Redistribution and Social Insurance

Mikhail Golosov   Maxim Troshkin   Aleh Tsyvinski*

Princeton   Cornell   Yale

July 2015

Abstract

We study optimal redistribution and insurance in a lifecycle economy with privately observed idiosyncratic shocks. We characterize Pareto optima, show the forces that determine the optimal labor distortions, and derive closed form expressions for their limiting behavior. The labor distortions for high-productivity shocks are determined by the labor elasticity and the higher moments of the shock process; the labor distortions for low-productivity shocks are determined by the autocorrelation of the shock process, redistributive objectives, and past distortions. We calibrate our model using newly available estimates of idiosyncratic shocks. The optimal labor distortions are U-shaped and the optimal savings distortions are generally increasing in current earnings. The constrained optimum has 2 to 4 percent higher welfare than equilibria with affine taxes.

We study a lifecycle economy with individuals who are ex ante heterogeneous in their abilities and experience idiosyncratic shocks to their skills over time. We derive a novel decomposition that allows us to isolate key economic forces determining the optimal labor distortions in lifecycle economies with unobservable idiosyncratic shocks and to provide their characterization. We also compute the optimal labor and savings distortions in a model calibrated to match moments of the labor earnings process from a newly available high-quality U.S. administrative data. The data allow us to estimate the higher moments of the stochastic process for skills, such as kurtosis, which emerge from our analysis as key parameters determining the properties of the optimum.

Most of our analysis focuses on characterizing the properties of the optimal labor distortions, or wedges, between marginal utilities of consumption and leisure. We show that the labor distortion in a given period is driven by two components: an *intra*temporal component that provides insurance against new shocks in that period, and an *inter*temporal component that relaxes incentive constraints and reduces the costs of insurance provision against shocks in the previous periods. The intratemporal component depends on the elasticity of labor supply, the hazard rate of the current period shock conditional on past information, and the welfare gain from providing insurance against that shock. The intertemporal component depends on past distortions, a specific form of a likelihood ratio of the shock realization, and the marginal utility of consumption.

We characterize the behavior of each component in the tails, for high and

low realizations of idiosyncratic shocks in the current period. Our benchmark specification focuses on separable preferences and shocks drawn from a commonly used family of stochastic processes that include lognormal, mixtures of lognormals, and Pareto-lognormal distributions. We show that for such specifications the distortions in the right tail are determined by the intratemporal component and derive a simple formula for their asymptotic behavior. This behavior depends on the elasticity of labor supply and the tail hazard rate of shocks and is independent of age, past history, or Pareto weights of the planner. The distortions in the left tail depend asymptotically only on the intertemporal component and are given by a formula that consists of the autocorrelation of the shock process, past labor distortions, and consumption growth rates. They depend on past history and Pareto weights and generally increase with age. We also explain how the degree of the progressivity of the labor distortions depends on the higher moments of the shock distribution, such as kurtosis, and extend our results to non-separable preferences.

We then use newly available high-quality administrative data on labor earnings (see Guvenen, Ozkan and Song (2013) and Guvenen et al. (2013)) and the U.S. tax code to estimate the stochastic process for skills and quantify the implications for the optimal distortions. Similar to the earnings, the process for the shocks is highly persistent and leptokurtic. The optimal labor distortions are approximately U-shaped as a function of current labor earnings, with the dip in the distortions around the level of earnings in the previous period. The optimal savings distortions generally increase in labor earnings. The distortions are fairly large in magnitude, especially in the right tail: the labor

distortions approach 75 percent, while savings distortions approach 2 percent of gross (i.e., interest plus principal) return to savings. We provide a detailed quantitative decomposition of the labor distortions into the intertemporal and intratemporal components. Finally, we show that the welfare losses from using affine policies instead of the optimal policy are around 2 to 4 percent of consumption. Moreover, the optimal labor distortions differ significantly from those in a model with the lognormal shocks, both qualitatively and quantitatively, and imply higher welfare gains from non-linear, history-dependent policies. The key feature of the data that drives these differences is the high kurtosis emphasized by Guvenen et al. (2013).

More broadly, we view the contribution of our paper as a step for the dynamic optimal taxation literature, using the mechanism-design approach, to connect more closely to applied work that studies design of social insurance programs. Eligibility rules for welfare programs, rates of phase out of transfers, the degree of progressivity of the statutory tax rates all introduce effective labor and savings distortions. The mechanism design approach provides an upper bound on welfare that can be achieved with such programs. We characterize labor and savings distortions in a model with rich and realistic processes for idiosyncratic shocks that are emphasized in the empirical labor literature. These insights can be used as guidance in designing specific insurance programs in applied settings to maximize welfare gains.

A number of papers are related to our work. Our theoretical and quantitative analyses are built on the recursive approach developed in Kapička (2013) and Pavan, Segal and Toikka (2014). Golosov, Kocherlakota and Tsyvinski

3

(2003), Kocherlakota (2005), Golosov and Tsyvinski (2006), Werning (2009) are some of the examples of the theoretical work examining different properties of the optimal distortions and their relationships to taxes. Our quantitative analysis is also related to a number of studies. Albanesi and Sleet (2006) provide a comprehensive numerical and theoretical study of optimal capital and labor taxes in a dynamic economy with i.i.d. shocks. Golosov, Tsyvinski and Werning (2006) is a two-period numerical study of the determinants of the dynamic optimal taxation in the spirit of Tuomala (1990). Ales and Maziero (2007) numerically solve a version of a lifecycle economy with i.i.d. shocks drawn from a discrete, two-type distribution, and find that the labor distortions are lower earlier in life. Weinzierl (2011) and Fukushima (2010) numerically solve the optimal labor and savings distortions in dynamic economies. Conesa, Kitao and Krueger (2009), Heathcote, Storesletten and Violante (2014), and Kindermann and Krueger (2014) characterize optimal policies using rich but restricted tax instruments.

An important contribution of Farhi and Werning (2013) characterizes the dynamics of labor distortions in lifecycle settings similar to ours. Most of their analysis focuses on time-series properties of labor distortions and shows that the stochastic process for labor distortions has autocorrelation equal to that of the shock process and a positive trend. In a numerical exercise they use lognormal shocks and show that affine taxes capture most of the welfare gains from the optimal policies. In contrast, our analysis focuses on how the labor distortions depend on earnings realization, determining the degree of optimal progressivity of the distortions in different parts of the earnings distribution.

4

Our decomposition shows the main economic trade-offs and highlights how the hazard of the shock process plays important qualitative and quantitative roles in the shape of the distortions. The main insights - the expressions for the asymptotic behavior of distortions, the observation that redistributive object-ives and past history affect distortions only in the left tail, and the analysis of the effects of higher moments of shocks on the labor distortions - are all new. Our analysis is also the first attempt, to the best of our knowledge, to estimate the effects of higher moments using available data on earnings and the tax code. The main insights - the U-shaped labor distortions, their magnitudes, and large welfare gains from the optimal non-linear, history de-pendent policies - differ substantially from the results that can be obtained with lognormal shocks.

The rest of the paper is organized as follows. Section 1 describes the envir-onment. Section 2 provides the theoretical analysis. Section 3 quantitatively analyzes the calibrated life-cycle model. Section 4 concludes.

1   Environment

We consider an economy that lasts $T + 1$ periods, denoted by $t = 0, ..., T$. Each agent's preferences are described by a time separable utility function over consumption c_t and labor l_t,

$$\mathbb{E}_0 \sum_{t=0}^{T} \beta^t U(c_t, l_t), \tag{1}$$

where $\beta \in (0,1)$ is a discount factor, $\mathbb{E}_0$ is a period 0 expectation operator, and $U : \mathbb{R}_+^2 \to \mathbb{R}$.

In period $t = 0$, agents draw their initial type (skill), θ_0, from a distribution $F_0(\theta)$. For $t \geq 1$, 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 $f_t(\theta|\theta_{t-1})$. For parts of the analysis it will be convenient to assume that people retire at some period $\hat{T}$, in which case $F_t(0|\theta) = 1$ for all θ and all $t \geq \hat{T}$. Skills are non-negative: $\theta_t \in \Theta = \mathbb{R}_+$ for all t. The set of possible histories up to period t is denoted by Θ^t.

Assumption 1. *For all $t < \hat{T}$, density f_t is differentiable in both arguments with $f_t' \equiv \frac{\partial f_t}{\partial \theta}$ and $f_{2,t} \equiv \frac{\partial f_t}{\partial \theta_{t-1}}$. For all θ_{t-1}, $\varphi_t(\theta|\theta_{t-1}) \equiv \frac{\theta_{t-1} \int_\theta^\infty f_{2,t}(x|\theta_{t-1})dx}{\theta f_t(\theta|\theta_{t-1})}$ is bounded for all θ and $\lim_{\theta \to \infty} \frac{1-F_t(\theta|\theta_{t-1})}{\theta f_t(\theta|\theta_{t-1})}$ is finite.*

The function φ_t defined in this assumption is bounded for many commonly used stochastic processes; for AR(1) lognormal shocks it is equal to the auto-correlation of the shock process for all θ.

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 are privately observed by the agent. Output y_t and consumption c_t are publicly observed. In period t, the agent knows his skill realization only for the first t periods $\theta^t = (\theta_0, ..., \theta_t)$. Denote by $c_t(\theta^t)$: $\Theta^t \to \mathbb{R}_+$ the agent's allocation of consumption and by $y_t(\theta^t)$: $\Theta^t \to \mathbb{R}_+$ the agent's allocation of output in period t. Denote by $\sigma^t(\theta^t)$: $\Theta^t \to \Theta^t$ the agent's report in period t. Let Σ^t be the set of all such reporting strategies in period t. Resources can be transferred between periods at a rate $R > 0$. The observability of consumption implies that all savings are publicly observable.

The social planner evaluates welfare using Pareto weights $\alpha : \Theta \to \mathbb{R}_+$, where $\alpha(\theta)$ is a weight assigned to an agent born in period 0 with type θ. We assume that α is non-negative and normalize $\int_0^\infty \alpha(\theta) \, dF_0(\theta) = 1$. Social welfare is given by $\int_0^\infty \alpha(\theta) \left(\mathbb{E}_0 \sum_{t=0}^T \beta^t U(c_t, l_t) \right) dF_0(\theta)$.

We denote partial derivatives of U with respect to c and l as U_c and U_l and define all second derivatives and cross-partials accordingly. Similarly, U_y and U_θ denote derivatives of $U\left(c, \frac{y}{\theta}\right)$ with respect to y and θ. We make the following assumptions about U.

Assumption 2. *U is twice continuously differentiable in both arguments, satisfies $U_c > 0, U_l < 0, U_{cc} \leq 0, U_{ll} \leq 0$, and $\frac{\partial}{\partial \theta} \frac{U_y(c,y/\theta)}{U_c(c,y/\theta)} \geq 0$.*

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

$$\max_{\left\{ c_t\left(\theta^t\right), y_t\left(\theta^t\right) \right\}_{\theta^t \in \Theta^t; t=0,..,T}} \int_0^\infty \alpha(\theta) \left(\mathbb{E}_0 \left\{ \left. \sum_{t=0}^T \beta^t U\left(c_t\left(\theta^t\right), y_t\left(\theta^t\right)/\theta_t\right) \right| \theta \right\} \right) dF_0(\theta)$$

$$(2)$$

subject to the incentive compatibility constraint:

$$\mathbb{E}_0 \left\{ \left. \sum_{t=0}^T \beta^t U\left(c_t\left(\theta^t\right), y_t\left(\theta^t\right)/\theta_t\right) \right| \theta \right\}$$

$$\geq \mathbb{E}_0 \left\{ \left. \sum_{t=0}^T \beta^t U\left(c_t\left(\sigma^t\left(\theta^t\right)\right), y_t\left(\sigma^t\left(\theta^t\right)\right)/\theta_t\right) \right| \theta \right\}, \forall \sigma^T \in \Sigma^T, \sigma^t \in \sigma^T, \theta \in \Theta$$

$$(3)$$

and the feasibility constraint:

$$\int_0^\infty \mathbb{E}_0 \left\{ \left. \sum_{t=0}^T R^{-t} c_t \left(\theta^t \right) \right| \theta \right\} dF_0(\theta) \leq \int_0^\infty \mathbb{E}_0 \left\{ \left. \sum_{t=0}^T R^{-t} y_t \left(\theta^t \right) \right| \theta \right\} dF_0(\theta). \tag{4}$$

We follow Fernandes and Phelan (2000) and Kapička (2013) to write the problem recursively. Here we sketch the main steps and refer to the two papers for technical details. Constraint (3) can be written recursively as

$$U \left(c_t \left(\theta^t \right), y_t \left(\theta^t \right) / \theta_t \right) + \beta \omega_{t+1} \left(\theta^t | \theta_t \right)$$
$$\geq U \left(c_t \left(\theta^{t-1}, \hat{\theta} \right), y_t \left(\theta^{t-1}, \hat{\theta} \right) / \theta_t \right) + \beta \omega_{t+1} \left(\theta^{t-1}, \hat{\theta} | \theta_t \right), \ \forall \hat{\theta}, \theta \in \Theta, \forall t \tag{5}$$

and

$$\omega_{t+1} \left(\theta^{t-1}, \hat{\theta} | \theta_t \right) = \mathbb{E}_t \left\{ \left. \sum_{s=t+1}^T \beta^{s-t-1} U \left(c_s \left(\hat{\theta}^s \right), y_s \left(\hat{\theta}^s \right) / \theta_s \right) \right| \theta_t \right\},$$

where $\hat{\theta}^s = \left(\theta_0 ..., \theta_{t-1}, \hat{\theta}, \theta_{t+1}, ..., \theta_s \right)$ are all the histories in which the agent misreports his type once in the history θ^s. It is possible to write the problem recursively using $\omega \left(\hat{\theta} | \theta \right)$ as a state variable following the methods developed by Fernandes and Phelan (2000). The problem, however, is intractable since $\omega \left(\hat{\theta} | \theta \right)$ is a function of $\left(\hat{\theta}, \theta \right)$ and thus the state space becomes infinite dimensional. Kapička (2013) and Pavan, Segal and Toikka (2014) further simplify the problem by replacing global incentive constraints (5) with their local analogue, the first-order conditions, to obtain a more manageable recursive formulation. When non-local constraints do not bind one needs to keep track of only on-the-path promised utility $w(\theta) = \omega(\theta | \theta)$ and the utility from a local

8

deviation $w_2(\theta) = \omega_2(\theta|\theta)$, where $\omega_2(\theta|\theta)$ is the derivative of ω with respect to its second argument evaluated at $(\theta|\theta)$. The maximization problem (2) can then be written recursively for $t \geq 1$ as

$$V_t(\hat{w}, \hat{w}_2, \theta_-) = \min_{c,y,u,w,w_2} \int_0^\infty \left(c(\theta) - y(\theta) + R^{-1}V_{t+1}\left(w(\theta), w_2(\theta), \theta \right) \right) f_t(\theta|\theta_-) \, d\theta \tag{6}$$

subject to

$$\dot{u}(\theta) = U_\theta(c(\theta), y(\theta)/\theta) + \beta w_2(\theta), \tag{7}$$

$$\hat{w} = \int_0^\infty u(\theta) f_t(\theta|\theta_-) \, d\theta, \tag{8}$$

$$\hat{w}_2 = \int_0^\infty u(\theta) f_{2,t}(\theta|\theta_-) \, d\theta, \tag{9}$$

$$u(\theta) = U(c(\theta), y(\theta)/\theta) + \beta w(\theta). \tag{10}$$

The value function V_{T+1} as well as w and w_2 disappear from this formulation in the last period.[1] The value function V_0 in period $t = 0$ takes the form

$$V_0(\hat{w}_0) = \min_{c,y,u,w,w_2} \int_0^\infty \left(c(\theta) - y(\theta) + R^{-1}V_1\left(w(\theta), w_2(\theta), \theta \right) \right) f_0(\theta) \, d\theta \tag{11}$$

subject to (7), (10) and

$$\hat{w}_0 = \int_0^\infty \alpha(\theta) u(\theta) f_0(\theta) \, d\theta. \tag{12}$$

---

[1] This discussion is given for the case without retirement. If there are retirement periods, the value function $V_{\hat{T}}(\hat{w})$ is equal to the present value of resources needed to provide $\hat{w}$ utils to a retired agent between periods $\hat{T}$ and T. In this case the choice variable w_2 disappears from the recursive formulation in period $\hat{T} - 1$. The rest of the formulation is unchanged.

There are four state variables in this recursive formulation: $\hat{w}$ is the promised utility associated with the promise-keeping constraint (8); $\hat{w}_2$ is the state variable associated with the threat-keeping constraint (9); θ_- is the reported type in period $t-1$; and age t. The initial value $\hat{w}_0$ is the largest solution to the equation $V_0(\hat{w}_0) = 0$.[2]

The first-order approach is valid only if at the optimum the local constraints (7) are sufficient to guarantee that global incentive constraints (5) are satisfied. It is well known that there are no general conditions either in the static mechanism design problem with multiple goods (see, e.g., Mirrlees (1976)) or in dynamic models (see, e.g., Kapička (2013)) which guarantee that only local incentive constraints bind. It is possible, however, to solve the relaxed problem (6) and (11) and verify whether the solution to that problem satisfies global incentive constraints (5). If it does, it is also a solution to the original problem (2).

Assumption 3. *In the optimum $c\left(\cdot\right)$ and $\omega\left(\cdot|\theta\right)$ are piecewise $\mathcal{C}^1$ and increasing for all θ; the derivative of $\omega\left(\hat{\theta}|\theta\right)$ with respect to $\hat{\theta}$ (when exists), $\omega_1\left(\hat{\theta}|\theta\right)$, is increasing in θ for all $\hat{\theta}$; $U_{cl} \geq 0$.*

Lemma 1. *If Assumptions 2 and 3 are satisfied, then (7) implies (5).*

The focus of our analysis is on the qualitative and quantitative characterization of the optimal labor and savings distortions, or wedges. For an agent

---

[2]If we add exogenous government expenditures to our model, then $\hat{w}_0$ should satisfy $V_0(\hat{w}_0) = -G$ where G is the present value of such expenditures.

with the history of shocks θ^t at time t, we define a labor distortion, $\tau_t^y\left(\theta^t\right)$, as

$$1 - \tau_t^y\left(\theta^t\right) \equiv \frac{-U_l\left(c_t\left(\theta^t\right), y_t\left(\theta^t\right)/\theta_t\right)}{\theta_t U_c\left(c_t\left(\theta^t\right), y_t\left(\theta^t\right)/\theta_t\right)} \tag{13}$$

and a savings distortion, $\tau_t^s\left(\theta^t\right)$, as

$$1 - \tau_t^s\left(\theta^t\right) \equiv \left(\frac{1}{\beta R}\right) \frac{U_c\left(c_t\left(\theta^t\right), y_t\left(\theta^t\right)/\theta_t\right)}{\mathbb{E}_t\left\{U_c\left(c_{t+1}\left(\theta^{t+1}\right), y_{t+1}\left(\theta^{t+1}\right)/\theta_{t+1}\right)\right\}}. \tag{14}$$

2 Characterization of distortions

In this section, we characterize the properties of the optimal distortions in the solution to the planning problems (6) and (11). These distortions are generally history dependent. To describe the properties of the solution, we fix any past history θ^{t-1} and characterize the behavior of the optimal distortions as a function of period-t shock θ_t. To simplify notation, we omit explicit dependence on θ^{t-1}. Thus, whenever it does not cause confusion, a notation $z_t\left(\theta\right)$ denotes the value of a random variable z_t at a history $\left(\theta^{t-1}, \theta\right)$ in the solution of the planning problem; z_{t-1} denotes $z_{t-1}\left(\theta^{t-1}\right)$.

2.1 Separable preferences

We start with the analysis of the optimal labor distortions when preferences are separable between consumption and labor. Let

$$\varepsilon_t\left(\theta\right) \equiv \frac{U_{ll,t}\left(\theta\right) l_t\left(\theta\right)}{U_{l,t}\left(\theta\right)}, \quad \sigma_t\left(\theta\right) \equiv -\frac{U_{cc,t}\left(\theta\right) c_t\left(\theta\right)}{U_{c,t}\left(\theta\right)}. \tag{15}$$

$\varepsilon_t(\theta)$ and $\sigma_t(\theta)$ are the inverses of the Frisch elasticity of labor supply and the elasticity of the intertemporal substitution (EIS) respectively. It is more convenient to work with the inverses of the elasticities since it allows us to easily incorporate the limiting cases of infinite elasticities. These elasticities are, in general, endogenous. Isoelastic preferences

$$U(c,l) = \frac{c^{1-\sigma} - 1}{1 - \sigma} - \frac{l^{1+\varepsilon}}{1 + \varepsilon} \tag{16}$$

provide one useful benchmark that keeps both elasticities constant.

