem (2).[8] We solve the planner's maximization problem using an implementation of the interior-point optimization algorithm with the conjugate gradient iteration to compute the optimization step. The implementation we use is KNITRO.[9] 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 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

[8] A 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 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 (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).

[9] To streamline the application of KNITRO we use a modelling language AMPL.

the problem is correctly scaled, we observe quadratic convergence to a local maximum. Our globalization strategy is to explore multiple (1000) feasible starting points. Once we compute 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 results are presented in Figure 2.

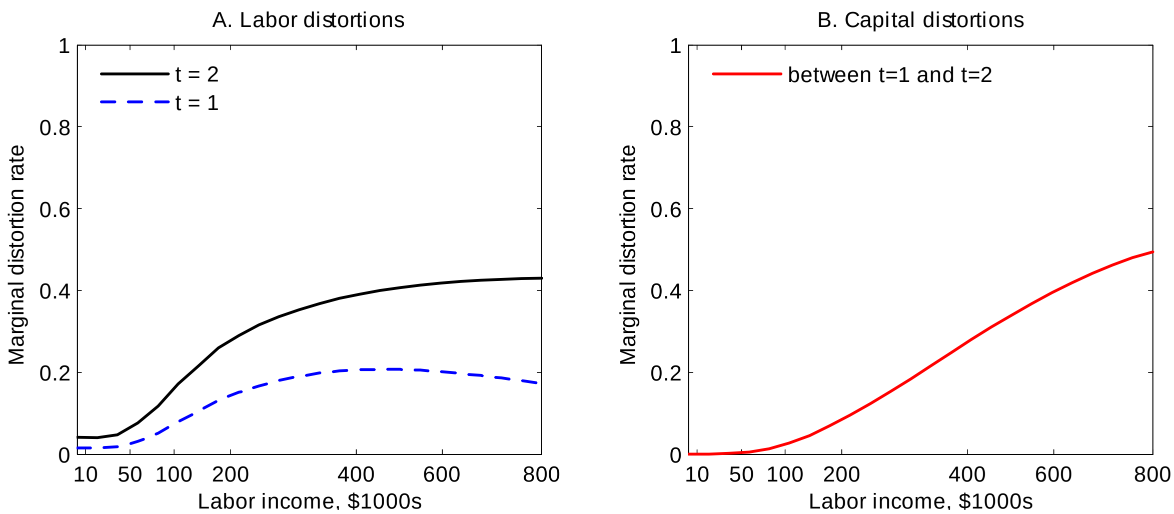

Figure 2: Optimal labor and capital distortions in the illustrative example

First, consider the optimal marginal labor distortions in period 2 – higher, solid line in panel A of Figure 2. The distortions coincide with those in a static economy and are similar to those in Diamond (1998) and Saez (2001). For lower incomes, the optimal distortions exhibit somewhat of a shallow U-shaped pattern starting from 4.1% and decreasing to 4.0% at $20,000 income before increasing for all incomes above. The distortion rapidly increases from 8% at $50,000 to 39% at $500,000 eventually reaching 43% at $800,000 and tending toward the analytic limit given by $T'_{D,2}/(1-T'_{D,2})$ increasing and converging to γ/a as discussed above.

Next, consider the optimal marginal labor distortions in period 1 – lower, dashed line in panel A of Figure 2. These distortions are lower than those in period 2 at all income levels. This is consistent with the discussion of the differences between formulas (24) and (25) above. Importantly, in contrast to period 2 the labor distortions in period 1 start to decline around annual income of $500,000. Consistent with the theoretical results above, they tend to zero at incomes above $2 million.

