Taxation under Learning-by-Doing: Supplementary Material

Miltiadis Makris
University of Kent

Alessandro Pavan
Northwestern University

November 2018

Abstract

Section A in this supplement contains the proofs of various results in the main body omitted from the manuscript. Section B contains comparative statics results of optimal tax codes with respect to the agents' degree of risk aversion, the Frisch elasticity of the agents' labor supply, and the planner's preferences for redistribution. Section C describes the computational methods used in the main body and in Section B in the present supplement to establish all the numerical results.

A Omitted Proofs

This section has three parts. Subsection A.1 formally establishes the analytical results about the effects of LBD on the level, dynamics, and progressivity of the wedges in the benchmark economy with Rawlsian preferences for redistribution and risk-neutral agents, as reported in Proposition 2 in the main body. Subsection A.2 contains all the proofs for the results relating wedges to optimal tax rates. Subsection A.3 contains various results establishing the equivalence between the 40-period economy of Section 5 in the main body and the 2-period economy in Section 2 of the main body.

A.1 Wedges under Rawlsian preferences for redistribution and risk-neutral agents

Proof of Proposition 2 in the main body. From the analysis in the main text, we have that, when the disutility of labor is given by

$$\psi(y_t, \theta_t) = \frac{1}{1+\phi}\left(\frac{y_t}{\theta_t}\right)^{1+\phi}$$

and period-2 productivity is given by $\theta_2 = \theta_1^{\rho} y_1^{\zeta} \varepsilon_2$, in the absence of LBD, the period-t relative wedge is given by

$$\hat{W}_t^{NOLBD}(\theta^t) = \rho^{t-1}\frac{1+\phi}{\theta_1 \gamma_1(\theta_1)}, \tag{A.1}$$

where $\gamma_1(\theta_1) \equiv \frac{f_1(\theta_1)}{1-F_1(\theta_1)}$.

Now let $y_1(\theta_1)$ be the unique solution to the following equation

$$\left[1+\hat{W}_1^{NOLBD}(\theta_1)\right]^{-1}\theta_1^{1+\phi}+\delta\zeta\overline{\varepsilon}(\phi)\theta_1^{\frac{(1+\phi)^2}{\phi}}\left[\frac{1+\rho\hat{W}_1^{NOLBD}(\theta_1)}{1+\hat{W}_1^{NOLBD}(\theta_1)}\right]\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]^{-\frac{1+\phi}{\phi}}y_1^{\frac{\zeta(1+\phi)-\phi}{\phi}} = y_1^{\phi} \tag{A.2}$$

where $\overline{\varepsilon}(\phi)\equiv\mathbb{E}\left[\tilde{\varepsilon}_2^{\frac{1+\phi}{\phi}}\right]$. Observe that the assumption that $\zeta\leq\phi/(1+\phi)$ implies that the left-hand-side of (A.2) is strictly decreasing in y_1. In turn, this implies that the unique solution $y_1(\theta_1)$ to (A.2) is nondecreasing in θ_1 whenever $\hat{W}_1^{NOLBD}(\theta_1)$ is nonincreasing.

Then, let

$$\epsilon^{\hat{W}_1^{NOLBD}}(\theta_1)\equiv\frac{d\hat{W}_1^{NOLBD}(\theta_1)}{d\theta_1}\frac{\theta_1}{\hat{W}_1^{NOLBD}(\theta_1)},$$

and

$$\epsilon^{y_1}(\theta_1)\equiv\frac{dy_1(\theta_1)}{d\theta_1}\frac{\theta_1}{y_1(\theta_1)}.$$

The proof proceeds in four steps. Step 1 shows that the period-1 wedge can be expressed as

$$\hat{W}_1(\theta_1)=\hat{W}_1^{NOLBD}(\theta_1)$$
$$+\hat{W}_1^{NOLBD}(\theta_1)\left\{\left[\frac{1+\rho\hat{W}_1^{NOLBD}(\theta_1)}{1+\hat{W}_1^{NOLBD}(\theta_1)}\right]^{-1}\frac{\frac{\rho}{\phi}\theta_1^{\frac{1+\phi}{\phi}}y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}{\frac{1}{\delta\zeta\overline{\varepsilon}(\phi)}\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]^{\frac{1}{\phi}}+\theta_1^{\frac{1+\phi}{\phi}}y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}\right\} \tag{A.3}$$

with $y_1(\theta_1)$ defined by the unique solution to equation (A.2). Given that, at the optimum, $y_1(\theta_1)>0$, it is then immediate that LBD contributes to a higher period-1 wedge for all θ_1 and to a difference between first-period and second-period wedges that is higher than in the absence of LBD (that is, $\hat{W}_1(\theta_1)-\hat{W}_2(\theta)>\hat{W}_1^{NOLBD}(\theta_1)-\hat{W}_2^{NOLBD}(\theta)$ all θ). These properties prove parts (i) and (ii) in the proposition. Step 2, in turn, proves existence of a function $J(\theta_1)$ such that LBD contributes to a higher progressivity of the period-1 wedge if and only $J(\theta_1)\geq 0$, which is always the case when F_1 is Pareto. Step 3 establishes the result in part (iii). Finally, Step 4 establishes part (iv) by showing that, when $\hat{W}_1^{NOLBD}(\theta_1)$ is nonincreasing, the policies (y_1,y_2) that solve the relaxed program are such that $y_2(\theta_1,.)$ is nondecreasing and the integral monotonicity conditions

$$\int_{\hat{\theta}_1}^{\theta_1}\left\{\psi_\theta(y_1(s),s)+\delta\mathbb{E}^{\lambda[\chi]|s,y_1(s)}\left[I_1^2(\tilde{\theta},y_1(s))\psi_\theta(y_2(s,\tilde{\theta}_2),\tilde{\theta}_2)\right]\right\}ds \tag{A.4}$$
$$\leq\int_{\hat{\theta}_1}^{\theta_1}\left\{\psi_\theta(y_1(\hat{\theta}_1),s)+\delta\mathbb{E}^{\lambda[\chi]|s,y_1(\hat{\theta}_1)}\left[I_1^2(\tilde{\theta},y_1(\hat{\theta}_1))\psi_\theta(y_2(\hat{\theta}_1,\tilde{\theta}_2),\tilde{\theta}_2)\right]\right\}ds$$

are satisfied. As explained in the main text, these properties imply that the solution to the relaxed program also solves the full program.

Step 1. Recall from the analysis in the main text that the effects of LBD on the period-1 wedge are summarized by the term

$$\Omega(\theta_1)=\frac{\delta\rho}{\psi_y(y_1(\theta_1),\theta_1)}\hat{W}_1^{NOLBD}(\theta_1)\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1,y_1(\theta_1)}\left[\psi(y_2(\tilde{\theta}),\tilde{\theta}_2)\right]. \tag{A.5}$$

Next, use the first-order condition for period-1 output

$$\psi_y(y_1(\theta_1),\theta_1)-\frac{1}{\gamma_1(\theta_1)}\psi_{\theta y}(y_1(\theta_1),\theta_1)-\delta\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1,y_1(\theta_1)}\left[\frac{I_1^2(\tilde{\theta},y_1(\theta_1))}{\gamma_1(\theta_1)}\psi_\theta(y_2(\tilde{\theta}),\tilde{\theta}_2)\right]=1+LD_1^\chi(\theta_1),$$

established in the main text, along with the fact that

$$-\theta_2\psi_\theta(y_2(\theta),\theta_2)=(1+\phi)\psi(y_2(\theta),\theta_2)=-\theta_1\frac{\psi_{\theta y}(y_1(\theta_1),\theta_1)}{\psi_y(y_1(\theta_1),\theta_1)}\psi(y_2(\theta),\theta_2),$$

to verify that $y_1(\theta_1)$ must solve the first-order-condition

$$1+LD_1^\chi(\theta_1)-\delta\rho\hat{W}_1^{NOLBD}(\theta_1)\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1,y_1(\theta_1)}\left[\psi(y_2(\tilde{\theta}),\tilde{\theta}_2)\right]=\psi_y(y_1(\theta_1),\theta_1)\left[1+\hat{W}_1^{NOLBD}(\theta_1)\right], \tag{A.6}$$

where recall that

$$LD_1^\chi(\theta_1)=\delta\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1,y_1(\theta_1)}\left[y_2(\tilde{\theta})-\psi(y_2(\tilde{\theta}),\tilde{\theta}_2)\right]. \tag{A.7}$$

Next, use the definition of $\hat{W}_1^{NOLBD}(\theta_1)$ to rewrite the first-order condition for period-2 output as

$$1=\psi_y(y_2(\theta),\theta_2)\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right].$$

When ψ is isoelastic, the last condition can be rewritten as

$$y_2(\theta)=(1+\phi)\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]\psi(y_2(\theta),\theta_2). \tag{A.8}$$

Replacing (A.8) into (A.7), we have that

$$LD_1^\chi(\theta_1)=\delta\left\{(1+\phi)\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]-1\right\}\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1,y_1(\theta_1)}\left[\psi(y_2(\tilde{\theta}),\tilde{\theta}_2)\right]$$

and hence that

$$\begin{aligned}&LD_1^\chi(\theta_1)-\delta\rho\hat{W}_1^{NOLBD}(\theta_1)\tfrac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1,y_1(\theta_1)}\left[\psi(y_2(\tilde{\theta}),\tilde{\theta}_2)\right]\\&=\delta\phi\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]\tfrac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1,y_1(\theta_1)}\left[\psi(y_2(\tilde{\theta}),\tilde{\theta}_2)\right].\end{aligned} \tag{A.9}$$

Using (A.6), in turn we have that

$$\frac{1}{\psi_y(y_1(\theta_1),\theta_1)}=\frac{1+\hat{W}_1^{NOLBD}(\theta_1)}{1+\delta\phi\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1,y_1(\theta_1)}\left[\psi(y_2(\tilde{\theta}),\tilde{\theta}_2)\right]}. \tag{A.10}$$

Replacing (A.10) into (A.5), we have that $\Omega_1(\theta_1)$ can be rewritten as

$$\Omega_1(\theta_1)=\frac{\delta\rho\hat{W}_1^{NOLBD}(\theta_1)\left[1+\hat{W}_1^{NOLBD}(\theta_1)\right]\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1,y_1(\theta_1)}\left[\psi\left(y_2(\tilde{\theta}),\tilde{\theta}_2\right)\right]}{1+\delta\phi\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1,y_1(\theta_1)}\left[\psi(y_2(\tilde{\theta}),\tilde{\theta}_2)\right]},$$

or, equivalently,

$$\Omega_1(\theta_1) = \frac{\delta\rho\hat{W}_1^{NOLBD}(\theta_1)\left[1+\hat{W}_1^{NOLBD}(\theta_1)\right]\Lambda^{\chi}(\theta_1, y_1(\theta_1))}{1+\delta\phi\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]\Lambda^{\chi}(\theta_1, y_1(\theta_1))},$$

where we used the shortcut notation

$$\Lambda^{\chi}(\theta_1, y_1(\theta_1)) \equiv \frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)}\left[\psi(y_2(\tilde{\theta}), \tilde{\theta}_2)\right]. \tag{A.11}$$

Next, use (A.8) to observe that, when ψ is isoelastic,

$$y_2(\theta) = \theta_2^{\frac{1+\phi}{\phi}}\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]^{-\frac{1}{\phi}} \tag{A.12}$$

and hence

$$\psi\left(y_2(\theta), \theta_2\right) = \frac{1}{1+\phi}\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]^{-\frac{1+\phi}{\phi}} \times \theta_2^{\frac{1+\phi}{\phi}}. \tag{A.13}$$

It follows that

$$\Lambda^{\chi}(\theta_1, y_1(\theta_1)) = \frac{1}{1+\phi}\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]^{-\frac{1+\phi}{\phi}}\frac{\partial}{\partial y_1}\mathbb{E}\left[\tilde{\theta}_2^{\frac{1+\phi}{\phi}}|\theta_1, y_1(\theta_1)\right].$$

Using $\overline{\varepsilon}(\phi) \equiv \mathbb{E}\left[\tilde{\varepsilon}_2^{\frac{1+\phi}{\phi}}\right]$, we then have that[1]

$$\begin{aligned}
\frac{\partial}{\partial y_1}\left\{\mathbb{E}\left[\tilde{\theta}_2^{\frac{1+\phi}{\phi}}|\theta_1, y_1\right]\right\} &= \frac{1+\phi}{\phi}\mathbb{E}\left[\tilde{\theta}_2^{\frac{1}{\phi}}\left(-\frac{\frac{\partial F_2(\tilde{\theta}_2|\theta_1,y_1)}{\partial y_1}}{f_2(\tilde{\theta}_2|\theta_1, y_1)}\right)|\theta_1, y_1\right] \\
&= \frac{1+\phi}{\phi}\mathbb{E}\left[\tilde{\theta}_2^{\frac{1}{\phi}}\left(\frac{\tilde{\theta}_2\zeta}{y_1}\right)|\theta_1, y_1\right] = \frac{1+\phi}{\phi}\mathbb{E}\left[\tilde{\theta}_2^{\frac{1+\phi}{\phi}}\left(\frac{\zeta}{y_1}\right)|\theta_1, y_1\right] \\
&= \frac{\zeta(1+\phi)}{\phi}\frac{1}{y_1}(\theta_1 y_1^{\zeta})^{\frac{1+\phi}{\phi}}\overline{\varepsilon}(\phi).
\end{aligned}$$

This implies that

$$\Lambda^{\chi}(\theta_1, y_1(\theta_1)) = \left\{\frac{1}{1+\phi}\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]^{-\frac{1+\phi}{\phi}}\frac{\zeta(1+\phi)}{\phi}\overline{\varepsilon}(\phi)\theta_1^{\frac{1+\phi}{\phi}}\right\} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}. \tag{A.14}$$

Replacing the formula for $\Lambda^{\chi}(\theta_1, y_1(\theta_1))$ into the formula for $\Omega_1(\theta_1)$ above, we then have that the latter can be expressed as

$$\Omega_1(\theta_1) = \left[\frac{1+\rho\hat{W}_1^{NOLBD}(\theta_1)}{1+\hat{W}_1^{NOLBD}(\theta_1)}\right]^{-1}\frac{\frac{\rho}{\phi}\hat{W}_1^{NOLBD}(\theta_1)\theta_1^{\frac{1+\phi}{\phi}}y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}{\frac{1}{\delta\zeta\overline{\varepsilon}(\phi)}\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]^{\frac{1}{\phi}}+\theta_1^{\frac{1+\phi}{\phi}}y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}. \tag{A.15}$$

---

[1] Observe that, given any Lipschitz continuous function $J(\theta_2)$, and any kernel $F_2(\theta_2|\theta_1, y_1)$, $\frac{\partial}{\partial y_1}\mathbb{E}\left[J(\tilde{\theta}_2)|\theta_1, y_1\right] = \mathbb{E}\left[\left(-\frac{\frac{\partial F_2(\tilde{\theta}_2|\theta_1,y_1)}{\partial y_1}}{f_2(\tilde{\theta}_2|\theta_1,y_1)}\right)\frac{\partial J(\tilde{\theta}_2)}{\partial\theta_2}|\theta_1, y_1\right]$.

Replacing (A.15) into the formula for the period-1 relative wedge

$$\hat{W}_1(\theta_1) = \hat{W}_1^{NOLBD}(\theta_1) + \Omega(\theta_1)$$

derived in the main text permits us to establish the formula for $\hat{W}_1(\theta_1)$ in (A.3).

We conclude this step by showing that $y_1(\theta_1)$ is implicitly given by equation (A.2). This follows from combining (A.6), (A.9) and (A.11) with (A.14).

Step 2. Differentiating $\Omega_1(\theta_1)$ in (A.15), and simplifying the derivative using the fact that, at the optimum, $y_1(\theta_1) > 0$, we have that $\Omega_1(\theta_1)$ is increasing in θ_1 if and only if the following function

$$J(\theta_1) \equiv$$

$$\left[\frac{\epsilon^{\hat{W}_1^{NOLBD}}(\theta_1)\hat{W}_1^{NOLBD}(\theta_1)}{\theta_1} \frac{(1-\rho)}{\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]^2} \frac{\frac{\rho}{\phi}\hat{W}_1^{NOLBD}(\theta_1)\theta_1^{\frac{1+\phi}{\phi}} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}{\frac{1}{\delta\zeta\overline{\varepsilon}(\phi)}\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]^{\frac{1}{\phi}} + \theta_1^{\frac{1+\phi}{\phi}} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}} \right] \times$$

$$\left[\frac{\left[\frac{1+\rho\hat{W}_1^{NOLBD}(\theta_1)}{1+\hat{W}_1^{NOLBD}(\theta_1)}\right]^{-1} \frac{\hat{W}_1^{NOLBD}(\theta_1)}{\theta_1} \frac{\rho}{\phi}\theta_1^{\frac{1+\phi}{\phi}} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}{\left[\frac{1}{\delta\zeta\overline{\varepsilon}(\phi)}\left[1+\rho\hat{W}_1^{NOLBD}(\theta_1)\right]^{\frac{1}{\phi}} + \theta_1^{\frac{1+\phi}{\phi}} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}\right]^2} \right]^{-1} +$$

$$\epsilon^{\hat{W}_1^{NOLBD}}(\theta_1) \left\{ \frac{1}{\delta\zeta\overline{\varepsilon}(\phi)} \left[1 + \rho\hat{W}_1^{NOLBD}(\theta_1)\right]^{\frac{1}{\phi}} \left[1 - \frac{\rho\hat{W}_1^{NOLBD}(\theta_1)}{\phi(1+\rho\hat{W}_1^{NOLBD}(\theta_1))})\right] + \theta_1^{\frac{1+\phi}{\phi}} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}} \right\} +$$

$$\left[\frac{1+\phi}{\phi} + \frac{\zeta(1+\phi)-\phi}{\phi}\epsilon^{y_1}(\theta_1)\right] \frac{1}{\delta\zeta\overline{\varepsilon}(\phi)} \left[1 + \rho\hat{W}_1^{NOLBD}(\theta_1)\right]^{\frac{1}{\phi}}$$

is non-negative.