The optimal labor distortions are determined by several economic forces that have distinct behavior. To separate these forces, we define

$$
\begin{aligned}
A_t(\theta) &= 1 + \varepsilon_t(\theta), \\
B_t(\theta) &= \frac{1 - F_t(\theta)}{\theta f_t(\theta)}, \\
C_t(\theta) &= \int_\theta^\infty \exp\left(\int_\theta^x \sigma_t(\tilde{x}) \frac{\dot{c}_t(\tilde{x})}{c_t(\tilde{x})} d\tilde{x}\right) (1 - \lambda_{1,t}\bar{\alpha}_t(x) U_{c,t}(x)) \frac{f_t(x)\,dx}{1 - F_t(\theta)}, \\
D_t(\theta) &= \frac{A_t(\theta)}{A_{t-1}} \frac{U_{c,t}(\theta)}{U_{c,t-1}} \varphi_t(\theta) \ \text{ for } t > 0, \ D_0(\theta) = 0,
\end{aligned}
$$

where

$$\lambda_{1,t} = \int_0^\infty \frac{f_t(x)}{U_{c,t}(x)} dx, \quad \bar{\alpha}_t(\theta) = \begin{cases} \alpha(\theta) \ \text{ if } t = 0, \\ 1 \ \text{ if } t > 0. \end{cases}$$

Functions A_t, B_t, C_t, and D_t define the four main forces characterizing the optimal labor distortions. In the online appendix we show that applying

optimal control techniques one can derive the following expression:

$$\underbrace{\frac{\tau_t^y(\theta)}{1 - \tau_t^y(\theta)}}_{} = \underbrace{A_t(\theta)\, B_t(\theta)\, C_t(\theta)}_{\text{intratemporal component}} + \underbrace{\beta R \frac{\tau_{t-1}^y}{1 - \tau_{t-1}^y} D_t(\theta)}_{\text{intertemporal component}}. \qquad (17)$$

Equation (17) shows that the optimal labor distortion is a sum of two components. The first component, $A_t B_t C_t$, takes a form that can be obtained by manipulating the optimality conditions in the static model of Mirrlees (1971). We call it the *intra*temporal component. The second component, to which we refer as the *inter*temporal component, is specific to dynamic models. Before characterizing how functions A_t, B_t, C_t, and D_t depend on the realization of the shock θ_t it is instructive to briefly discuss the economic intuition behind these forces.

The intratemporal component captures the costs and benefits of labor distortions in providing insurance against period-t shocks. These costs and benefits have analogues in static models, such as Diamond (1998) and Saez (2001), although dynamics introduce additional considerations. To see the intuition for these terms, observe that a labor distortion for type θ discourages that type's labor supply. The behavioral response of labor supply is captured by type θ's Frisch elasticity of labor supply, summarized by $A_t(\theta)$. A higher labor distortion for type θ lowers total output in proportion to $\theta f_t(\theta)$ but allows the planner to relax the incentive constraints for all types above θ. This trade-off is summarized by the hazard ratio defined in $B_t(\theta)$. Since the intratemporal term captures distortions arising from insurance against *new* shocks, the term B_t is a hazard of period-t shocks conditional on a given history θ^{t-1}. Finally,

the relaxed incentive constraints allow the planner to extract more resources from individuals with skills above θ and transfer them to all agents. The social value of that transfer depends on the ratio of the Pareto-weighted marginal utility of consumptions of agents with skills above θ, $\int_\theta^\infty \bar{\alpha}_t(x) U_{c,t}(x) f_t(x) \, dx$, to the average marginal utility, summarized by $\lambda_{1,t}$. This trade-off is captured by the term $C_t(\theta)$.[3] The redistributive component C_t has Pareto weights only in period 0 because efficiency requires that the planner maximizes Pareto-weighted *lifetime* utilities of agents. This implies that all future idiosyncratic shocks are weighted with agent's marginal utility of consumption irrespective of the lifetime Pareto weights.

The intertemporal component captures how the planner uses distortions in the current period t to provide incentives for information revelation in earlier periods. The likelihood $\varphi_t(\theta|\theta_{t-1})$ that appears in D_t summarizes the information that period t shock θ carries about θ_{t-1}. To see this effect, note that $\int_\theta^\infty f_{2,t}(x|\theta_{t-1}) \, dx$ measures the difference in the probability of receiving any shock greater than θ in period t between an agent with skill slightly above θ_{t-1} and an agent with skill θ_{t-1}. When $\int_\theta^\infty f_{2,t}(x|\theta_{t-1}) \, dx > 0$, a labor distortion in period t in a history (θ^{t-1}, θ) is less likely to affect type θ_{t-1} than a type above. Therefore a positive labor distortion in period t allows to relax the incentive constraint in history θ^{t-1}. The opposite argument holds for $\int_\theta^\infty f_{2,t}(x) \, dx < 0$. The term D_t also depends on $\frac{A_t(\theta)}{A_{t-1}}$ and $\frac{U_{c,t}(\theta)}{U_{c,t-1}}$, which capture the fact that it is cheaper to provide incentives in those states in which the elasticity of labor

---

[3] The extraction of resources from types above θ also has an income effect on labor supply of those types, which is captured by the expression $\exp\left(\int_\theta^x \sigma_t(\tilde{x}) \frac{\dot{c}_t(\tilde{x})}{c_t(\tilde{x})} d\tilde{x}\right)$ in the definition of C_t.

supply is low and the marginal utility of consumption is high.

The sharpest characterization of the optimal labor distortions can be obtained in the tails as θ goes to 0 or to infinity. We focus on the situations in which the solution is well-behaved, as summarized by the following assumption.

Assumption 4. *(a)* $\lim_{c \to 0, \infty} \frac{U_{cc}c}{U_c}$, $\lim_{c \to 0, \infty} \frac{U_{ll}l}{U_l}$ *are finite and non-zero;* $\alpha(\cdot)$ *is bounded with a finite* $\lim_{\theta \to 0, \infty} \alpha(\theta)$.

(b) $c_t(\theta)$, $l_t(\theta)$, $\frac{\dot{c}_t(\theta)/c_t(\theta)}{\dot{y}_t(\theta)/y_t(\theta)}$ *have limits;* $\frac{c_t(\theta)}{y_t(\theta)}$ *has a finite, non-zero limit;* $\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta)}$ *has a finite limit as* $\theta \to \infty$*;* $l_t(\theta)$ *has a limit;* $U_{c,t}(\theta)$ *has a finite limit as* $\theta \to 0$.

The main purpose of this assumption is to rule out two singular cases: that distortions fluctuate periodically in the tails without settling to a limit and that they diverge to $+1$ or $-\infty$. We are not aware of any examples in which distortions do not settle to a limit. The optimal distortions may diverge to 1 in some cases[4] and abstracting from them streamlines our discussion. We discuss relaxing this assumption after presenting our main results. We call U generic if it satisfies Assumption 4(a) and $\lim_{c \to \infty} \frac{-U_{cc}c}{U_c} \neq 1$.

Proposition 1. *Suppose Assumptions 1 and 4 are satisfied and preferences are separable. Then there are* $k_1, k_2, k_3, k_4 \in \mathbb{R}$ *such that*[5]

$$A_t(\theta) B_t(\theta) C_t(\theta) \sim k_3 \frac{1 - F_t(\theta)}{\theta f_t(\theta)}, \ D_t(\theta) = o\left(\frac{1}{\theta^{k_4}}\right) \ (\theta \to \infty),$$

---

[4]For example, Mirrlees (1971) shows that labor distortions can only converge to 1 for a class of preferences that imply that $\varepsilon(\theta) \to \infty$.

[5]For any functions h, g and $c \in \bar{\mathbb{R}}$, $h(x) \sim g(x)$ $(x \to c)$ if $\lim_{x \to c} h(x)/g(x) = 1$, $h(x) = o(g(x))$ $(x \to c)$ if $\lim_{x \to c} h(x)/g(x) = 0$, and $h(x) = O(g(x))$ $(x \to c)$ if there is a constant K such that $|h(x)| \le K|g(x)|$ for all x in a neighborhood of c.

$$A_t\left(\theta\right)B_t\left(\theta\right)C_t\left(\theta\right)\sim k_1\frac{F_t\left(\theta\right)}{\theta f_t\left(\theta\right)},\ \ D_t\left(\theta\right)\sim k_2\varphi_t\left(\theta\right)\ \ \left(\theta\to0\right).$$

$k_3 > 0$ *depends generically only on* U *and* f_t *and* $k_4 > 0$ *depends generically only on* U, k_1 *and* k_2 *generally depend on the past history of shocks.*

This proposition offers two insights about the economic forces that determine the labor distortions in the right and left tails. First, it shows the asymptotic behavior of each component in the tails. As we shall see, these results are very informative about the behavior of the labor distortions and their components. The second insight is that the labor distortions in the right tail depend only on the functional form of U and the tail behavior of the hazard; the history of past shocks, redistribution objectives or any other property of the optimum do not affect those parameters.

To illustrate the intuition for this result, assume that preferences are isoelastic and first consider the distortions in the right tail, as $\theta \to \infty$. We have $l\left(\theta\right)^{\varepsilon} = \left(1 - \tau^y\left(\theta\right)\right)\theta c\left(\theta\right)^{-\sigma}$ and by Assumption 4 $c\left(\theta\right) \propto \left(1 - \tau^y\left(\theta\right)\right)\theta l\left(\theta\right)$ in the limit. Since $1 - \tau^y\left(\theta\right)$ converges to a non-zero limit, $c_t\left(\theta\right) \propto \theta^{\frac{1+\varepsilon}{\sigma+\varepsilon}}$, which implies that the marginal utility of consumption declines at a geometric rate, $U_{c,t}\left(\theta\right) \propto \theta^{\frac{-(1+\varepsilon)\sigma}{\sigma+\varepsilon}}$. This has two implications for the behavior of the labor distortions in the right tail. The first implication is that D_t declines at a geometric rate that does not depend on the past history as $\theta \to \infty$. The second implication is that $\bar{\alpha}_t\lambda_t U_{c,t}$ drops out of the expression for C_t, indicating that asymptotically the planner maximizes the extraction of resources from the right tail of the distribution. The expression for the peak of the Laffer curve for the labor distortion can be obtained in a closed form and it depends only on the hazard rate B_t and the income and substitution effects summarized by

σ and ε. This provides an explanation for the asymptotic equivalence result for the intratemporal term in the right tail.

The asymptotic behavior of the intratemporal component in the left tail is shaped by the tension between the hazard B_t and the redistributive term C_t. The two forces affect the labor distortion in the opposite directions. The hazard B_t favors high labor distortions because low θ types are not very productive and distorting their labor supply has little effect on output. It is easy to see from the definition of B_t that $B_t \sim \frac{1}{\theta f_t} \, (\theta \to 0)$. The redistributive term C_t favors low labor distortions in the left tail because the marginal utility of consumption of those agents is low. We show in the online appendix that $C_t \sim \hat{k} F_t \, (\theta \to 0)$, where $\hat{k} = 1 - \bar{\alpha}_t (0) \lambda_t U_{c,t} (0)$. These two observations imply the asymptotic equivalence result for the intratemporal component in the left tail. The behavior of intertemporal component, particularly of the term D_t, can be seen directly from its definition with $k_2 = \frac{A_t(0)}{A_{t-1}} \frac{U_{c,t}(0)}{U_{c,t-1}}$. The optimal distortions in the left tail are typically history-dependent since $U_{c,t} (0)$, λ_t, and $A_t (0)$ all generally depend on the past realizations of the shock.

Proposition 1 also shows a link between the optimal labor distortions in dynamic lifecycle models and static environments built on Mirrlees (1971). In particular, Diamond (1998) first used the decomposition similar to our intratemporal component to analyze the behavior of optimal distortions in a static model with quasi-linear preferences. Our analysis of the intratemporal component is a generalization of his approach to more general preferences and shock distributions, which also applies to static settings. Since Proposition 1 shows that the dynamic component disappears in the right tail of the distri-

bution, the economic forces that determine the optimal labor distortions for high shocks are similar in static and dynamic settings. We further discuss this connection in specific examples below.

Proposition 1 shows that the hazard rate of productivity shocks plays an important role in shaping the optimal labor distortion. To gain further insight into that behavior we focus on a family of stochastic processes frequently used in applied labor and public finance literatures.[6]

Assumption 5. θ_t *satisfies*

$$\ln \theta_t = b_t + \rho \ln \theta_{t-1} + \epsilon_t,$$

where ϵ_t is drawn from one of the three distributions:

(a) lognormal: $\epsilon_t \sim N\left(0, \upsilon\right);$

(b) Pareto-lognormal: $\epsilon_t \sim NE\left(\mu, \upsilon, a\right)$, where NE is a normal-exponential distribution;

(c) mixture of lognormals: $\epsilon_t \sim N\left(\mu_i, \upsilon_i\right)$ with probability p_i for $i = 1, ..., I$; let $\upsilon = \max_i \upsilon_i$.

The log-normal distribution (a) is a special case of the mixture of lognormals (c). It is useful to keep in mind that if shocks are log-normal then ϵ_t has skewness of 0 and kurtosis of 3 (or excess kurtosis of 0), while the mixture distribution allows to construct ϵ_t with other values of these moments. We can use the tail properties of these distributions (see the online appendix for

---

[6]For example, Storesletten, Telmer and Yaron (2004) and Farhi and Werning (2013) use lognormal distributions, Badel and Huggett (2014) and Lockwood, Nathanson and Weyl (2014) use Pareto-lognormal distributions, Geweke and Keane (2000) and Guvenen et al. (2013) use mixtures of lognormals.

details) to prove the following corollary to Proposition 1.

Corollary 1. *Suppose that Assumptions 4 and 5 are satisfied and preferences are separable. Then there are constants $\bar{\sigma} > 0, \bar{\varepsilon} > 0$ such as*

$$\lim_{\theta \to \infty} C_t(\theta) = 1 + \frac{\bar{\sigma}}{\bar{\sigma} + \bar{\varepsilon}} \lim_{\theta \to \infty} \frac{\tau_t^y(\theta)}{1 - \tau_t^y(\theta)}. \tag{18}$$

Moreover, $\bar{\sigma} = \lim_{c \to \infty} \frac{-U_{cc}c}{U_c}$, $\bar{\varepsilon} = \lim_{l \to \infty} \frac{U_{ll}l}{U_l}$ if $\bar{\sigma} < 1$, $\bar{\varepsilon} = \lim_{l \to 0} \frac{U_{ll}l}{U_l}$ if $\bar{\sigma} > 1$.

Asymptotically as $\theta \to \infty$

$$\frac{\tau_t^y(\theta)}{1 - \tau_t^y(\theta)} \sim A_t(\theta) B_t(\theta) C_t(\theta)$$

$$\sim \begin{cases} \left[a \frac{1}{1+\bar{\varepsilon}} - \frac{\bar{\sigma}}{\bar{\sigma}+\bar{\varepsilon}} \right]^{-1} & \textit{if } f_t \textit{ is Pareto-lognormal, } a \frac{1}{1+\bar{\varepsilon}} - \frac{\bar{\sigma}}{\bar{\sigma}+\bar{\varepsilon}} > 0 \\ \left[\frac{\ln \theta}{v^2} \frac{1}{1+\bar{\varepsilon}} \right]^{-1} & \textit{if } f_t \textit{ is lognormal/mixture.} \end{cases} \tag{19}$$

Asymptotically as $\theta \to 0$, as long as $\rho \tau_{t-1}^y \neq 0$,

$$\frac{\tau_t^y(\theta)}{1 - \tau_t^y(\theta)} \sim \beta R \frac{\tau_{t-1}^y}{1 - \tau_{t-1}^y} D_t(\theta) \sim \beta R \frac{\tau_{t-1}^y}{1 - \tau_{t-1}^y} \rho \frac{U_{c,t}(0)}{U_{c,t-1}} \frac{A_t(0)}{A_{t-1}}. \tag{20}$$

Although the three classes of the distributions of shocks have substantial differences, they share some common implications. All of them imply that the optimal labor distortions are determined by the intratemporal forces in the right tail and by the intertemporal forces in the left tail. The optimal labor distortions in the right tail do not depend on the history of the shocks and are pinned down by the two elasticities defined in Corollary 1 and the tail behavior of the hazard rate: $B_t \sim a^{-1}$ in the Pareto-lognormal case, $B_t \sim \left[\frac{\ln \theta}{v^2} \right]^{-1}$ in the lognormal/mixture case as $\theta \to \infty$. The labor distortions in the left tail

depend on the autocorrelation of the shock process, past labor distortions, and the ratios of the marginal utilities of consumption and the Frisch elasticities of labor supply in periods t and $t-1$.

We next discuss the intuition for Corollary 1 and make some additional observations about its implications. The first result of the corollary, equation (18), characterizes properties of the redistributive component C_t in the right tail. It is a sum of two terms. The number 1 comes from the fact that the marginal utility of the highly skilled converges to zero and the planner would like to extract all the surplus from those agents. The second term on the right-hand side of (18) captures the income effect of the labor supply from the marginal labor distortions on type θ as $\theta \to \infty$.[7] The size of the income effect is proportional to the limiting tax rate.

The second part of Corollary 1 characterizes labor distortions in the right tail. The fact that they are determined by the intratemporal forces follows from Proposition 1 and Assumption 5. We know from our decomposition (17) that the optimal distortions are the sum of the intertemporal and the intratemporal components. The intertemporal component always converges to 0 at a geometric rate by Proposition 1. The intratemporal component, when f_t satisfies Assumption 5, either does not converge to zero at all or converges to zero at a slower rate of $\left[\frac{\ln \theta}{v^2}\right]^{-1}$. Hence, the intratemporal forces eventually dominate the intertemporal forces.

Note that when shocks are drawn from a mixture distribution, v is the highest standard deviation in the mixture. In many applications (see, e.g.

---

[7] In static models the income effect emerges because a higher marginal labor tax on type θ increases average taxes on all types above θ and induces them to increase labor supply.

Guvenen et al. (2013)) this parameter is chosen to capture kurtosis of the shock process. Hence, stochastic processes with higher kurtosis, holding variance fixed, imply higher labor distortions in the right tail. The intuition is as follows. If kurtosis of the shock process is high, the hazard ratio $\frac{1-F_t(\theta)}{\theta f_t(\theta)}$ is large for high θ. This implies that any given marginal labor distortion has a smaller output loss than the same distortion with lognormal shocks. Also note that even though $\frac{v^2}{\ln \theta}$ converges to zero, this rate of convergence is very slow.

As we noted in the discussion of Proposition 1, the intratemporal component in the right tail depends only on the hazard B_t, the elasticity ε and the income effect, which with separable preferences is summarized by $\frac{\bar{\sigma}}{\bar{\sigma}+\bar{\varepsilon}}$ in the limit. The income effect becomes second order if the labor distortions go to zero, which explains why it disappears in the asymptotic formulas in the lognormal/mixture case in (19); it affects the limiting labor distortions if the shocks are fat-tailed.

Expressions (19) generalize those derived by Mirrlees (1971), Diamond (1998), and Saez (2001) for the optimal behavior of labor distortions in static models. Corollary 1 thus shows that their insights continue to hold in dynamic environments in the right tail of the distribution. The restriction $a\frac{1}{1+\bar{\varepsilon}} - \frac{\bar{\sigma}}{\bar{\sigma}+\bar{\varepsilon}} > 0$ is needed in (19) to make sure that the limiting value of $\frac{\tau_t^y}{1-\tau_t^y}$ is finite. When this restriction is not satisfied, τ_t^y may converge to 1. Note that even in this case the general conclusion of Corollary 1 remains unchanged – the optimal labor distortions are still determined by the intratemporal component as $\theta \to \infty$ even if this component diverges to infinity.

The last part of Corollary 1 characterizes the behavior of labor distortions

in the left tail. This result also follows from Proposition 1 and Assumption 5. Under Assumption 5 the intratemporal component converges to zero, while the intertemporal component is non-zero as long as the shocks are not i.i.d.

Expression (20) further simplifies if preferences are isoelastic. In this case $\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta)} \sim \beta R \frac{\tau_{t-1}^y}{1-\tau_{t-1}^y} \rho \left(\frac{c_t(0)}{c_{t-1}} \right)^{-\sigma}$ as $\theta \to 0$. Thus, the marginal distortions depend on the autocorrelation of the shocks, past labor distortions, and consumption growth rate. The latter two forces generally depend on the agent's age t, the past history of shocks, and Pareto weights.

We can also use decomposition (17) to obtain additional insights about time-series properties of the optimal labor distortions studied by Farhi and Werning (2013). Observe that $\mathbb{E}_{t-1} \frac{1}{U_{c,t}} B_t C_t = cov_{t-1}\left(\ln \theta, \frac{1}{U_{c,t}} \right)$. If we assume isoelastic preferences, multiply (17) by $\frac{1}{U_{c,t}(\theta)}$ and integrate, we get

$$\mathbb{E}_{t-1}\left[\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta)} \frac{1}{U_{c,t}(\theta)} \right] = \rho \beta R \frac{\tau_{t-1}^y}{1-\tau_{t-1}^y} \frac{1}{U_{c,t-1}} + (1+\varepsilon) cov_{t-1}\left(\ln \theta, \frac{1}{U_{c,t}} \right).$$
$$(21)$$

This equation is one of the key results of Farhi and Werning (2013). In particular, they show that it implies that the marginal utility-adjusted labor distortions follow an AR(1) process with a drift. Persistence of that process is determined by the autocorrelation parameter ρ, and its drift is strictly positive since generally we should expect that $cov_{t-1}\left(\ln \theta, \frac{1}{U_{c,t}} \right) > 0$. Farhi and Werning (2013) conclude that the optimal labor distortions should increase with age. Corollary 1 qualifies this result by showing that this drift should be observed in the left but not right tails of shock realizations since the asymptotic behavior of the labor distortions in the right tail is independent of t by equation (19). The intuition for this result follows from our discussion of the underlying economic

forces that determine the optimal labor wedge.