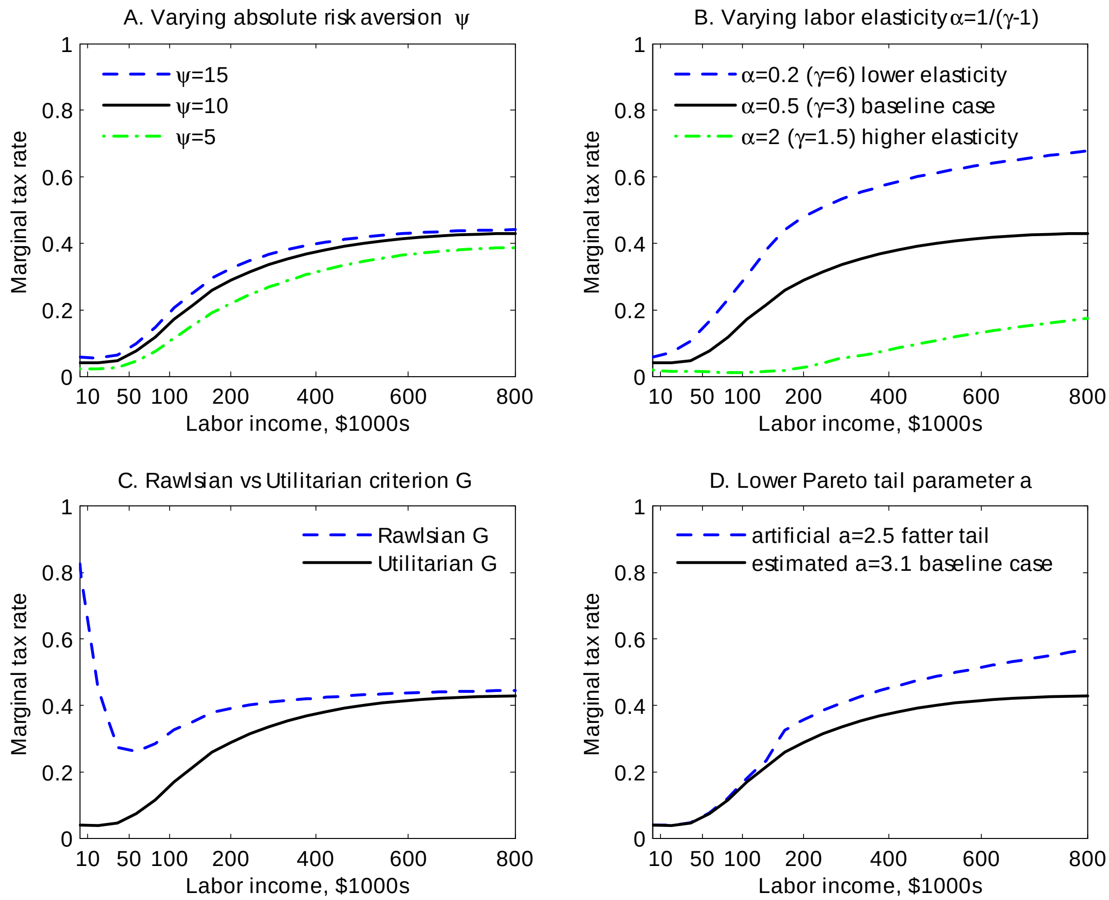

Figure 3: Comparative statics for optimal labor distortions

The savings distortions are represented by the solid line in panel B of Figure 2. The wedge is positive and increases for all income levels. It ranges from less than 0.1% for low incomes increasing very modestly up to 0.5% at $50,000 and increasing more rapidly after that reaching 49% at $800,000. 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 higher skilled agents.

Finally, we study how the results are affected by parameter choices and the main forces identified in the analytical discussion above. Figure 3 presents a series of comparative statics for the optimal labor distortions. The figure displays the labor distortions in a static setting, which, as we have seen, can be equivalently interpreted as the second period of the illustrative example of this section. The solid line in all four panels of Figure 3 represents our baseline case and is identical to the solid line in panel A of Figure 2.

Panel A of Figure 3 presents the optimal labor distortions for three values of absolute risk aversion parameter ψ. While the effects on the labor distortions appear to be modest, higher values of risk aversion result in higher optimal labor distortions at all income levels.

Panels B through D of Figure 3 present comparative statics with respect to the three main forces identified in the theoretical analysis above: labor elasticity, degree of redistribution in the social welfare function, and the distribution of skills or more specifically the tail ratio of the distribution. Consider panel B first. As expected, less elastic labor supply results in optimal labor distortions significantly higher (dashed line) than the baseline case, especially for high income levels. At $\$800,000$ income the distortion reaches 66% compared to 43% in the baseline case, once again tending to the analytic limit. More elastic labor has the opposite effect of lowering optimal labor distortions (dot-dashed line) reaching just below 20% at $\$800,000$. Next, consider panel C. Rawlsian criterion. A much more redistributive social welfare function like Rawlsian results in generally higher distortions (dashed line in panel C). The U-shape at low incomes becomes more pronounced and the level of distortion for the lowest income reaches 82% before declining to 26% at $\$60,000$, rising for incomes above, and reaching 45% at $\$800,000$. Finally, in panel D, we alter our estimated skill distribution by artificially making the right tail fatter until the maximum likelihood estimate of the Pareto parameter for the tail reached 2.5. The labor distortions with this new distribution are represented by the dashed line in panel D. As expected, the higher incomes are the ones affected most and the optimal distorion is increased. The difference is larger for higher income levels.

4 General case of persistent shocks

We now return to the general problem stated in (22) with T periods and general skill processes. We show how the results derived in the illustrative example in Section 3 extend to the general environment. In the analysis below, we need to make two additional assumptions – that V_t is concave and that only the downward incentive compatibility constraints bind.[10]

[10] Both of these conditions are satisfied in the numerical simulations we performed. These conditions are essentially technical and are only used in the part of the proof that rules out that the $T_{D,t}$ does not converge to 1.

4.1 Characterizing optimal wedges

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

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

Part 1. The optimal labor distortion in period t satisfies

$$\frac{T'_{D,t}(\theta)}{1-T'_{D,t}(\theta)} = \gamma \frac{1-\tilde{F}_t(\theta|\theta_-)}{\theta \tilde{f}_t(\theta|\theta_-)} \int_\theta^\infty \left(1 - \frac{\alpha_t(x)\bar{U}_{c,t}(x)}{\lambda_t}\right) \frac{\tilde{f}_t(x|\theta_-)\,dx}{1-\tilde{F}_t(\theta|\theta_-)} \tag{27}$$

where

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

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

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

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

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

$$1-\tau_{S,t}(\theta) = z_t(\theta)\left(1 - \frac{\psi_t(\theta)}{\gamma} T'_{D,t}(\theta) y_t(\theta)\right) \tag{28}$$

where

$$z_t(\theta) = \frac{\int_0^\infty \left(1 - \tilde{\zeta}_{t+1} \frac{f_{2,t+1}(x|\theta_-)}{f_{t+1}(x|\theta_-)}\right) \bar{U}_{c,t+1}(x)\, \tilde{f}_{t+1}(x|\theta) dx}{\int_0^\infty \bar{U}_{c,t+1}(x)\, f_{t+1}(x|\theta) dx}. \tag{29}$$

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 (27) to (25) there are several differences. First, marginal utility $\bar{U}_c(x)$ in (27) 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 period 1 in equation (25) as we assumed that the 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 $\Psi'(\theta) = -\beta\psi(\theta)\frac{\omega_{1,t}(\theta|\theta)}{\bar{U}_{c,t}(\theta)}\Psi(\theta)$. Moreover, if $\omega_1(\theta|\theta) > 0$, then $\Psi'(\theta) \leq 0$ for all θ and the distribution F has a property:

$$\frac{1 - \tilde{F}_t(\theta|\theta_-)}{\theta\tilde{f}_t(\theta|\theta_-)} \leq \frac{1 - F_t(\theta|\theta_-)}{\theta f_t(\theta|\theta_-)}, \tag{30}$$

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 (28). 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_-). For the analysis below we assume that only downward IC bind and that the value functions in iid case are concave which we verify in numerical simulations. 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 $f_{t+1}^{\varepsilon}(\theta|\theta_{-})$ *with a property that* $\lim_{\varepsilon\to 0} f_{2,t+1}^{\varepsilon}(\theta|\theta_{-}) \to 0$ *uniformly. Suppose the optimal* $\left(w^{\varepsilon}, w_2^{\varepsilon}, T_{D,t}^{\varepsilon\prime}\right)$ *are bounded. Then there exists* $\xi'' > 0$ *s.t. for all* $|\varepsilon| < \xi''$ $T_{D,t}^{\varepsilon\prime}(\theta) \to 0$ *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 that the 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.