Step 3. Now observe that, when F_1 is Pareto, $\gamma_1(\theta_1)\theta_1 = M$, in which case

$$\hat{W}_1^{NOLBD}(\theta_1) = \frac{1+\phi}{M} \text{ all } \theta,$$

and equation (A.2) reduces to

$$P_1(\theta_1) + P_2(\theta_1)y_1^{\frac{\zeta(1+\phi)-\phi}{\phi}} - y_1^{\phi} = 0, \tag{A.16}$$

where

$$P_1(\theta_1) \quad \equiv \quad \left[1 + \frac{1+\phi}{M}\right]^{-1} \theta_1^{1+\phi}$$

and

$$P_2(\theta_1) \equiv \delta\zeta\overline{\varepsilon}(\phi)\theta_1^{\frac{(1+\phi)^2}{\phi}} \left[\frac{1+\rho\frac{1+\phi}{M}}{1+\frac{1+\phi}{M}}\right] \left[1 + \rho\frac{1+\phi}{M}\right]^{-\frac{1+\phi}{\phi}}.$$

Furthermore, in this case,

$$\delta\zeta J(\theta_1) = \frac{1}{\overline{\varepsilon}(\phi)} \left[1 + \rho\frac{1+\phi}{M}\right]^{\frac{1}{\phi}} \frac{1+\phi}{\phi} + \frac{\zeta(1+\phi)-\phi}{\phi} \frac{1}{\overline{\varepsilon}(\phi)} \left[1 + \rho\frac{1+\phi}{M}\right]^{\frac{1}{\phi}} \epsilon^{y_1}(\theta_1).$$

Now use (A.16) to obtain that

$$\frac{dy_1(\theta_1)}{d\theta_1} = -\frac{\frac{dP_1(\theta_1)}{d\theta_1} + \frac{dP_2(\theta_1)}{d\theta_1} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}{\frac{\zeta(1+\phi)-\phi}{\phi} P_2(\theta_1) y_1(\theta_1)^{\frac{\zeta(1+\phi)-2\phi}{\phi}} - \phi y_1(\theta_1)^{\phi-1}}. \tag{A.17}$$

Using the fact that, for all θ_1, $y_1(\theta_1) > 0$, we then have that

$$\frac{dy_1(\theta_1)}{d\theta_1} = -\frac{\frac{dP_1(\theta_1)}{d\theta_1} y_1(\theta_1) + \frac{dP_2(\theta_1)}{d\theta_1} y_1(\theta_1)^{\frac{\zeta(1+\phi)}{\phi}}}{\frac{\zeta(1+\phi)-\phi}{\phi} P_2(\theta_1) y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}} - \phi y_1(\theta_1)^{\phi}}. \tag{A.18}$$

Substituting

$$y_1(\theta_1)^{\phi} = P_1(\theta_1) + P_2(\theta_1) y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}$$

into (A.18), we then have that

$$\frac{dy_1(\theta_1)}{d\theta_1} = -\frac{\frac{dP_1(\theta_1)}{d\theta_1} y_1(\theta_1) + \frac{dP_2(\theta_1)}{d\theta_1} y_1(\theta_1)^{\frac{\zeta(1+\phi)}{\phi}}}{\frac{\zeta(1+\phi)-\phi}{\phi} P_2(\theta_1) y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}} - \phi \left[P_1(\theta_1) + P_2(\theta_1) y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}\right]}.$$

Rearranging, we have that

$$\frac{dy_1(\theta_1)}{d\theta_1} = -\frac{\frac{dP_1(\theta_1)}{d\theta_1} y_1(\theta_1) + \frac{dP_2(\theta_1)}{d\theta_1} y_1(\theta_1)^{\frac{\zeta(1+\phi)}{\phi}}}{\frac{(\zeta-\phi)(1+\phi)}{\phi} P_2(\theta_1) y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}} - \phi P_1(\theta_1)}. \tag{A.19}$$

Now note that

$$\begin{aligned}
\frac{dP_1(\theta_1)}{d\theta_1} &= (1+\phi)\frac{P_1(\theta_1)}{\theta_1}, \\
P_2(\theta_1) &= \delta\zeta\bar{\varepsilon}(\phi)\theta_1^{\frac{(1+\phi)}{\phi}} \left[1 + \rho\frac{1+\phi}{M}\right]^{-\frac{1}{\phi}} P_1(\theta_1), \\
\frac{dP_2(\theta_1)}{d\theta_1} &= \frac{(1+\phi)^2}{\phi}\frac{P_2(\theta_1)}{\theta_1}.
\end{aligned}$$

Replacing these functions into (A.19), and letting $n(\theta_1) \equiv \delta\zeta\bar{\varepsilon}(\phi)\theta_1^{\frac{(1+\phi)}{\phi}} \left[1 + \rho\frac{1+\phi}{M}\right]^{-\frac{1}{\phi}}$, we then have that

$$\epsilon^{y_1}(\theta_1) = -\frac{1 + \phi + \frac{(1+\phi)^2}{\phi} n(\theta_1) y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}{\frac{(\zeta-\phi)(1+\phi)}{\phi} n(\theta_1) y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}} - \phi}. \tag{A.20}$$

It follows that

$$\delta\zeta J(\theta_1) = \frac{1}{\bar{\varepsilon}(\phi)}\left[1 + \rho\frac{1+\phi}{M}\right]^{\frac{1}{\phi}} \frac{1+\phi}{\phi}\left\{1 + \left(\frac{\phi}{1+\phi} - \zeta\right)\left[\frac{1 + \phi + \frac{(1+\phi)^2}{\phi} n(\theta_1) y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}{\frac{(\zeta-\phi)(1+\phi)}{\phi} n(\theta_1) y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}} - \phi}\right]\right\}.$$

Hence $J(\theta_1) > 0$ if and only if

$$1 + \left(\frac{\phi}{1+\phi} - \zeta\right)\left[\frac{1 + \phi + \frac{(1+\phi)^2}{\phi} n(\theta_1) y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}{\frac{(\zeta-\phi)(1+\phi)}{\phi} n(\theta_1) y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}} - \phi}\right] > 0. \tag{A.21}$$

Now fix θ_1 and observe that the left-hand side of (A.21) is nondecreasing in $y_1(\theta_1)$. A sufficient condition for $J(\theta_1) > 0$ is thus that the inequality in (A.21) holds when $y_1(\theta_1) = 0$. It is easy to see that, when $y_1(\theta_1) = 0$, the left-hand side of (A.21) reduces to $\zeta(1+\phi)/\phi$ which is obviously positive. The result in part (iii) then follows from the property above, along with the result in Step 2.

Step 4. First use (A.12) to observe that, for any θ_1, $y_2(\theta_1, \theta_2)$ is nondecreasing in θ_2. Next note that (A.4) is equivalent to

$$\int_{\hat{\theta}_1}^{\theta_1} \left[\frac{y_1(s)^{1+\phi}}{s^{2+\phi}} + \frac{\delta\rho}{s} \int_0^{+\infty} \left(\frac{y_2(s,z)}{z} \right)^{1+\phi} f_2(z \mid s, y_1(s)) dz \right] ds \geq$$

$$\int_{\hat{\theta}_1}^{\theta_1} \left[\frac{y_1\left(\hat{\theta}_1\right)^{1+\phi}}{s^{2+\phi}} + \frac{\delta\rho}{s} \int_0^{+\infty} \left(\frac{y_2\left(\hat{\theta}_1, z\right)}{z} \right)^{1+\phi} f_2(z \mid s, y_1(\hat{\theta}_1)) dz \right] ds$$

with $\hat{\theta}_1 < \theta_1$. Now, define the variable $e_2(s, \varepsilon)$ according to

$$e_2(s, \varepsilon) = \frac{y_2(s, s^\rho y_1(s)^\zeta \varepsilon)}{s^\rho y_1(s)^\zeta \varepsilon}.$$

Using this definition, the change of variables $z = s^\rho y_1(s)^\zeta \varepsilon$ in the left integral, the change of variables $z = s^\rho y_1(\hat{\theta}_1)^\zeta \varepsilon$ in the right integral, and the fact that

$$e_2\left(\hat{\theta}_1, \left(\frac{s}{\hat{\theta}_1}\right)^\rho \varepsilon\right) = \frac{y_2(\hat{\theta}_1, s^\rho y_1(\hat{\theta}_1)^\zeta \varepsilon)}{s^\rho y_1(\hat{\theta}_1)^\zeta \varepsilon},$$

we have that the above inequality can be rewritten as

$$\int_{\hat{\theta}_1}^{\theta_1} \left[\frac{y_1(s)^{1+\phi}}{s^{2+\phi}} + \frac{\delta\rho}{s} \int_0^{+\infty} \left[e_2(s, \varepsilon)^{1+\phi} \right] g(\varepsilon) d\varepsilon \right] ds \geq$$

$$\int_{\hat{\theta}_1}^{\theta_1} \left[\frac{y_1\left(\hat{\theta}_1\right)^{1+\phi}}{s^{2+\phi}} + \frac{\delta\rho}{s} \int_0^{+\infty} \left[e_2\left(\hat{\theta}_1, \left(\frac{s}{\hat{\theta}_1}\right)^\rho \varepsilon\right)^{1+\phi} \right] g(\varepsilon) d\varepsilon \right] ds.$$

Clearly, the inequality above is satisfied if for all $\theta_1, \hat{\theta}_1 \in \Theta_1$, $\hat{\theta}_1 < \theta_1$, and ε, both 1 and 2 below hold:

1. $y_1(\theta_1)$ is nondecreasing;
2. $e_2(s, \varepsilon) \geq e_2\left(\hat{\theta}_1, \left(\frac{s}{\hat{\theta}_1}\right)^\rho \varepsilon\right)$ for all $\hat{\theta}_1 \leq s \leq \theta_1$.

Using the definitions of $e_2(s, \varepsilon)$ and $e_2\left(\hat{\theta}_1, \left(\frac{s}{\hat{\theta}_1}\right)^\rho \varepsilon\right)$, we have that the inequality in part 2 above can be expressed as

$$\frac{y_2(s, s^\rho y_1(s)^\zeta \varepsilon)}{y_1(s)^\zeta} \geq \frac{y_2(\hat{\theta}_1, s^\rho y_1(\hat{\theta}_1)^\zeta \varepsilon)}{y_1(\hat{\theta}_1)^\zeta}.$$

Using again (A.12), we have that

$$\frac{y_2(\theta_1, s^\rho y_1(\theta_1)^\zeta \varepsilon)}{y_1(\theta_1)^\zeta} = [s^\rho \varepsilon]^{\frac{1+\phi}{\phi}} \left[1 + \rho \hat{W}_1^{NOLBD}(\theta_1)\right]^{-\frac{1}{\phi}} y_1(\theta_1)^{\frac{\zeta}{\phi}}.$$

Properties 1 and 2 above are thus satisfied if, for all $\theta_1, \hat{\theta}_1 \in \Theta_1$, $\hat{\theta}_1 < \theta_1$, and ε, both (a) and (b) below hold:
(a) $y_1(\theta_1)$ is nondecreasing;
(b) $\left[1 + \rho \hat{W}_1^{NOLBD}(\theta_1)\right]^{-\frac{1}{\phi}} y_1(\theta_1)^{\frac{\zeta}{\phi}}$ is nondecreasing.
The result in part (iv) then follows from the fact that $y_1(\theta_1)$, which is given by the unique solution to (A.2), is nondecreasing whenever $\hat{W}_1^{NOLBD}(\theta_1)$ is nonincreasing. Q.E.D.

A.2 Wedges, Taxes and Sufficient Statistics

This subsection contains all the results relating wedges to tax rates claimed in the main text. In particular, Sub-subsection A.2.1 establishes the relation between wedges and tax rates under arbitrary tax codes. Sub-subsection A.2.2 contains the proofs of Propositions 3 and 4 in the main body establishing optimality conditions for period-1 and period-2 taxes in terms of sufficient statistics. Finally, Sub-subsection A.2.3 establishes equivalence between the formulas for the wedges obtained under the allocation approach and the perturbation approach.

A.2.1 Relationship between taxes and wedges under arbitrary tax codes

In this sub-subsection, we formally establish the relationship between taxes and wedges alluded to in Section 4 in the main body. Consider an arbitrary tax code $\mathcal{T}$. The problem of a worker with period-1 productivity θ_1 consists of choosing a period-1 income y_1 and a contingent period-2 income schedule $\overline{y}_2(\theta_2; y_1)$ so as to maximize[2]

$$y_1 - \mathcal{T}_1(y_1) - \psi(y_1, \theta_1) + \delta \int \left[\overline{y}_2(\theta_2; y_1) - \mathcal{T}_2(y_1, \overline{y}_2(\theta_2; y_1)) - \psi(\overline{y}_2(\theta_2; y_1), \theta_2)\right] dF_2(\theta_2|\theta_1, y_1).$$

The corresponding first-order conditions (FOCs) for y_1 and $\overline{y}_2(\theta_2; y_1)$ are

$$\begin{aligned} 1 - \tau_1(y_1) &= \Gamma(y_1, \theta_1) \\ &\equiv \psi_y(y_1, \theta_1) + \delta \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1, \overline{y}_2(\theta_2; y_1)) dF_2(\theta_2|\theta_1, y_1) \\ &\quad - \delta \frac{\partial}{\partial y_1} \int \left[\overline{y}_2(\theta_2; y_1) - \psi(\overline{y}_2(\theta_2; y_1), \theta_2)\right] dF_2(\theta_2|\theta_1, y_1) \end{aligned} \tag{A.22}$$

and

$$1 - \tau_2(y_1, \overline{y}_2(\theta_2; y_1)) = \psi_y(\overline{y}_2(\theta_2; y_1), \theta_2). \tag{A.23}$$

[2]That the agent is risk neutral, along with the fact that the after-capital-income-tax gross interest rate is equal to the inverse of the discount factor imply that the agent is indifferent as to the specific consumption path consistent with the income choices y_1 and $\overline{y}_2(\theta_2; y_1)$.

Note that the derivatives of $\int \mathcal{T}_2(y_1, \overline{y}_2(\theta_2; y_1))dF_2(\theta_2|\theta_1, y_1)$ and

$$\int \left[\overline{y}_2(\theta_2; y_1) - \psi(\overline{y}_2(\theta_2; y_1), \theta_2)\right] dF_2(\theta_2|\theta_1, y_1)$$

in (A.22) are computed holding the optimal period-2 income schedule constant by usual envelope arguments. The solution to the above system of FOCs yields the policies $y_1(\theta_1)$ and $y_2(\theta) = \overline{y}_2(\theta_2; y_1(\theta_1))$, where the dependence of such policies on the tax code $\mathcal{T}$ is dropped to ease the notation.[3]

Using the above FOCs, we then have that

$$W_1(\theta_1) \equiv 1 - \frac{\psi_y(y_1(\theta_1), \theta_1)}{1 + LD_1^{\chi}(\theta_1)} = \frac{\tau_1(y_1(\theta_1)) + \delta \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1(\theta_1), y_2(\theta))dF_2(\theta_2|\theta_1, y_1(\theta_1))}{1 + \delta \frac{\partial}{\partial y_1} \int \left[y_2(\theta) - \psi(y_2(\theta), \theta_2)\right] dF_2(\theta_2|\theta_1, y_1(\theta_1))} \tag{A.24}$$

and

$$W_2(\theta) \equiv 1 - \psi_y(y_2(\theta), \theta_2) = \tau_2(y_1(\theta_1), y_2(\theta)),$$

where we used the fact that

$$LD_1^{\chi}(\theta_1) \equiv \delta \frac{\partial}{\partial y_1} \int \left[y_2(\theta) - \psi(y_2(\theta), \theta_2)\right] dF_2(\theta_2|\theta_1, y_1(\theta_1)).$$

Furthermore, using the fact that, for any θ_1, $1 - \tau_1(y_1(\theta_1)) = \Gamma(y_1(\theta_1), \theta_1)$, we have that the relative wedges can be expressed in terms of the underlying tax code $\mathcal{T}$ according to the formulas in the main text. Q.E.D.

A.2.2 Optimal tax formulas: Sufficient statistics approach

This sub-subsection contains the proofs of Propositions 3 and 4 in the main body establishing optimality conditions for period-1 and period-2 taxes in terms of sufficient statistics.

Proof of Proposition 3 in the main body. Consider the perturbation of the tax code $\mathcal{T}$ whereby the slope of the period-1 income schedule is increased by $d\tau_1$ for all earnings in the bracket $[y_1, y_1 + dy_1)$, where y_1 is an arbitrary income level generated by some type θ_1 under the original tax code $\mathcal{T}$. Note that, under the perturbed tax code, for any period-1 income $y_1' \in [y_1, y_1 + dy_1)$, the marginal tax rate is $\tau_1(y_1') + d\tau_1$. The above perturbation comes with three effects on the government's objective.