In the analysis above we restricted our attention to the preference specific-ations for which the Frisch elasticity and the EIS are finite. It is often possible to obtain simpler closed form expressions when this assumption is relaxed. These expressions, although special, can illustrate some key trade-offs in a transparent way. Assume, for example, that preferences are isoelastic with $\sigma = 0$. In this case we obtain from (17) for $t = 0$

$$\frac{\tau_0^y(\theta)}{1 - \tau_0^y(\theta)} = \underbrace{(1 + \varepsilon)}_{A_0(\theta)} \underbrace{\frac{1 - F_0(\theta)}{\theta f_0(\theta)}}_{B_0(\theta)} \underbrace{\int_\theta^\infty (1 - \alpha(x)) \frac{f_0(x)\, dx}{1 - F_0(\theta)}}_{C_0(\theta)}$$

and for $t > 0$

$$\frac{\tau_t^y(\theta)}{1 - \tau_t^y(\theta)} = \beta R \rho \frac{\tau_{t-1}^y}{1 - \tau_{t-1}^y}.$$

The quasi-linear case is special since it sets both the risk-aversion and the income effect to zero. Since agents are risk-neutral, they require no insurance against lifecycle shocks and therefore the intratemporal components are zero for all $t > 0$. Persistence of the shock process determines how initial hetero-geneity affects labor distortions in those periods because $\varphi_t(\theta) = \rho$ under any of the three stochastic processes in Assumption 5.

The absence of income effects allows us to illustrate transparently the trade-off between the redistribution and the minimization of output losses (i.e. "effi-ciency") in period 0. Suppose that α monotonically decreases and converges to zero, so that the planner favors redistribution from the more productive types. In this case the redistributive component C_0 monotonically increases from 0 to 1, reflecting higher gains of redistribution from higher types. The hazard

rate B_0 starts at ∞ and decreases (monotonically in the case of lognormal and Pareto-lognormal f_0) to its long-run finite value as $\theta \to \infty$, reflecting the fact that labor distortions for more productive types generate higher output losses.

Figure 1 illustrates how the shape and the size of the labor distortions depend on the hazard rate. We consider the three types of distributions from Assumption 5 and choose the parameters of these distributions so that $\ln \theta$ has mean and variance of 0 and 1 respectively in all cases. The Pareto-lognormal distribution has a tail parameter of 2.5. The mixture is drawn from two mean-zero normal distributions chosen so that excess kurtosis of $\ln \theta$ is equal to 10.[8] We set $\varepsilon = 2$ and $\alpha(\theta) \propto \exp(-\theta)$.

This figure shows several general principles that, as we shall see in Section 3, carry through to calibrated economies with risk-aversion. Panels A, B, and C show that the redistributive component C_t converges quickly to its limiting value of 1 as $\theta \to \infty$, while the hazard rate B_t converges to its right limit much slower. This implies that the shape of the optimal labor distortions resembles the shape of the hazard rate as long as θ is not too low. The hazard rates are slowly decreasing when shocks are lognormal or Pareto-lognormal, and are first U-shaped and then slowly decreasing when shocks are drawn from a mixture of lognormal. The optimal labor distortion τ_0^y (solid lines in Panels D, E, and F), which is a monotonic transformation of $\frac{\tau_0^y}{1-\tau_0^y}$, follows the same patterns.

Panels A, B, and C of Figure 1 also show that hazard rates in lognormal/mixture cases converge to their right limit of 0 slowly. At $\theta = 20$, which is about 3 standard deviations above the mean, both the hazard rate B_t and

---

[8]There are multiple ways to generate excess kurtosis of 10 and variance of 1 from the mixture of normal distributions. Figure 1 shows a representative pattern of distortions.
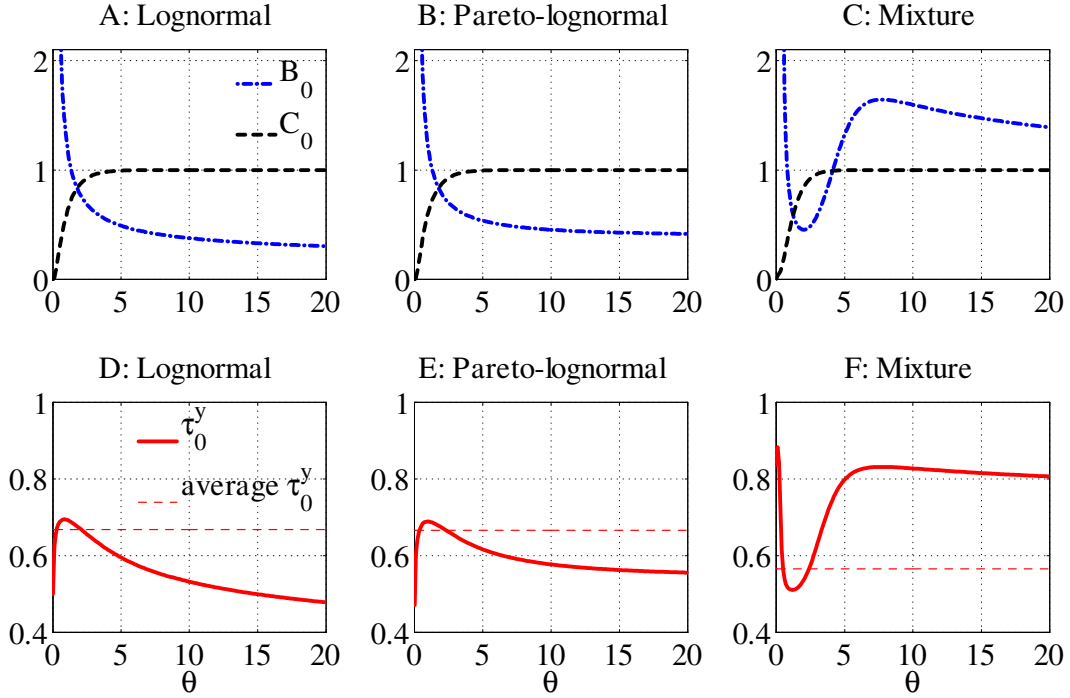
Figure 1: Optimal labor distortions in period 0 and their components for three distributions of shocks with quasi-linear preferences.

the optimal labor distortions with both lognormal and mixture shocks are substantially above 0. Even at $\theta = 22,000$, which is 10 standard deviations above the mean, the optimal labor distortion in the mixture case is equal to 0.62, both well above its limit value of zero and the limit value of the thick-tailed Pareto-lognormal shocks.

Panels D, E, and F of Figure 1 show that two commonly used summary statistics of the shock process – variance and the fatness of the tail – do not provide sufficient information to determine the size of the distortion or whether the optimal distortions should be progressive, even in the tail. The dashed line in Panels D, E, and F is the average labor distortion defined as $\int_0^\infty \tau_0^y(\theta) \, dF_0$.

The average labor distortions are almost 10 percentage points lower in the mixture of lognormals case. The reason for it is that, due to the high kurtosis of that distribution, most of the time individuals receive small shocks that require little insurance. On the other hand, medium size shocks occur with a much higher probability in the mixture case and hence the labor distortions for such shocks are high. Lognormal and Pareto-lognormal shocks imply very similar labor distortions for most of the shocks even though the former distribution has a thin tail while the latter has a thick Pareto tail. Figure 1 also contradicts the view that the optimal labor distortions should be progressive for high types if shocks are fat-tailed.[9] The optimal labor distortions are progressive in the right tail if the hazard rate B_0 either converges to its long run value from below or converges from above at a faster rate than the redistributive component C_0 converges to 1. The opposite result holds with Pareto-lognormal shocks for a wide range of Pareto weights α.

2.2   Non-separable preferences

We discuss next the extensions of our analysis to the case when utility is not separable in consumption and labor. We show that many principles discussed in the previous section continue to hold, although with some caveats. We also discuss the optimal savings distortions.

Let $\gamma_t(\theta) \equiv \frac{U_{cl,t}(\theta) l_t(\theta)}{U_{c,t}(\theta)}$ be the degree of complementary between consumption and labor and $\bar{X} = \lim_{\theta \to \infty} \frac{c_t(\theta)}{\left(1 - \tau_t^y(\theta)\right) y_t(\theta)}$ be the marginal propensity of consume out of the after-tax income in the right tail of the distribution. We

---

[9]See, for example, Diamond (1998) and Diamond and Saez (2011).

continue to make Assumption 4 with an additional extension that $\sigma_t(\theta), \varepsilon_t(\theta)$ and $\gamma_t(\theta)$ have finite limits denoted by $\bar{\sigma}, \bar{\varepsilon}, \bar{\gamma}$ as $\theta \to \infty$.

The decomposition of the labor distortions (17) still holds in the non-separable case, with the following modifications:

$$A_t(\theta) = 1 + \varepsilon_t(\theta) - \gamma_t(\theta),$$

$$C_t(\theta) = \int_\theta^\infty \exp\left(\int_\theta^x \left[\sigma_t(\tilde{x})\frac{\dot{c}_t(\tilde{x})}{c_t(\tilde{x})} - \gamma_t(\tilde{x})\frac{\dot{y}_t(\tilde{x})}{y_t(\tilde{x})}\right] d\tilde{x}\right)(1 - \lambda_{1,t}\bar{\alpha}_t(x) U_{c,t}(x)) \frac{f_t(x)\,dx}{1 - F_t(\theta)},$$

and

$$D_t(\theta) = \frac{A_t(\theta)}{A_{t-1}} \frac{U_{c,t}(\theta)}{U_{c,t-1}} \frac{\theta_{t-1} \int_\theta^\infty \exp\left(-\int_\theta^x \gamma_t(\tilde{x})\frac{d\tilde{x}}{\tilde{x}}\right) f_{2,t}(x)\,dx}{\theta f_t(\theta)}.$$

One difference with the separable case is in the intertemporal component and term D_t. When preferences are non-separable, the marginal utility of consumption is no longer the sufficient statistic for the relative costs of providing incentives in periods t and $t-1$ and γ_t enters into the expression for D_t. If $U_{cl} \geq 0^{10}$ and $\rho \geq 0$, then much of the previous analysis of the intertemporal component still applies because $D_t(\theta)$ is bounded and both $U_{c,t}(\theta)$ and $D_t(\theta)$ decline to zero at a geometric rate as $\theta \to \infty$. In this case the asymptotic behavior of labor distortions in the right tail, assuming shocks satisfy Assumption

---

[10]Empirical labor literature often finds that consumption and labor are complements (Browning, Hansen and Heckman (1999)), although some authors recently challenged that conclusion (Blundell, Pistaferri and Saporta-Eksten (2014)).

[5](#), is driven by the intratemporal component. That is, as $\theta \to \infty$,

$$\frac{\tau_t^y(\theta)}{1 - \tau_t^y(\theta)} \sim A_t(\theta) B_t(\theta) C_t(\theta)$$

$$\sim \begin{cases} \left[a \frac{1}{1+\bar{\varepsilon}-\bar{\gamma}} - \frac{\bar{\sigma}-\bar{\gamma}}{\bar{\sigma}+\bar{\varepsilon}-\bar{\gamma}(\bar{X}+1)} \right]^{-1} & \text{if } f_t \text{ is Pareto-lognormal, } a \frac{1}{1+\bar{\varepsilon}-\bar{\gamma}} - \frac{\bar{\sigma}-\bar{\gamma}}{\bar{\sigma}+\bar{\varepsilon}-\bar{\gamma}(\bar{X}+1)} > 0 \\ \left[\frac{\ln \theta}{v} \frac{1}{1+\bar{\varepsilon}-\bar{\gamma}} \right]^{-1} & \text{if } f_t \text{ is lognormal/mixture.} \end{cases}$$

$$(22)$$

A more substantive difference with the separable case is that the limiting values $\bar{\varepsilon}$, $\bar{\sigma}$, $\bar{\gamma}$ and $\bar{X}$ are endogenous and depend on the way incentives are provided intertemporally. To illustrate the key economic mechanism, it is convenient to re-write A_t and C_t not in terms of structural parameters ε_t, σ_t and γ_t but in terms of income and substitution effects. In particular, let $\zeta_t^u(\theta)$ and $\zeta_t^c(\theta)$ be the uncompensated and compensated elasticities of labor supply, holding savings fixed, and $\eta_t(\theta)$ be the income effect holding savings fixed defined by the Slutsky equation $\eta_t(\theta) = \zeta_t^u(\theta) - \zeta_t^c(\theta)$. Then A_t and C_t can be written as

$$A_t(\theta) = \frac{1 + \zeta_t^u(\theta)}{\zeta_t^c(\theta)}$$

$$C_t(\theta) = \int_\theta^\infty \exp(g_t(x;\theta)) (1 - \lambda_{1,t} \bar{\alpha}_t(x) U_{c,t}(x)) \frac{f_t(x)\,dx}{1 - F_t(\theta)},$$

where

$$g_t(x;\theta) = \int_\theta^x \left\{ \frac{-\eta_t(\tilde{x})}{\zeta_t^c(\tilde{x})} \frac{\dot{y}_t}{y_t} \tilde{x} - \sigma_t(\tilde{x}) \frac{(1 - \tau_t^y(\tilde{x})) \dot{y}_t - \dot{c}_t}{c_t} \tilde{x} \right\} d\tilde{x}.$$

The dependence of $A_t(\theta)$ on the elasticities is standard and appears in the

28

same way as in the static models (see Saez (2001)). The term g_t measures the income effect on labor supply. It consists of two parts. The first one determines the income effect on labor supply holding savings fixed, which is also analogous to the equivalent term in the static models. In dynamic models relaxed incentive constraints allow the planner to redistribute resources not only in the current period but also in the future. This dynamic income effect is captured by the second term in function g_t. It depends on the elasticity of intertemporal substitution, σ_t, and the difference between the after-tax income and consumption in period t. This term is positive if and only if reporting a higher θ makes the consumers better off in the future, $\omega_{1,t+1}(\theta|\theta) \geq 0$.[11] In this case the intertemporal provision of incentives lowers the effective income effect on labor supply.

To get the intuition for the behavior of the optimal labor distortions we consider commonly used GHH preferences (see Greenwood, Hercowitz and Huffman (1988)):

$$U(c,l) = \frac{1}{1-\nu}\left(c - \frac{1}{1+1/\zeta}l^{1+1/\zeta}\right)^{1-\nu} \tag{23}$$

for some $\nu, \zeta > 0$. For such preferences $A_t(\theta) = 1 + 1/\zeta$, $\eta_t(\theta) = 0$, $\gamma_t(\theta) \geq 0$ and $D_t(\theta)$ converges to zero at a geometric rate as $\theta \to \infty$. Therefore many of the arguments used to prove Corollary 1 continue to apply. In particular, as long as f_t satisfies Assumption 5, labor distortions are asymptotically equivalent to the intratemporal term in the right tail. If the shocks are mixture/lognormal, then the income effects are of second order and $\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta)} \sim$

---

[11]This condition holds if Assumption 3 is satisfied.

$$\left[\frac{\ln\theta}{v}\frac{1}{1+1/\zeta}\right]^{-1}\ (\theta\to\infty).$$

When the tails of the shock process are Pareto, the income effects are no longer of the second order. In this case the redistributive component $C_t(\theta)$ depends in the limit both on the marginal propensity to consume and the limiting value of labor distortions,

$$\lim_{\theta\to\infty}C_t(\theta)=1+\zeta\frac{-\bar{\sigma}\left(1-\bar{X}\right)}{\bar{X}}\lim_{\theta\to\infty}\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta)}. \tag{24}$$

The limiting value of labor distortions is then given by

$$\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta)}\sim\left[a\frac{1}{1+1/\zeta}-\zeta\frac{-\bar{\sigma}\left(1-\bar{X}\right)}{\bar{X}}\right]^{-1}\ (\theta\to\infty),$$

provided that the expression on the right hand side is positive. Unlike the separable case, the dynamic provision of incentives, summarized by $\bar{X}$, affects the value of labor distortions in the limit. If the marginal propensity to consume converges to 1 for high θ, as it is the case in static models, then this formula reduces to the one obtained by Saez (2001). This labor distortion is strictly lower than the static limit if reporting higher type in period t improves utility in the future, since $\omega_{1,t+1}(\theta|\theta)\geq0\ (\theta\to\infty)$ if and only if $\bar{X}\leq1$ (see the online appendix).

We can obtain starker results if we replace the power utility function in (23) with any functional form that bounds U''/U' away from zero (which effectively implies that $\sigma_t(\theta)\to\infty$ as $\theta\to\infty$, while keeping $\gamma_t(\theta)>0$). In this case it can be shown that the marginal labor distortions converge to 0 independently of the thickness of the Pareto tail (see Golosov, Troshkin and

Tsyvinski (2011)) or properties of $\bar{X}$. See Lemma 9 in the online appendix for the formal statement of this result and its proof.

We conclude this section with a general result about the optimality of savings distortions. When preferences are separable, it is well known (see, e.g., Golosov, Kocherlakota and Tsyvinski (2003)) that savings distortions are positive as long as $var_t (c_{t+1}) > 0$. We show that a weaker version of this result holds in the non-separable case. Let $\tilde{\tau}_t^s$ be a life-time saving distortion defined as

$$1 - \tilde{\tau}_t^s \left(\theta^t \right) \equiv \left(\frac{1}{\beta R} \right)^{T-t} \frac{U_c \left(c_t \left(\theta^t \right), y_t \left(\theta^t \right) / \theta_t \right)}{\mathbb{E}_t \left\{ U_c \left(c_T \left(\theta^T \right), y_T \left(\theta^T \right) / \theta_T \right) \right\}}.$$

Proposition 2. *Suppose Assumption 2 is satisfied, $U_{cl} \geq 0$, and $F_T \left(0 | \theta \right) = 1$ for all θ. Then $\tau_t^y \left(\theta^t \right) \geq 0$ implies $\tilde{\tau}_t^s \left(\theta^t \right) \geq 0$ with strict inequality if variance of consumption in period T conditional on information in θ^t is positive, $var_t \left(c_T \right) > 0$.*

Note that $\tilde{\tau}_t^s \left(\theta^t \right) > 0$ implies that some savings distortions following history θ^t must be strictly positive. By the law of iterated expectations

$$\frac{1}{1 - \tilde{\tau}_t^s} = \mathbb{E}_t \frac{1}{1 - \tau_t^s} \times ... \times \frac{1}{1 - \tau_{T-1}^s},$$

therefore, $\tilde{\tau}_t^s > 0$ if there is a positive saving distortion in at least some states in the future.

The intuition for this result comes from the observation made by Mirrlees (1976) that in a static, multi-good economy it is optimal to have a positive distortion on the consumption of goods that are complementary with leisure, assuming the optimal labor tax is positive. In our dynamic economy the

assumption that $\gamma \geq 0$ implies that the future consumption is more complementary with leisure and hence a positive wedge is desirable. This wedge, however, cannot be interpreted as a distortion in the Euler equation of the consumer, since this is a distortion conditional on providing optimal insurance in the future. Therefore an extra unit of savings does not increase future utility by $\beta R \mathbb{E}_t U_{c,t+1}$ as in the standard incomplete market models and this relationship in general is more nuanced.[12] The optimal provision of incentives implies that if in any period $\hat{T}$ the labor supply becomes constant (as it happens if individuals retire in that period), an extra unit of savings generates $\frac{1}{\beta R^{\hat{T}-t}} \mathbb{E}_t \frac{1}{U_{c,\hat{T}}}$ utils in the future, which is an extension of the Inverse Euler equation obtained in the separable case. Then the combination of arguments in Mirrlees (1976) and Golosov, Kocherlakota and Tsyvinski (2003) leads to Proposition 2.

3   Quantitative analysis

We now turn to the quantitative analysis of the model calibrated to the U.S. administrative data. We study a 65-period lifecycle in which agents work for the first 40 periods, from 25 to 64 years old, and then retire for the remaining 15 years. For a baseline calibration we use isoelastic preferences (16) with $\sigma = 1$ and $\varepsilon = 2$ and choose $\beta = R^{-1} = 0.98$ and utilitarian Pareto weights. We provide comparisons where the baseline calibrated stochastic process is

---

[12]Golosov, Troshkin and Tsyvinski (2011) discuss in detail the mapping between our recursive mechanism design problem and a static optimal tax problem with multiple goods. We refer the reader to that paper for the intuition on how distortions driven by complementarities with leisure map into distortions in the Euler equation.

replaced with a lognormal process with the same mean and variance, as well as robustness checks in the online appendix.

Our analysis above emphasizes the stochastic process for skills as a crucial determinant of the key features of the optimal distortions. Figure 1 shows that higher moments play an important role in determining their patterns. Such moments are difficult to estimate reliably using easily accessible panel data sets such as the U.S. Panel Study of Income Dynamics due to the small sample size and top coding. To overcome this problem we use the findings of Guvenen, Ozkan and Song (2013) and Guvenen et al. (2013), who study newly available high-quality administrative data from the U.S. Social Security Administration based on a nationally representative panel containing 10 percent of the U.S. male taxpayers from 1978 to 2011.

Guvenen, Ozkan and Song (2013) and Guvenen et al. (2013) document that the stochastic process for annual log labor earnings is highly leptokurtic, negatively skewed, and is not well approximated by a lognormal distribution. They also show that the empirical shock process can be approximated well by a mixture of three lognormal distributions, shocks from two of which are drawn with low probabilities. The high-probability distribution controls the variance of the shocks, while the two low-probability distributions control their skewness and kurtosis.

Guvenen et al. (2013) report statistics for the stochastic process for labor earnings, which correspond to y_t in our model. To calibrate the stochastic process for skills θ_t we use the following procedure. We assume that the initial θ_0 is drawn from a three-parameter Pareto-lognormal distribution, analyzed in

the previous section, and that for all $t > 0$ the stochastic process for θ_t follows

a mixture of lognormals[13]

$$\ln \theta_t = \ln \theta_{t-1} + \epsilon_t,$$

where

$$\epsilon_t = \begin{cases} \epsilon_{1,t} \sim \mathcal{N}\left(\mu_1, \upsilon_1\right) & \text{w.p. } p_1, \\ \epsilon_{2,t} \sim \mathcal{N}\left(\mu_2, \upsilon_2\right) & \text{w.p. } p_2, \\ \epsilon_{3,t} \sim \mathcal{N}\left(\mu_3, \upsilon_3\right) & \text{w.p. } p_3. \end{cases}$$

We impose $p_3 = 1 - p_1 - p_2$, $\upsilon_3 = \upsilon_1$, $\mu_2 = 0$. The individuals, whose skills
are drawn from the stochastic process, choose their optimal labor and savings
given a tax function $\mathcal{T}(y)$. We follow Heathcote, Storesletten and Violante
(2014), who find that a good fit to the effective earnings taxes in the U.S. is
given by $\mathcal{T}(y) = y - \lambda y^{1-\tau}$, where the progressivity parameter τ is equal to
0.151.[14] We choose the six parameters of the stochastic process and the three
parameters of the initial distribution to balance the government budget and to
minimize the sum of the least absolute deviations of nine simulated moments
of the earnings process in the model from the nine moments in the data in
Guvenen et al. (2013) and Guvenen, Ozkan and Song (2013).