4.2 Quantitative analysis

We now numerically study a calibrated general case with empirical persistent skill shocks based on the U.S. micro level data.

We model 40 years of working life, i.e., individuals of ages 25 to 65. We use all of the PSID waves from 1990 to 2006. Since the PSID waves come in two year intervals and we estimate conditional distributions (transition probabilities) from individual transitions between these waves, we let one period in the model correspond to two years in the data, i.e., $T = 20$. We check that our results are robust when the number of periods is doubled to $T = 40$.

The calculations are done with the same parameters as in the baseline case of the illustrative example in Section 3: the coefficient of *absolute* risk aversion, ψ, is equal to 10, the discount factor is $\beta = 0.9852$, and the marginal rate of transformation across periods is $\delta = 1.015$ so that the social planner at the solution of the optimal program chooses not to transfer resources between periods.[11]

[11] When we take one period to be 2 years the discount factor is β^2 and the marginal rate of tranformation between periods is δ^2.

As in Section 3, we first construct a dataset of implied individual skills θ_i from the individual first order conditions.[12] Once again, we assume exponential preferences of the form (23). Since there are no income effects, the individual labor supply decision is unaffected by the individual savings choice and thus the implied skills can be determined from the static consumption-labor margin. We use the data on the individual labor incomes observed in the PSID, Y_i, and the effective marginal tax rate that the individual was facing when she earned her labor income, $T'(Y_i)$, estimated using TAXSIM.

We start by estimating the initial unconditional distribution of implied skills among the 25 year olds, $f_1(\theta)$. We consider the 25 year olds from all of the PSID waves to obtain a samle of about 8,200. We estimate $f_1(\theta)$ non-parametrically using a kernel density estimation with the normal kernel function and $\mathbb{R}_+$ as the support. The resulting distribution is shown as a dot-dashed line in Figure 4. A comparison with Figure 1 reveals this to be a less unequal distribution with a thinner right tail. The unconditional expected income is approximately \$25,000.

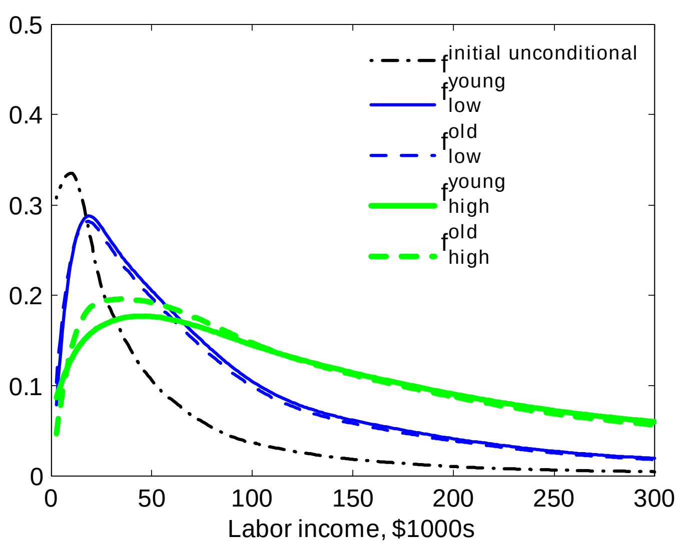

Figure 4: Initial unconditional and some conditional distributions

Next, we exploit the longitudinal feature of the PSID to estimate transition probabilities, i.e., conditional distributions $f(\theta|\theta\_)$. We consider all individual transitions between adja-

[12]See Section 3 for details.

cent PSID waves to obtain above $27,000$ transitions.[13] We break these transitions into two subsamples – when the individual is younger than 45 at the beginning of the transition (about $15,000$ 'young' transitions) and when the individual is 45 or older (about $12,000$ 'old' transitions). We estimate two separate conditional distributions $f^{young}(\theta|\theta\_)$ and $f^{old}(\theta|\theta\_)$. Thus we assume age dependence between the groups and age-independent transitions within each group. In other words, we allow younger individuals to experience different transition probabilities than older individuals; within each age group, we assume age-independent transition probabilities.[14] We estimate each conditional distribution non-parametrically using a kernel density estimation with the normal kernel function and $\mathbb{R}_+$ as the support. Figure 4 displays four examples of the estimated conditional distributions: one young distribution conditional on $50,000 income ('low', thiner solid line), one conditional on $150,000 income ('high', thicker solid line), and the same pair for old (thinner and thicker dashed lines respectively). The conditional expected incomes are as follows (rounded to the nearest thousand): $60,000 for young low, $58,000 for old low, $124,000 for young high, and $110,000 for old high.

To be able to numerically solve the problem of this size and complexity (i.e., with multitude of periods and correlated shocks) we exploit the recursive structure of the dual formulation of the planner's problem analyzed in Section 2. We proceed in three stages.

The first stage is a value function iteration. We start from period T and proceed by backward induction. First, we solve period $t = T$ problem for a fixed set of values of the state vector and compute V_T for each 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}. We continue until we compute V_1. With the exponential preferences we can show that

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

and in particular

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

[13] Note that the PSID is not a balanced panel. An individual may appear in a wave, stay for several waves, and then disappear. We consider all wave to wave transitions as separate data points.

[14] We stop at just two groups to have sufficient number of data points to estimate all conditional distributions. There is nothing in our computational solution method that would stop us from having a different transition matrix for each period, provided that we had enough data to obtain those transition matrices.