First, all individuals with first-period earnings (weakly) higher than $y_1 + dy_1$ pay higher taxes (for given earnings), by an amount of $d\tau_1 dy_1$. Assuming dy_1 is small, this mechanical effect is equal to

$$d\tau_1 dy_1[1 - H_Y(y_1)] \tag{A.25}$$

where $H_Y(y_1)$ is the cumulative distribution of period-1 earnings under the original tax schedule, $\mathcal{T}$.

Second, all individuals with first-period earnings $y_1' \in [y_1, y_1 + dy_1)$ reduce their earnings by $\frac{\partial \widehat{y}_1(1-\tau_1(y_1'), \theta_1(y_1'))}{\partial(1-\tau_1)} d\tau_1$, where $\theta_1(y_1')$ is the period-1 productivity of all agents whose period-1 income

[3] When we find it useful to highlight this dependence, we will do it by denoting the optimal income policies by $y_1(\theta_1; \mathcal{T})$, and $y_2(\theta; \mathcal{T}) = \overline{y}_2(\theta_2; y_1(\theta_1; \mathcal{T}), \mathcal{T})$.

under the original tax code $\mathcal{T}$ is y_1'.[4] Using again the fact that dy_1 is small, we have that the total reduction in first-period tax revenues from these individuals is equal to

$$-d\tau_1 \frac{y_1}{1-\tau_1(y_1)} \hat{E}_1(y_1)\tau_1(y_1)\hat{h}_Y(y_1)dy_1, \tag{A.26}$$

where recall that (a) $\hat{h}_Y$ is the density of period-1 earnings in the fictitious economy in which the original tax code $\mathcal{T}$ is replaced with the tax code $\hat{\mathcal{T}} \equiv (\hat{\mathcal{T}}_1, \mathcal{T}_2)$ where $\hat{\mathcal{T}}_1$ is a linear tax schedule with constant marginal tax rate equal to $\tau_1(y_1)$, and (b)

$$\hat{E}_1(y_1) \equiv \frac{1-\tau_1(y_1)}{y_1} \frac{\partial \widehat{y}_1(1-\tau_1(y_1), \theta_1(y_1))}{\partial(1-\tau_1)}.$$

Third, a change in the period-1 marginal tax rate, by triggering a change in the period-1 earnings of those individuals generating income $y_1' \in [y_1, y_1 + dy)$, also induces a variation in the period-2 tax revenues. This variation combines the fact that the period-2 tax schedule $\mathcal{T}_2(y_1, y_2)$ depends directly on period-1 income, along with the fact that the distribution of the period-2 productivity changes in response to variations in period-1 incomes, due to LBD. This period-2 behavioral effect (expressed in terms of period-1 tax revenues) is equal to

$$-d\tau \frac{y_1}{1-\tau_1(y_1)} \hat{E}_1(y_1) \left[\delta \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1, y_2) dH_O(y_2|y_1) \right] \hat{h}_Y(y_1)dy_1, \tag{A.27}$$

where $H_O(y_2|y_1)$ is the cumulative distribution of period-2 earnings of those individuals generating period-1 earnings equal to y_1, under the original code $\mathcal{T}$ (which is given by $H_O(y_2|y_1) = F_2(\theta_2(y_1, y_2), |\theta_1(y_1), y_1)$ with $\theta_2(y_1, y_2)$ implicitly defined by $y_2(\theta_1(y_1), \theta_2(y_1, y_2)) = y_2$ or, equivalently, by $1-\tau_2(y_1, y_2) = \psi_y(y_2, \theta_2)$). Note that the derivative with respect to y_1 in (A.27) combines the direct effect of a change in period-2 taxes for a given distribution of period-2 incomes with the indirect effect due to a variation in the distribution of period-2 income for a given period-2 tax schedule $\mathcal{T}_2(y_1, \cdot)$. Importantly, when differentiating the distribution $H_O(y_2|y_1)$ with respect to y_1, the derivative must be computed holding fixed the agent's period-1 productivity at $\theta_1(y_1)$.[5]

For the tax code $\mathcal{T} = (\mathcal{T}_1, \mathcal{T}_2)$ to be optimal, the sum of the above behavioral and mechanical effects must be zero. This is the case for all income levels in the support of the income distribution only if, for any y_1 in the range of the period-1 income schedule, the condition stated in Proposition 3 in the main body holds. Q.E.D.

Proof of Proposition 4 in the main body. Consider the reform of the tax schedule described in the main text. Recall that this reform consists of three parts: (a) an increase by $d\tau_2$ of the period-2 marginal tax rate over the bracket $[y_2, y_2 + dy_2)$ for those individuals generating period-1 earnings

[4] The function $\theta_1(y_1')$ is implicitly defined by the FOC $1-\tau_1(y_1') = \Gamma(y_1', \theta_1)$.

[5] Formally, let $h_O(y_2|y_1) \equiv f_2(\theta_2(y_1, y_2), |\theta_1(y_1), y_1) \frac{\partial \theta_2(y_1, y_2)}{\partial y_2}$ denote the density of the distribution of period-2 incomes among those individuals generating period-1 income equal to y_1, under the original code $\mathcal{T}$. Then

$$\frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1, y_2) dH_O(y_2|y_1) = \int \frac{\partial \mathcal{T}_2(y_1, y_2)}{\partial y_1} dH_O(y_2|y_1) + \int \mathcal{T}_2(y_1, y_2) \frac{\partial h_O(y_2|y_1)}{\partial y_1} dy_2.$$

In computing $\partial h_O(y_2|y_1)/\partial y_1$ one must hold $\theta_1(y_1)$ fixed.

in the bracket $[y_1, y_1 + dy_1)$, (b) an increase in the period-1 marginal tax rate at any income level $y_1' \in [y_1, y_1 + dy_1)$ by

$$\delta \left(\frac{\partial \widetilde{H}_O(y_2|y_1')}{\partial y_1} - \frac{\partial H_O(y_2|y_1')}{\partial y_1} \right) d\tau_2 dy_2,$$

and (c) an income-contingent period-1 subsidy equal to $S(y_1') \equiv \delta[1 - H_O(y_2|y_1')]d\tau_2 dy_2$ to all individuals with period-1 income $y_1' \in [y_1, y_1 + dy_1)$.

This perturbation yields two effects in terms of total tax revenues. The first effect is the usual static period-2 behavioral effect, originating from the fact that all individuals who, prior to the reform, would have generated period-1 earnings $y_1' \in [y_1, y_1 + dy_1)$ and period-2 earnings $y_2' \in [y_2, y_2 + dy_2)$, reduce their period-2 earnings by $-\frac{\partial \widehat{y}_2(1-\tau_2(y_1',y_2'),\theta_2(y_1',y_2'))}{\partial(1-\tau_2)} d\tau_2$, where $\widehat{y}_2(1 - \tau_2, \theta_2)$ denotes the optimal period-2 income choice of an individual of period-2 productivity θ_2 facing a linear period-2 tax schedule with constant marginal tax rate equal to τ_2.

For small dy_1 and dy_2, these behavioral responses imply a total loss in period-2 tax revenues (from all individuals who would have generated period-1 earnings $y_1' \in [y_1, y_1 + dy_1)$ and period-2 earnings $y_2' \in [y_2, y_2 + dy_2)$) equal to

$$-d\tau_2 \frac{y_2}{1 - \tau_2(y_1, y_2)} \hat{E}_2(y_1, y_2)\tau_2(y_1, y_2)dy_2 dy_1 \hat{h}_O(y_2|y_1)h_Y(y_1),$$

where $\hat{h}_O(y_2|y_1)$ is the conditional density of period-2 earnings in a fictitious economy in which the period-2 non-linear tax schedule $\mathcal{T}_2(y_1, \cdot)$ is replaced with the linear tax schedule with constant marginal tax rate $\tau_2 = \tau_2(y_1, y_2)$, and where

$$\hat{E}_2(y_1, y_2) \equiv \frac{1 - \tau_2(y_1, y_2)}{y_2} \frac{\partial \widehat{y}_2(1 - \tau_2(y_1, y_2), \theta_2(y_1, y_2))}{\partial(1 - \tau_2)}.$$

In terms of period-1 tax dollars, the total behavioral effect of the proposed reform is thus equal to

$$-\delta \left[d\tau_2 \frac{y_2}{1 - \tau_2(y_1, y_2)} \hat{E}_2(y_1, y_2)\tau_2(y_1, y_2)dy_2 dy_1 \hat{h}_O(y_2|y_1)h_Y(y_1) \right]. \tag{A.28}$$

The second effect is a mechanical effect and originates from the fact that all individuals with period-1 earnings $y_1' \in [y_1, y_1 + dy_1)$ and period-2 earnings $y_2' \geq y_2 + dy_2$ pay higher taxes (for given earnings in both periods) by an amount of $d\tau_2 dy_2$. When dy_2 is small, this means that, from the perspective of period 1, all individuals generating period-1 earnings $y_1' \in [y_1, y_1 + dy_1)$ expect to pay $[1 - F_2(\theta_2(y_1', y_2)|\theta_1(y_1'), y_1')]d\tau_2 dy_2$ more in period 2. Combined with the other parts of the reform, this means that any individual with period-1 productivity $\theta_1(y_1')$ and period-1 income (prior to the reform) $y_1' \in [y_1, y_1 + dy_1)$ expects a net increase in his lifetime taxes equal to

$$\begin{gathered} \delta[1 - F_2(\theta_2(y_1', y_2)|\theta_1(y_1'), y_1')]d\tau_2 dy_2 + \delta d\tau_2 dy_2 \int_{y_1}^{y_1'} \left(\frac{\partial \widetilde{H}_O(y_2|s)}{\partial y_1} - \frac{\partial H_O(y_2|s)}{\partial y_1} \right) ds \\ -\delta[1 - H_O(y_2|y_1')]d\tau_2 dy_2. \end{gathered} \tag{A.29}$$

Importantly, note that, as mentioned in the main text, under the reform, the optimal period-1 income choice for any such individual remains the same as prior to the reform.[6] That is, the reform in question neutralizes the impact of the variation in the period-2 marginal tax rate on first-period earnings. Crucially, however, under this reform and when dy_2 is small, all individuals with period-1 earnings (weakly) higher than $y_1 + dy_1$ pay higher taxes in period 1 by an amount of $dy_1 d\tau_2 dy_2 \delta \left[\frac{\partial \widetilde{H}_O(y_2|y_1)}{\partial y_1} - \frac{\partial H_O(y_2|y_1)}{\partial y_1}\right]$. Integrating over all period-1 incomes above y_1, we then have that, when dy_1 is small, the reform yields an increase in the period-1 tax revenues equal to

$$dy_1 d\tau_2 dy_2 \delta \left[\frac{\partial \widetilde{H}_O(y_2|y_1)}{\partial y_1} - \frac{\partial H_O(y_2|y_1)}{\partial y_1}\right][1 - H_Y(y_1)], \tag{A.30}$$

where recall that $H_Y(\cdot)$ is the cumulative income distribution among young workers.

Under any optimal tax code, the sum of the above behavioral and mechanical effects on the net present value of intertemporal tax revenues must be equal to zero.[7] For this to be the case, it must be that, for any (y_1, y_2) in the support of the induced income distribution, the condition in Proposition 4 in the main body holds. Q.E.D.

A.2.3 Equivalence between allocation and perturbation approach

This sub-subsection establishes the equivalence between the formulas for the wedges under optimal tax codes derived in the main body with the direct approach and the corresponding formulas derived with the sufficient statistics approach.

Period-1 relative wedges Changing the variables of integration by letting $y_1 = y_1(\theta_1)$ and noting that, when earnings are monotone in productivities, as assumed in the literature, $H_Y(y_1(\theta_1)) = F_1(\theta_1)$, we have that the formula for the optimal period-1 marginal tax rates in Proposition 3 in the paper can be rewritten as

$$\frac{\tau_1(y_1(\theta_1))}{1 - \tau_1(y_1(\theta_1))} = \frac{1 - F_1(\theta_1)}{y_1(\theta_1)\hat{h}_Y(y_1(\theta_1))} \frac{1}{\hat{E}_1(y_1(\theta_1))} \left[\frac{1}{1 + \frac{\delta \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1(\theta_1), y_2) dH_O(y_2|y_1(\theta_1))}{\tau_1(y_1(\theta_1))}}\right]. \tag{A.31}$$

Next, use the that fact that $\widehat{y}_1(1 - \tau_1(y_1(\theta_1)), \theta_1) = y_1(\theta_1)$ to rewrite $\hat{E}_1(y_1(\theta_1))$ as

$$\hat{E}_1(y_1(\theta_1)) = \frac{1 - \tau_1(y_1(\theta_1))}{y_1(\theta_1)} \frac{\partial \widehat{y}_1(1 - \tau_1(y_1(\theta_1)), \theta_1)}{\partial(1 - \tau_1)}.$$

Then use the fact that $\widehat{y}_1(1 - \tau_1, \theta_1)$ is implicitly defined by $1 - \tau_1 = \Gamma(y_1, \theta_1)$, where Γ is the function defined in (A.22), to note that

$$\frac{\partial \widehat{y}_1(1 - \tau_1(y_1(\theta_1)), \theta_1)}{\partial(1 - \tau_1)} = \frac{1}{\Gamma_y(\widehat{y}_1(1 - \tau_1(y_1(\theta_1)), \theta_1), \theta_1)} = \frac{1}{\Gamma_y(y_1(\theta_1), \theta_1)}$$

[6] This is a direct consequence of the fact that the derivative of the term in (A.29) with respect to y_1' is zero. To see this, note that (a) $H_O(y_2|y_1') \equiv F_2(\theta_2(y_1', y_2)|\theta_1(y_1'), y_1')$, (b) $\partial \widetilde{H}_O(y_2|y_1')/\partial y_1$ is the derivative of $H_O(y_2|y_1')$ with respect to y_1, holding constant θ_1 at $\theta_1(y_1')$ and evaluated at $y_1 = y_1'$, (c) $\partial H_O(y_2|y_1')/\partial y_1$ is the derivative of $H_O(y_2|y_1')$ with respect to y_1, evaluated at $y_1 = y_1'$, which includes the effect of a variation in y_1 on $\theta_1(y_1)$, (d) the derivative of the first term in (A.29) must be computed holding θ_1 fixed at $\theta_1(y_1')$.

[7] Note that, for $dy_1 \to 0$, for any $y_1' \in [y_1, y_1 + dy_1)$, $\delta d\tau_2 dy_2 \int_{y_1}^{y_1'} \left[\frac{\partial \widetilde{H}_O(y_2|s)}{\partial y_1} - \frac{\partial H_O(y_2|s)}{\partial y_1}\right] ds + \delta d\tau_2 dy_2 [1 - H_O(y_2|y_1')] - S(y_1') \to 0$ at the optimum.

and hence

$$\hat{E}_1(y_1(\theta_1)) = \frac{\Gamma(y_1(\theta_1),\theta_1)}{y_1(\theta_1)\Gamma_y(y_1(\theta_1),\theta_1)}. \tag{A.32}$$

Now use the fact that, for any (τ_1, θ_1), $\hat{H}_Y(\widehat{y}_1(1-\tau_1,\theta_1)) = F_1(\theta_1)$ to write

$$\hat{h}_Y(\widehat{y}_1(1-\tau_1,\theta_1))\frac{\partial \widehat{y}_1(1-\tau_1,\theta_1)}{\partial \theta_1} = f_1(\theta_1). \tag{A.33}$$

When evaluated at $\tau_1 = \tau_1(y_1(\theta_1))$, Condition (A.33) implies that

$$\hat{h}_Y(y_1(\theta_1)) = \frac{f_1(\theta_1)}{\frac{\partial \widehat{y}_1(1-\tau_1(y_1(\theta_1)),\theta_1)}{\partial \theta_1}}.$$

Now use again the fact that $\widehat{y}_1(1-\tau_1,\theta_1)$ is implicitly defined by $1-\tau_1 = \Gamma(y_1,\theta_1)$ to observe that

$$\frac{\partial \widehat{y}_1(1-\tau_1,\theta_1)}{\partial \theta_1} = -\frac{\Gamma_\theta(\widehat{y}_1(1-\tau_1,\theta_1),\theta_1)}{\Gamma_y(\widehat{y}_1(1-\tau_1,\theta_1),\theta_1)}.$$

Because $\widehat{y}_1(1-\tau_1(y_1(\theta_1)),\theta_1) = y_1(\theta_1)$, this means that

$$\hat{h}_Y(y_1(\theta_1)) = \frac{f_1(\theta_1)}{\frac{\partial \widehat{y}_1(1-\tau_1(y_1(\theta_1)),\theta_1)}{\partial \theta_1}} = -\frac{f_1(\theta_1)\Gamma_y(y_1(\theta_1),\theta_1)}{\Gamma_\theta(y_1(\theta_1),\theta_1)}. \tag{A.34}$$

Combining (A.32) with (A.34), we thus have that, under the optimal tax code,

$$\frac{\tau_1(y_1(\theta_1))}{1-\tau_1(y_1(\theta_1))} = \frac{1}{\theta_1\gamma_1(\theta_1)}\left[\frac{\left(-\theta_1\frac{\Gamma_\theta(y_1(\theta_1),\theta_1)}{\Gamma(y_1(\theta_1),\theta_1)}\right)}{1+\frac{\delta\frac{\partial}{\partial y_1}\int \mathcal{T}_2(y_1(\theta_1),y_2)dH_O(y_2|y_1(\theta_1))}{\tau_1(y_1(\theta_1))}}\right].$$

Next, observe that, in the iso-elastic case,

$$-\frac{\theta_1\psi_{y\theta}(y_1,\theta_1)}{\psi_y(y_1,\theta_1)} = 1+\phi = -\frac{\theta_2\psi_\theta(y_2,\theta_2)}{\psi(y_2,\theta_2)}. \tag{A.35}$$