Table 1 reports the calibrated parameters, the simulated moments, and

---

[13] Guvenen et al. (2013) find that the persistence of the stochastic process for earnings is
very close to one. We set $\rho = 1$ in our calibration of the shock process and later discuss the
differences between the earnings process and the shock process in the model.

[14] The marginal labor distortions in the model correspond to the effective marginal labor
tax rates in the data, which is a combination of the statutory tax rate (which is generally
progressive) and the rate of the phase out of welfare transfers (which is generally regressive).
In the U.S., there is heterogeneity in the shapes of the effective tax rates as a function of
income as they vary by state, family status, age, type of residence, etc. Some typical
patterns of the effective marginal rates in the U.S. data are progressive, U-shaped, and
inverted S-shaped (see CBO (2007) and Maag et al. (2012)).

Table 1: Calibrated parameters of the shock process, simulated moments, and the target moments in the data.

Calibrated Shock Parameters								
μ_1	μ_3	v_1	v_2	p_1	p_2	μ	υ	a
0.03	-0.47	0.22	2.64	0.71	0.15	0.17	5.59	2.73

Moments of Distributions								
Stochastic process						Initial distribution		
Mean	SD	Kurtosis	Kelly's Skewness	P10	P90	P50	P90	P99
Simulated shock moments (θ_t):								
0.010	0.46	10.15	-0.24	-0.47	0.49	10.41	11.13	12.07
Simulated equilibrium earnings moments (y_t):								
0.008	0.51	11.30	-0.20	-0.45	0.44	10.39	11.06	11.94
Data earnings targets (y_t^{data}):								
0.009	0.52	11.31	-0.21	-0.44	0.47	10.06	10.76	11.71

the data targets.[15] Table A.4 in Guvenen, Ozkan and Song (2013) provides the 50th, 90th and 99th percentiles of the earnings of the 25 year old in their base sample that we use as data targets for period-0 distribution of earnings in the model and report as the last three numbers in the bottom row of Table 1. Guvenen et al. (2013) report in Table II, Specification 3, their estimation results for the stochastic process of earnings in the data, which we use to generate the other six data targets reported in the bottom row of Table 1.[16]

3.1 Computational approach

We use the recursive formulation of the planning problems (6) and (11). Here we provide a summary of our approach while the online appendix contains

---

[15] Kelly's skewness is defined as $\frac{(P90-P50)-(P50-P10)}{(P90-P10)}$, where Pz is the zth percentile growth rate.

[16] We take unconditional moments. Guvenen et al. (2013) and Guvenen, Ozkan and Song (2013) also report how they change with age, with income level, and over the business cycle. This can be incorporated with age-dependent parameters that depend also on past shock realizations and on an aggregate shock.

further details.

The main problem is a finite-horizon discrete-time dynamic programming problem with a three-dimensional continuous state space. We solve it by value function iteration starting from the period before retirement, $\hat{T} - 1$. The present value of the resources required to provide promised utility over the remaining $T - \hat{T} + 1$ periods of retirement is added to the value function in period $\hat{T} - 1$. We approximate each value function with tensor products of orthogonal polynomials evaluated at their root nodes and proceed by backward induction. To solve each node's minimization sub-problem efficiently, we use an implementation of interior-point algorithm with a trust-region method to solve barrier problems and an l_1 barrier penalty function. Assumption 2 is satisfied trivially for the preferences and parameter values we chose above. We verify the increasing properties in Assumption 3 numerically. We compute $\hat{w}_0$ such that $V_0(\hat{w}_0) = 0$ and compute the optimal allocations reported below by forward induction. The optimal labor and savings distortions are then computed from the policy functions using definitions (13) and (14).

3.2 Results

We first discuss the optimal labor and savings distortions in the calibrated economy. Figure 2 shows typical distortions for representative histories. Each thick line in Panel A plots $\tau_t^y \left(\bar{\theta}^{t-1}, \theta_t \right)$ at a given t for a history of past shocks $\bar{\theta}^{t-1} = \left(\bar{\theta}, ..., \bar{\theta} \right)$. We chose $\bar{\theta}$ for Panel A so that an individual with a lifetime of $\bar{\theta}$ shocks will have the average lifetime earnings, $\frac{1}{\hat{T}} \sum_{t=0}^{\hat{T}-1} y_t \left(\bar{\theta}^t \right)$ where $y_t \left(\bar{\theta}^t \right) = \bar{\theta} l_t \left(\bar{\theta}^t \right)$, approximately equal to the average U.S. male earnings in
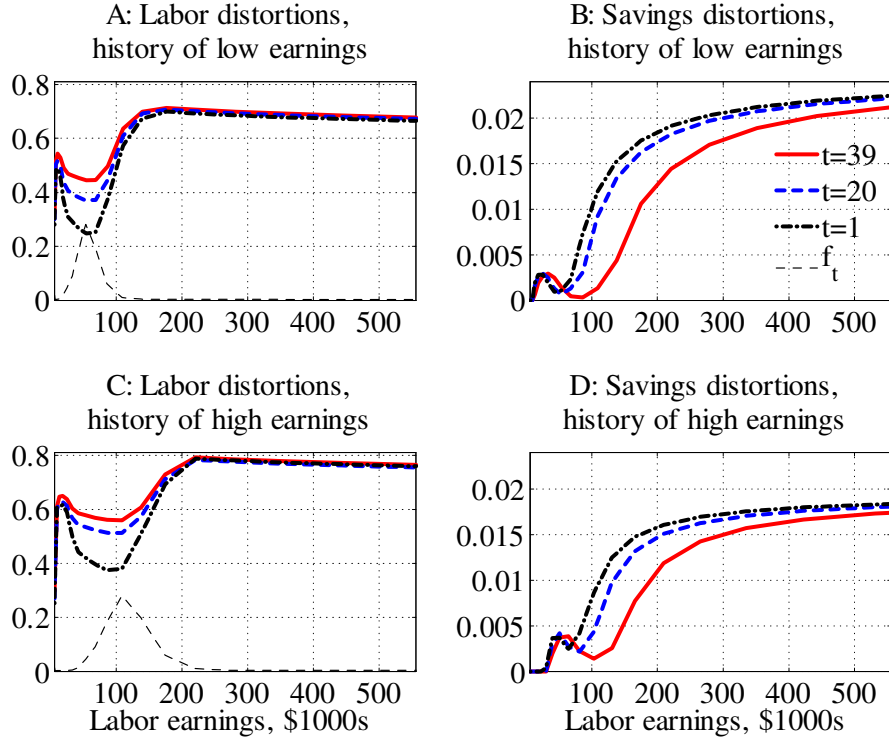
Figure 2: Optimal distortions at selected periods: Panels A and B have a history of $\bar{\theta}$ shocks chosen so that an individual with a lifetime of $\bar{\theta}$ shocks will have the average lifetime earnings approximately equal to the average U.S. male earnings in 2005; Panels C and D are the analogues with $\bar{\theta}$ chosen so that the average lifetime earnings approximately equal twice the U.S. average.

2005; Panel C is the analogue with $\bar{\theta}$ chosen so that the average lifetime earnings approximately equal twice the U.S. average.[17] The distortions are plotted against current earnings, $y_t\left(\bar{\theta}^{t-1}, \theta_t\right) = \theta_t l_t\left(\bar{\theta}^{t-1}, \theta_t\right)$, measured on the horizontal axis in 1,000s of real 2005 dollars. The lines in Panels B and D plot the corresponding values for $\tau_t^s\left(\bar{\theta}^{t-1}, \theta_t\right)$. The thin lines in Panels A and C

---

[17]The average lifetime earnings are \$53,934 for the history in Panel A and \$108,990 in Panel C. According to the U.S. Census, the average male earnings in 2005 were \$54,170 (see U.S. Census Bureau, Historical Income Tables, Table P-12 at https://www.census.gov/hhes/www/income/data/historical/people/)

display $f\left(\theta|\bar{\theta}\right)$.

Several insights emerge from examining the distortions in Figure 2. First, the optimal labor distortions are highly non-linear, with pronounced U-shape patterns. The U-shapes are centered around the expected realization of the shock conditional on past earnings, as indicated by the peaks of the conditional distributions. The individuals who experienced higher realizations of the shocks in period $t-1$ are expected to have higher productivity in period t and the U-shape of their labor distortions is shifted to the right. Since the individuals in Panel C have a history of higher earnings than the individuals in Panel A, the U-shapes in Panel C are centered around higher earnings than those in Panel A. The optimal savings distortions, Panels B and D, are similarly non-linear and non-monotone but the non-monotonicities are much less pronounced than in labor distortions.

Proposition 1 and Corollary 1 show that an understanding of the economic forces behind these observations can be gained by examining our decomposition (17). Figure 3 illustrates the decomposition for the histories shown in Figure 2. The intratemporal terms B_t and C_t are shown in Panels A and C (A_t is constant given the preferences); Panels B and D show the intertemporal terms D_t. Many of the insights that emerge from Figure 3 can be understood from our analysis in Section 2. The intertemporal term D_t converges to zero at a geometric rate as labor earnings increase (cf. Proposition 1). The hazard term B_t first follows a U-shape and then declines to zero but at a much slower rate (see Corollary 1), while C_t increases. The U-shaped pattern of the hazard term is driven by the high kurtosis of the calibrated shock process, implied by
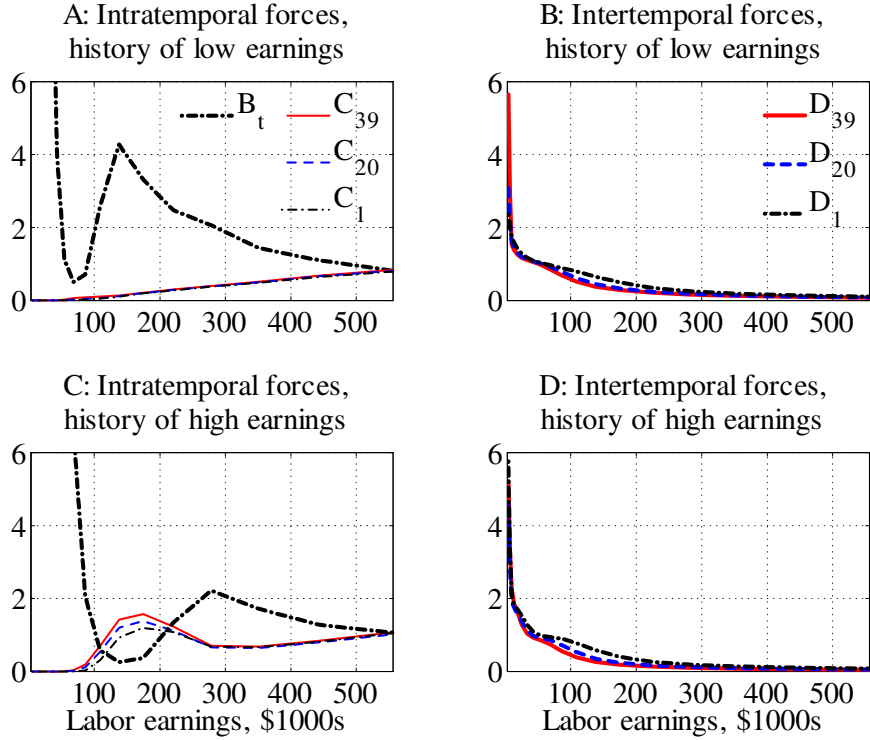
Figure 3: The decomposition of optimal labor distortions: Panels A and B have a history of $\bar{\theta}$ shocks chosen so that an individual with a lifetime of $\bar{\theta}$ shocks will have the average lifetime earnings approximately equal to the average U.S. male earnings in 2005; Panels C and D are the analogues with $\bar{\theta}$ chosen so that the average lifetime earnings approximately equal twice the U.S. average.

the high kurtosis in the labor earnings in the data. The behavior of terms B_t and C_t and their implications for the optimal labor distortions are very similar to the quasi-linear example in Figure 1, Panels C and F, with the exception that C_t is not necessarily monotone. The sum of the intratemporal component $(1 + \varepsilon) B_t C_t$ and the intertemporal component $\frac{\tau_{t-1}}{1-\tau_{t-1}} D_t$ implies the U-shaped patterns of the labor distortions in Figure 2. Finally, note that all three terms B_t, C_t, and D_t depend little on individual age t and are mainly driven by the past realization of the shock. In the online appendix we provide additional

illustrations of the decompositions.

Figure 2 also shows that the labor distortions increase with age at low and medium labor earnings but do not depend on age at high labor earnings. Farhi and Werning (2013) showed that it is optimal for labor distortions to increase with age on average (see also our discussion around equation (21)), while our Corollary 1 qualifies this insight by showing that the increase happens only for shocks in the left tail.

The second insight that emerges from examining distortions in Figure 2 is that their quantitative magnitude is relatively high. The labor distortions for high shocks often exceed 70 percent. Savings distortions are defined as a wedge in the gross return to capital (i.e., interest return plus principle) and for high realizations of shocks can be as high as 2 percent. We could equivalently define savings distortions on the net capital return $R - 1$; given our parametrization of R the net savings distortion is approximately 50 times the gross savings distortion. In the online appendix we report robustness checks for the re-calibrated economy with $\varepsilon = 1$ and $\varepsilon = 4$. The labor distortions remain high, especially in the tails.

To examine the magnitudes of the optimal distortions more systematically we compute a weighted average of labor distortions that a person with a realization of a shock θ_t experiences in period t. In particular, we define average distortions as $\bar{\tau}_t^i(\theta_t) \equiv \int_{\Theta^{t-1}} \tau_t^i(\theta^{t-1}, \theta_t) \, dF(\theta^{t-1})$ for $i \in \{y, s\}$. In Figure 4 we show these distortions plotted against labor earnings $\bar{y}_t(\theta_t) = \theta_t \bar{l}_t(\theta_t)$, where $\bar{l}_t(\theta_t)$ is the weighted average across the simulated histories for a given t. At high earnings these average labor distortions are about 75 percent and

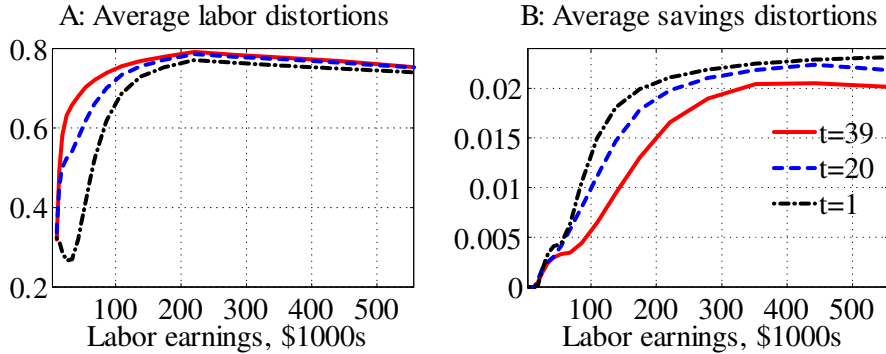A: Average labor distortions   B: Average savings distortions

Figure 4: Optimal average labor (Panel A) and savings distortions (Panel B) as functions of current earnings at selected periods.

are virtually independent of t. At average earnings they vary from about 25 percent early in life to about 65 percent late in life. Average savings distortions range from about 0.3 percent at average labor earnings to 2-2.5 percent at high earnings.

It is instructive to compare the quantitative predictions about the size of the optimal labor distortions with the distortions that arise in a static model. Saez (2001) calibrated the distribution of skills in a static model using data on the cross-sectional distribution of labor earnings. The specification that is closest to ours is his Figure 5, Utilitarian criterion, utility type II. He finds that the optimal labor distortions are U-shaped, with the distortions at average earnings about 40-55 percent and at high earnings about 65-80 percent, depending on the chosen elasticity of labor supply. The cross-sectional distribution of labor earnings in our data and the magnitude of the average distortions in Figure 4 are similar.[18] In our dynamic model these distortions

---

[18]Saez (2001) uses preference specification $\ln(c) - \ln\left(1 + \frac{l^{1+\kappa}}{1+\kappa}\right)$ and targets the compensated elasticity of labor supply $1/\kappa$ rather than the Frisch elasticity that we use in our

are history dependent and are similar to the distortions in the static model only on average. As we showed in Figure 2, the U-shapes in the dynamic economy are centered around the expected realization of earnings conditional on past earnings, while in the static model they are centered around the cross-sectional average labor earnings. In the dynamic economy, the planner also conditions the average labor distortions on age and uses savings distortions.

The third insight that emerges from our analysis is that higher moments of the stochastic process for idiosyncratic shocks, such as kurtosis, have an important effect on both the shape and the size of the optimal distortions. To illustrate their effect, we compare our baseline simulations with the simulations in the economy where we set the shock process to be lognormal with the same mean and variance as our baseline. Figure 5 compares the distortions with the lognormal shocks (thick lines) to the baseline mixture case (thin lines), for a history of low earnings in Panels A and B and for the average distortions in Panels C and D.

Since the baseline uses a mixture of lognormals, the hazard ratios and the labor distortions with both log-normal and mixture distributions are proportional to $1/\ln\theta$ in the right tail. Away from their asymptotic limit, the labor distortions behave very differently in the two cases. While the labor distortions are U-shaped in the mixture case, they are mildly regressive in the lognormal case. This implies different responses to earnings shocks: the labor distortions typically increase in response to a positive earnings shock in the baseline economy, while they decrease in the economy with lognormal

_____

analysis. Saez (2001) reports optimal taxes for the compensated elasticities of 0.25 and 0.5. Our preference parametrization implies the compensated elasticity of 0.33 in the static model.

A: Labor distortions,
history of low earnings

B: Savings distortions,
history of low earnings

C: Average labor distortions

D: Average savings distortions
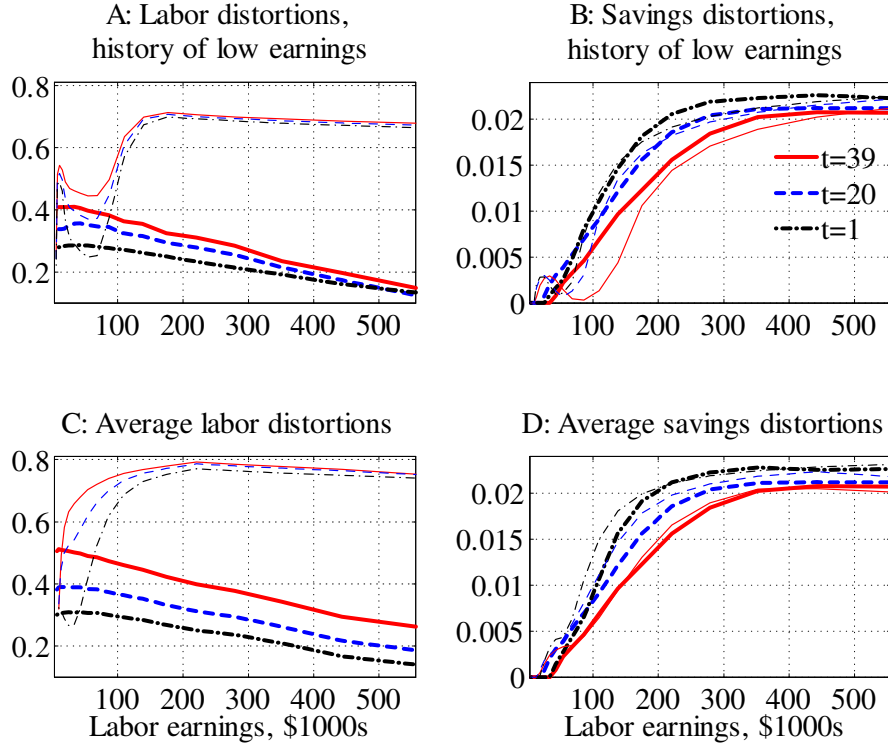
Labor earnings, \$1000s

Labor earnings, \$1000s

Figure 5: Optimal distortions with and without higher moments: the lognormal process (thick lines) has the same mean and variance as the mixture (thin lines). Panels A and B have a history of $\bar{\theta}$ shocks chosen so that an individual with a lifetime of $\bar{\theta}$ shocks will have the average lifetime earnings approximately equal to the average U.S. male earnings in 2005; Panels C and D are average distortions.

shocks. The magnitudes of the distortions are also different, for example, at the annual labor earnings of \$500,000 the average labor distortion is almost four times as large as in the lognormal case. The intuition for these findings follows directly from our discussion of Figure 1. The differences in savings distortions are much less significant in the two cases, as are the differences in lifetime average distortions, $\frac{1}{\hat{T}} \sum_{t=0}^{\hat{T}-1} \int_{\Theta^t} \tau_t^i \left(\theta^t \right) dF \left(\theta^t \right)$ for $i \in \{y, s\}$: the average labor distortions are 42.7 percent in the mixture case and 40.6 percent in the lognormal case; the average savings distortions are 0.6 and 0.5 percent

respectively. In the online appendix we illustrate the corresponding changes to earnings and consumption moments.