This means two things for our computations. First, after discretizing the type space Θ, we only need to consider w and $\frac{w_2}{w}$ as the state variables. That is, our state space is discretized in skill dimention and is continuous in the other two dimentions. 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 the approximation of V_t. We approximate a_t's using a shape-preserving LAD method with Chebyshev polynomials. The evaluation nodes are chosen as the roots of Chebyshev polynomials.[15]

To compute the full constrained optimal allocation, we find w_0 such that $V_1(w_0) = 0$. This is the second stage. Given V_1 computed in the first stage, we efficiently search for an interval containing zero by quadratically increasing jumps. 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 policy fnctions that we also approximated during the first stage. 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 $t = T$.[16] Finally, we verify that the first order approach is valid by veryfying sufficient conditions ex post.

Figure 5 presents the results of our numerical simulations. Consider first panels A and B of Figure 5. Panel A presents the labor distortions at ages 25, 35, 45, 55, and 65 (with generally lower lines representing younger ages) for the agent with a history of shocks up to that period such that in each previous period she had income of \$50,000. Panel B displays the labor distortions at ages 25, 35, 45, 55, and 65 (with generally lower lines representing younger ages) for the agent with a history of shocks up to that period such that in each previous period he had income of \$150,000. The lowest line in each panel is the unconditional labor wedge at the initial age of 25, which is identical in both panels. There are three key features of interest with the labor wedge results.

First, both for the agent with the history of \$50,000 incomes and for the agent with the history of \$150,000 incomes, the average conditional labor wedges are increasing with age.

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

[16] 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 in Section 3. Note also that we can use our direct optimization approach to the primal planner's problem in the illustratve example as a check.

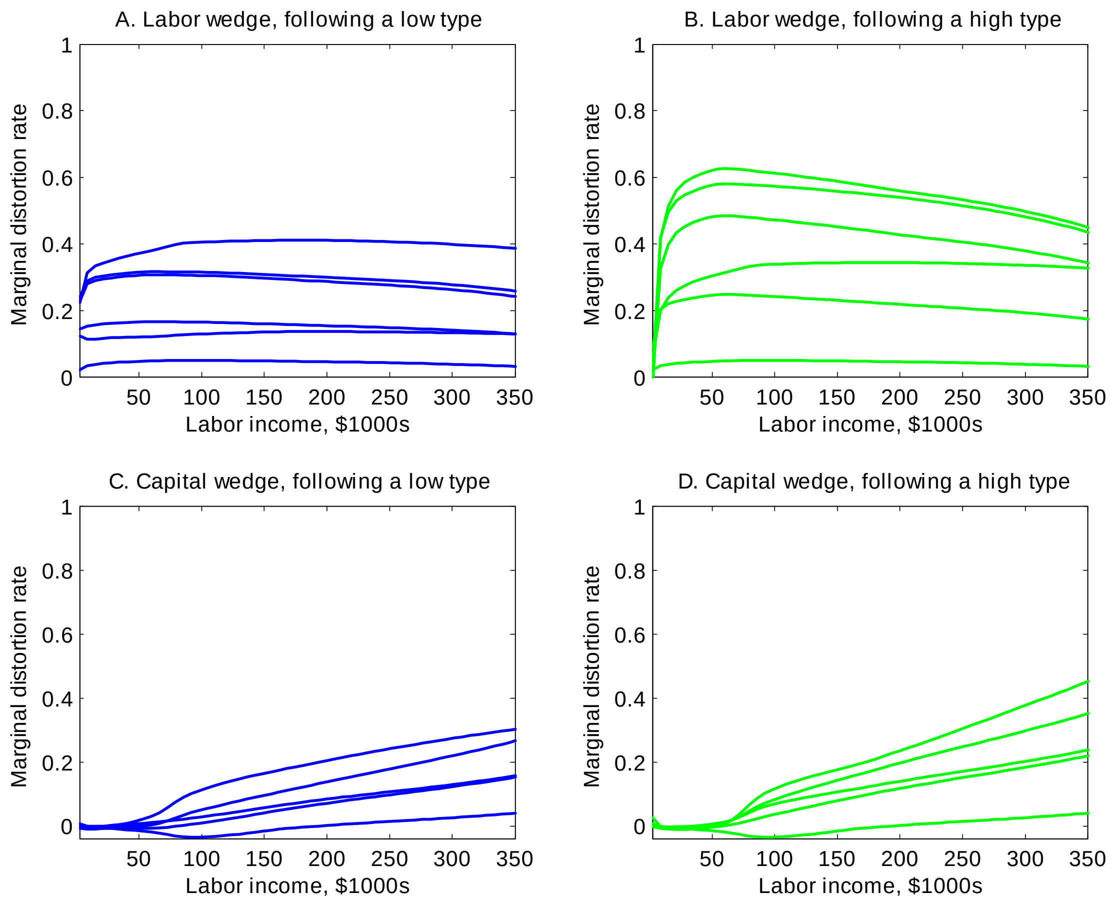

Figure 5: Labor and capital distortions with persistent shocks

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 an agent with a history of $\$50,000$ incomes are generally lower than those for an agent of the same age with a history of incomes of $\$150,000$. There are two forces driving the differences in taxes for these two agents that follow from the discussion of Proposition 1: (i) the additional redistribution implied by the term $\left(\zeta - \frac{f_2(x|\theta_-)}{f(x|\theta_-)}\right)$ in equation (27) and (ii) the differences between conditional and unconditional distributions of skills as well as the differences in conditional distributions among agents, specifically between those with a history of relatively low incomes and those with a history of relatively high incomes as is evident from the discussion of Figure 4 above.

Third, consistent with Proposition 1, the labor wedge decreases for the high enough incomes at every age for any history. 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 in panels C and D of Figure 5. As in the case of labor wedges, panel C presents the savings distortions at ages 25, 35, 45, 55, and 65 (with generally lower lines representing younger ages) 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 distortions at ages 25, 35, 45, 55, and 65 (with generally lower lines representing younger ages) for the agent with a history of shocks up to that period such that in each previous period he had income of $\$150,000$. In both cases, the conditional savings distortions are generally increasing in current period realization of income: they are close to zero for current incomes below $\$100,000$ and increase up to 30% and 45% at $\$300,000$ income for the agents with a history of $\$50,000$ and $\$150,000$ incomes respectively.