Hence, using the definition of the function Γ in (A.22), we have that, under this specification,

$$-\theta_1\Gamma_\theta(y_1(\theta_1),\theta_1) =$$

$$(1+\phi)\left[\begin{array}{c}\psi_y(y_1(\theta_1),\theta_1)\\ -\frac{\psi_y(y_1(\theta_1),\theta_1)}{\psi_{y\theta}(y_1(\theta_1),\theta_1)}\delta\frac{\partial^2}{\partial\theta_1\partial y_1}\int\left[y_2(\theta)-\mathcal{T}_2(y_1(\theta_1),y_2(\theta))-\psi(y_2(\theta),\theta_2)\right]dF_2(\theta_2|\theta_1,y_1(\theta_1))\end{array}\right] =$$

$$(1+\phi)\left[\psi_y(y_1(\theta_1),\theta_1)+\frac{\psi_y(y_1(\theta_1),\theta_1)}{\psi_{y\theta}(y_1(\theta_1),\theta_1)}\delta\rho\frac{\partial}{\partial y_1}\int\left[\psi_\theta(y_2(\theta),\theta_2)\frac{\theta_2}{\theta_1}\right]dF_2(\theta_2|\theta_1,y_1(\theta_1))\right] =$$

$$(1+\phi)\left[\psi_y(y_1(\theta_1),\theta_1)+\frac{\delta\rho}{1+\phi}\frac{\partial}{\partial y_1}\int\left[(1+\phi)\psi(y_2(\theta),\theta_2)\right]dF_2(\theta_2|\theta_1,y_1(\theta_1))\right] =$$

$$(1+\phi)\psi_y(y_1(\theta_1),\theta_1)\left[1+\frac{\delta\rho}{\psi_y(y_1(\theta_1),\theta_1)}\frac{\partial}{\partial y_1}\int\left[\psi(y_2(\theta),\theta_2)\right]dF_2(\theta_2|\theta_1,y_1(\theta_1))\right].$$

Note that, for the second equality, we used (a) the fact that, given $(\theta_1, y_1(\theta_1))$, for any θ_2, $\bar{y}_2(\theta_2; y_1(\theta_1)) = y_2(\theta)$, along with (b) the fact that, for any θ_2,

$$\frac{\partial}{\partial y_2}\left\{\bar{y}_2(\theta_2; y_1(\theta_1)) - \mathcal{T}_2(y_1(\theta_1), \bar{y}_2(\theta_2; y_1(\theta_1))) - \psi(\bar{y}_2(\theta_2; y_1(\theta_1)), \theta_2)\right\} = 0,$$

and (c) the fact that, for any Lipschitz continuous function $g(\theta_2)$,

$$\frac{\partial}{\partial \theta_1}\int g(\theta_2)dF_2(\theta_2|\theta_1, y_1) = \int g'(\theta_2)I_1^2(\theta_1, \theta_2, y_1)dF_2(\theta_2|\theta_1, y_1),$$

where

$$I_1^2(\theta_1, \theta_2, y_1) = \rho\frac{\theta_2}{\theta_1}$$

under the technology $\theta_2 = \theta_1^{\rho} y_1^{\zeta}\varepsilon_2$ relating period-2 productivity to period-1 productivity assumed in the paper. For the third equality, we used (A.35).

Accordingly, we have that, under the optimal tax code,

$$\frac{\tau_1(y_1(\theta_1))}{1-\tau_1(y_1(\theta_1))} =$$
$$\frac{(1+\phi)}{\theta_1\gamma_1(\theta_1)}\left[1+\frac{\delta\rho}{\psi_y(y_1(\theta_1), \theta_1)}\frac{\partial}{\partial y_1}\int\left[\psi(y_2(\theta), \theta_2)\right]dF_2(\theta_2|\theta_1, y_1(\theta_1))\right]\times$$
$$\left[\frac{\frac{\psi_y(y_1(\theta_1),\theta_1)}{\Gamma(y_1(\theta_1),\theta_1;\tau_2)}}{1+\frac{\delta}{\tau_1(y_1(\theta_1))}\frac{\partial}{\partial y_1}\int\mathcal{T}_2(y_1(\theta_1), y_2)dH_O(y_2|y_1(\theta_1))}\right].$$

Using the fact that, when $y_1 = y_1(\theta_1)$, $\bar{y}_2(\theta_2; y_1(\theta_1)) = y_2(\theta)$ and the fact that[8]

$$\frac{\partial}{\partial y_1}\int\mathcal{T}_2(y_1(\theta_1), y_2)dH_O(y_2|y_1(\theta_1)) = \frac{\partial}{\partial y_1}\int\mathcal{T}_2(y_1(\theta_1), y_2(\theta))dF_2(\theta_2|\theta_1, y_1(\theta_1)),$$

we thus have that

$$\frac{\tau_1(y_1(\theta_1))}{1-\tau_1(y_1(\theta_1))} =$$
$$\frac{(1+\phi)}{\theta_1\gamma_1(\theta_1)}\left[1+\frac{\delta\rho}{\psi_y(y_1(\theta_1), \theta_1)}\frac{\partial}{\partial y_1}\int\left[\psi(y_2(\theta), \theta_2)\right]dF_2(\theta_2|\theta_1, y_1(\theta_1))\right]\times \tag{A.36}$$
$$\left[\frac{\frac{\psi_y(y_1(\theta_1),\theta_1)}{\Gamma(y_1(\theta_1),\theta_1)}}{1+\delta\frac{\partial}{\partial y_1}\int\mathcal{T}_2(y_1(\theta_1), y_2(\theta))dF_2(\theta_2|\theta_1, y_1(\theta_1))}\right].$$

Now use the formula relating the tax code to the wedges established in (A.24) along with the fact that $1-\tau_1(y_1(\theta_1)) = \Gamma(y_1(\theta_1), \theta_1)$ to rewrite the relative wedge as

$$\widehat{W}_1(\theta_1) \equiv \frac{W_1(\theta_1)}{1-W_1(\theta_1)}$$
$$= \left[\frac{\tau_1(y_1(\theta_1))}{1-\tau_1(y_1(\theta_1))}\right]\left[\frac{1+\frac{\delta}{\tau_1(y_1(\theta_1))}\frac{\partial}{\partial y_1}\int\mathcal{T}_2(y_1(\theta_1),y_2(\theta))dF_2(\theta_2|\theta_1,y_1(\theta_1))}{\frac{\psi_y(y_1(\theta_1),\theta_1)}{\Gamma(y_1(\theta_1),\theta_1)}}\right]. \tag{A.37}$$

[8]Recall that, when computing the derivative of the measure $H_O(y_2|y_1(\theta_1))$ with respect to y_1, one holds $\theta_1(y_1)$ constant.

Substituting the formula in (A.36) into (A.37), we then have that, under the optimal tax code, the relative wedge at any productivity level θ_1 is equal to

$$\widehat{W}_1(\theta_1) =$$

$$\frac{(1+\phi)}{\theta_1\gamma_1(\theta_1)}\left[1+\frac{\delta\rho}{\psi_y(y_1(\theta_1),\theta_1)}\frac{\partial}{\partial y_1}\int[\psi(y_2(\theta),\theta_2)]\,dF_2(\theta_2|\theta_1,y_1(\theta_1))\right],$$

which is the same formula for the optimal period-1 wedge derived through the allocation approach.[9] Q.E.D.

Period-2 relative wedges Observe that

$$\frac{\partial\tilde{H}_O(y_2|y_1)}{\partial y_1}-\frac{\partial H_O(y_2|y_1)}{\partial y_1}=-\frac{\frac{\partial F_2(\theta_2(y_1,y_2)|\theta_1(y_1),y_1)}{\partial\theta_1}}{\frac{\partial y_1(\theta_1(y_1))}{\partial\theta_1}}=\frac{\rho\theta_2(y_1,y_2)f_2(\theta_2(y_1,y_2)|\theta_1(y_1),y_1)}{\theta_1(y_1)\frac{\partial y_1(\theta_1(y_1))}{\partial\theta_1}}. \quad \text{(A.38)}$$

Note that the first equality uses the fact that $H_O(y_2|y_1)=F_2(\theta_2(y_1,y_2)|\theta_1(y_1),y_1)$, along with the fact that, by definition, $\partial\tilde{H}_O(y_2|y_1)/dy_1$ is computed holding θ_1 fixed at $\theta_1=\theta_1(y_1)$. The second equality uses the fact that

$$\frac{\partial}{\partial\theta_1}[1-F_2(\theta_2(y_1,y_2)|\theta_1(y_1),y_1)]=I_1^2(\theta_1(y_1),\theta_2(y_1,y_2),y_1)f_2(\theta_2(y_1,y_2)|\theta_1(y_1),y_1),$$

along with the fact that, when $\theta_2=\theta_1^{\rho}y_1^{\zeta}\varepsilon_2$,

$$I_1^2(\theta_1,\theta_2,y_1)=\rho\frac{\theta_2}{\theta_1}.$$

Replacing (A.38) into the formula for the wedge in Proposition 4 in the main text, we have that, at the income history $(y_1(\theta_1),y_2(\theta_1,\theta_2))$ induced under the optimal tax code, the period-2 relative wedges are equal to

$$\widehat{W}_2(\theta_1,\theta_2)=\frac{\tau_2(y_1(\theta_1),y_2(\theta_1,\theta_2))}{1-\tau_2(y_1(\theta_1),y_2(\theta_1,\theta_2))}$$

$$=\frac{\left[\rho\frac{\theta_2 f_2(\theta_2|\theta_1,y_1(\theta_1))}{\theta_1\frac{\partial y_1(\theta_1(y_1))}{\partial\theta_1}}\right][1-H_Y(y_1(\theta_1))]}{h_Y(y_1(\theta_1))y_2(\theta_1,\theta_2)\hat{h}_O(y_2(\theta_1,\theta_2)|y_1(\theta_1))\hat{E}_2(y_1(\theta_1),y_2(\theta_1,\theta_2))}, \quad \text{(A.39)}$$

where we used the fact that $y_1=y_1(\theta_1)$ and $y_2=y_2(\theta_1,\theta_2)$.

Next, use the fact that, for any θ_1,

$$H_Y(y_1(\theta_1))=F_1(\theta_1) \quad \text{(A.40)}$$

to note that

$$h_Y(y_1(\theta_1))\frac{\partial y_1(\theta_1(y_1))}{\partial\theta_1}=f_1(\theta_1) \quad \text{(A.41)}$$

[9] To see this, recall that the formula for the period-1 wedges derived with the allocation approach is $\widehat{W}_1(\theta_1)=\widehat{W}_1^{NOLBD}(\theta_1)+\Omega(\theta_1)$, where $\widehat{W}_1^{NOLBD}(\theta_1)$ is given by (A.1) and $\Omega(\theta_1)$ is given by (A.5).

and, similarly, the fact that, for any (θ_1, θ_2),

$$\hat{H}_O(y_2(\theta_1,\theta_2)|y_1(\theta_1)) = \hat{H}_O(\hat{y}_2(1-\tau_2(y_1(\theta_1), y_2(\theta_1,\theta_2)), \theta_2)|y_1(\theta_1)) = F_2(\theta_2|\theta_1, y_1(\theta_1))$$

to note that

$$\hat{h}_O(y_2(\theta_1,\theta_2)|y_1(\theta_1)) = \frac{f_2(\theta_2|\theta_1, y_1(\theta_1))}{\frac{\partial \hat{y}_2(1-\tau_2(y_1(\theta_1),y_2(\theta_1,\theta_2)),\theta_2)}{\partial\theta_2}}. \tag{A.42}$$

Finally, use the fact that

$$\hat{E}_2(y_1, y_2) = \frac{1-\tau_2(y_1,y_2)}{\widehat{y}_2(1-\tau_2(y_1,y_2), \theta_2(y_1,y_2))} \frac{\partial \widehat{y}_2(1-\tau_2(y_1,y_2), \theta_2(y_1,y_2))}{\partial(1-\tau_2)},$$

along with the fact that $\widehat{y}_2(1-\tau_2, \theta_2)$ is implicitly defined by

$$\psi_y(\widehat{y}_2(1-\tau_2,\theta_2),\theta_2) = 1-\tau_2$$

and hence

$$\frac{\partial \widehat{y}_2(1-\tau_2,\theta_2)}{\partial\theta_2} = -\frac{\psi_{y\theta}(\widehat{y}_2(1-\tau_2,\theta_2),\theta_2)}{\psi_{yy}(\widehat{y}_2(1-\tau_2,\theta_2),\theta_2)}$$

and

$$\frac{\partial \widehat{y}_2(1-\tau_2,\theta_2)}{\partial(1-\tau_2)} = \frac{1}{\psi_{yy}(\widehat{y}_2(1-\tau_2,\theta_2),\theta_2)},$$

to write[10]

$$\hat{E}_2(y_1(\theta_1), y_2(\theta_1,\theta_2)) = -\frac{\psi_y(y_2(\theta_1,\theta_2),\theta_2)}{y_2(\theta_1,\theta_2)\psi_{y\theta}(y_2(\theta_1,\theta_2),\theta_2)} \frac{\partial \widehat{y}_2(1-\tau_2(y_1(\theta_1), y_2(\theta_1,\theta_2)),\theta_2)}{\partial\theta_2}. \tag{A.43}$$

Also recall that, under the assumed iso-elastic specification

$$-\theta_2 \frac{\psi_{y\theta}(y_2,\theta_2)}{\psi_y(y_2,\theta_2)} = 1+\phi. \tag{A.44}$$

Substituting (A.40)-(A.43) into (A.39), and using (A.44), we have that

$$\hat{W}_2(\theta_1,\theta_2) = \frac{\tau_2(y_1(\theta_1), y_2(\theta_1,\theta_2))}{1-\tau_2(y_1(\theta_1), y_2(\theta_1,\theta_2))} = \frac{(1+\phi)\rho}{\gamma_1(\theta_1)\theta_1},$$

which is the same formula obtained through the direct allocation approach.[11] Q.E.D.

A.3 Equivalence between 40-period economy and 2-period model

In this subsection, we establish the equivalence between the 40-period economy used in the quantitative analysis in Section 5 in the main body and the 2-period economy used in Sections 2-4 in the main body.

[10]Note that we also used the fact that $\hat{y}_2(1-\tau_2(y_1(\theta_1), y_2(\theta_1,\theta_2)),\theta_2) = y_2(\theta_1,\theta_2)$.

[11]To see this, recall that the formula for the period-2 wedges derived with the allocation approach is $\widehat{W}_2(\theta_1,\theta_2) = \widehat{W}_2^{NOLBD}(\theta_1,\theta_2)$ where $\widehat{W}_2^{NOLBD}(\theta_1,\theta_2)$ is given by (A.1).

A.3.1 The 40-period economy

Suppose each worker works for $T = 2\widehat{T}$ periods and discounts the future with the discount factor β. Productivity is constant within each of the two blocks of a worker's life, with each block comprising $\widehat{T}$ periods. In the quantitative analysis in the main text $\widehat{T} = 20$ with each period corresponding to a year. Let θ_1 denote productivity in the first $\widehat{T}$ periods and θ_2 denote productivity in the second $\widehat{T}$ periods. Moreover, assume that

$$\theta_2 = h_2 \theta_1^{\rho} \left(\frac{\sum_{s=1}^{\widehat{T}} \hat{\beta}_s y_s}{\sum_{s=1}^{\widehat{T}} \hat{\beta}_s} \right)^{\zeta} \varepsilon_2$$

for some vector of weights $(\hat{\beta}_s)$ nonincreasing in s. Note that LBD is active in each of the first $\widehat{T}$ periods. Also note that the above representation implies that LBD is stronger in earlier years than in later ones. Assume that $(\widehat{\beta}_1, \widehat{\beta}_2, ..., \widehat{\beta}_{\widehat{T}})$ is proportional to $(1, \beta, ..., \beta^{\widehat{T}-1})$, in which case $\frac{\widehat{\beta}_t}{\sum_{s=1}^{\widehat{T}} \widehat{\beta}_s} = \frac{\beta^{t-1}}{\sum_{s=1}^{\widehat{T}} \beta^{s-1}}$.

Next let

$$\overline{y}(\theta_1) \equiv \frac{\sum_{s=1}^{\widehat{T}} \beta^{s-1} y_s(\theta_1)}{\sum_{s=1}^{\widehat{T}} \beta^{s-1}}$$

denote the average income generated by an agent of initial productivity θ_1 over the first $\widehat{T}$ periods. The expected lifetime utility of a worker of period-1 productivity equal to θ_1 is given by

$$V_1(\theta_1) \equiv \textstyle\sum_{t=1}^{\widehat{T}} \beta^{t-1} [v(c_t(\theta_1)) - \psi(y_t(\theta_1), \theta_1)]$$

$$+\beta^{\widehat{T}} \textstyle\int \left\{ \sum_{t=1}^{\widehat{T}} \beta^{t-1} [v(c_{\widehat{T}+t}(\theta_1, \theta_2)) - \psi(y_{\widehat{T}+t}(\theta_1, \theta_2), \theta_2)] \right\} dF_2(\theta_2 | \theta_1, \overline{y}(\theta_1)).$$

With an abuse of notation, hereafter, let $\theta^s = \theta_1$ and $\theta_s = \theta_1$ for all $s = 1, ..., \widehat{T}$, and $\theta^s = (\theta_1, \theta_2)$ and $\theta_s = \theta_2$ for all $s = \widehat{T} + 1, ..., 2\widehat{T}$.