Finally, we quantify the importance of nonlinearities and history dependence emphasized above by computing welfare losses from using simpler, affine tax functions. We consider an equilibrium in the economy with linear taxes on capital and labor income, reimbursed lump-sum to all agents. In the first experiment the tax rates are the same for all ages and are chosen to maximize ex-ante welfare. In the second experiment we allow tax rates to depend on t and set them to the age-t average constrained-optimal labor and savings distortions, $\int_{\Theta^t} \tau_t^i \left(\theta^t \right) dF \left(\theta^t \right)$ for $i \in \{y, s\}$. In each case, we compute consumption equivalent welfare loss, Δ, from using a simple policy instead of the constrained optimal policies, given by $\mathbb{E}_{-1} \sum_{t=0}^{T} \beta^t U \left(c_t^{ce} \times \Delta, l_t^{ce} \right) = \mathbb{E}_{-1} \sum_{t=0}^{T} \beta^t U \left(c_t, l_t \right)$ where (c, l) are constrained-optimal allocations and (c^{ce}, l^{ce}) are equilibrium choices given the simple policy.

In the baseline mixture case, the policy of age-independent taxes leads to the welfare loss of 3.64 percent of consumption, with the labor tax of 43.1 percent, quite close to the lifetime average, and the capital tax of 0.05 percent. The age-dependent tax rates reduce the welfare loss to 1.81 percent. Higher moments of the shock process have a significant impact on the losses. Repeating the same two experiments in the lognormal case, the welfare losses from age-independent policies are 0.51 percent, with the labor tax of 41.2 percent and the capital tax of 0.07 percent, while the age-dependent policies reduce the loss to 0.30 percent. The smaller welfare changes with lognormals shocks are perhaps not surprising in light of the analysis of Figure 5 where linear

44

taxes appear to be better approximations for the optimal distortions.

4   Conclusion

This paper takes a step toward the characterization of the optimal labor and savings distortions in a lifecycle model. Our analysis focuses on the distortions in fully optimal allocations, restricted only by the information constraint. The optimal allocations and distortions can be implemented as a competitive equilibrium with non-linear taxes that depend on the current and past choices of labor supply and savings. Our approach is complementary to that of Conesa, Kitao and Krueger (2009), Heathcote, Storesletten and Violante (2014) or Kindermann and Krueger (2014) and others, who restrict attention to a-priori chosen functional forms of tax rates as a function of income and optimize within that class. Informationally constrained optimum that we study provides an upper bound on welfare that can be attained with such taxes. The properties of the distortions in the constrained optimum can serve as a guidance in choosing simple functional forms for taxes that capture most of the possible welfare gains.

References

Abraham, Arpad, and Nicola Pavoni. 2008. "Efficient Allocations with Moral Hazard and Hidden Borrowing and Lending: A Recursive Formulation." *Review of Economic Dynamics*, 11(4): 781–803.

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

Ales, Laurence, and Pricila Maziero. 2007. "Accounting for private information." *Working paper.*

Badel, Alejandro, and Mark Huggett. 2014. "Taxing top earners: a human capital perspective." *Working paper.*

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten. 2014. "Consumption inequality and family labor supply." European Central Bank Working Paper Series 1656.

Browning, Martin, Lars Peter Hansen, and James J. Heckman. 1999. "Micro data and general equilibrium models." In *Handbook of Macroeconomics*. Vol. 1 of *Handbook of Macroeconomics*, , ed. J. B. Taylor and M. Woodford, Chapter 8, 543–633. Elsevier.

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

Colombi, Roberto. 1990. "A New Model Of Income Distribution: The Pareto-Lognormal Distribution." In *Income and Wealth Distribution, Inequality and Poverty. Studies in Contemporary Economics*, , ed. Camilo Dagum and Michele Zenga, 18–32. Springer Berlin Heidelberg.

Conesa, Juan Carlos, Sagiri Kitao, and Dirk Krueger. 2009. "Taxing capital? Not a bad idea after all!" *American Economic Review*, 99(1): 25–48.

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

Diamond, Peter, and Emmanuel Saez. 2011. "The Case for a Progressive Tax: From Basic Research to Policy Recommendations." *Journal of Economic Perspectives*, 25(4): 165–90.

Farhi, Emmanuel, and Ivan Werning. 2013. "Insurance and Taxation over the Life Cycle." *Review of Economic Studies*, 80(2): 596–635.

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

Fukushima, Kenichi. 2010. "Quantifying the welfare gains from flexible dynamic income tax systems." *Working paper*.

Geweke, John, and Michael Keane. 2000. "An empirical analysis of earnings dynamics among men in the PSID: 1968-1989." *Journal of Econometrics*, 96(2): 293–356.

Golosov, Mikhail, Aleh Tsyvinski, and Iván Werning. 2006. "New dynamic public finance: A user's guide." *NBER Macroeconomics Annual*, 21: 317–363.

Golosov, Mikhail, and Aleh Tsyvinski. 2006. "Designing optimal disability insurance: A case for asset testing." *Journal of Political Economy*, 114(2): 257–279.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski. 2011. "Optimal Dynamic Taxes." *NBER Working Paper 17642.*

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

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W Huffman. 1988. "Investment, Capacity Utilization, and the Real Business Cycle." *American Economic Review,* 78(3): 402–17.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song. 2013. "What Do Data on Millions of U.S. Workers Say About Labor Income Risk?" *Working paper.*

Guvenen, Fatih, Serdar Ozkan, and Jae Song. 2013. "The Nature of Countercyclical Income Risk." *Working paper.*

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. 2014. "Optimal Tax Progressivity: An Analytical Framework." *NBER Working Paper 19899.*

Kapička, Marek. 2013. "Efficient allocations in dynamic private information economies with persistent shocks: A first order approach." *Review of Economic Studies,* 80(3): 1027–1054.

Kindermann, Fabian, and Dirk Krueger. 2014. "High Marginal Tax Rates on the Top 1%? Lessons from a Life Cycle Model with Idiosyncratic Income Risk." *NBER Working Paper 20601.*

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

Lockwood, Benjamin B., Charles G. Nathanson, and E. Glen Weyl. 2014. "Taxation and the Allocation of Talent." *Working paper.*

Maag, Elaine, C. Eugene Steuerle, Ritadhi Chakravarti, and Caleb Quakenbush. 2012. "How Marginal Tax Rates Affect Families at Various Levels of Poverty." *National Tax Journal*, 65(4): 759–82.

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

Mirrlees, James. 1976. "Optimal tax theory: A synthesis." *Journal of Public Economics*, 6(4): 327–358.

Pavan, Alessandro, Ilya Segal, and Juuso Toikka. 2014. "Dynamic Mechanism Design: A Myersonian Approach." *Econometrica*, 82(2): 601–653.

Reed, William J., and Murray Jorgensen. 2004. "The Double Pareto-Lognormal Distribution - A New Parametric Model for Size Distributions." *Communications in Statistics - Theory and Methods*, 33(8): 1733–1753.

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

Storesletten, Kjetil, Christopher I. Telmer, and Amir Yaron. 2004. "Consumption and risksharing over the life cycle." *Journal of Monetary Economics*, 51: 609–633.

Su, Che-Lin., and Kenneth L. Judd. 2007. "Computation of Moral-Hazard Problems." *Working paper.*

Tuomala, Matti. 1990. *Optimal income tax and redistribution.* Oxford University Press, USA.

Weinzierl, Matthew. 2011. "The Surprising Power of Age-Dependent Taxes." *Review of Economic Studies*, 78(4): 1490–1518.

Werning, Iván. 2009. "Nonlinear capital taxation." *Working paper.*

A Online Appendix

A.1 Proof of Lemma 1

Given any solution $u^*(\theta)$, following a sequence of reports $\left(\theta^{t-1}, \hat{\theta}\right)$, to maximization problem (6) and (11), we can construct

$$\omega\left(\hat{\theta}|\theta\right) = \int_0^\infty u^*\left(\theta^{t-1}, \hat{\theta}, s\right) f_{t+1}(s|\theta)\, ds.$$

We can re-write (5) as

$$\max_{\hat{\theta}} \mathcal{V}\left(\hat{\theta}; \theta\right) \equiv \max_{\hat{\theta}} U\left(c(\hat{\theta}), y(\hat{\theta}); \theta\right) + \beta\omega(\hat{\theta}|\theta).$$

Since $c(\cdot)$ and $\omega(\cdot|\theta)$ are piecewise $\mathcal{C}^1$, they are differentiable except at a finite number of points. Then for all θ where they are differentiable,

$$U_c\left(c(\theta), y(\theta); \theta\right) \dot{c}(\theta) + U_y\left(c(\theta), y(\theta); \theta\right) \dot{y}(\theta) + \beta\omega_1(\theta|\theta) = 0, \qquad (25)$$

where $\dot{c}$ and $\dot{y}$ are derivatives of c and y. Optimality requires that $y(\cdot)$ and $\mathcal{V}(\cdot; \theta)$ are piecewise $\mathcal{C}^1$ and $c(\cdot)$ and $\omega(\cdot|\theta)$ are.

Suppose that the global incentive constraint is violated, i.e. $\mathcal{V}\left(\hat{\theta}; \theta\right) - \mathcal{V}(\theta; \theta) > 0$ for some $\hat{\theta}$. Suppose $\hat{\theta} > \theta$ is a point of differentiability. Then

$$
\begin{aligned}
0 &< \int_\theta^{\hat{\theta}} \frac{\partial \mathcal{V}(x; \theta)}{\partial x} dx \\
&= \int_\theta^{\hat{\theta}} \left[U_c(x; \theta)\dot{c}(x) + U_y(x; \theta)\dot{y}(x) + \beta\frac{d\omega(x|\theta)}{dx} \right] dx.
\end{aligned}
$$

i

Since all of the objects under the integral are piecewise differentiable, it can be represented as a finite sum of the terms

$$\int_{\theta_j}^{\theta_{j+1}} U_c\left(x;\theta\right)\left[\dot{c}\left(x\right)+\dot{y}\left(x\right)\frac{U_y\left(x;\theta\right)}{U_c\left(x;\theta\right)}+\beta\frac{\omega_1\left(x|\theta\right)}{U_c\left(x;\theta\right)}\right]dx$$

for some finite number of intervals $\left(\theta_j,\theta_{j+1}\right)$.

If $x>\theta$, $\frac{U_y(x;\theta)}{U_c(x;\theta)}\leq\frac{U_y(x;x)}{U_c(x;x)}$ and $U_c\left(x;\theta\right)\geq U_c\left(x;x\right)$ (from the single crossing property in Assumption 2 and $U_{cl}\geq0$ in Assumption 3) and $\omega_1\left(x|x\right)\geq\omega_1\left(x|\theta\right)$ from Assumption 3. Therefore

$$\int_{\theta_j}^{\theta_{j+1}} U_c\left(x;\theta\right)\left[\dot{c}\left(x\right)+\dot{y}\left(x\right)\frac{U_y\left(x;\theta\right)}{U_c\left(x;\theta\right)}+\beta\frac{\omega_1\left(x|\theta\right)}{U_c\left(x;\theta\right)}\right]dx$$
$$\leq\int_{\theta_j}^{\theta_{j+1}} U_c\left(x;\theta\right)\left[\dot{c}\left(x\right)+\dot{y}\left(x\right)\frac{U_y\left(x;x\right)}{U_c\left(x;x\right)}+\beta\frac{\omega_1\left(x|x\right)}{U_c\left(x;x\right)}\right]dx$$
$$=\quad 0$$

where the last equality follows from (25). Therefore, $\int_{\theta}^{\hat{\theta}}\frac{\partial\mathcal{V}(x;\theta)}{\partial x}dx\leq0$, a contradiction. If $\hat{\theta}<\theta$ the arguments are analogous. Finally, since $\mathcal{V}\left(\hat{\theta};\theta\right)$ is continuous in $\hat{\theta}$, taking limits establishes that $\mathcal{V}\left(\hat{\theta};\theta\right)\leq\mathcal{V}\left(\theta;\theta\right)$ at the points of non-differentiability.

A.2 Decomposition in equation (17)

We omit explicit time subscripts t whenever it does not lead to confusion. The Hamiltonian to problem (6) and (11) is

$$
\begin{aligned}
H \; = \; & \left(c - \theta l + R^{-1} V_{t+1} \left(w, w_2, \theta \right) \right) f_t + \psi \left[-U_l(c, l) \frac{l}{\theta} + \beta w_2 \right] \\
& - \lambda_1 \bar{\alpha}_t u \left(\theta \right) f_t + \lambda_2 u \left(\theta \right) f_{2,t} + \varphi \left[u - U(c, l) - \beta w \right],
\end{aligned}
$$

where $f_{2,t} = 0$ if $t = 0$. The envelope conditions are

$$
\frac{\partial V_t}{\partial \hat{w}} = \lambda_1, \frac{\partial V_t}{\partial \hat{w}_2} = -\lambda_2. \tag{26}
$$

The first-order conditions are

$$
\varphi - \lambda_1 \bar{\alpha}_t f + \lambda_2 f_2 \; = \; -\dot{\psi} \tag{27}
$$

$$
-U_l \varphi - \theta f \; = \; -\frac{1}{\theta} \psi \left[\frac{U_{ll} l + U_l}{U_l} \right] (-U_l) \tag{28}
$$

$$
f - \psi U_{cl} \frac{l}{\theta} \; = \; \varphi U_c \tag{29}
$$

$$
\frac{1}{R} \frac{\partial V_{t+1}}{\partial w} f \; = \; \varphi \beta \tag{30}
$$

$$
\frac{1}{R} \frac{\partial V_{t+1}}{\partial w_2} f \; = \; -\psi \beta
$$

Use (29) to substitute away for φ

$$\frac{1}{U_c}f - \lambda_1\bar{\alpha}_t f + \lambda_2 f_2 - \frac{\psi}{\theta}\frac{U_{cl}l}{U_c} = -\dot{\psi} \tag{31}$$

$$-\frac{U_l}{U_c}f - \psi\frac{U_{cl}l}{U_c}\frac{(-U_l)}{\theta} - \theta f = -\frac{1}{\theta}\psi\left[\frac{U_{ll}l + U_l}{U_l}\right](-U_l) \tag{32}$$

$$\frac{1}{\beta R}\frac{\partial V_{t+1}}{\partial w} = \frac{1}{U_c} - \frac{\psi}{\theta f}\frac{U_{cl}l}{U_c} \tag{33}$$

$$\frac{1}{\beta R}\frac{\partial V_{t+1}}{\partial w_2} = -\frac{\psi}{f} \tag{34}$$

Use definitions of ε, γ to write (32) as

$$\left(\frac{U_l}{\theta U_c} + 1\right)\theta f = \frac{1}{\theta}\psi\left(1 + \varepsilon - \gamma\right)(-U_l).$$

Since $\tau^y = 1 + \frac{U_l}{\theta U_c}$ this can be equivalently written as

$$\frac{\tau^y}{1 - \tau^y} = \frac{\psi U_c}{\theta f}\left(1 + \varepsilon - \gamma\right). \tag{35}$$

This expression together with (34) implies

$$\lambda_{2,t+1} = -\frac{\partial V_{t+1}}{\partial w_2} = \beta R\frac{\tau_t^y(\theta)}{1 - \tau_t^y(\theta)}\frac{\theta}{U_{c,t}(\theta)}\left(1 + \varepsilon_t(\theta) - \gamma_t(\theta)\right)^{-1}. \tag{36}$$

To find ψ we integrate (31)

$$\psi(\theta) = \int_\theta^\infty \exp\left(-\int_\theta^x \gamma(\tilde{x})\frac{d\tilde{x}}{\tilde{x}}\right)\left(\frac{1}{U_c(x)}f(x) - \lambda_1\bar{\alpha}_t(x)f(x) + \lambda_2 f_2(x)\right)dx.$$

iv

From boundary condition $\psi(0) = 0$ we get

$$\lambda_{1,t} = \frac{\int_0^\infty \exp\left(-\int_0^x \gamma_t(\tilde{x}) \frac{d\tilde{x}}{\tilde{x}}\right) \left(\frac{1}{U_{c,t}(x)} f_t(x) + \lambda_{2,t} f_{2,t}(x)\right) dx}{\int_0^\infty \exp\left(-\int_0^x \gamma_t(\tilde{x}) \frac{d\tilde{x}}{\tilde{x}}\right) \bar{\alpha}_t(x) f_t(x) dx} \tag{37}$$

and $\lambda_{2,t}$ is given by (36). If U is separable, then $\gamma = 0$ and from our assumption on Pareto weights that implies that $\int_0^\infty \bar{\alpha}_t(x) f_t(x) dx = 1$ for all t, we get $\lambda_{1,t} = \int_0^\infty \frac{f_t(x) dx}{U_{c,t}(x)}$ for all t.

Use the expression for $\psi(\theta)$ and (36) for $t-1$ to substitute into (35):

$$\frac{\tau_t^y(\theta)}{1 - \tau_t^y(\theta)}$$

$$= (1 + \varepsilon_t(\theta) - \gamma_t(\theta)) \frac{1}{\theta_t f_t(\theta)} \int_\theta^\infty \frac{U_{c,t}(\theta_t)}{U_{c,t}(x)} \exp\left(-\int_\theta^x \gamma_t(\tilde{x}) \frac{d\tilde{x}}{\tilde{x}}\right) (1 - \lambda_{1,t} \bar{\alpha}_t U_{c,t}(x)) f_t(x) dx$$

$$+ \beta R \frac{\tau_{t-1}^y}{1 - \tau_{t-1}^y} \frac{1 + \varepsilon_t(\theta) - \gamma_t(\theta)}{1 + \varepsilon_{t-1} - \gamma_{t-1}} \frac{U_{c,t}(\theta)}{U_{c,t-1}} \frac{\theta_{t-1}}{\theta f_t(\theta)} \int_\theta^\infty \exp\left(-\int_\theta^x \gamma_t(\tilde{x}) \frac{d\tilde{x}}{\tilde{x}}\right) f_{2,t}(x) dx.$$

Finally note that

$$\frac{U_{c,t}(\theta_t)}{U_{c,t}(x)} \exp\left(-\int_\theta^x \gamma_t(\tilde{x}) \frac{d\tilde{x}}{\tilde{x}}\right) = \exp\left(\ln \frac{U_{c,t}(\theta_t)}{U_{c,t}(x)} - \int_\theta^x \gamma_t(\tilde{x}) \frac{d\tilde{x}}{\tilde{x}}\right) \tag{38}$$

$$= \exp\left(-\int_\theta^x \frac{dU_{c,t}(\tilde{x})}{U_{c,t}(\tilde{x})} - \int_\theta^x \gamma_t(\tilde{x}) \frac{d\tilde{x}}{\tilde{x}}\right)$$

$$= \exp\left(\int_\theta^x \left(\sigma_t(\tilde{x}) \frac{\dot{c}_t(\tilde{x})}{c_t(\tilde{x})} - \gamma_t(\tilde{x}) \frac{\dot{y}_t(\tilde{x})}{y_t(\tilde{x})}\right) d\tilde{x}\right)$$

which is the same expression as (17) in the general, non-separable case.

A.3 Proofs of Proposition 1, Corollary 1, equation (22)

A.3.1 Preliminary results

We first prove some preliminary results about the speed of convergence of $c_t(\theta), y_t(\theta)$, and $l_t(\theta)$, provided that limits exist, distortions remain finite, and elasticities are bounded. These arguments are the same for both separable and non-separable preferences, so we present them for the general case.

Let U be a utility function that satisfies Assumption 2, let σ, ε be as defined in (15) and $\gamma \equiv \frac{U_{cl}l}{U_c}$. Preferences are separable if $\gamma = 0$ for all (c, l). Preferences are GHH if $U(c, l) = \frac{1}{1-\nu}\left(c - \frac{1}{1+1/\zeta}l^{1+1/\zeta}\right)^{1-\nu}$ for $\zeta, \nu > 0$.

We use notation $x_t(\theta)$ to represent the optimal value of variable $x_t\left(\theta^{t-1}, \theta\right)$ for a given θ^{t-1}. We make the following assumption.

Assumption 6. $\varepsilon_t(\theta), \sigma_t(\theta), \gamma_t(\theta), \frac{c_t(\theta)}{y_t(\theta)}$ have finite, non-zero limits; $\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta)}$ has a finite limit; $\frac{\dot{c}_t(\theta)/c_t(\theta)}{\dot{y}_t(\theta)/y_t(\theta)}$ has a limit as $\theta \to \infty$. $\alpha(\theta)$ is bounded and $\alpha(\theta), \varepsilon_t(\theta), U_{c,t}(\theta)$ have finite limits as $\theta \to 0$.

Note in particular that when preferences are separable, then Assumption 4 implies Assumption 6. Let $\sigma_t(\theta) \to \bar{\sigma}, \gamma_t(\theta) \to \bar{\gamma}, \varepsilon_t(\theta) \to \bar{\varepsilon}, \tau_t(\theta)/(1-\tau_t(\theta)) \to \bar{\tau}/(1-\bar{\tau})$ for some $\bar{\sigma}, \bar{\gamma}, \bar{\varepsilon}, \bar{\tau}/(1-\bar{\tau})$, and let $X_t(\theta) \equiv \frac{c_t(\theta)}{(1-\tau_t(\theta))\theta l_t(\theta)} \to \bar{X}$ as $\theta \to \infty$. If Assumption 6 is satisfied, these limits are well defined, finite and, with the exception of $\bar{\gamma}$ and $\bar{\tau}/(1-\bar{\tau})$, are non-zero.