Finally, we perform the calculations of the welfare gains of using the optimal dynamic non-linear policy. 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 0.9% of consumption. The optimal age-*in*dependent labor distortions increase the welfare loss to 1.6%. 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 4.3% of consumption compared to the constrained optimum. The optimal age-*in*dependent labor distortion yields a welfare loss of 5%. 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. In particular, an important insight is that the dynamic optimal taxation models (in a recursive formulation) are similar to the static Mirrlees models with two consumption goods. In the dynamic case, the goods are today's consumption and suitably defined future promises. This argument is the main reason that relates the analysis and characterization of the static Mirrlees models as in Diamond (1998) and Saez (2001) to our dynamic setting.

First, we focus on the case in which 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

$$\bar{U}\left(c - \frac{l^{\gamma}}{\gamma}\right) + h(w)^{17}. \tag{31}$$

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 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 $dy\left(\theta^*\right) d\tau$. The increase in tax revenues is given by $M = d\tau dy\left(\theta^*\right) \int_{\theta^*}^{\infty} f(\theta) d\theta$. It can be shown that $dy(\theta^*) = (1+\zeta) \frac{y(\theta^*)}{\theta^*} d\theta^*$, so that

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

where ζ is the elasticity of labor supply.[18]

[17] Recall that with iid shocks the value function $V_t(w, w_2, \theta_-)$ simplifies to $V_t(w)$. In this case $h(w) = V_t^{-1}(w)$.
[18] In particular, $\zeta = 1/\left(\gamma - 1\right)$.

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

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

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 (31). In this case, there is an additional tax on good w, $P(w|y)$, which we assume to converge to P' for high types (that is, as $\theta^* \to \infty$). 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 $d\tau P' \int_{\theta^*}^{\infty} \frac{dw(\theta)}{d\tau} f(\theta) d\theta$. If $P' > 0$, this effect decreases tax revenues and results in lower optimal marginal tax 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 our 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 and $-l$. To make comparisons with the previous result more straightforward, we assume that $U(c, l)$ satisfies

$$U_{cl} \frac{U_l}{U_c} = U_{cl} \frac{U_{cl}}{U_{cc}}. \tag{32}$$

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 $U(c,l)$ satisfies (32). Then*

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

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

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

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 (32), 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 $\frac{1-\tilde{F}_t(\theta|\theta_-)}{\theta\tilde{f}_t(\theta|\theta_-)}$ thicker and increases the optimal labor wedges. When $U_{cl} > 0$, there is a force that makes the tail ratio thinner and decreases 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

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

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 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.

7 Appendix

7.1 Proof of Lemma 1

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

$$\begin{aligned} H &= -\left(c\left(\theta\right)-y\left(\theta\right)+\delta V_{t+1}\left(\beta^{-1}\left(u(\theta)-U(c(\theta),y(\theta)/\theta)\right),w_2(\theta),\theta\right)\right)f_t\left(\theta|\theta_-\right) \\ &\quad +\mu(\theta)\left[U_l(c(\theta),y(\theta)/\theta)\left(-\frac{y\left(\theta\right)}{\theta^2}\right)+\beta w_2\left(\theta\right)\right] \\ &\quad -pu\left(\theta\right)f\left(\theta|\theta_-\right)-p_2u\left(\theta\right)f_2(\theta|\theta_-) \\ &= -\left(c\left(\theta\right)-y\left(\theta\right)+\delta V_{t+1}\left(\beta^{-1}\left(u(\theta)-U(c(\theta),y(\theta)/\theta)\right),w_2(\theta),\theta\right)\right)f_t\left(\theta|\theta_-\right) \\ &\quad +\mu(\theta)\left[U_l(c(\theta),y(\theta)/\theta)\left(-\frac{y\left(\theta\right)}{\theta^2}\right)+\beta w_2\left(\theta\right)\right] \\ &\quad +\left(-\frac{p}{p_2}-\frac{f_2(\theta|\theta_-)}{f\left(\theta|\theta_-\right)}\right)p_2u\left(\theta\right)f\left(\theta|\theta_-\right) \end{aligned}$$

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).

7.2 Proof of Proposition 1

Form the Hamiltonian to (22):

$$\begin{aligned} \mathcal{H} &= -\left(\frac{1}{\gamma}l\left(\theta\right)^{\gamma}+\bar{U}^{-1}\left(\left[u\left(\theta\right)-\beta w\left(\theta\right)\right]\right)-\theta l\left(\theta\right)+\delta V_{t+1}\left(w\left(\theta\right),w_2\left(\theta\right),\theta\right)\right)f\left(\theta|\theta_-\right) \\ &\quad +p_2u\left(\theta\right)f\left(\theta|\theta_-\right)\left(p-\frac{f_2(\theta|\theta_-)}{f\left(\theta|\theta_-\right)}\right)+\mu(\theta)\left(m\left(u\left(\theta\right)-\beta w\left(\theta\right)\right)\frac{l\left(\theta\right)^{\gamma}}{\theta}+\beta w_2\left(\theta\right)\right) \end{aligned}$$

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

First we characterize the labor wedge (27). Define $r(\theta) = c(\theta) - \frac{1}{\gamma}l(\theta)^{\gamma}$. The first order condition with respect to $u\left(\theta\right)$ is

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

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

$\mu(\theta) =$

$$\int_{\theta}^{\infty} -\frac{1}{\bar{U}_c(r(x))} \exp\left(\int_{\theta}^{x} m'\left(u\left(\tilde{\theta}\right) - \beta w\left(\tilde{\theta}\right)\right) \frac{l\left(\tilde{\theta}\right)^{\gamma}}{\tilde{\theta}} d\tilde{\theta}\right)\left(1 - p_2\left(p - \frac{f_2(\theta|\theta_-)}{f(\theta|\theta_-)}\right)\bar{U}_c(r(x))\right) dF(x|\theta_-) . \tag{33}$$