A.3.2 Optimal allocations in the 40-period economy

We now show that the allocations under the optimal tax code in the 40-period economy coincide with those in the two-period model in the main text. That is, for any $s = 1, ..., \hat{T}$, $(c_s(\theta_1), y_s(\theta_1)) = (c_1^{2pm}(\theta_1), y_1^{2pm}(\theta_1))$, and, for any $s = \hat{T}+1, ..., 2\hat{T}$, $(c_s(\theta_1, \theta_2), y_s(\theta_1, \theta_2)) = (c_2^{2pm}(\theta_1, \theta_2), y_2^{2pm}(\theta_1, \theta_2))$, where $(c_1^{2pm}(\theta_1), y_1^{2pm}(\theta_1))$ and $(c_2^{2pm}(\theta_1, \theta_2), y_2^{2pm}(\theta_1, \theta_2))$ are the optimal policies in the two-period model.

Preliminary Analysis For any $t = 1, ..., \widehat{T}$, let

$$V_t(\theta^t) \equiv \sum_{s=t}^{\widehat{T}} \beta^{s-t} [v(c_s(\theta^s)) - \psi(y_s(\theta^s), \theta_s)] +$$

$$\beta^{\widehat{T}+1-t} \int \left\{ \sum_{s=1}^{\widehat{T}} \beta^{s-1} [v(c_{\widehat{T}+s}(\theta_1, \theta_2)) - \psi(y_{\widehat{T}+s}(\theta_1, \theta_2), \theta_2)] \right\} dF_2(\theta_2 | \theta_1, \overline{y}(\theta_1))$$

denote the continuation expected utility of each worker whose productivity in the first block is equal to θ_1. Observe that

$$V_t(\theta^t) = v(c_t(\theta_1)) - \psi(y_t(\theta_1), \theta_1) + \beta\Pi_{t+1}(\theta^t),$$

where, for any $t = 1, ..., \widehat{T}$,

$$\Pi_{t+1}(\theta^t) \equiv \sum_{s=t+1}^{\widehat{T}} \beta^{s-t-1}[v(c_s(\theta_1)) - \psi(y_s(\theta_1), \theta_1)] +$$

$$\beta^{\widehat{T}-t} \int \left\{ \sum_{s=1}^{\widehat{T}} \beta^{s-1}[v(c_{\widehat{T}+s}(\theta_1, \theta_2)) - \psi(y_{\widehat{T}+s}(\theta_1, \theta_2), \theta_2)] \right\} dF_2(\theta_2|\theta_1, \overline{y}(\theta_1))$$

denotes the utility expected from period $t+1$ onwards.

Likewise, for any $t = \widehat{T}+1, ..., 2\widehat{T}$, let

$$V_t(\theta^t) \equiv \sum_{s=t}^{2\widehat{T}} \beta^{s-t}[v(c_t(\theta^t)) - \psi(y_t(\theta^t), \theta_t)]$$

denote the continuation expected utility of each worker whose productivity in the first block is equal to θ_1 and whose productivity in the second block is equal to θ_2. Observe that, for any $t = \widehat{T}+1, ..., 2\widehat{T}$,

$$V_t(\theta^t) = v(c_t(\theta_1, \theta_2)) - \psi(y_t(\theta_1, \theta_2), \theta_2) + \beta\Pi_{t+1}(\theta^t),$$

where, for any $t = \widehat{T}+1, ..., 2\widehat{T}-1$,

$$\Pi_{t+1}(\theta^t) \equiv \sum_{s=t+1}^{2\widehat{T}} \beta^{s-t-1}[v(c_t(\theta^t)) - \psi(y_t(\theta^t)), \theta_t)] = V_{t+1}(\theta^{t+1}),$$

whereas, for $t = 2\widehat{T}$,

$$\Pi_{2\widehat{T}+1}(\theta^{2\widehat{T}}) \equiv 0.$$

Also observe that, for any $t = 1, ..., \widehat{T}-1$,

$$\Pi_{t+1}(\theta^t) = V_{t+1}(\theta^{t+1}),$$

whereas, for $t = \widehat{T}$,

$$\Pi_{\widehat{T}+1}(\theta^t) = \int V_{\widehat{T}+1}(\theta_1, \theta_2) dF_2(\theta_2|\theta_1, \overline{y}(\theta_1)).$$

The local IC conditions for this 40-period economy are

$$\frac{\partial V_1(\theta_1)}{\partial \theta_1} = -\sum_{s=1}^{\widehat{T}} \beta^{s-1}\psi_\theta(y_s(\theta_1), \theta_1) - \beta^{\widehat{T}}\mathbb{E}^{\lambda[\chi]|\theta_1} \left[I_1^2(\tilde{\theta}, \overline{y}(\theta_1)) \sum_{s=1}^{\widehat{T}} \beta^{s-1}\psi_\theta(y_{\widehat{T}+s}(\theta_1, \tilde{\theta}_2), \tilde{\theta}_2) \right]$$

almost all θ_1, and

$$\frac{\partial V_{\widehat{T}+1}(\theta_1, \theta_2)}{\partial \theta_2} = -\sum_{s=\widehat{T}+1}^{2\widehat{T}} \beta^{s-\widehat{T}-1}\psi_\theta(y_s(\theta_1, \theta_2), \theta_2) = -\sum_{s=1}^{\widehat{T}} \beta^{s-1}\psi_\theta(y_{\widehat{T}+s}(\theta_1, \theta_2), \theta_2)$$

all θ_1, almost all $\theta_2 \in Supp[F_2(\cdot|\theta_1, \overline{y}(\theta_1)]$. Note that $\lambda[\chi]|\theta_1$ is the probability measure over $\theta \equiv (\theta_1, \theta_2)$ given θ_1 and $\overline{y}(\theta_1)$.

Relaxed program in the 40-period economy Denoting by $C(x) \equiv v^{-1}(x)$, we have that the planner's relaxed problem can be described as follows:

$$\max_{\substack{y_t(\cdot), V_t(\cdot), \Pi_{\widehat{T}+1}(\cdot), \\ Z_{\widehat{T}+1}(\cdot), t=1,...,\widehat{T}}} \int \{ \sum_{s=1}^{\widehat{T}-1} \beta^{s-1} \left[y_s(\theta_1) - C\left(V_s(\theta_1) + \psi(y_s(\theta_1), \theta_1) - \beta V_{s+1}(\theta_1) \right) \right]$$

$$+ \beta^{\widehat{T}-1} \left[y_{\widehat{T}}(\theta_1) - C\left(V_{\widehat{T}}(\theta_1) + \psi(y_{\widehat{T}}(\theta_1), \theta_1) - \beta \Pi_{\widehat{T}+1}(\theta_1) \right) \right] + \beta^{\widehat{T}} Q_2(\theta_1, \overline{y}(\theta_1), \Pi_{\widehat{T}+1}(\theta_1), Z_{\widehat{T}+1}(\theta_1)) \} dF_1(\theta_1)$$

subject to

$$\int V_1(s) q(s) dF_1(s) - \kappa = 0 \tag{A.45}$$

and

$$\frac{\partial V_1(\theta_1)}{\partial \theta_1} = - \sum_{s=1}^{\widehat{T}} \beta^{s-1} \psi_\theta(y_s(\theta_1), \theta_1) + \beta^{\widehat{T}} Z_{\widehat{T}+1}(\theta_1), \tag{A.46}$$

where

$$Q_2(\theta_1, \overline{y}(\theta_1), \Pi_{\widehat{T}+1}(\theta_1), Z_{\widehat{T}+1}(\theta_1)) \equiv$$

$$\max_{\substack{y_{\widehat{T}+t}(\theta_1, \cdot), \\ V_{\widehat{T}+t}(\theta_1, \cdot), \\ t=1,...,\widehat{T}}} \int \sum_{s=1}^{\widehat{T}} \beta^{s-1} \left\{ y_{\widehat{T}+s}(\theta) - C(V_{\widehat{T}+s}(\theta) + \psi(y_{\widehat{T}+s}(\theta), \theta_2) - \beta V_{\widehat{T}+s+1}(\theta) \right\} dF_2(\theta_2 \mid \theta_1, \overline{y}(\theta_1))$$

subject to

$$V_{2\widehat{T}+1}(\theta) = 0,$$

$$\Pi_{\widehat{T}+1}(\theta_1) = \int V_{\widehat{T}+1}(\theta) dF_2(\theta_2 \mid \theta_1, \overline{y}(\theta_1)), \tag{A.47}$$

$$Z_{\widehat{T}+1}(\theta_1) = -\mathbb{E}^{\lambda[\chi]|\theta_1} \left[I_1^2(\tilde{\theta}, \overline{y}(\theta_1)) \sum_{s=1}^{\widehat{T}} \beta^{s-1} \psi_\theta(y_{\widehat{T}+s}(\theta_1, \tilde{\theta}_2), \tilde{\theta}_2) \right], \tag{A.48}$$

and

$$\frac{\partial V_{\widehat{T}+1}(\theta_1, \theta_2)}{\partial \theta_2} = - \sum_{s=1}^{\widehat{T}} \beta^{s-1} \psi_\theta(y_{\widehat{T}+s}(\theta_1, \theta_2), \theta_2). \tag{A.49}$$

The above is a two-stage optimal control problem. In the first problem, the controls are $(y_1(\theta_1), ..., y_{\widehat{T}}(\theta_1), V_2(\theta_1), ..., V_{\widehat{T}}(\theta_1), \Pi_{\widehat{T}+1}(\theta_1), Z_{\widehat{T}+1}(\theta_1))$, while the state variable is $V_1(\theta_1)$. In the second problem, the controls are $(y_{\widehat{T}+1}(\theta), ..., y_{2\widehat{T}}(\theta), V_{\widehat{T}+2}(\theta), ..., V_{2\widehat{T}}(\theta))$ while the state variable is $V_{\widehat{T}+1}(\theta)$.

Also note that the first-best allocations solve a similar problem but without the local IC constraints. Thus, the FB allocations can be read from the SB allocations characterized below by setting the costate variables associated with these constraints to zero.

Solution to the relaxed program in the 40-period economy As usual, we proceed backwards, by solving first for the policies that correspond to the second block. Let $\pi_2(\theta_1)$ and $\xi_2(\theta_1)$ be the multipliers associated with the two integral constraints (A.47) and (A.48) and $\mu_2(\theta_1, \theta_2)$ the costate variable for the law of motion of $V_{\widehat{T}+1}(\theta_1, \theta_2)$.

Along with (A.47), (A.48) and (A.49), the following are necessary optimality conditions:

$$\frac{1}{v'(c_{\widehat{T}+s}(\theta))} = \frac{1}{v'(c_{\widehat{T}+s+1}(\theta))}, \text{ for all } s = 1, ..., \widehat{T} - 1, \tag{A.50}$$

$$1 - \frac{\psi_y(y_{\widehat{T}+s}(\theta), \theta_2)}{v'(c_{\widehat{T}+s}(\theta))} - \mu_2(\theta)\frac{\psi_{\theta y}(y_{\widehat{T}+s}(\theta), \theta_2)}{f_2(\theta_2 \mid \theta_2, \overline{y}(\theta_1))} + \xi_2(\theta_1) I_1^2(\theta, \overline{y}(\theta_1))\psi_{\theta y}(y_{\widehat{T}+s}(\theta), \theta_2) = 0, \text{for all } s = 1, ..., \widehat{T}, \tag{A.51}$$

$$\frac{\partial \mu_2(\theta)}{\partial \theta_2} = f_2(\theta_2 \mid \theta_1, \overline{y}(\theta_1)) \cdot \left\{ \frac{1}{v'(c_{\widehat{T}+1}(\theta))} + \pi_2(\theta_1) \right\}, \tag{A.52}$$

along with the boundary conditions

$$\mu_2(\theta_1, \underline{\theta}_2) = 0, \tag{A.53}$$

$$\mu_2(\theta_1, \overline{\theta}_2) = 0, \tag{A.54}$$

where, for $s = 1, ..., \widehat{T}$,

$$c_{\widehat{T}+s}(\theta) = C(V_{\widehat{T}+s}(\theta) + \psi(y_{\widehat{T}+s}(\theta), \theta_2) - \beta V_{\widehat{T}+s+1}(\theta))$$

and

$$V_{2\widehat{T}+1}(\theta) = 0.$$

Conditions (A.50) and (A.51) imply that $c_{\widehat{T}+1} = c_{\widehat{T}+2} = ... = c_{2\widehat{T}}$ and, $y_{\widehat{T}+1} = y_{\widehat{T}+2} = ... = y_{2\widehat{T}}$. It is then immediate to see that the policies that solve the above conditions coincide with the period-2 policies that solve the relaxed program in the two-period model. That is, for any $s = 1, ..., \hat{T}$, $(c_{\widehat{T}+s}(\theta_1, \theta_2), y_{\widehat{T}+s}(\theta_1, \theta_2)) = (c_2^{2pm}(\theta_1, \theta_2), y_2^{2pm}(\theta_1, \theta_2))$. Furthermore, the continuation utility at the beginning of period $\widehat{T} + 1$ satisfies

$$V_{\widehat{T}+1}(\theta) = (1 + \beta + ... + \beta^{\widehat{T}-1})V_2^{2pm}(\theta), \tag{A.55}$$

where $V_2^{2pm}(\theta)$ is the continuation utility in the two-period model.

Next, consider the choice of the policies for the first block. Let $\mu_1(\theta_1)$ be the costate variable associated with the constraint (A.46) and π_1 the multiplier associated with the redistribution constraint (A.45). In addition to (A.45) and (A.46), the following optimality conditions must hold:

$$\frac{1}{v'(c_s(\theta_1))} = \frac{1}{v'(c_{s+1}(\theta_1))}, \text{ for all } s = 1, ..., \widehat{T} - 1, \tag{A.56}$$

$$1 - \tfrac{\psi_y(y_s(\theta_1), \theta_1)}{v'(c_s(\theta_1))} + \beta^{\widehat{T}} \tfrac{\partial}{\partial \overline{y}} \int \left\{ y_{\widehat{T}+1}(\theta) - c_{\widehat{T}+1}(\theta) - \pi_2(\theta_1)\tfrac{V_{\widehat{T}+1}(\theta)}{\sum_{s=1}^{\widehat{T}} \beta^{s-1}} \right\} dF_2(\theta_2 \mid \theta_1, \overline{y}(\theta_1))$$

$$+\beta^{\widehat{T}} \xi_2(\theta_1) \tfrac{\partial}{\partial \overline{y}} \int I_1^2(\theta, \overline{y}(\theta_1))\psi_\theta(y_{\widehat{T}+1}(\theta), \theta_2) dF_2(\theta_2 \mid \theta_1, \overline{y}(\theta_1)) - \mu_1(\theta_1) \tfrac{\psi_{\theta y}(y_s(\theta_1), \theta_1)}{f_1(\theta_1)} = 0, \text{ for } s = 1, ..., \widehat{T}, \tag{A.57}$$

$$\frac{\partial \mu_1(\theta_1)}{\partial \theta_1} = f_1(\theta_1) \cdot \left\{ \frac{1}{v'(c_1(\theta_1))} - \pi_1 q(\theta_1) \right\}, \tag{A.58}$$

$$\frac{1}{v'(c_1(\theta_1))} + \pi_2(\theta_1) = 0, \tag{A.59}$$

$$\mu_1(\theta_1) + \xi_2(\theta_1) f_1(\theta_1) = 0, \tag{A.60}$$

along with the boundary conditions

$$\mu_1(\underline{\theta}_1) = 0, \tag{A.61}$$

and

$$\mu_1(\overline{\theta}_1) = 0, \tag{A.62}$$

where, for $s = 1, ..., \widehat{T}$,

$$c_s(\theta_1) = C\left(V_s(\theta_1) + \psi(y_s(\theta_1), \theta_1) - \beta V_{s+1}(\theta_1)\right).$$

Note that, when writing the FOCs with respect to $\Pi_{\widehat{T}+1}(\theta_1)$, $Z_{\widehat{T}+1}(\theta_1)$, and $y_s(\theta_1)$, we have used the properties that $\frac{\partial Q_2}{\partial \Pi_{\widehat{T}+1}} = \pi_2(\theta_1)$ and $\frac{\partial Q_2}{\partial Z_{\widehat{T}+1}} = \xi_2(\theta_1)$, along with the fact that

$$\begin{array}{c} \frac{\partial Q_2}{\partial \overline{y}} = \frac{\partial}{\partial \overline{y}} \int \sum_{s=1}^{\widehat{T}} \beta^{s-1} \left\{ y_{\widehat{T}+s}(\theta) - c_{\widehat{T}+s}(\theta) \right\} dF_2(\theta_2 \mid \theta_1, \overline{y}(\theta_1)) \\ -\pi_2(\theta_1) \frac{\partial}{\partial \overline{y}} \int V_{\widehat{T}+1}(\theta) dF_2(\theta_2 \mid \theta_1, \overline{y}(\theta_1)) \\ +\xi_2(\theta_1) \frac{\partial}{\partial \overline{y}} \int I_1^2(\theta, \overline{y}(\theta_1)) \sum_{s=1}^{\widehat{T}} \beta^{s-1} \psi_\theta(y_{\widehat{T}+s}(\theta), \theta_2) dF_2(\theta_2 \mid \theta_1, \overline{y}(\theta_1)) \end{array}$$

and

$$\frac{\partial \overline{y}}{\partial y_s} = \frac{\beta^{s-1}}{1 + \beta + ... + \beta^{\widehat{T}-1}}.$$

We also used the property we identified above that consumption and earnings allocations are constant over the second block.