Lemma 2. *Suppose that Assumption 6 is satisfied. If $\lim_{\theta \to \infty} \theta \dot{l}_t/l_t$ and $\lim_{\theta \to \infty} \theta \dot{c}_t/c_t$*

are finite, then

$$\lim_{\theta\to\infty}\frac{\dot{l}_t}{l_t}\theta = \frac{1-\bar{\sigma}+\bar{\gamma}\bar{X}}{\bar{\sigma}+\bar{\varepsilon}-\bar{\gamma}\left(\bar{X}+1\right)}, \quad \lim_{\theta\to\infty}\frac{\dot{y}_t}{y_t}\theta = \lim_{\theta\to\infty}\frac{\dot{c}_t}{c_t}\theta = \frac{1+\bar{\varepsilon}-\bar{\gamma}}{\bar{\sigma}+\bar{\varepsilon}-\bar{\gamma}\left(\bar{X}+1\right)}.$$
(39)

If U is separable or GHH, then these limits exist and finite. In separable case, $\bar{\sigma},\bar{\varepsilon},\bar{\gamma}$ generically depend only on U : $\bar{\gamma}=0$, $\bar{\sigma}=\lim_{c\to\infty}\frac{-U_{cc}c}{U_c}$, $\bar{\varepsilon}=\lim_{l\to\infty}\frac{U_{ll}l}{U_l}$ if $\bar{\sigma}<1$, $\bar{\varepsilon}=\lim_{l\to 0}\frac{U_{ll}l}{U_l}$ if $\bar{\sigma}>1$. In GHH case, $\lim_{\theta\to\infty}\frac{\dot{l}_t}{l_t}\theta=\zeta$, $\lim_{\theta\to\infty}\frac{\dot{y}_t}{y_t}\theta = \lim_{\theta\to\infty}\frac{\dot{c}_t}{c_t}\theta = 1+\zeta$.

Proof. Since $\frac{\dot{c}_t}{\dot{y}_t}=\frac{\dot{c}_t/c_t}{\dot{y}_t/y_t}\frac{c_t}{y_t}$ and the limit of the right hand side exists as $\theta\to\infty$, $\lim_{\theta\to\infty}\frac{\dot{c}_t}{\dot{y}_t}$ exists. We must have $c_t\left(\theta\right),y_t\left(\theta\right)\to\infty$ as $\theta\to\infty$, otherwise $1-\tau_t^y\left(\theta\right)=\frac{-U_{l,t}(\theta)}{\theta U_{c,t}(\theta)}\to 0$, contradicting the assumption that $\lim_{\theta\to\infty}\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta)}<\infty$. Therefore the L'Hospital's rule implies

$$\lim_{\theta\to\infty}\frac{c_t\left(\theta\right)}{y_t\left(\theta\right)}=\lim_{\theta\to\infty}\frac{\dot{c}_t\left(\theta\right)}{\dot{y}_t\left(\theta\right)}=\lim_{\theta\to\infty}\frac{\dot{c}_t\left(\theta\right)/c_t\left(\theta\right)}{\dot{y}_t\left(\theta\right)/y_t\left(\theta\right)}\frac{c_t\left(\theta\right)}{y_t\left(\theta\right)}$$

or

$$1=\lim_{\theta\to\infty}\frac{\frac{\dot{c}_t}{c_t}\theta}{\frac{\dot{y}_t}{y_t}\theta}=\lim_{\theta\to\infty}\frac{\frac{\dot{c}_t}{c_t}\theta}{1+\frac{\dot{l}_t}{l_t}\theta}.$$
(40)

Since $\bar{\tau}<1$, applying L'Hospital's rule,

$$1=\frac{\lim_{\theta\to\infty}\frac{U_{l,t}(\theta)}{\theta U_{c,t}(\theta)}}{-\left(1-\bar{\tau}\right)}=\lim_{\theta\to\infty}\frac{\varepsilon_t\left(\theta\right)\frac{\dot{l}_t}{l_t}\theta-\gamma_t\left(\theta\right)X_t\left(\theta\right)\frac{\dot{c}_t}{c_t}\theta}{1-\sigma_t\left(\theta\right)\frac{\dot{c}_t}{c_t}\theta+\gamma_t\left(\theta\right)\frac{\dot{l}_t}{l_t}\theta}.$$
(41)

When $\lim_{\theta\to\infty}\theta\dot{l}_t/l_t$, $\lim_{\theta\to\infty}\theta\dot{c}_t/c_t$ are finite, we can use (40) and (41) to get (39).

We verify that $\lim_{\theta\to\infty}\theta\dot{l}_t/l_t$, $\lim_{\theta\to\infty}\theta\dot{c}_t/c_t$ are finite when preferences are

separable or GHH. If either limit is infinite, then $\lim_{\theta \to \infty} \left(\theta \frac{\dot{c}_t}{c_t} \right) / \left(\theta \frac{\dot{l}_t}{l_t} \right) = 1$ by (40). Suppose $\lim_{\theta \to \infty} \left| \theta \dot{l}_t / l_t \right| = \infty$. Consider GHH preferences first, in which case (41) is $1 = \frac{1}{\zeta} \lim_{\theta \to \infty} \theta \dot{l}_t / l_t = \pm \infty$, a contradiction. With separable preferences $\gamma_t (\theta) = 0$ for all θ and (41) implies that $1 = -\bar{\varepsilon} / \bar{\sigma} < 0$, a contradiction.

Since $\lim_{\theta \to \infty} c_t (\theta) = \infty$, if preferences are separable then $\bar{\sigma} = -\lim_{c \to \infty} U_{cc} c / U_c$. By (41), this implies that $\lim_{\theta \to \infty} l_t (\theta) = \infty$ if $\bar{\sigma} < 1$ and $\lim_{\theta \to 0} l_t (\theta) = 0$ if $\bar{\sigma} > 1$. This justifies the definition of $\bar{\varepsilon}$. Note that if $\bar{\sigma} = 1$ then $\lim_{\theta \to \infty} \theta \dot{l}_t / l_t = 0$ and $\lim_{\theta \to \infty} \theta \dot{c}_t / c_t = 1$.

Finally, note that with GHH preferences (40) simplifies to $\lim_{\theta \to \infty} \theta \dot{l}_t / l_t = \zeta$, $\lim_{\theta \to \infty} \theta \dot{c}_t / c_t = 1 + \zeta$. $\qquad \square$

Lemma 3. *Suppose that Assumption 6 is satisfied. Then* $c_t (\theta) = o \left(\theta^{\hat{k}} \right)$ *($\theta \to \infty$) for any* $\hat{k} > \frac{1 + \bar{\varepsilon} - \bar{\gamma}}{\bar{\sigma} + \bar{\varepsilon} - \bar{\gamma} (\bar{X} + 1)}$ *and there exists* $\kappa > 0$ *such that* $U_{c,t} = o \left(\theta^{-\kappa} \right)$ *($\theta \to \infty$). If preferences are separable, this holds for any* $\kappa < \frac{(1 + \bar{\varepsilon}) \bar{\sigma}}{\bar{\sigma} + \bar{\varepsilon}}$.

Proof. We first show that for any $\hat{k} > \frac{1 + \bar{\varepsilon} - \bar{\gamma}}{\bar{\sigma} + \bar{\varepsilon} - \bar{\gamma} (\bar{X} + 1)}$ there exist $\hat{K}, \hat{\theta}$ such that $c_t (\theta) \leq \hat{K} \theta^{\hat{k}}$ for all $\theta \geq \hat{\theta}$. By Lemma 2 for any $\hat{k} > \frac{1 + \bar{\varepsilon} - \bar{\gamma}}{\bar{\sigma} + \bar{\varepsilon} - \bar{\gamma} (\bar{X} + 1)}$ we can pick $\hat{\theta}$ such that $\theta \dot{c}_t / c_t < \hat{k}$ for all $\theta \geq \hat{\theta}$. Let $\hat{K} = c_t \left(\hat{\theta} \right) / \hat{\theta}^{\hat{k}}$. Consider a function $G (\theta) \equiv \hat{K} \theta^{\hat{k}} - c_t (\theta)$, which is continuous for $\theta \geq \hat{\theta}$ with $G \left(\hat{\theta} \right) = 0$. For any $\theta > \hat{\theta}$ we have

$$G (\theta) = \int_{\hat{\theta}}^{\theta} G' (x) \, dx = \int_{\hat{\theta}}^{\theta} \left[\hat{K} \hat{k} x^{\hat{k}} - \frac{\dot{c}_t (x) x}{c_t (x)} c_t (x) \right] \frac{dx}{x}.$$

If $G (\theta) = 0$ for some $\theta \geq \hat{\theta}$, then $G' (\theta) = \left[\hat{K} \hat{k} \theta^{\hat{k}} - \frac{\dot{c}_t (\theta) \theta}{c_t (\theta)} c_t (\theta) \right] \frac{1}{\theta} > \frac{1}{\theta} \frac{\dot{c}_t (\theta) \theta}{c_t (\theta)} G (\theta) = 0$. Since $G \left(\hat{\theta} \right) = 0$, this implies that for all $\theta \geq \hat{\theta}$, $G (\theta)$ never crosses zero

from above and is weakly positive. This establishes that $c_t(\theta) = O\left(\theta^{\hat{k}}\right)$.

Since $c_t(\theta) = O\left(\theta^k\right)$ for any $k \in \left(\frac{1+\bar{\varepsilon}-\bar{\gamma}}{\bar{\sigma}+\bar{\varepsilon}-\bar{\gamma}(\bar{X}+1)}, \hat{k}\right)$ and $\theta^k = o\left(\theta^{\hat{k}}\right)$ it also implies that $c_t(\theta) = o\left(\theta^{\hat{k}}\right)$.

If preferences are separable, we can use the same arguments to show that $U_c(c) = o\left(c^{-\tilde{k}}\right)$ for any $\tilde{k} < \bar{\sigma}$. We then define $\kappa = \tilde{k}\hat{k}$ to show that $U_{c,t} = o\left(\theta^{-\kappa}\right)$ for any $\kappa < \frac{(1+\bar{\varepsilon})\bar{\sigma}}{\bar{\sigma}+\bar{\varepsilon}}$. For all other preferences

$$U_c(c_t(\theta), l_t(\theta)) - U_c\left(c_t\left(\hat{\theta}\right), l_t\left(\hat{\theta}\right)\right) = \int_{\hat{\theta}}^{\theta} \left[-\sigma_t(x)\frac{\dot{c}_t(x)\,x}{c_t(x)} + \gamma_t(x)\frac{\dot{l}_t(x)\,x}{l_t(x)}\right]\frac{U_{c,t}(x)\,dx}{x}$$

and the bounds are established analogously to the bounds for $c_t(\theta)$. □

Lemma 4. *Suppose that Assumptions 1, 2 and 6 are satisfied and* $\lim_{\theta\to\infty} D_t(\theta) = 0$. *Then* $C_t(\theta) \geq 0$ *for sufficiently large* θ *and*

$$\lim_{\theta\to\infty} C_t(\theta) = 1 + \frac{\bar{\tau}}{1-\bar{\tau}}\frac{\bar{\sigma}-\bar{\gamma}}{\bar{\sigma}+\bar{\varepsilon}-\bar{\gamma}(\bar{X}+1)}. \tag{42}$$

Assumption 6 is satisfied only if equation

$$\frac{\bar{\tau}}{1-\bar{\tau}} = (1+\bar{\varepsilon})\left(1 + \frac{\bar{\tau}}{1-\bar{\tau}}\frac{\bar{\sigma}-\bar{\gamma}}{\bar{\sigma}+\bar{\varepsilon}-\bar{\gamma}(\bar{X}+1)}\right)\lim_{\theta\to\infty}\frac{1-F_t(\theta)}{\theta f_t(\theta)} \tag{43}$$

holds for a non-negative $\frac{\bar{\tau}}{1-\bar{\tau}}$.

Proof. Let $g_t(\tilde{x}) \equiv \left[\sigma_t(\tilde{x})\frac{\dot{c}_t}{c_t}\tilde{x} - \gamma_t(\tilde{x})\left(\frac{\dot{l}_t}{l_t}\tilde{x}+1\right)\right]$ and re-write C_t as

$$C_t(\theta) = \frac{\exp\left(-\int_0^{\theta}\frac{g_t(\tilde{x})}{\tilde{x}}d\tilde{x}\right)\int_{\theta}^{\infty}\exp\left(\int_0^x\frac{g_t(\tilde{x})}{\tilde{x}}d\tilde{x}\right)(1-\lambda_{1,t}\bar{\alpha}_t(x)U_{c,t}(x))f_t(x)\,dx}{1-F_t(\theta)}. \tag{44}$$

ix

Since $\frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta)}$, $A_t(\theta)$, $B_t(\theta)$ and $D_t(\theta)$ all tend to finite limits as $\theta \to \infty$ by Assumptions 1 and 6, equation (17) implies that the limit of $C_t(\theta)$ also exists and is finite. Since $U_{c,t}(\theta) \to 0$ $(\theta \to \infty)$ from Lemma 3 and $\bar{\alpha}_t(\theta)$ is bounded, $C_t(\theta)$ is positive for sufficiently high θ.

Apply L'Hospital's rule and substitute for $C_t(\theta)$ from (17)

$$
\begin{aligned}
\lim_{\theta \to \infty} C_t(\theta) &= \lim_{\theta \to \infty} \frac{-(1-\lambda_{1,t}\bar{\alpha}_t(\theta)U_{c,t}(\theta))f_t(\theta)}{-f_t(\theta)} + \lim_{\theta \to \infty} \frac{-g_t(\theta)C_t(\theta)(1-F_t(\theta))}{-\theta f_t(\theta)} \qquad (45)\\
&= 1 + \lim_{\theta \to \infty} g_t(\theta) \left\{ \left[\frac{\tau_t(\theta)}{1-\tau_t(\theta)} - \beta R \frac{\tau_{t-1}}{1-\tau_{t-1}} D_t(\theta) \right] \frac{1}{A_t(\theta)} \right\}\\
&= 1 + \lim_{\theta \to \infty} \left\{ \left[\sigma_t(\theta)\frac{\dot{c}_t}{c_t}\theta - \gamma_t(\theta)\frac{\dot{y}_t}{y_t}\theta \right] \left[\frac{\tau_t(\theta)}{1-\tau_t(\theta)} - \beta R \frac{\tau_{t-1}}{1-\tau_{t-1}} D_t(\theta) \right] \frac{1}{A_t(\theta)} \right\}.
\end{aligned}
$$

Equation (42) follows from substituting (39) and $\lim_{\theta \to \infty} D_t(\theta) = 0$ into the expression above.

Since U satisfies Assumption 2, $A_t(\theta) \geq 0$ for all θ and therefore $\lim_{\theta \to \infty} \frac{\tau_t^y(\theta)}{1-\tau_t^y(\theta)} = \lim_{\theta \to \infty} A_t(\theta) B_t(\theta) C_t(\theta) \geq 0$. Therefore equation (43) should be satisfied for a non-negative $\frac{\bar{\tau}}{1-\bar{\tau}}$. $\qquad \square$

A.3.2 Proof of Proposition 1

Proof. We first show that there are real k_1, k_2 such that $A_t(\theta) B_t(\theta) C_t(\theta) \sim k_1 \frac{F_t(\theta)}{\theta f_t(\theta)}$, $D_t(\theta) \sim k_2 \varphi_t(\theta)$ $(\theta \to 0)$. Use (38) to write C_t as

$$
C_t(\theta) = \int_\theta^\infty \frac{U_{c,t}(\theta)}{U_{c,t}(x)} (1-\lambda_{1,t}\bar{\alpha}_t(x)U_{c,t}(x)) \frac{f_t(x)\,dx}{1-F_t(\theta)}.
$$

x

Note that since $U_{c,t}(0)$ is well-defined and finite by Assumption 4,

$$\lim_{\theta \to 0} C_t(\theta) = U_{c,t}(0) \int_0^\infty \left(\frac{1}{U_{c,t}(x)} - \lambda_{1,t} \bar{\alpha}_t(x) \right) f_t(x) \, dx$$

$$= U_{c,t}(0) \left[\int_0^\infty \frac{1}{U_{c,t}(x)} f_t(x) \, dx - \lambda_{1,t} \right] = 0$$

from the definition of $\lambda_{1,t}$ and the fact that $\int_0^\infty \bar{\alpha}_t(x) f_t(x) \, dx = 1$ for all t. Applying L'Hospital's rule,

$$\lim_{\theta \to 0} \frac{C_t(\theta) \frac{1 - F_t(\theta)}{U_{c,t}(\theta)}}{F_t(\theta)} = -\left(\frac{1}{U_{c,t}(0)} - \lambda_{1,t} \bar{\alpha}_t(0) \right),$$

since limits of $U_{c,t}(\theta)$ and $\bar{\alpha}_t(\theta)$ are well defined. Let $k_1 = -(1 + \varepsilon_t(0))(1 - \lambda_{1,t}\bar{\alpha}_t(0) U_{c,t}(0))$, which is well-defined by Assumption 4. We have

$$\lim_{\theta \to 0} \frac{A_t(\theta) B_t(\theta) C_t(\theta)}{k_1 F_t(\theta) / \theta f_t(\theta)} = \lim_{\theta \to 0} \frac{1}{k_1} A_t(\theta) U_{c,t}(\theta) \frac{C_t(\theta) \frac{1 - F_t(\theta)}{U_{c,t}(\theta)}}{F_t(\theta)} = 1.$$

The result for $D_t(\theta)$ follows immediately by setting $k_2 = \frac{A_t(0)}{A_{t-1}} \frac{U_{c,t}(0)}{U_{c,t-1}}$, which is well-defined by Assumption 4.

We next show that $D_t(\theta) = o\left(\frac{1}{\theta^{k_4}}\right)$ $(\theta \to \infty)$ and $k_4 > 0$ generically depends only on U. Since $\varphi_t(\theta)$ is bounded by Assumption 1 and $A_t(\theta)$ is bounded for θ sufficiently high by Assumption 4, $|D_t(\theta)| \leq K_{t-1} U_{c,t}(\theta)$ for some $K_{t-1} > 0$. Lemma 3 yields the result.

Finally we show that $A_t(\theta) B_t(\theta) C_t(\theta) \sim k_3 \frac{1 - F_t(\theta)}{\theta f_t(\theta)}$ as $\theta \to \infty$ and $k_3 > 0$

depends generically on U and f. Using Lemma 4,

$$\frac{\bar{\tau}}{1-\bar{\tau}} = \lim_{\theta\to\infty} A_t(\theta) B_t(\theta) C_t(\theta) = (1+\bar{\varepsilon})\left(1 + \frac{\bar{\tau}}{1-\bar{\tau}}\frac{\bar{\sigma}}{\bar{\sigma}+\bar{\varepsilon}}\right)\lim_{\theta\to\infty}\frac{1-F_t(\theta)}{\theta f_t(\theta)}.$$
(46)

If $\lim_{\theta\to\infty}\frac{1-F_t(\theta)}{\theta f_t(\theta)} = 0$, then $A_t(\theta) B_t(\theta) C_t(\theta) \sim (1+\bar{\varepsilon})\frac{1-F_t(\theta)}{\theta f_t(\theta)}$ $(\theta \to \infty)$. If $\lim_{\theta\to\infty}\frac{1-F_t(\theta)}{\theta f_t(\theta)} > 0$, then (46) defines $\frac{\bar{\tau}}{1-\bar{\tau}}$ as a function of $f_t, \bar{\varepsilon}, \bar{\sigma}$. Then setting $k_3 = (1+\bar{\varepsilon})\left(1 + \frac{\bar{\tau}}{1-\bar{\tau}}\frac{\bar{\sigma}}{\bar{\sigma}+\bar{\varepsilon}}\right)$ we obtain the result for $\theta \to \infty$. Note that $\bar{\varepsilon}, \bar{\sigma}$ generically depend only on U by Lemma 2. $\qquad\square$

A.3.3 Proof of Corollary 1

We first prove a preliminary lemma about the properties of f.

Lemma 5. *Suppose f_t satisfies Assumption 5. Then*

i $\varphi_t(\theta) = \rho$ *for all θ; If $\rho \geq 0$ then there is $\hat{\theta}$ such that $f_{2,t}(\theta) \geq 0$ for all $\theta \geq \hat{\theta}$.*

ii $\frac{F_t(\theta)}{\theta f_t} \sim \frac{f_t}{\theta f_t'} \sim \frac{v^2}{-\ln\theta}$ $(\theta \to 0)$;

iii *If f_t is lognormal/mixture, then $\frac{1-F_t(\theta)}{\theta f_t} \sim \frac{f_t}{\theta f_t'} \sim \frac{v^2}{\ln\theta}$ $(\theta \to \infty)$; if f_t is Pareto-lognormal then $\lim_{\theta\to\infty}\frac{1-F_t(\theta)}{\theta f_t} = \frac{1}{a}$ and $\lim_{\theta\to\infty}\frac{f_t}{\theta f_t'} = -\frac{1}{a+1}$.*

Proof. Let $\Phi(\cdot), \phi(\cdot)$ be standard normal cdf and pdf. Direct calculations yield

$$\lim_{x\to\infty}\frac{\phi(x)}{\Phi(x)} = 0, \quad \lim_{x\to-\infty}\frac{\phi(x)}{\Phi(x)} = \infty, \quad \lim_{x\to-\infty}\frac{\phi(x)}{-x\Phi(x)} = 1.$$
(47)

When f_t is lognormal, it is given by $f_t(\theta) = \frac{1}{\theta v}\phi\left(\frac{\ln\theta - \hat{\mu}_t}{v}\right)$ where $\hat{\mu}_t = b_t + \rho\ln\theta_{t-1}$; when f_t is a mixture then $f_t(\theta) = \sum_{i=1}^{I}\frac{p_i}{\theta v_i}\phi\left(\frac{\ln\theta - \hat{\mu}_{i,t}}{v_i}\right)$ where $\hat{\mu}_t =$

xii

$b_t + \mu_i + \rho \ln \theta_{t-1}$; when f_t is Pareto-lognormal then (see Colombi (1990), Reed and Jorgensen (2004)) $f_t(\theta) = \hat{A}_t \theta^{-a-1} \Phi\left(\frac{\ln \theta - \hat{\mu}_t - av^2}{v}\right)$ where $\hat{A}_t = \exp\left(a\hat{\mu}_t + a^2v^2/2\right)$, $\hat{\mu}_t = \rho \ln \theta_{t-1} + b_t - \frac{1}{a}$, in which case $\mathbb{E}\left[\ln \theta | \ln \theta_{t-1}\right] = \rho \ln \theta_{t-1} + b_t$.