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

$$\left[l(\theta)^{\gamma-1} - \theta\right] f(\theta|\theta_-) = \gamma\mu(\theta) m\left(u(\theta) - \beta w(\theta)\right) \frac{l(\theta)^{\gamma-1}}{\theta}. \tag{34}$$

>From the definition of labor distortion (16),

$$T'_D(\theta) = 1 - \frac{l(\theta)^{\gamma-1}}{\theta}, \tag{35}$$

which implies together with (34) that

$$\begin{aligned} \frac{T'_D(\theta)}{1 - T'_D(\theta)} &= -\frac{\gamma\mu(\theta) m\left(u(\theta) - \beta w(\theta)\right)}{\theta f(\theta|\theta_-)} \\ &= -\mu(\theta)\bar{U}_c(r(\theta))\frac{\gamma}{\theta f(\theta|\theta_-)}. \end{aligned} \tag{36}$$

We now proceed to characterize $\mu(\theta)\bar{U}_c(r(\theta))$ by substituting the expression for $\mu(\theta)$ from (33) and using the fact that $m'(\theta) = \frac{\bar{U}_{cc}(r(\theta))}{\bar{U}_c(r(\theta))}$.

$$\mu(\theta)\bar{U}_c(r(\theta)) = \int_{\theta}^{\infty} -\frac{\bar{U}_c(r(\theta))}{\bar{U}_c(r(x))} \exp\left(\int_{\theta}^{x} \frac{\bar{U}_{cc}(r(\tilde{\theta}))}{\bar{U}_c(r(\tilde{\theta}))} \frac{l\left(\tilde{\theta}\right)^{\gamma}}{\tilde{\theta}} d\tilde{\theta}\right)\left(1 - p_2\left(p - \frac{f_2(x|\theta_-)}{f(x|\theta_-)}\right)\bar{U}_c(r(x))\right) f(x|\theta_-)\, dx. \tag{37}$$

Observe that

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

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

$$\frac{\bar{U}_c(r(\theta))}{\bar{U}_c(r(x))} = \exp\left(-\int_{\theta}^{x} \frac{\bar{U}_{cc}(r(\tilde{\theta}))}{\bar{U}_c(r(\tilde{\theta}))} \frac{u'(\tilde{\theta}) - \beta w'(\tilde{\theta})}{\bar{U}_c(r(\tilde{\theta}))} d\tilde{\theta}\right). \tag{38}$$

Substitute expression for $u'(\tilde{\theta})$ from (20) to (38) and observe that since $w(\theta) = \omega(\theta|\theta)$ and $w_2(\theta) = \omega_2(\theta|\theta)$, then

$$\begin{aligned} w'(\theta) &= \omega_1(\theta|\theta) + \omega_2(\theta|\theta) \\ &= \omega_1(\theta|\theta) + w_2(\theta). \end{aligned} \tag{39}$$

Then expression (38) 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 (38) into (37)

$$\mu(\theta)\bar{U}_c(r(\theta)) = \int_\theta^\infty -\exp\left(\int_\theta^x \frac{\bar{U}_{cc}(r(\tilde{\theta}))}{\bar{U}_c(r(\tilde{\theta}))}\frac{\beta\omega_1\left(\tilde{\theta}|\tilde{\theta}\right)}{\bar{U}_c(r(\tilde{\theta}))}d\tilde{\theta}\right)\left(1 - p_2\left(p - \frac{f_2(x|\theta_-)}{f\left(x|\theta_-\right)}\right)\bar{U}_c(r(x))\right)f\left(x|\theta_-\right)dx. \tag{40}$$

Substitute (40) into (36)

$$\begin{aligned} \frac{T_D'\left(\theta\right)}{1 - T_D'\left(\theta\right)} &= \frac{\gamma}{\theta f\left(\theta|\theta_-\right)} \\ &\quad\times \int_\theta^\infty \exp\left(\int_\theta^x \frac{\bar{U}_{cc}(r(\tilde{\theta}))}{\bar{U}_c(r(\tilde{\theta}))}\frac{\beta\omega_1\left(\tilde{\theta}|\tilde{\theta}\right)}{\bar{U}_c(r(\tilde{\theta}))}d\tilde{\theta}\right)\left(1 - p_2\left(p - \frac{f_2(x|\theta_-)}{f\left(x|\theta_-\right)}\right)\bar{U}_c(r(x))\right)f\left(x|\theta_-\right)dx \\ &= \frac{\gamma}{\theta f\left(\theta|\theta_-\right)\exp\left(\int_0^\theta \frac{\bar{U}_{cc}(r(\tilde{\theta}))}{\bar{U}_c(r(\tilde{\theta}))}\frac{\beta\omega_1\left(\tilde{\theta}|\tilde{\theta}\right)}{\bar{U}_c(r(\tilde{\theta}))}d\tilde{\theta}\right)} \\ &\quad\times \int_\theta^\infty \exp\left(\int_0^x \frac{\bar{U}_{cc}(r(\tilde{\theta}))}{\bar{U}_c(r(\tilde{\theta}))}\frac{\beta\omega_1\left(\tilde{\theta}|\tilde{\theta}\right)}{\bar{U}_c(r(\tilde{\theta}))}d\tilde{\theta}\right)\left(1 - p_2\left(p - \frac{f_2(x|\theta_-)}{f\left(x|\theta_-\right)}\right)\bar{U}_c(r(x))\right)f\left(x|\theta_-\right)dx \end{aligned} \tag{41}$$