Clearly, (A.56) and (A.57) imply that $c_1 = c_2 = ... = c_{\widehat{T}}$ and $y_1 = y_2 = ... = y_{\widehat{T}}$. Given this property and (A.55) above, we have that the necessary conditions in the above program reduce to the same conditions for the period-1 policies in the relaxed program in the two-period model, with $\delta = \beta^{\widehat{T}}$. We conclude that consumption and earnings in the first block of the 40-period economy are given by the period-1 consumption and earnings polices in the two-period model. That is, for any $s = 1, ..., \hat{T}$, $(c_s(\theta_1), y_s(\theta_1)) = (c_1^{2pm}(\theta_1), y_1^{2pm}(\theta_1))$. Furthermore, the lifetime expected utility of each worker with productivity equal to θ_1 in the first block is given by

$$V_1(\theta_1) = (1 + \beta + ... + \beta^{\widehat{T}-1}) V_1^{2pm}(\theta_1),$$

where $V_1^{2pm}(\theta_1)$ is the lifetime expected utility in the two-period model.

Sufficiency in the 40-period economy We conclude by showing that, when the solution to the relaxed program in the two-period model satisfies all the integral monotonicity constraints of the two-period model, then the solution to the relaxed program in the 40-period model (which, by virtue of the results above, consists of the repetition over each of the two blocks of the corresponding policies in the two-period model) also satisfies all the corresponding integral-monotonicity conditions in the 40-period model.

Let $(c_t^{2pm}(\theta), y_t^{2pm}(\theta))_{t=1,2}$ denote the solution to the relaxed program in the two-period model. Assume these policies satisfy the following integral monotonicity constraints: for any pair $\theta_2, \hat{\theta}_2 \in \Theta_2$ and any pair $\theta_1, \hat{\theta}_1 \in \Theta_1$,

$$\int_{\hat{\theta}_2}^{\theta_2} \psi_\theta(y_2^{2pm}(\theta_1, r), r)dr \leq \int_{\hat{\theta}_2}^{\theta_2} \psi_\theta(y_2^{2pm}(\theta_1, \hat{\theta}_2), r)dr \tag{A.63}$$

and

$$\begin{aligned} &\int_{\hat{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1^{2pm}(r), r) + \delta \int \left[I_1^2((r, \theta_2), y_1^{2pm}(r))\psi_\theta(y_2^{2pm}(r, \theta_2), \theta_2) \right] dF_2(\theta_2 | r, y_1^{2pm}(r)) \right\} dr \\ \leq \; &\int_{\hat{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1^{2pm}(\hat{\theta}_1), r) + \delta \int \left[I_1^2((r, \theta_2), y_1^{2pm}(\hat{\theta}_1))\psi_\theta(y_2^{2pm}(\hat{\theta}_1, \theta_2), \theta_2) \right] dF_2(\theta_2 | r, y_1^{2pm}(\hat{\theta}_1)) \right\} dr. \end{aligned}$$

Next observe that the integral monotonicity conditions in the 40-period model require that, for any pair $\theta_2, \hat{\theta}_2 \in \Theta_2$ and any pair $\theta_1, \hat{\theta}_1 \in \Theta_1$,

$$\int_{\hat{\theta}_2}^{\theta_2} \sum_{s=1}^{\widehat{T}} \beta^{s-1} \psi_\theta(y_{\widehat{T}+s}(\theta_1, r), r)dr \leq \int_{\hat{\theta}_2}^{\theta_2} \sum_{s=1}^{\widehat{T}} \beta^{s-1} \psi_\theta(y_{\widehat{T}+s}(\theta_1, \hat{\theta}_2), r)dr,$$

and

$$\int_{\hat{\theta}_1}^{\theta_1} \left\{ \sum_{s=1}^{\widehat{T}} \beta^{s-1}\psi_\theta(y_s(r), r) + \beta^{\widehat{T}} \int I_1^2((r, \theta_2), \overline{y}(r)) \sum_{s=1}^{\widehat{T}} \beta^{s-1}\psi_\theta(y_{\widehat{T}+s}(r, \theta_2), \theta_2) dF_2(\theta_2 | r, \overline{y}(r)) \right\} dr$$

$$\leq \int_{\hat{\theta}_1}^{\theta_1} \left\{ \sum_{s=1}^{\widehat{T}} \beta^{s-1}\psi_\theta(y_s(\hat{\theta}_1), r) + \beta^{\widehat{T}} \int I_1^2((r, \theta_2), \overline{y}(\hat{\theta}_1)) \sum_{s=1}^{\widehat{T}} \beta^{s-1}\psi_\theta(y_{\widehat{T}+s}(\hat{\theta}_1, \theta_2), \theta_2) dF_2(\theta_2 | r, \overline{y}(\hat{\theta}_1)) \right\} dr.$$

It is easy to see that when, for any $s = 1, ..., \hat{T}$, $(c_s(\theta_1), y_s(\theta_1)) = (c_1^{2pm}(\theta_1), y_1^{2pm}(\theta_1))$ and $(c_{\widehat{T}+s}(\theta_1, \theta_2), y_{\widehat{T}+s}(\theta_1, \theta_2)) = (c_2^{2pm}(\theta_1, \theta_2), y_2^{2pm}(\theta_1, \theta_2))$, and $\overline{y}(\theta_1) = y_1^{2pm}(\theta_1)$ and $\delta = \beta^{\widehat{T}}$, the above integral monotonicity conditions reduce to their counterparts in the two-period economy. Hence, when the solution to the relaxed program also solves the full program in the two-period economy, the same is true in the 40-period economy.

A.3.3 Allocations under history-independent tax codes in the 40-period economy

Finally we show that, when agents face tax schedules which may depend on the age block but are invariant to past income levels (which is the case when the tax code is the one that approximates the current US code, the quasi-optimal code, or the linear code discussed in the main body), consumption and labor supply are constant over each of the two blocks and coincide with the corresponding levels in the two-period version of the same economy when the discount factor is given by $\delta = \beta^{\hat{T}}$.

To see this, suppose that, in each period $s = 1, ..., \widehat{T}$, workers face a tax schedule $\mathcal{T}_1(y_s)$, whereas in periods $s = \widehat{T} + 1, ..., 2\widehat{T}$, they face a tax schedule $\mathcal{T}_2(y_s)$. They then choose consumption and earnings in each period to maximize their expected continuation utility subject to the budget constraint

$$c_s = y_s - \mathcal{T}_t(y_s) + \frac{S_s}{\beta} - S_{s+1},$$

where $\mathcal{T}_t(y_s) = \mathcal{T}_1(y_s)$ if $s \leq \widehat{T}$, and $\mathcal{T}_t(y_s) = \mathcal{T}_2(y_s)$ if $\widehat{T} < s \leq 2\widehat{T}$. The variable S_{s+1} represents the balance in the worker's savings account at the end of period s, with S_1 pre-determined and

$$S_{2\widehat{T}+1} = 0.$$

Importantly, note that the (after-tax) return on savings is equal to the inverse of the annual discount factor, that is, $(1 + r(1 - \tau_{capital})) = 1/\beta$, where $\tau_{capital}$ denotes the capital tax rate.

The optimal allocations then solve the following necessary conditions:

$$v'(c_s(\theta^s)) = v'(c_{s+1}(\theta^s)), \text{ for all } s = 1, ..., \widehat{T} - 1, \widehat{T} + 1, ..., 2\widehat{T} - 1,$$

$$v'(c_{\widehat{T}}(\theta_1)) = \int v'(c_{\widehat{T}+1}(\theta_1, \theta_2)) dF_2(\theta_2|\theta_1, \overline{y}(\theta_1)), \tag{A.64}$$

$$1 - \mathcal{T}_2'(y_s(\theta^s)) = \frac{\psi_y(y_s(\theta^s), \theta_s)}{v'(c_s(\theta^s))}, \text{ for all } s = \widehat{T} + 1, ..., 2\widehat{T}, \tag{A.65}$$

and

$$\begin{aligned} &v'(c_s(\theta^s))[1 - \mathcal{T}_1'(y_s(\theta^s))] \\ +&\frac{\beta^{\widehat{T}}}{\sum_{s=1}^{\widehat{T}} \beta^{s-1}} \frac{\partial}{\partial \overline{y}(\theta_1)} \int \left\{ \sum_{t=1}^{\widehat{T}} \beta^{t-1} [v(c_{\widehat{T}+t}(\theta_1, \theta_2)) - \psi(y_{\widehat{T}+t}(\theta_1, \theta_2), \theta_2)] \right\} dF_2(\theta_2|\theta_1, \overline{y}(\theta_1)) \\ &= \psi_y(y_s(\theta^s), \theta_s), \text{ for all } s = 1, ..., \widehat{T}. \end{aligned} \tag{A.66}$$

Note that Condition (A.66) uses the fact that

$$\frac{\partial \overline{y}(\theta_1)}{\partial y_s} = \frac{\beta^{s-1}}{\sum_{s=1}^{\widehat{T}} \beta^{s-1}}.$$

Clearly, workers choose the same consumption within each block of their working life, with consumption across the two blocks satisfying the standard Euler condition (A.64). Given this, Condition (A.65) implies that output is constant in the second block. In turn, Condition (A.66) implies that output is also constant over the first block. Given the above observations, it is then immediate to see that consumption and output decisions in this multi-period economy coincide with their counterparts in the two-period version of the same economy in which $\delta = \beta^{\hat{T}}$.

B Comparative Statics

B.1 Risk aversion

To facilitate the comparison with the calibrated economy in the paper, consider an economy where (a) the function $v(c)$ describing the agents' preferences for consumption smoothing is CRRA with

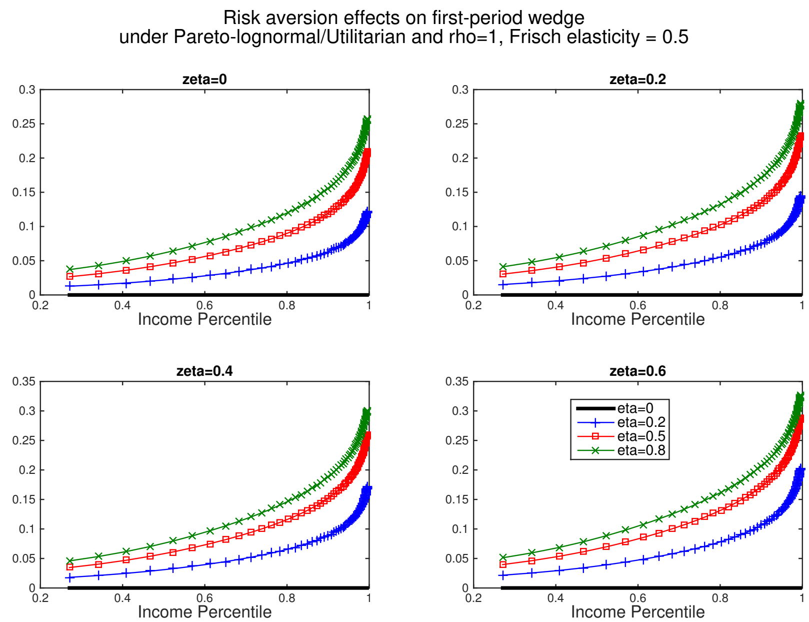

Figure B.1: Period-1 wedges in the Utilitarian Pareto-lognormal case

coefficient of relative-risk aversion η, (b) the planner's preferences for redistribution take the familiar Utilitarian specification $q(\theta_1) = 1$ for all θ_1, and (c) both the period-1 productivity θ_1 and the period-2 shock ε_2 are drawn from a Pareto-lognormal distribution.

Figure B.1 (which is the same as Figure 2 in the main body) depicts the wedges for four different levels of the coefficient of relative risk aversion, namely for $\eta = 0$, $\eta = 0.2$, $\eta = 0.5$, and $\eta = 0.8$ and for four different levels of LBD, namely for $\zeta = 0$, $\zeta = 0.2$, $\zeta = 0.4$, and $\zeta = 0.6$.[12] As is well known, under a Utilitarian welfare objective, in the absence of LBD, wedges are identically equal to zero when $\eta = 0$, i.e., when agents are risk neutral. The same remains true with LBD. Interestingly, as the figure shows, higher degrees of risk aversion contribute to higher and more progressive period-1 wedges, across all intensities of LBD. As explained in the main text, the reason why wedges increase with the degree of risk aversion η is that the cost of asking type-θ_1 agents for higher effort, accounting for the effects that this increase has on the rents of types $\theta_1' > \theta_1$, is higher the lower the marginal utility of consumptions of types $\theta_1' > \theta_1$ relative to the marginal utility of type θ_1. That the progressivity of the period-1 wedges also increases with the degree of risk aversion seems to originate in the combination of the specific distribution from which the productivity shocks are drawn along with the property that the increase in the cost of providing rents to types $\theta_1' > \theta_1$ due to the higher degree of risk aversion is stronger the higher θ_1 is. This in turn follows from the fact that the conditional average inverse marginal utility of consumption of types $\theta_1' > \theta_1$, that is, $\int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} \frac{dF_1(s)}{1-F_1(\theta_1)}$, is increasing in θ_1.

In the absence of LBD, the reason why risk aversion contributes to higher period-1 wedges is

[12]Section C in this supplement contains details about the numerical computations behind the figures below as well as those in Section 4 of the paper.

that, when the agents are risk averse, the extra compensation the planner must provide to all types above θ_1 when the latter type is asked to work more is higher than when the agents are risk neutral. This is because all types above θ_1 consume more than type θ_1 and hence have a lower marginal utility of consumption. The extra compensation the planner must provide to all types above θ_1 to dissuade them from mimicking type θ_1 is thus higher than under risk neutrality. As a result, the planner further distorts downwards type θ_1's labor supply. Furthermore, the stronger the agents' risk aversion, the larger the distortions.

Under LBD, the effects of risk aversion on period-1 wedges are more sophisticated, as one has to account also for the effects of risk aversion on period-2 policies. To gain more insights about the interaction of risk aversion with LBD in the determination of the optimal wedges, Figures B.2, B.3, and B.4 focus on the components of the period-1 wedges that are specific to LBD. In particular, Figure B.2 illustrates the effect of risk aversion on the Ω term. To understand the figure, recall that Ω measures the variation in the expected net present value of future information rents triggered by a variation in period-1 labor supply, due to LBD, in the benchmark economy with risk-neutral agents and Rawlsian preferences for redistribution. A higher degree of risk aversion contributes to higher costs to the planner of incentivizing period-2 labor supply. As a result, the planner optimally responds to higher degrees of risk aversion by reducing the agents' period-2 labor supply. In turn, this reduces the expected net present value of period-2 rents. The benefit of shifting the distribution of period-2 productivity towards lower levels to reduce the expected net present value of period-2 rents may thus decrease with the agents' risk aversion. Higher degrees of risk aversion may thus contribute to lower levels of Ω. Figure B.2 shows that this is indeed the case under the assumed specification. Furthermore, because highly productive agents have a lower marginal utility of consumption than less productive ones, the reduction in period-2 labor supply can be most pronounced at the top of the period-2 distribution. Because productivity is serially correlated, in turn this implies that the reduction in Ω can be stronger for high income percentiles. Risk aversion may thus also contribute to a reduction in the progressivity of Ω, as can be seen from Figure B.2.

Next, consider the effects of risk aversion on the correction term

$$RA(\theta_1) - D(\theta_1) \equiv v'(c_1(\theta_1)) \left[\int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} \frac{dF_1(s)}{1-F_1(\theta_1)} - \int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} dF_1(s) \right].$$

As discussed in the paper, the term

$$RA(\theta_1) = v'(c_1(\theta_1)) \int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} \frac{dF_1(s)}{1-F_1(\theta_1)}$$

controls for the effects of the heterogeneity in the agents' marginal utility of consumption on the planner's costs of increasing the compensation provided to the highly productive period-1 agents to dissuade them from mimicking the less productive ones. The term

$$D(\theta_1) \equiv v'(c_1(\theta_1)) \int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} dF_1(s),$$

on the other hand, controls for the benefits of increasing the agents' lifetime expected utility through the effects that this increase has on the redistribution constraint. Recall from the discussion in the

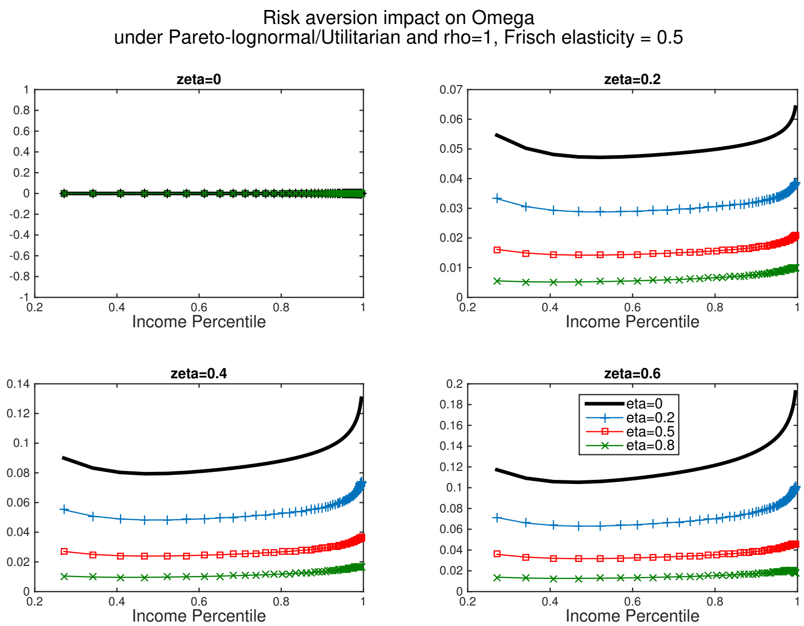

Figure B.2: The LBD term Ω in the Utilitarian Pareto-lognormal case

main text that the term $\int_{\underline{\theta}_1}^{\overline{\theta}_1} \frac{1}{v'(c_1(s))} dF_1(s)$ is the shadow benefit of providing all agents with a higher util, in consumption terms. The higher the agents' degree of risk aversion, the higher the amount of fiscal resources the planner can save by providing the extra util while respecting the redistribution constraint.