(i). Suppose f_t is lognormal. Then $\theta_{t-1} f_{2,t} = -\frac{\rho}{v^2\theta}\phi' = -\frac{\rho}{v}\frac{\partial\phi\left(\frac{\ln\theta-\hat{\mu}_t}{v}\right)}{\partial\theta}$. Therefore $\varphi_t(\theta) = \frac{-\rho\int_\theta^\infty \partial\phi\left(\frac{\ln x-\hat{\mu}_t}{v}\right)/\partial\theta\,dx}{\phi\left(\frac{\ln\theta-\hat{\mu}_t}{v}\right)} = \rho$. The same argument applies for the mixtures of lognormal. If f_t is Pareto-lognormal, then

$$\theta_{t-1}f_{2,t}(\theta) = -\hat{A}_t\frac{\rho}{v}\theta^{-a-1}\phi\left(\frac{\ln\theta-\hat{\mu}_t-av^2}{v}\right) + a\rho f_t(\theta).$$

Note that using integration by parts

$$
\begin{aligned}
\int_\theta^\infty \hat{A}_t\frac{\rho}{v}x^{-a-1}\phi\left(\frac{\ln x-\hat{\mu}_t-av^2}{v}\right)dx &= \rho a\int_\theta^\infty \hat{A}_t x^{-a-1}\Phi\left(\frac{\ln x-\hat{\mu}_t-av^2}{v}\right)dx \\
&\quad -\rho\hat{A}_t\theta^{-a}\Phi\left(\frac{\ln\theta-\hat{\mu}_t-av^2}{v}\right) \\
&= a\rho\left(1-F_t(\theta)\right) - \rho\theta f_t(\theta).
\end{aligned}
$$

Therefore

$$\int_\theta^\infty \theta_{t-1}f_{2,t}(x)\,dx = -a\rho\left(1-F_t(\theta)\right) + \rho\theta f_t(\theta) + a\rho\left(1-F_t(\theta)\right) = \rho\theta f_t(\theta)$$

and hence $\varphi_t(\theta) = \rho$ for all θ. The second part of (i) follows by inspection of expressions for $f_{2,t}(\theta)$.

(ii) and (iii). Suppose f_t is lognormal. Then $\theta f'_t = -f_t\left(1 + \frac{\ln\theta-\hat{\mu}_t}{v^2}\right)$, and

therefore $\theta f_t'/f_t \sim -\frac{\ln\theta}{v^2}$ $(\theta \to 0, \infty)$. By L'Hospital's rule,

$$\lim_{\theta\to 0} \frac{F_t(\theta)}{\theta f_t(\theta)} = \lim_{\theta\to 0} \frac{1}{\theta f_t'(\theta)/f_t(\theta)+1} = 0,$$

$$\lim_{\theta\to\infty} \frac{1-F_t(\theta)}{\theta f_t(\theta)} = \lim_{\theta\to\infty} \frac{-1}{\theta f_t'(\theta)/f_t(\theta)+1} = 0,$$

and

$$\lim_{\theta\to 0} \frac{-F_t\ln\theta/v^2}{\theta f_t} = \lim_{\theta\to 0} \frac{-\ln\theta/v^2 - F_t/(\theta f_t v^2)}{(\theta f_t'/f_t+1)} = 1,$$

$$\lim_{\theta\to\infty} \frac{(1-F_t)\ln\theta/v^2}{\theta f_t} = \lim_{\theta\to\infty} \frac{-\ln\theta/v^2 + (1-F_t)/(\theta f_t v^2)}{(\theta f_t'/f_t+1)} = 1.$$

This implies that $F_t(\theta)/\theta f_t \sim \frac{v^2}{-\ln\theta}$ $(\theta\to 0)$ and $(1-F_t(\theta))/\theta f_t \sim \frac{v^2}{\ln\theta}$ $(\theta\to\infty)$.

If f_t is a mixture, assume without loss of generality that $v_1 \geq v_i$ for all i. Then

$$\frac{\theta f_t'}{f_t} = -\left(\frac{\ln\theta - \hat\mu_{1,t}}{v_1^2}+1\right)\frac{\frac{p_1}{v_1} + \sum_{i=2}^{I}\frac{p_i}{v_i}\frac{\phi\left(\frac{\ln\theta-\hat\mu_{i,t}}{v_i}\right)}{\phi\left(\frac{\ln\theta-\hat\mu_{1,t}}{v_1}\right)}\frac{\frac{\ln\theta-\hat\mu_{i,t}}{v_i^2}+1}{\frac{\ln\theta-\hat\mu_{1,t}}{v_1^2}+1}}{\frac{p_1}{v_1} + \sum_{i=2}^{I}\frac{p_i}{v_i}\frac{\phi\left(\frac{\ln\theta-\hat\mu_{i,t}}{v_i}\right)}{\phi\left(\frac{\ln\theta-\hat\mu_{1,t}}{v_1}\right)}}.$$

Since $v_1 \geq v_i$, $\phi\left(\frac{\ln\theta-\hat\mu_{i,t}}{v_i}\right)\Big/\phi\left(\frac{\ln\theta-\hat\mu_{1,t}}{v_1}\right) \to 0$ as $\ln\theta \to \pm\infty$ and therefore the last term in the expression above converges to 0 as $\ln\theta \to \pm\infty$. This implies that $\theta f_t'/f_t \sim -\ln\theta/v_1^2$ $(\theta\to 0,\infty)$. The rest follows by analogy with the lognormal case.

If f_t is Pareto-lognormal, then $\theta f_t' = (-a-1)f_t + f_t\frac{1}{v}\phi\left(\frac{\ln\theta-\hat\mu}{v}\right)\Big/\Phi\left(\frac{\ln\theta-\hat\mu}{v}\right)$,

which immediately implies that $\theta f_t' / f_t \rightarrow -(a+1) \, (\theta \rightarrow \infty)$. Also

$$\lim_{\theta \rightarrow 0} \frac{\theta f_t'}{f_t} \frac{\upsilon}{-\ln \theta} = \lim_{\theta \rightarrow 0} \frac{\phi \left(\frac{\ln \theta - \hat{\mu}}{\upsilon} \right)}{-\ln \theta \Phi \left(\frac{\ln \theta - \hat{\mu}}{\upsilon} \right)}$$

From (47), $\phi \left(\frac{\ln \theta - \hat{\mu}}{\upsilon} \right) / \Phi \left(\frac{\ln \theta - \hat{\mu}}{\upsilon} \right) \sim -\frac{\ln \theta - \mu}{\upsilon} \left(\frac{\ln \theta - \mu}{\upsilon} \rightarrow -\infty \right)$, therefore $\theta f_t' / f_t \sim$ $-\frac{\ln \theta - \mu}{\upsilon^2} \sim -\frac{\ln \theta}{\upsilon^2} \, (\theta \rightarrow 0)$. The rest follows by analogy with the lognormal case. $\qquad \square$

With this lemma we can prove Corollary 1.

Proof (of Corollary 1). If f_t satisfies assumption 5, then Lemma 5 and Proposition 1 show that $A_t(\theta) \, B_t(\theta) \, C_t(\theta) \sim k_1 \frac{\upsilon^2}{-\ln \theta} \, (\theta \rightarrow 0)$, $D_t(0) = \rho \frac{U_{c,t}(0)}{U_{c,t-1}} > 0$. This establishes (20). They also establish that $\lim_{\theta \rightarrow \infty} D_t(\theta) = 0$. Therefore from Lemma 4 it follows that $\lim_{\theta \rightarrow \infty} C_t(\theta) = 1 + \frac{\bar{\sigma}}{\bar{\sigma} + \bar{\varepsilon}} \frac{\bar{\tau}}{1 - \bar{\tau}}$ and the expressions for $\bar{\sigma}$ and $\bar{\varepsilon}$ in terms of limits of $\frac{U_{cc}c}{U_c}$ and $\frac{U_{ll}l}{U_l}$ follows from Lemma 2. This establishes (18).

Finally, to show (19) we first suppose that f_t is Pareto-lognormal. Then from Lemma 5 $\lim_{\theta \rightarrow \infty} B_t(\theta) = a^{-1}$, and taking limits of (17) yields

$$\frac{\bar{\tau}}{1 - \bar{\tau}} = \frac{1 + \bar{\varepsilon}}{a} \left[1 + \frac{\bar{\tau}}{1 - \bar{\tau}} \frac{\bar{\sigma}}{\bar{\sigma} + \bar{\varepsilon}} \right].$$

Re-arranging the terms, we obtain $\frac{\bar{\tau}}{1 - \bar{\tau}} = \left[a \frac{1}{1 + \bar{\varepsilon}} - \frac{\bar{\sigma}}{\bar{\sigma} + \bar{\varepsilon}} \right]^{-1}$. By Lemma 4 this limit must be non-negative, therefore a necessary condition for the distortions to be finite is that $a \frac{1}{1 + \bar{\varepsilon}} - \frac{\bar{\sigma}}{\bar{\sigma} + \bar{\varepsilon}} > 0$.

If f_t is lognormal/mixture, then Lemma 5 and Proposition 1 imply that $B_t(\theta) \sim \left[\frac{\ln \theta}{\upsilon^2} \right]^{-1}$, $D_t(\theta) = o \left(\left[\frac{\ln \theta}{\upsilon^2} \right]^{-1} \right) \, (\theta \rightarrow \infty)$, the latter follows from the

fact that $\lim_{\theta \to \infty} \theta^{-\kappa} \ln \theta = 0$ for any $\kappa > 0$. Since $\lim_{\theta \to \infty} A_t (\theta) = 1 + \bar{\varepsilon}$ and $C_t (\theta)$ are bounded, this implies that $\lim_{\theta \to \infty} \frac{\tau_t^y (\theta)}{1 - \tau_t^y (\theta)} = 1$ and therefore $\lim_{\theta \to \infty} C_t (\theta) = 1$. Therefore $\frac{\tau_t^y (\theta)}{1 - \tau_t^y (\theta)} \sim A_t (\theta) B_t (\theta) C_t (\theta) \sim \left[\frac{\ln \theta}{v^2} \frac{1}{1 + \bar{\varepsilon}} \right]^{-1}$.

$\square$

A.3.4 Proofs in Section 2.2

We first show equation (22).

Lemma 6. *Suppose that Assumptions 5 and 6 are satisfied, $\rho \geq 0$, $U_{cl} \geq 0$. Then (22) holds.*

Proof. We first show that $D_t (\theta) = o \left(\theta^{-\kappa} \right)$ $(\theta \to \infty)$ where κ as defined in Lemma 3. If f_t satisfies Assumption 5 and $\rho \geq 0$, then by Lemma 5 there exists $\hat{\theta}$ such that $f_{2,t} (\theta) \geq 0$ for all $\theta \geq \hat{\theta}$. Therefore if $U_{cl} \geq 0$ then $\gamma \geq 0$ and $\exp \left(- \int_\theta^x \gamma (\tilde{x}) \frac{d\tilde{x}}{\tilde{x}} \right) f_{2,t} (x) \leq f_{2,t} (x)$ for all x, θ such that $x \geq \theta \geq \hat{\theta}$. Using Lemma 5,

$$\frac{\theta_{t-1} \int_\theta^\infty \exp \left(- \int_\theta^x \gamma (\tilde{x}) \frac{d\tilde{x}}{\tilde{x}} \right) f_{2,t} (x) \, dx}{\theta f_t (\theta)} \leq \rho \text{ for all } \theta \geq \hat{\theta}.$$

Therefore $D_t (\theta) \leq K_{t-1} U_{c,t} (\theta)$ for some K_{t-1} and then Lemma 3 yields the result that $D_t (\theta) = o \left(\theta^{-\kappa} \right)$. Since $\lim_{\theta \to \infty} D_t (\theta) = 0$, Lemma 4 implies that the limit $\bar{\tau}$ satisfies (43). The rest of the steps are identical to the proof of Corollary 1.

$\square$

We now show the remaining results discussed in Section 2.2.

Lemma 7. *Suppose that Assumption 6 is satisfies. Then*

$$\frac{\beta w_{1,t}\left(\theta\right)\theta}{U_{c,t}\left(\theta\right)c_{t}\left(\theta\right)} = \frac{\left(1 - \tau_{t}^{y}\left(\theta\right)\right)\dot{y}_{t} - \dot{c}_{t}}{c_{t}}\theta. \tag{48}$$

In the limit

$$\lim_{\theta\to\infty}\frac{\beta w_{1,t}\left(\theta\right)\theta}{U_{c,t}\left(\theta\right)c_{t}\left(\theta\right)} = \frac{1 - \bar{X}}{\bar{X}}\lim_{\theta\to\infty}\frac{\dot{y}_{t}}{y_{t}}\theta. \tag{49}$$

Proof. Differentiating (10), we get $\dot{u}_{t}\left(\theta\right) = U_{c,t}\left(\theta\right)\dot{c}_{t}\left(\theta\right) + U_{l,t}\left(\theta\right)\dot{l}_{t}\left(\theta\right) + \beta\left(w_{1,t}\left(\theta\right) + w_{2,t}\left(\theta\right)\right).$ Substitute into (7) to get

$$U_{c,t}\left(\theta\right)\dot{c}_{t}\left(\theta\right) + U_{l,t}\left(\theta\right)\dot{l}_{t}\left(\theta\right) + \beta w_{1,t}\left(\theta\right) = -U_{l,t}\left(\theta\right)\frac{l_{t}\left(\theta\right)}{\theta}. \tag{50}$$

Re-arrange to get (48). Note that $\theta\dot{y}_{t}/y_{t} = 1 + \theta\dot{l}_{t}/l_{t}$. Then use (39) to obtain the limit. □

Compensated and uncompensated elasticities holding savings fixed coincide with compensated and uncompensated elasticities in the static model, where they are given by (see p. 227 in Saez (2001))

$$\begin{aligned}
\zeta^{u} &= \frac{U_{l}/l - \left(U_{l}/U_{c}\right)^{2}U_{cc} + \left(U_{l}/U_{c}\right)U_{cl}}{U_{ll} + \left(U_{l}/U_{c}\right)^{2}U_{cc} - 2\left(U_{l}/U_{c}\right)U_{cl}}, \\
\zeta^{c} &= \frac{U_{l}/l}{U_{ll} + \left(U_{l}/U_{c}\right)^{2}U_{cc} - 2\left(U_{l}/U_{c}\right)U_{cl}}, \\
-\eta &= \zeta^{c} - \zeta^{u}.
\end{aligned}$$

Note that normality of leisure implies $\eta < 0$. We use $\zeta_{t}^{u}\left(\theta\right), \zeta_{t}^{c}\left(\theta\right), \eta_{t}\left(\theta\right)$ to denote the elasticities evaluated at the optimum and $\bar{\zeta}^{u}, \bar{\zeta}^{c}, \bar{\eta}$ their limits as $\theta\to\infty$.

Lemma 8. $A_t(\theta)$ and $C_t(\theta)$ can be written as

$$A_t(\theta) = \frac{1 + \zeta_t^u(\theta)}{\zeta_t^c(\theta)},$$

$$C_t(\theta) = \frac{\int_\theta^\infty \exp\left(\left\{\int_\theta^x \frac{-\eta_t(\tilde{x})}{\zeta_t^c(\tilde{x})}\frac{\dot{y}_t}{y_t} + \sigma_t(\tilde{x})\frac{\left(1-\tau_t^y(\tilde{x})\right)\dot{y}_t - \dot{c}_t}{c_t}\right\}\tilde{x}d\tilde{x}\right)\left(1 - \lambda_{1,t}\bar{\alpha}_t U_{c,t}(x)\right)f_t(x)\,dx}{1 - F_t(\theta)}.$$

If preferences are GHH, then (24) holds.

Proof. The proof for $A_t(\theta)$ follows from the definition of elasticities. To rewrite $C_t(\theta)$ let $g_t(\cdot)$ be as defined in the proof of Lemma 4. Using (48) it can be written as

$$g_t(\theta) = -\sigma_t(\theta)\frac{\left(1 - \tau_t^y(\theta)\right)\dot{y}_t - \dot{c}_t}{c_t}\theta - \frac{\eta_t(\theta)}{\zeta_t^c(\theta)}\frac{\dot{y}_t}{y_t}\theta.$$

Substitute into (44) to get the expression for C_t. When preferences are GHH, $\eta_t(\theta) = 0$ and $\lim_{\theta\to\infty}\frac{\dot{y}_t}{y_t}\theta = 1 + \zeta$ by Lemma 2. Use this fact together with Lemma 7 to show that $g_t(\theta) \to \frac{1-\bar{X}}{\bar{X}}(1+\zeta)$ ($\theta \to \infty$) in this case. Since $A_t(\theta) = (1+\zeta)/\zeta$, this together with (45) implies (24). $\square$

Proof (of Proposition 2). We can express φ using (30) rather than (29), in which case the differential equation for ψ, (31), becomes $\frac{1}{\beta R}\frac{\partial V_{t+1}}{\partial w}f - \lambda_1 f + \lambda_2 f_2 = -\dot{\psi}$. Integrate this expression from 0 to infinity, use the boundary conditions $\psi(0) = \psi(\infty) = 0$ and $\int_0^\infty f_2 dx = 0$ to obtain[19]

$$\lambda_{1,t} = \frac{1}{\beta R}\int_0^\infty \frac{\partial V_{t+1}}{\partial w}(x)f(x)\,dx.$$

---

[19] To see that $\int_0^\infty f_2(x|\theta_-)\,dx = 0$ for all θ_-, differentiate both sides of $\int_0^\infty f(x|\theta_-)\,dx = 1$ with respect to θ_-

Combine this expression with (26) to get $\frac{\partial V_t}{\partial w} = \frac{1}{\beta R} \mathbb{E}_t \frac{\partial V_{t+1}}{\partial w}$ and by the law of iterated expectations

$$\frac{\partial V_t}{\partial w} = \left(\frac{1}{\beta R}\right)^{T-t-1} \mathbb{E}_t \frac{\partial V_T}{\partial w}. \tag{51}$$

When $F_T(0|\theta) = 1$ for all θ, $\frac{\partial V_T}{\partial w}(\theta^T) = \frac{1}{U_c(\theta^T)}$, which, from (51), implies

$$\lambda_{1,t} = \frac{\partial V_t}{\partial w} > 0 \text{ for all } t \tag{52}$$

and, in combination with (33),

$$\frac{1}{U_c(\theta^t)} - \frac{\psi}{\theta f}\gamma = \left(\frac{1}{\beta R}\right)^{T-t} \mathbb{E}_t \frac{1}{U_c(\theta^T)}.$$

Note that $\frac{\partial}{\partial \theta} \frac{U_y(c,y/\theta)}{U_c(c,y/\theta)} \geq 0$ from Assumption 2 implies that $1 + \varepsilon - \gamma \geq 0$, therefore from equation (35) the sign of ψ is equal to the sign of τ^y. Thus if $\tau^y \geq 0$ then

$$\frac{1}{U_c(\theta^t)} \geq \left(\frac{1}{\beta R}\right)^{T-t} \mathbb{E}_t \frac{1}{U_c(\theta^T)} \geq \left(\frac{1}{\beta R}\right)^{T-t} \frac{1}{\mathbb{E}_t U_c(\theta^T)},$$

where the last expression follows from Jensen's inequality. This expression implies that $\tilde{\tau}_t^s(\theta^t) \geq 0$. This inequality is strict if $var_{\theta^t}(c_T) > 0$. $\qquad \square$

Lemma 9. *Suppose that preferences are* $U\left(c - \frac{1}{1+1/\varsigma}l^{1+1/\varsigma}\right)$*, where* U *is concave,* U''/U' *is bounded away from zero,* f_t *satisfies Assumption 5 with* $\rho \geq 0$ *and* $F_T(0|\theta) = 1$ *for all* θ*. If* $\tau_t^y(\theta)$ *is positive and bounded away from 1 for high* θ*, then* $\tau_t^y(\theta) \to 0$ *as* $\theta \to \infty$*. A sufficient condition for* $\tau_t^y(\theta)$ *to be*

bounded is that U is exponential: $U(x) = -\exp\left(-\hat{k}x\right)$ for some $\hat{k} > 0$.

Proof. The first order conditions (29) and (30) can be written as

$$\frac{1}{U_c} - \frac{\psi}{\theta f}\frac{U_{cl}l}{U_c} = \frac{1}{\beta R}\frac{\partial V_{t+1}}{\partial w}.$$

From (35) we have

$$\frac{\psi}{\theta f} = \frac{\tau^y}{1 - \tau^y}\left(1 + \varepsilon - \gamma\right)^{-1}\frac{1}{U_c}$$

which implies

$$\frac{1}{\beta R}\frac{\partial V_{t+1}(w)}{\partial w}U_c = 1 - \frac{\tau^y}{1 - \tau^y}\left(1 + 1/\zeta\right)^{-1}\frac{U_{cl}l}{U_c} = 1 - \left(1 + 1/\zeta\right)^{-1}\frac{-U''}{U'}\tau^y\theta l$$

(53)

Since $F_T(0|\theta) = 1$, both sides of this expression must be positive by (52).