Define

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

and

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

Then (41) can be re-written as

$$\frac{T_D'\left(\theta\right)}{1 - T_D'\left(\theta\right)} = \gamma\frac{1 - \tilde{F}(\theta|\theta_-)}{\theta\tilde{f}\left(\theta|\theta_-\right)}\int_\theta^\infty\left(1 - p_2\left(p - \frac{f_2(x|\theta_-)}{f\left(x|\theta_-\right)}\right)\bar{U}_c\left(x\right)\right)\frac{\tilde{f}\left(x|\theta_-\right)dx}{1 - \tilde{F}(\theta|\theta_-)} \tag{42}$$

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

$$0 = \int_0^\infty \left(1 - p_2 \left(p - \frac{f_2(x|\theta_-)}{f(x|\theta_-)}\right) \bar{U}_c(x)\right) \tilde{f}(x|\theta_-)\, dx. \tag{43}$$

Let $\lambda \equiv 1/p_2$. Then (43) implies that

$$\lambda = \frac{1}{p_2} = \int_0^\infty \left(p - \frac{f_2(x|\theta_-)}{f(x|\theta_-)}\right) \bar{U}_c(x)\, \tilde{f}(x|\theta_-)\, dx. \tag{44}$$

Substitute (44) into (42) to obtain (27).

We proceed to derive the marginal capital wedge (28). Take the first order condition of the Hamiltonian with respect to $w(\theta)$

$$\left(\frac{\beta}{\bar{U}_c(r_t(\theta))} - \delta \frac{\partial V_{t+1}(w(\theta), w_2(\theta), \theta)}{\partial w(\theta)}\right) f(\theta|\theta_-) = \beta \mu_t(\theta) \frac{\bar{U}_{cc}(r_t(\theta))}{\bar{U}_c(r_t(\theta))} \frac{l_t(\theta)^\gamma}{\theta}. \tag{45}$$

Rearrange (45)

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

Now use the expression (35) and (36)

$$-\frac{\delta}{\beta} \frac{\bar{U}_c(r_t(\theta))\, \partial V_{t+1}(w(\theta), w_2(\theta), \theta)}{\partial w(\theta)} = -1 + \psi(\theta) \frac{1}{\gamma} T'_{D,t}(\theta)\, y_t(\theta). \tag{46}$$

>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 (43) for period $t+1$

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

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 $w_2(\theta)$

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

7.3 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 increasing 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 (46). The left hand side of that expression is negative, since $\partial V_{t+1}(w_{t+1}, w_{2,t+1}, \theta)/\partial w_{t+1}$ is positive. 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 $\theta \geq \bar{\theta}$

$$\frac{1}{\gamma}\left(\frac{\bar{y}}{\theta}\right)^{\gamma} \leq \varepsilon \tag{47}$$

and

$$\bar{y} + \varepsilon \leq \frac{\gamma - 1}{\gamma}\theta^{\gamma/(\gamma-1)} \tag{48}$$

Such $\bar{\theta}$ exists because the left hand side of (48) 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

$$W_{\bar{\theta}}^{fb} = \max_{c,w} \int_{\bar{\theta}}^{\infty} \left(U(c(\theta), y^*(\theta)) + \beta w(\theta)\right) \frac{dF_t(\theta)}{1 - F_t(\bar{\theta})} \tag{49}$$

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

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

When 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

$$W_{\bar{\theta}}^{fb} \geq \int_{\bar{\theta}}^{\infty} u^*(\theta) \frac{dF_t(\theta)}{1 - F_t(\bar{\theta})}. \tag{50}$$

Since $u^*(\theta)$ is increasing in θ because of incentive compatibility, (50) implies

$$W_{\bar{\theta}}^{fb} \geq u^*(\bar{\theta}). \tag{51}$$

The first order conditions to (49) also imply that

$$c^{fb}(\theta) - \frac{1}{\gamma}\left(\frac{y^*(\theta)}{\theta}\right)^{\gamma} = c^{fb}(\theta') - \frac{1}{\gamma}\left(\frac{y^*(\theta')}{\theta'}\right)^{\gamma} \tag{52}$$

for all $\theta, \theta' \geq \bar{\theta}$. Therefore from (47) for all $\theta, \theta' \geq \bar{\theta}$

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

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

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

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 (48).

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

$$U\left(c^{aut}(\bar{\theta}), y^{aut}(\bar{\theta})/\bar{\theta}\right) + \beta w^{aut}(\bar{\theta}) > W_{\bar{\theta}}^{fb} \geq u^*(\bar{\theta}) \tag{53}$$

By construction

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

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

$$U\left(c^{aut}(\theta), \frac{y^{aut}(\theta)}{\theta}\right) + \beta w^{aut}(\theta) \geq U\left(c^{aut}(\theta'), \frac{y^{aut}(\theta')}{\theta}\right) + \beta w^{aut}(\theta') \text{ for all } \theta \geq \theta' \geq \bar{\theta} \tag{54}$$

and

$$U\left(c^{aut}(\theta), \frac{y^{aut}(\theta)}{\theta}\right) + \beta w^{aut}(\theta) \geq U\left(c^{*}(\theta'), \frac{y^{*}(\theta')}{\theta}\right) + \beta w^{*}(\theta') \text{ for all } \theta \geq \bar{\theta}, \theta' < \bar{\theta} \tag{55}$$

Equation (54) follows from construction of allocation $\left(c^{aut}, y^{aut}, w^{aut}\right)$.

Now consider equation (55). 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

$$U\left(c^{aut}(\bar{\theta}), \frac{y^{aut}(\bar{\theta})}{\bar{\theta}}\right) + \beta w^{aut}(\bar{\theta}) \geq U\left(c^{*}(\theta'), \frac{y^{*}(\theta')}{\bar{\theta}}\right) + \beta w^{*}(\theta') \text{ for } \theta' < \bar{\theta} \tag{56}$$

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 (56) is true because (53) 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 (55) 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.