Observe that the term $v'(c_1(\theta_1))$ in the expression for $RA-D$ controls for type θ_1's own marginal utility of consumption. The lower $v'(c_1(\theta_1))$ is, the smaller the benefit of shifting type θ_1's distribution of period-2 productivity towards levels that command lower period-2 rents. The term

$$\int_{\theta_1}^{\overline{\theta}_1} \frac{1}{v'(c_1(s))} \frac{dF_1(s)}{1-F_1(\theta_1)} - \int_{\underline{\theta}_1}^{\overline{\theta}_1} \frac{1}{v'(c_1(s))} dF_1(s), \tag{B.1}$$

instead, controls for the average inverse marginal utility of consumption among those workers whose compensation must be adjusted when the planner asks for higher output to type θ_1, relative to the unconditional average across all period-1 types. The larger this term, the larger the benefit of distorting type θ_1's period-1 labor supply so as to economize on future costs of incentives. In general, the aforementioned two parts move in opposite directions when the degree of risk aversion increases, or when the period-1 productivity threshold θ_1 increases. Depending on which of these effect prevails, the correction term $RA-D$ may thus contribute to an amplification or to a dampening of the LBD effects captured by the Ω term. Under the parameters' specification in the figure, the second channel prevails, and hence higher degrees of risk aversion contribute to a higher correction term. As Figure B.3 shows, this is true across all income percentiles and irrespective of the intensity of LBD. The figure also shows that the correction is monotone in the income percentile or, equivalently, in the agents' period-1 productivity. Again, this is because, under the parameters' configuration in the

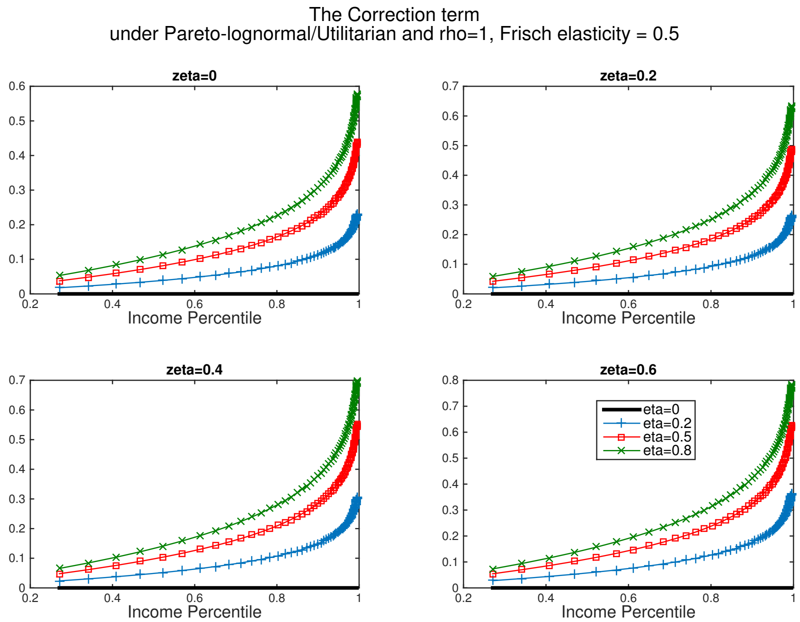

Figure B.3: The $RA - D$ correction term in the Utilitarian Pareto-lognormal case

figure, the second effect discussed above, as captured by the term (B.1) in $RA - D$, dominates over the effect that an increase in θ_1 has on $v'(c_1(\theta_1))$. As a result, under the considered specification, the correction term $RA - D$ contributes to an amplification of the LBD effects on the progressivity of the period-1 wedges.

Finally, Figure B.4 illustrates the net contribution $[RA - D]\Omega$ of LBD to the period-1 wedges. As the figure shows, the net effect of risk aversion on the corrected term $[RA - D]\Omega$ is in general ambiguous as it depends on which of the forces discussed above prevails.

B.2 Frisch elasticity

Next we provide some comparative statics with respect to the Frisch elasticity. Recall that the latter is captured by the inverse of the parameter ϕ in the agents' disutility of labor. We restrict attention again to economies where the productivity shocks are drawn from a Pareto-lognormal distribution, the degree of skill persistence is $\rho = 1$, the agents are risk averse with CRRA preferences for consumption with coefficient of relative-risk aversion $\eta = 0.5$, and the planner has Utilitarian preferences for redistribution.

Figure B.5 depicts the period-1 wedges for three different levels of the Frisch elasticity, namely, for $1/\phi = 0.3$, $1/\phi = 0.5$, and $1/\phi = 0.7$, and for four different levels of LBD, namely, for $\zeta = 0$, $\zeta = 0.2$, $\zeta = 0.4$, and $\zeta = 0.6$.[13] In the absence of LBD, wedges are decreasing in the Frisch elasticity. The reason is that, when the elasticity is high, small wedges suffice to induce a major

[13] Section C in this Supplementary Material contains details about the numerical computations behind the figures below.

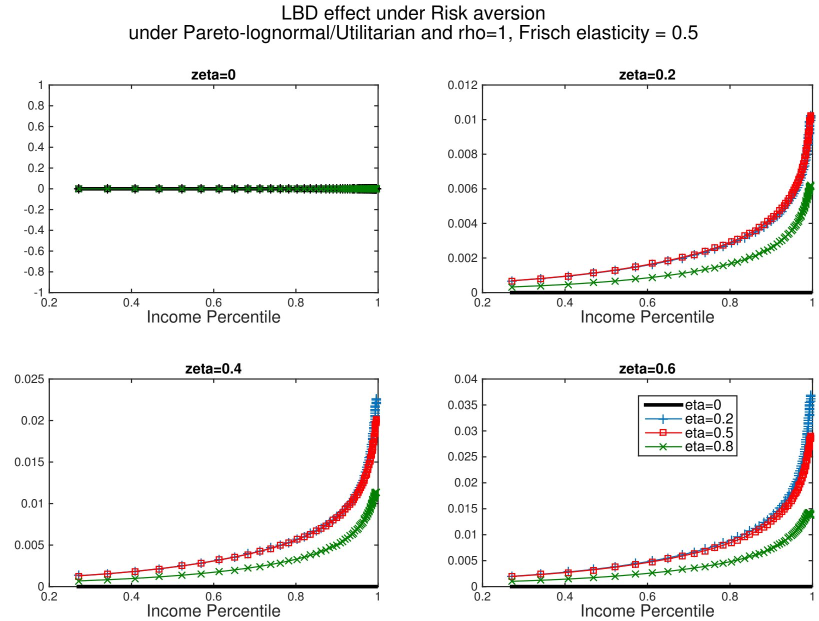

Figure B.4: Risk-adjusted LBD term $[RA - D] \cdot \Omega$ in the Utilitarian Pareto-lognormal case

reduction in the agents' labor supply, and hence in the information rents the planner must leave to highly productive types to induce them to reveal their private information.

Under LBD, the effects of variations in the Frisch elasticity on period-1 wedges are more sophisticated. To gain more insights into the interaction of the Frisch elasticity with LBD in the determination of the optimal wedges, we show how the various terms affected by LBD, Ω, $[RA-D]$, and $[RA-D]\cdot\Omega$, change with the Frisch elasticity.

We start with the term Ω. Observe that, for the functional forms used in the figures, the second-period handicaps are proportional to $-\theta_2\psi_\theta(y_2(\theta),\theta_2) = (1+\phi)\psi(y_2(\theta),\theta_2)$. An increase in the Frisch elasticity (i.e., a decrease in ϕ) has therefore the following two effects on the second-period handicaps (for any given period-1 productivity) and, thereby, on the LBD term Ω. First, holding constant the second-period output schedule, it lowers the second-period handicaps and hence it dampens the LBD effects captured by the Ω term. Second, for high levels of period-2 productivity, it increases period-2 effort, and hence it contributes to higher period-2 handicaps.[14] This second channel contributes to stronger LBD effects and hence to a larger Ω term. Figure B.6 shows that, for the parameters' specification considered in the figures, the second channel prevails, and hence a higher Frisch elasticity leads to higher Ω. We also note that an increase in the Frisch elasticity raises the period-2 output and, thereby, the expected period-2 handicaps for high levels of period-1

[14] To see this, note first that, for the functional forms used in these comparative statics, under the second-best allocations, $\frac{y_2(\theta)}{\theta_2} = \left(\frac{\theta_2}{1+\frac{\rho(1+\phi)}{\theta_1\gamma_1(\theta_1)}}\right)^{\frac{1}{\phi}}$. The term inside the bracket is increasing in θ_2 and in θ_1. Moreover, $\psi(\frac{y_2(\theta)}{\theta_2}) = \frac{1}{1+\phi}\left(\frac{\theta_2}{1+\frac{\rho(1+\phi)}{\theta_1\gamma_1(\theta_1)}}\right)^{\frac{1+\phi}{\phi}}$ which is decreasing in ϕ for high enough θ_2 and θ_1 (i.e. for θ_2 and θ_1 such that the term inside the bracket is greater or equal to one.)

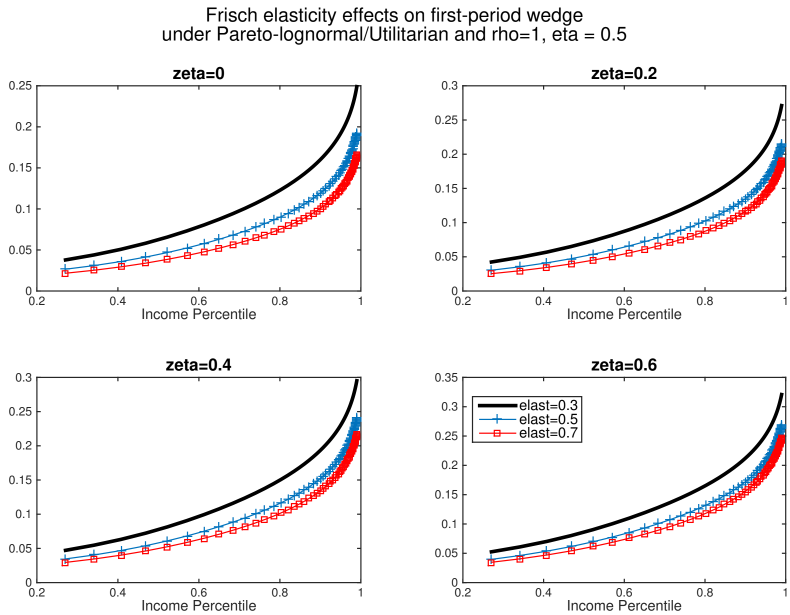

Figure B.5: Period-1 wedges in the Utilitarian Pareto-lognormal case

productivity.[15] This effect, together with the fact that period-2 output is increasing in period-2 productivity, implies that the effects of an increase in the Frisch elasticity on the LBD term Ω can be stronger for highly productive young workers. This is confirmed in Figure B.6.

The effects of an increase in the Frisch elasticity on the correction term $[RA - D]$ are more complicated as they arise through the impact of a higher Frisch elasticity on the profile of first-period consumption. One possible channel is the following. Fix some period-1 productivity θ_1. Under the parameters' specification, a higher Frisch elasticity lowers the second-period wedge, which in turn leads to higher consumption in the second period, for all histories (θ_1, θ_2). As is standard in dynamic models of asymmetric information like ours, consumption smoothing is optimal when agents are risk averse.[16] Thus, for workers with high period-1 productivity θ_1, a higher Frisch elasticity implies higher period-1 consumption and hence lower marginal utility of consumption. It follows then that, when the Frisch elasticity increases, the per-unit valuation of the LBD effect in utility terms decreases, which in turn pushes towards a lower correction term, all else equal. Figure B.7 indicates that this force is dominant: an increase in the Frisch elasticity leads to a dampening of the correction term.

Finally, Figure B.8 illustrates the effect of a variation in the Frisch elasticity on the corrected LBD term $[RA - D]\Omega$. As the figure shows, the comparative statics of the corrected term $[RA -$

[15]See Footnote 13.

[16]To be more precise, as we show in the proof of Proposition 5 in the paper, the following familiar Rogerson inverse-Euler condition holds under the second-best allocations:

$$\frac{1}{v'(c_1(\theta_1))} = \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{1}{v'(c_2(\theta_1, s))} dF_2(s|\theta_1, y_1(\theta_1)).$$

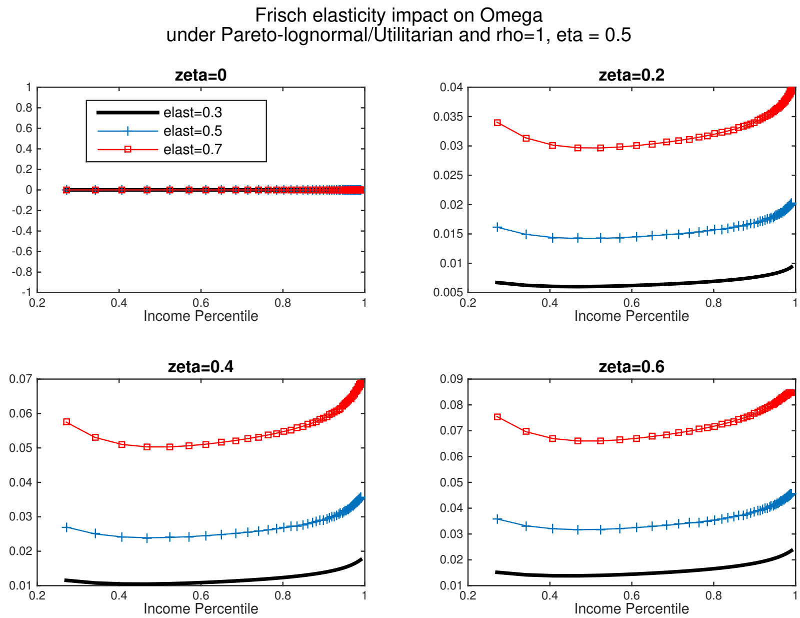

Figure B.6: Effects of Frisch elasticity on LBD term Ω in Utilitarian Pareto-lognormal case

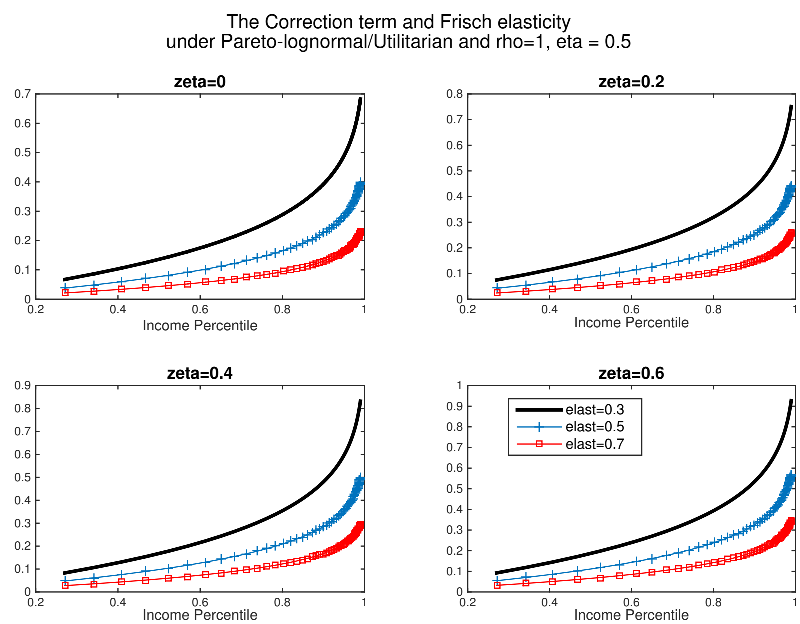

Figure B.7: Effects of Frisch elasticity on $RA - D$ term in Utilitarian Pareto-lognormal case

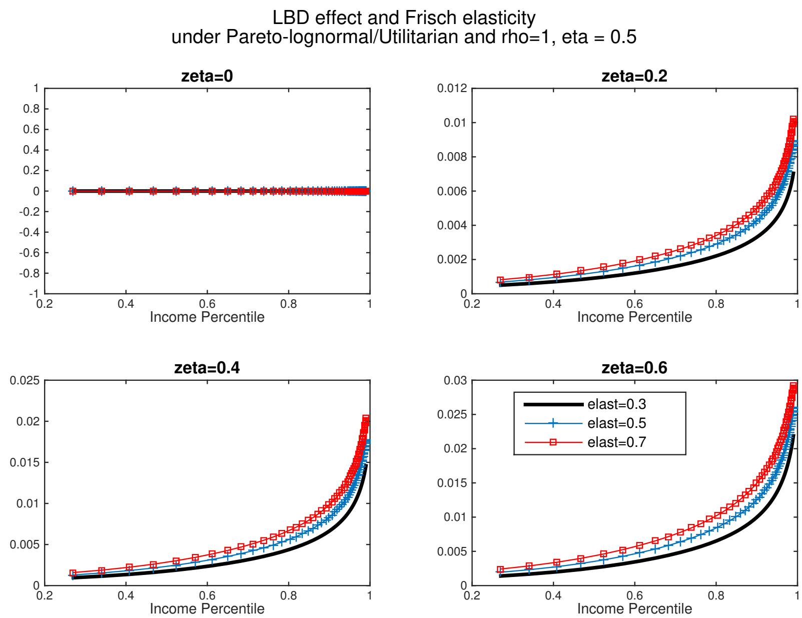

Figure B.8: Effects of Frisch elasticity on corrected LBD term $[RA - D]\Omega$ in Utilitarian Pareto-lognormal case

$D]\Omega$ parallel qualitatively those of the uncorrected term Ω, but are dampened by the opposite comparative statics of the correction term $RA - D$ discussed above.