Suppose that τ^y does not converge to 1. Take any sequence $\tau^y(\theta_n)$ and since $\tau^y(\theta_n) \in [0, 1]$ it must have a convergent subsequence. We will show that any such subsequence that does not converge to 1 must converge to 0. Suppose $\tau^y(\theta_n) \to \bar{\tau}^y < 1$. Then the FOCs $l^{1/\zeta} = \theta(1 - \tau^y)$ implies that $l \to \infty$ ($\theta \to \infty$) and, since $\frac{-U''}{U'}$ is bounded away from 0, the right hand side of (53) converges to $-\infty$. The left hand side is positive, leading to a contradiction.

Under our assumptions $\tau^y(\theta)$ diverges to 1 only if either $C_t(\theta)$ or $D_t(\theta)$ diverge to $+\infty$. Either of these cases would imply that $U_t(\theta) \to -\infty$ as $\theta \to \infty$. If U is exponential, it is bounded above in all periods, and therefore $U_t(\theta) \to -\infty$ in any t would violate the incentive compatibility. In particular, to see

that $C_t(\theta) \to \infty$ implies $U_t(\theta) \to -\infty$, note that with exponential U there is some $\hat{k} > 0$ so that

$$
\begin{aligned}
C_t(\theta) &= \int_\theta^\infty \exp\left(\int_\theta^x \beta \frac{U_t''(\tilde{x})}{U_t'(\tilde{x})} \dot{c}_t(\tilde{x}) \, d\tilde{x}\right) (1 - \lambda_{1,t}\bar{\alpha}_t(x) U_t'(x)) \frac{f_t(x)\, dx}{1 - F_t(\theta)} \\
&= \int_\theta^\infty \exp\left(\beta\hat{k}(c_t(\theta) - c_t(x))\right) (1 - \lambda_{1,t}\bar{\alpha}_t(x) U_t'(x)) \frac{f_t(x)\, dx}{1 - F_t(\theta)}.
\end{aligned}
$$

Since $\lambda_{1,t} > 0$ by (52), $1 - \lambda_{1,t}\bar{\alpha}_t(x) U_t'(x)$ is bounded from above and therefore $C_t(\theta)$ can diverge to infinity only if the exponent diverges to infinity, which is possible only if $c_t(x) \to -\infty$ and therefore $U_t(x) \to -\infty$. $\qquad\square$

A.4 Additional details for Section 3

We first describe further details of the analysis in Section 3 and then provide additional illustrations and robustness checks.

To make the numerical solution feasible we exploit the recursive structure of the dual formulation of the planning problem that we discussed in Section 1. The recursive problem is (6) together with (11) and $V_0(\hat{w}_0) = 0$, which is a finite-horizon discrete-time dynamic programming problem with a three-dimensional continuous state vector: $\hat{w}$ is the promised utility associated with the promise-keeping constraint (8); $\hat{w}_2$ is the state variable associated with the threat-keeping constraint (9); θ_- is the type in the preceding period. In the initial period the state is $\hat{w}_0$, given by the solution to $V_0(\hat{w}_0) = 0$.

We proceed in stages. First, we implement a value function iteration for problems (6) and (11). We start from the last working period, $\hat{T} - 1$, and proceed by backward induction. Since $F_t(0|\theta) = 1$ for all θ for $t \geq \hat{T}$, the
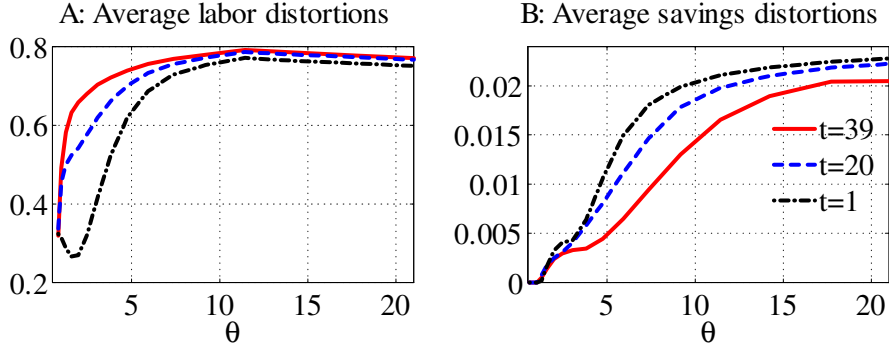
Figure 6: Optimal average labor (Panel A) and savings (Panel B) distortions as functions of current shock realization at selected periods.

planner sets $w_2(\theta) = 0$ for all θ in period $\hat{T} - 1$ and we replace the value function $V_{\hat{T}}(w(\theta), 0, \theta)$ in problem (6) for period $\hat{T} - 1$ with the discounted present value of resources required to provide promised utility w over the remaining $T - \hat{T} + 1$ periods.

We approximate value functions with tensor products of orthogonal polynomials evaluated over the state space. We use Chebyshev polynomials of degrees 1 through 10 and check in the baseline case that value function differences do not exceed 1 percent of original values after doubling the degrees to 20. The evaluation nodes are allocated over the state space at the roots of the polynomials, given by $r_n = -\cos(\pi(2n-1)/2N)$, where $n = 1, .., N$ indexes the nodes. This gives the roots on the interval $[-1, 1]$ and a change of variables is needed to adjust the root nodes. We let $N = 11$ for both the promise, $\hat{w}$, and for the threat, $\hat{w}_2$. For the skill, we set 30 logarithmically spaced nodes to better capture the more complex U-shapes in the left tail. The polynomial coefficients are computed by minimizing the sum of squared distances from the computed values at the nodes. The approximation provides each period-t

Table 2: Simulated earnings and consumption moments of the constrained optima and the earnings moments in the data.

Stochastic process						Initial distribution		
Mean	SD	Kurtosis	Kelly's Skewness	P10	P90	P50	P90	P99
Data earnings moments (y_t^{data}):								
0.009	0.52	11.31	-0.21	-0.44	0.47	10.06	10.76	11.71
Lognormal constrained-optimum earnings moments $(y_t^{\log normal})$:								
-0.005	1.09	2.51	-0.03	-1.02	0.90	13.46	13.79	14.48
Mixture constrained-optimum earnings moments $(y_t^{mixture})$:								
0.004	0.79	11.48	-0.37	-0.90	0.51	13.03	13.73	14.29
Lognormal constrained-optimum consumption moments $(c_t^{\log normal})$:								
0.001	0.75	3.95	-0.18	-0.13	1.07	10.52	11.92	12.66
Mixture constrained-optimum consumption moments $(c_t^{mixture})$:								
-0.001	0.13	19.07	0.15	-0.33	0.55	11.89	12.07	13.43

problem (6) with a continuously differentiable function approximating V_{t+1}. We use the trigonometric form of the polynomials in the evaluation of the tensor products, $P_d(r) = \cos(d \arccos(r))$, to be able to apply an implementation of algorithmic (chain rule) differentiation.

It is a familiar property of the state space in such problems that no constrained optimal allocations may exist for some nodes (see, e.g., the discussion in subsection 3.2. in Abraham and Pavoni (2008)). To deal with this while maintaining large enough number of computed nodes, we follow the procedure in Kapička (2013) in subsections 7.1 and 7.2.[20]

For computational feasibility it is essential to use an efficient and robust optimization algorithm for the minimization problems at each node. We use an implementation of the interior-point algorithm with conjugate gradient it-

---

[20] Generally one has a choice to implement a state space restriction procedure, to eliminate such nodes, or a procedure assigning sufficiently large penalties. For discussions of both and examples of implementation in closely related problem see, e.g., Abraham and Pavoni (2008) and Kapička (2013) and references therein.
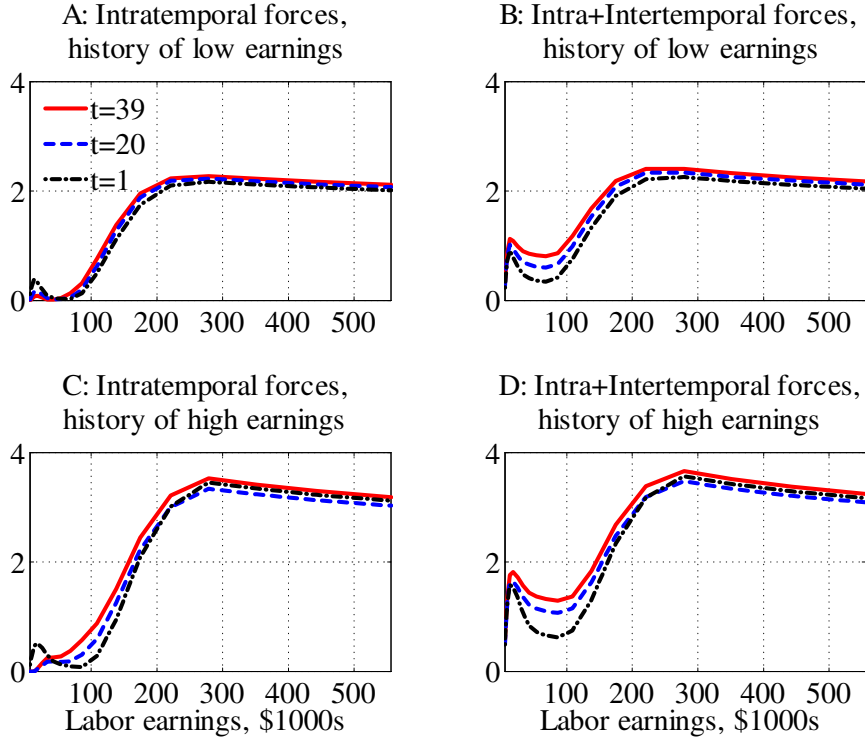
Figure 7: The decomposition of optimal labor distortions as functions of current earnings: only intratemporal forces $(A_t B_t C_t)$ in Panels A and C; both intra- and intertemporal forces $(\tau_t^y/1 - \tau_t^y)$ in Panels B and D. Panels A and B have a history of $\bar{\theta}$ shocks chosen so that an individual with a lifetime of $\bar{\theta}$ shocks will have the average lifetime earnings approximately equal to the average U.S. male earnings in 2005; Panels C and D are the analogues with $\bar{\theta}$ chosen so that the average lifetime earnings approximately equal twice the U.S. average.

eration to compute the optimization step.[21] It uses a trust-region method to solve barrier problems; the acceptance criterion is an l_1 barrier penalty function. To improve the accuracy of the solution estimates, including multipliers, we proceed to active-set iterations that use 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.

---
[21] See, for example, Su and Judd (2007).

Table 3: Calibrated parameters of the shock process for selected Frisch elasticity parameter values.

Stochastic process						Initial distribution		
μ_1	μ_3	υ_1	υ_2	p_1	p_2	μ	υ	a
The higher elasticity case of $\varepsilon = 1$:								
0.03	-0.45	0.23	2.76	0.71	0.14	0.17	5.49	2.74
The baseline case of $\varepsilon = 2$:								
0.03	-0.47	0.22	2.64	0.71	0.15	0.17	5.59	2.73
The lower elasticity case of $\varepsilon = 4$:								
0.02	-0.51	0.20	2.52	0.71	0.16	0.17	5.67	2.69

We check at this stage the increasing properties of Assumption 3 used in Lemma 1 (Assumption 2 is satisfied analytically given the choices of preferences). At each node, we compute relative forward differences in policies $c(\cdot)$, $\omega(\cdot|\theta)$, and $\omega_1\left(\hat{\theta}|\cdot\right)$, i.e. the differences in a policy at θ'' and at $\theta' < \theta''$ relative to the value of the policy at θ'. To verify the nodes with numerical errors (the largest relative error is one one-thousands of 1 percent of the policy at the lower type), we then follow the procedure in subsection 7.2.2 in Kapička (2013) as an additional check of global incentive constraints, which amounts to letting the agent re-optimize with respect to reported type given the policies and verifying that the true type is a solution.[22]

The next stage computes $\hat{w}_0$ such that $V_0(\hat{w}_0) = 0$ using binary search given V_0 computed in the first stage.

In the final stage, we simulate the optimal labor and savings distortions described in Section 3. Given V_t's computed in the first stage and $\hat{w}_0$ solved for in the second stage, we generate optimal allocations by forward induction, starting from policy functions produced by $V_0(\hat{w}_0)$ from (11). Optimal distor-

---

[22] Abraham and Pavoni (2008) in subsection 3.3 and Farhi and Werning (2013) in subsection 2.2.4 describe applications in related settings.

A: Labor distortions, t=1,
history of high earnings

B: Labor distortions, t=20,
history of high earnings

Labor earnings, $1000s

Labor earnings, $1000s

$\varepsilon=4$
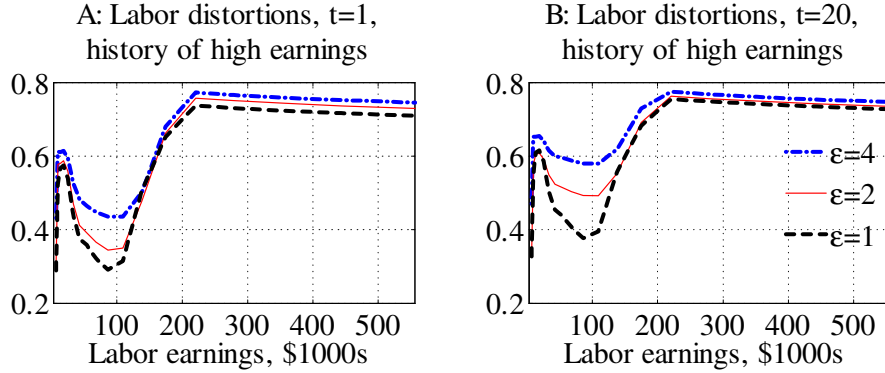
$\varepsilon=2$

$\varepsilon=1$

Figure 8: An illustration of the typical effects on the optimal labor distortions of the changes in the Frisch elasticity parameter.

tions can then be computed from the policy functions using definitions (13) and (14). To compute the average distortions in Section 3 we do 5×10^5 Monte Carlo simulations. As a robustness check, Figure 6 here provides the analogue of Figure 4 in the main text plotted against the shock realizations. In addition, Table 2 summarizes the changes in aggregate earnings and consumption moments in the simulations discussed in the main text.

At this stage we also compute the objects whose limiting behavior is required by Assumption 4: $U_{c,t}(\theta)$, $\frac{c_t(\theta)}{y_t(\theta)}$, and $\frac{\dot{c}_t(\theta)/c_t(\theta)}{\dot{y}_t(\theta)/y_t(\theta)}$. In the Monte Carlo histories we find that these expressions have finite numerical values of the same order of magnitude as the terms in Figure 3 in the main text, both in the left and right tails of the distribution. In a given period, the terms $\frac{c_t(\theta)}{y_t(\theta)}$ and $\frac{\dot{c}_t(\theta)/c_t(\theta)}{\dot{y}_t(\theta)/y_t(\theta)}$ asymptote fairly quickly as $\theta \to \infty$, to virtually constant values at earnings above \$300,000. Relatedly, Figure 7 here further quantifies the intertemporal forces in Figure 3 in the main text. For the history of low earnings, Panel A in Figure 7 isolates intratemporal forces, displaying them
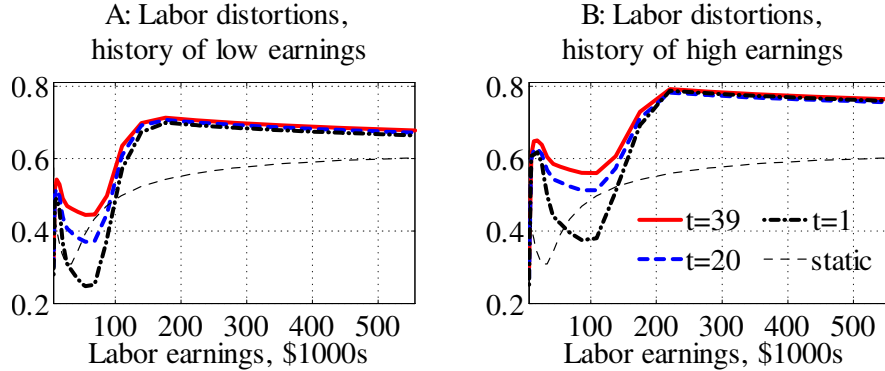
Figure 9: Optimal labor distortions as functions of current earnings at selected periods compared to an experiment with a static model with the distribution of shocks given by F_0. Panel A has a history of $\bar{\theta}$ shocks chosen so that an individual with a lifetime of $\bar{\theta}$ shocks will have the average lifetime earnings approximately equal to the average U.S. male earnings in 2005; Panel B is the analogue with $\bar{\theta}$ chosen so that the average lifetime earnings approximately equal twice the U.S. average.

without the intertemporal terms, and Panel B provides an illustration of the effect of including intertemporal forces; Panels C and D illustrate the same for the history of high earnings.

We provide several further robustness checks and additional illustrations. First, we summarize the robustness checks with respect to a key fundamental, the Frisch elasticity of labor supply. We follow the same procedure we described for the baseline case of parameter $\varepsilon = 2$ in the main text, calibrating the same setup except with $\varepsilon = 4$ and then with $\varepsilon = 1$, which correspond to Frisch elasticities of 0.25 and 1 respectively. Table 3 compares the calibrated parameters for the initial distribution and the stochastic process for the shock in the three cases. The parameters are chosen to match the moments from the data displayed in Table 1 in the main text. In particular, lower Frisch
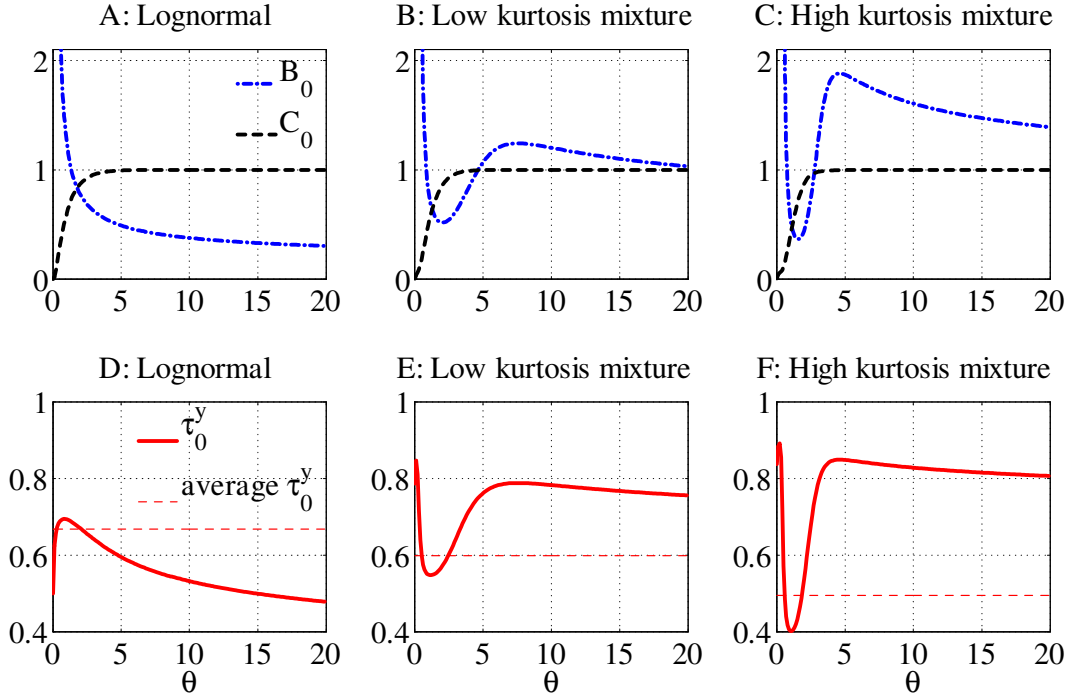
Figure 10: The effects of increasing kurtosis on optimal labor distortions in period 0 and their components with quasi-linear preferences.

elasticities of labor supply (which correspond to higher values of ε) require lower maximum variance in the mixture, but drawn with higher probability to match the same data moments we discussed in the main text, particularly the high kurtosis.

We simulate the optimal distortions in the economies with $\varepsilon = 4$ and with $\varepsilon = 1$ and compare them to the baseline distortions: Figure 8 displays the typical effects, shown here for a representative history of twice the average earnings. Lower elasticities result in generally higher distortions, especially for the left part of the earnings distribution around the U-shapes. The right tail of the distribution displays the same pattern but with smaller differences

because the effects of the higher parameter ε are offset by the effects of the lower maximum variance in the mixture.

Next, to supplement the comparison with the static results of Saez (2001) in the main text, we illustrate here an experiment where a static model is simulated with the shock distribution given by our calibrated initial distribution, F_0. Figure 9 reproduces the labor distortions from our baseline simulation, analyzed in the main text with Figure 2, and compares them to the static distortions in the experiment. It is important to keep in mind, however, that the static model in which shocks are drawn from an initial-period Pareto-lognormal distribution understates the actual cross-sectional dispersion of shocks and leads to lower distortions, as Figure 9 indicates.

Finally, we make here transparent the role of kurtosis explored in the main text and illustrated with Figures 1 and 5. Figure 10 here provides an analogue of Figure 1 where we vary the kurtosis in the mixture distribution. The three distribution examples in Figure 10 illustrate the effects of increasing the level of kurtosis from 3 in the case of normal shocks (reproduced in Panels A and D from Figure 1) to the kurtosis of 6 (Panels B and E) and finally to 12 (Panels C and F). The rest of the parameters are kept unchanged compared to Figure 1.