7.4 Proof of Proposition 3

We consider Hamiltonian to (10) where we substitute $y = \theta l$. Using Lemma 1, the Hamiltonian can be written as

$$H = -\left(\Upsilon\left(u - \beta w, l\right) - \theta l + \delta V_{t+1}(w, w_2, \theta)\right) f_t(\theta|\theta_{-}) + p\left[m\left(u - \beta w, l\right)\frac{l\left(\theta\right)}{\theta} + \beta w_2\right] + q\alpha_t(\theta)u(\theta)f(\theta|\theta_{-})$$

where we define a function

$$c = \Upsilon\left(u - \beta w, l\right)$$

implicitly from the equation $u - \beta w = U(c, l)$ and function m by

$$m\left(u - \beta w, l\right) = -U_l\left(\Upsilon\left(u - \beta w, l\right), l\right).$$

Note the relationships between derivatives of Υ and m and partial derivatives U_c and U_l :

$$\begin{aligned} U_c\Upsilon_1 &= 1 \\ U_c\Upsilon_2 + U_l &= 0 \end{aligned}$$

and

$$\begin{aligned} m_1 &= -U_{cl}\Upsilon_1 \\ m_2 &= -U_{cl}\Upsilon_2 - U_{ll} \end{aligned}$$

The optimality condition for u is

$$-\Upsilon_1 f(\theta|\theta_{-}) + q\alpha_t(\theta)f_t(\theta|\theta_{-}) + pm_1\frac{l\left(\theta\right)}{\theta} = -\dot{p},$$

which after integration implies

$$p = -\int_{\theta}^{\infty}\frac{1}{U_c(x)}\exp\left(\int_{\theta}^{x} m_1\left(\tilde{\theta}\right)\frac{l\left(\tilde{\theta}\right)}{\tilde{\theta}}d\tilde{\theta}\right)\left(1 - q\alpha_t(x)U_c(x)\right)f_t(x|\theta_{-})dx. \tag{57}$$

The optimality condition for l is

$$(1-\Upsilon_2/\theta)\,\theta f_t(\theta|\theta_-)+p\left(\frac{1}{\theta}m+\frac{l}{\theta}m_2\right)=0. \tag{58}$$

Using (16) we can show that

$$(1-T')=-\frac{U_l}{U_c}\frac{1}{\theta}=\Upsilon_2\frac{1}{\theta}$$

which after substituting into (58) implies that

$$T'=-\frac{p}{\theta f_t(\theta|\theta_-)}\frac{1-T'}{\Upsilon_2}\left(m+m_2 l\right).$$

Substitute expression for p from (57)

$$\begin{aligned}
\frac{T'}{1-T'} &= \frac{1-F_t(\theta|\theta_-)}{\theta f_t(\theta|\theta_-)}\int_\theta^\infty \frac{(m(\theta)+m_2(\theta)l(\theta))}{\Upsilon_2(\theta)U_c(x)}\exp\left(\int_\theta^x m_1\left(\tilde\theta\right)\frac{l\left(\tilde\theta\right)}{\tilde\theta}d\tilde\theta\right)\left(1-q\alpha_t(x)U_c(x)\right)\frac{f_t(x|\theta_-)dx}{1-F_t(\theta|\theta_-)}\\
&= \frac{1-F_t(\theta|\theta_-)}{\theta f_t(\theta|\theta_-)}\int_\theta^\infty\left[\begin{array}{c}\frac{(m(\theta)+m_2(\theta)l(\theta))}{\Upsilon_2(\theta)U_c(x)}\times\\ \times\exp\left(\int_\theta^x m_1\left(\tilde\theta\right)\left(\frac{u'(\tilde\theta)-\beta w_2(\tilde\theta)}{m(\tilde\theta)}\right)d\tilde\theta\right)\\ \times\left(1-q\alpha_t(x)U_c(x)\right)\frac{f_t(x|\theta_-)}{1-F_t(\theta|\theta_-)}\end{array}\right]dx\\
&= \frac{(m(\theta)+m_2(\theta)l(\theta))}{\Upsilon_2(\theta)U_c(\theta)}\frac{1-F_t(\theta|\theta_-)}{\theta f_t(\theta|\theta_-)}\times\\
&\quad\times\int_\theta^\infty\left[\begin{array}{c}\frac{U_c(\theta)}{U_c(x)}\exp\left(\int_\theta^x\frac{m_1(\tilde\theta)}{m(\tilde\theta)}\left(U'(\tilde\theta)+\beta\omega_1\left(\tilde\theta|\tilde\theta\right)\right)d\tilde\theta\right)\times\\ \times\left(1-q\alpha_t(x)U_c(x)\right)\frac{f_t(x|\theta_-)}{1-F_t(\theta|\theta_-)}\end{array}\right]dx
\end{aligned}$$

Observe that

$$\frac{m_1}{m}=\frac{U_{cl}\Upsilon_1}{U_l}=\frac{U_{cl}}{U_cU_l}$$

Finally, consider the first term:

$$\begin{aligned}
&\frac{(m(\theta)+m_2(\theta)l(\theta))}{\Upsilon_2(\theta)U_c(\theta)}\\
=&\ \frac{\left(-U_l+\left(-U_{cl}\left(-U_l/U_c\right)-U_{ll}\right)l\right)}{\left(-U_l/U_c\right)U_c}\\
=&\ \frac{\left(-U_l+\left(U_{cl}U_l/U_c-U_{ll}\right)l\right)}{-U_l}\\
=&\ 1+\frac{\left(U_{ll}-U_{cl}U_l/U_c\right)l}{U_l}
\end{aligned}$$

Frisch elasticity of labor supply is defined as

$$\eta^{Fr}=\frac{U_l}{\left(U_{ll}-U_{cl}^2/U_{cc}\right)l}. \tag{59}$$

Substitute (32) into (59) to show the result of Proposition 3.

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.

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.

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

Jacquet, L., E. Lehmann, and B. Van Der Linden (2010): "Optimal Redistributive Taxation with Both Extensive and Intensive Responses," *working paper*.

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

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

Kapicka, M. (2010): "Efficient allocations in dynamic private information economies with persistent shocks: A first order approach," *working paper*.

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

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," *working paper*.

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

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

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