B.3 Non-linear Pareto Weights

Next we investigate the comparative statics of the period-1 wedges with respect to the planner's preferences for redistribution. We assume the latter are captured by the following class of non-linear Pareto weights:

$$q(\theta_1) = \frac{e^{\mathbf{q}(\underline{\theta}-\theta_1)}}{\int_{\Theta_1} e^{\mathbf{q}(\underline{\theta}-s)} dF_1(s)},$$

where $\mathbf{q} \in \mathbb{R}_+$ is a non-negative scalar, with higher levels of $\mathbf{q}$ corresponding to higher preferences for redistribution. We restrict attention again to economies where the productivity shocks are drawn from a Pareto-lognormal distribution, the degree of skill persistence is $\rho = 1$, the Frisch elasticity is $1/\phi = 0.5$, and the agents are risk averse with CRRA preferences for consumption, with coefficient of relative-risk aversion $\eta = 0.5$.

Figure B.9 depicts the period-1 wedges for five different levels of $\mathbf{q}$, namely, $\mathbf{q} = 0$, $\mathbf{q} = 0.25$, $\mathbf{q} = 0.5$, $\mathbf{q} = 0.75$, and $\mathbf{q} = 1$, and four different levels of LBD, namely, $\zeta = 0$, $\zeta = 0.2$, $\zeta = 0.4$, and $\zeta = 0.6$.[17] As is well known, in the absence of LBD, wedges are higher and more progressive the stronger the planner's preferences for redistribution (i.e., the higher q is). The reason is that,

[17]Section C in this Supplementary Material contains details about the numerical computations behind the figures below.

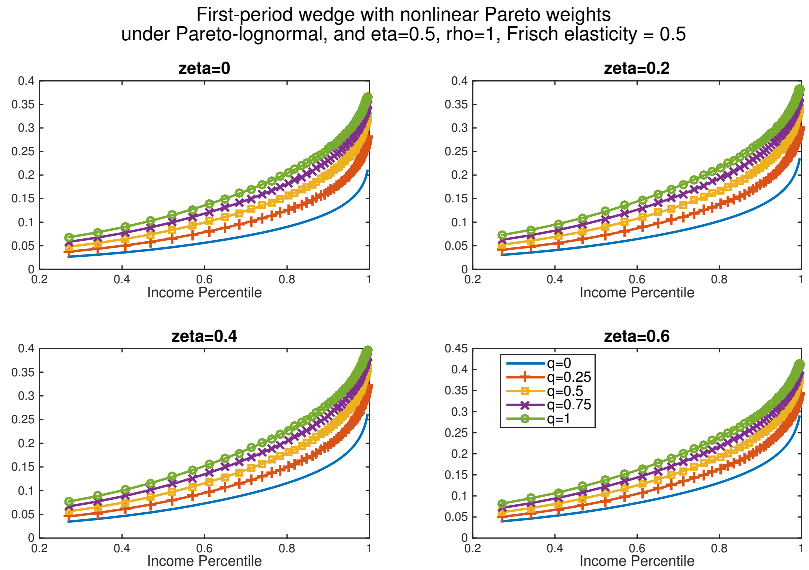

Figure B.9: Period-1 wedges in the Pareto-lognormal case

when q is high, the planner assigns a smaller value to the ex-ante expected utility of agents with high productivity.

Under LBD, the effects of $\mathbf{q}$ on period-1 wedges are more sophisticated. To gain more insights into the interaction of the planner's preferences for redistribution with LBD in the determination of the optimal period-1 wedges, we show how the various terms affected by LBD, Ω, $[RA - D]$, and $[RA - D] \cdot \Omega$, change with $\mathbf{q}$. Figures B.10-B.12 show that, for the parameters in question, the three terms of interest increase with $\mathbf{q}$. This reflects the fact that the average value $\int_{\theta_1}^{\overline{\theta}_1} q(s) \frac{dF_1(s)}{1-F_1(\theta_1)}$ assigned by the planner to an increase in the expected lifetime utility of types above θ_1 is decreasing in $\mathbf{q}$.

That stronger preferences for redistribution (equivalently, lower levels of $\int_{\theta_1}^{\overline{\theta}_1} q(s) \frac{dF_1(s)}{1-F_1(\theta_1)}$) lead, all else equal, to higher levels of the correction term $[RA - D]$ follows directly from the definition of the D term in the main text. That lower levels of $\int_{\theta_1}^{\overline{\theta}_1} q(s) \frac{dF_1(s)}{1-F_1(\theta_1)}$ also lead, all else equal, to higher levels of the LBD term Ω, is more difficult to explain. One can show that, under the specification considered in the simulations,

$$\Omega(\theta_1) = \delta \frac{\rho(1+\phi)}{\theta_1 \gamma_1(\theta_1)} \frac{\frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[\psi(y_2(\theta_1, \tilde{\theta}_2), \tilde{\theta}_2) \right]}{\psi_y(y_1(\theta_1), \theta_1)}.$$

On the one hand, a reduction in $\int_{\theta_1}^{\overline{\theta}_1} q(s) \frac{dF_1(s)}{1-F_1(\theta_1)}$ implies a reduction in period-2 labor supply. To understand why this is the case, recall that, when productivity is correlated over time, a reduction in the agents' period-2 labor supply permits the planner to reduce the rents she must provide to the period-1 types above θ_1 to induce them to reveal their private information. The benefit of reducing these rents is higher when the planner has stronger preferences for redistribution, i.e., when $\mathbf{q}$ is higher, or, equivalently, when $\int_{\theta_1}^{\overline{\theta}_1} q(s) \frac{dF_1(s)}{1-F_1(\theta_1)}$ is lower. The reduction in period-2 labor supply in

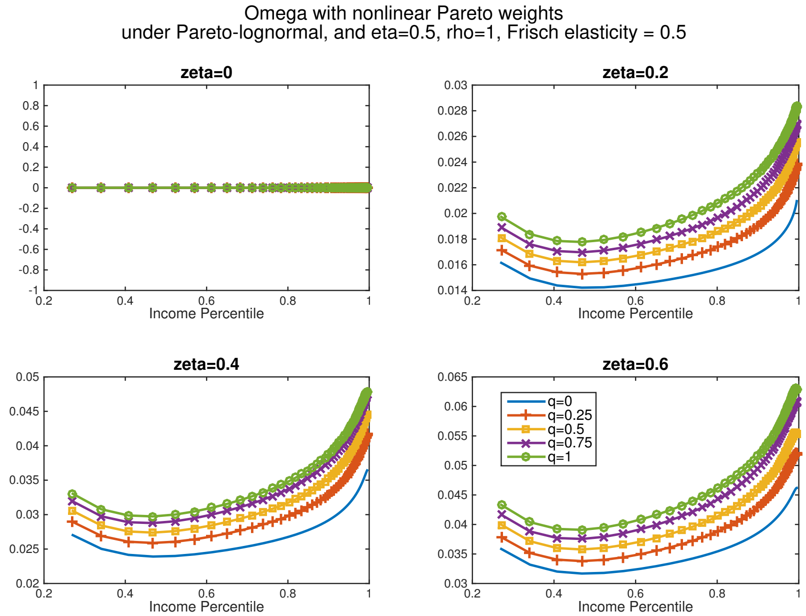

Figure B.10: Effects of intensity of preferences for redistribution on LBD term Ω in Pareto-lognormal case

turn implies a reduction in the expected period-2 handicaps. This reduction may reduce the benefits the planner assigns to shifting the distribution of the period-2 productivity towards lower levels to economize on period-2 rents, as captured by the numerator in $\Omega(\theta_1)$. On the other hand, a stronger preference for redistribution, equivalently, a smaller value of $\int_{\theta_1}^{\bar{\theta}_1} q(s)\frac{dF_1(s)}{1-F_1(\theta_1)}$, also implies a lower value of $y_1(\theta_1)$. To understand this, recall that a reduction in $y_1(\theta_1)$ also permits the planner to reduce the rent that she must provide to period-1 types above θ_1 to induce them to reveal their private information. This second channel is captured by the denominator in $\Omega(\theta_1)$. As Figure B.12 shows, this second channel dominates under the assumed specification, thus making Ω increase with the planner's preferences for redistribution.

C Computational Methods

C.1 Numerical results in Section 4 in the main body and in Section B in the Supplementary Material

For the numerical results in Section 4 in the main body, as well as for the comparative statics results in Section B in this Supplement, we numerically solve the planner's relaxed dual program, as described in Section 4 in the main body. We solve this program as follows. Using the identities $e_1(\theta_1) \equiv y_1(\theta_1)/\theta_1$, $e_2(\theta_1,\varepsilon_2) \equiv \frac{y_2(\theta_1,\theta_1^{\rho}y_1(\theta_1)^{\zeta}\varepsilon_2)}{\theta_1^{\rho}y_1(\theta_1)^{\zeta}\varepsilon}$, and $\hat{c}_2(\theta_1,\varepsilon_2) \equiv c_2(\theta_1,\theta_1^{\rho}y_1(\theta_1)^{\zeta}\varepsilon_2)$, we solve for the policies $e_1(\theta_1)$, $e_2(\theta_1,\varepsilon_2)$, $c_1(\theta_1)$ and $\hat{c}_2(\theta_1,\varepsilon_2)$ that maximize expected tax revenues subject to the redistribution constraint.

For the environments corresponding to Figure 1 in the main body, because the agents' preferences are linear in consumption, the level of promised utility κ in the redistribution constraint, as well as

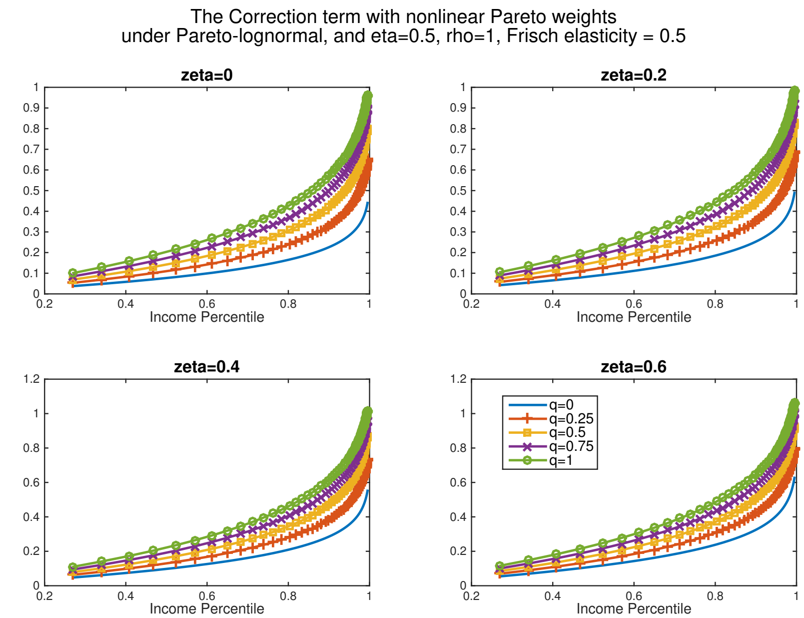

Figure B.11: Effects of intensity of preferences for redistribution on correction term $[RA - D]$ in Pareto-lognormal case

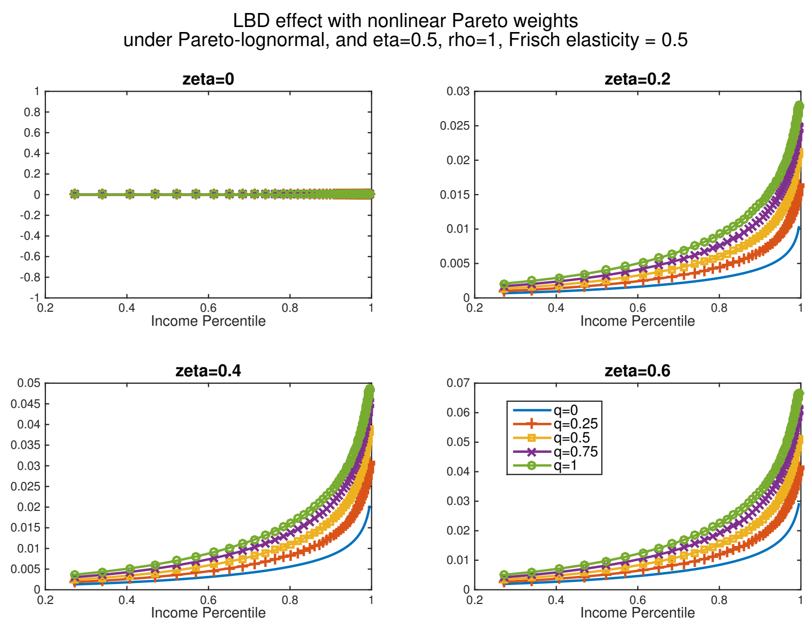

Figure B.12: Effects of intensity of preferences for redistribution on corrected LBD term $[RA - D]\Omega$ in Pareto-lognormal case the consumption policies $c_1(\theta_1)$ and $\hat{c}_2(\theta_1, \varepsilon_2)$, play no role in the determination of the optimal effort policies $e_1(\theta_1)$ and $e_2(\theta_1, \varepsilon_2)$, and hence play no role in the determination of the optimal wedges depicted in the figure. We thus proceed as follows. Let $c_1(\theta_1)$ be an arbitrary positive constant and set $\hat{c}_2(\theta_1, \varepsilon_2)$ as a residual of all other policies by using the envelope conditions, the definition of the flow and intertemporal utility functions, and the redistribution constraint.

For the results depicted in Figure 1, we have the following. When there are no LBD effects ($\zeta = 0$), the policies $e_1(\theta_1)$ and $e_2(\theta_1, \varepsilon_2)$ can be derived exactly from the optimality conditions with respect to period-1 and period-2 earnings, as described in the proof of Proposition 2 in Section A.1 in this Supplement.[18] Namely, $e_1(\theta_1)$ and $e_2(\theta_1, \varepsilon_2)$ are given by the solution to the system of the following two equations: $e_1(\theta_1) = \left\{ \left[1 + \frac{1+\phi}{\theta_1 \gamma_1(\theta_1)} \right]^{-1} \theta_1 \right\}^{\frac{1}{\phi}}$ and $e_2(\theta_1, \varepsilon_2) = \left\{ \left[1 + \rho \frac{1+\phi}{\theta_1 \gamma_1(\theta_1)} \right]^{-1} \theta_1^{\rho} (e_1(\theta_1)\theta_1)^{\zeta} \varepsilon_2 \right\}^{\frac{1}{\phi}}$. In the presence of LBD effects (i.e., when $\zeta > 0$), we interpolate the policy $e_1(\theta_1)$ (the interpolation is described below) and then replace the interpolation in the above formula for $e_2(\theta_1, \varepsilon_2)$. We use the interpolation coefficients for the policy $e_1(\theta_1)$ as control variables in the numerical optimization problem obtained from the original dual problem described above by letting the policies $e_2(\theta_1, \varepsilon_2)$, $c_1(\theta_1)$ and $\hat{c}_2(\theta_1, \varepsilon_2)$ be the ones described above. The interpolation coefficients defining the optimal policy $e_1(\theta_1)$ are then obtained by using the routine lsqnonlin.m in MATLAB_R2014b. Finally, throughout the different parameter specifications covered in the figures, the promised utility level κ is maintained fixed at the level that guarantees that, in the absence of LBD effects (i.e., when $\zeta = 0$), expected tax revenues under the policies solving the above problem are equal to zero.

Given the policies $e_1(\theta_1)$ and $e_2(\theta_1, \varepsilon_2)$, we then compute the policies $y_1(\theta_1)$ and $y_2(\theta_1, \theta_2)$ by letting $y_1(\theta_1) = \theta_1 e_1(\theta_1)$ and $y_2(\theta_1, \theta_2) = \theta_2 e_2 \left(\theta_1, \left(\frac{\theta_2}{\theta_1^{\rho} y_1(\theta_1)^{\zeta}} \right) \right)$. That the policies $y_1(\theta_1)$, $y_2(\theta_1, \theta_2)$, $c_1(\theta_1)$, and $c_2(\theta_1, \theta_2) = \hat{c}_2 \left(\theta_1, \left(\frac{\theta_2}{\theta_1^{\rho} y_1(\theta_1)^{\zeta}} \right) \right)$ jointly define an incentive-compatible mechanism and thus constitute a solution to the planner's full problem follows from the result in Proposition 2 in the main body.

The interpolation of the policy $e_1(\theta_1)$ for the results that correspond to the environment in the left panel of Figure 1 in the main body is done by considering a uniform grid of 75 productivity shocks with the last point in the grid corresponding to the 99.995^{th} percentile of the corresponding Pareto productivity distribution. The interpolation of the policy $e_1(\theta_1)$ for the results corresponding to the right panel in Figure 1 in the main body is done by considering a uniform grid of 150 productivity shocks with the first and last points in the grid corresponding to the 0.05^{th} and 99.95^{th} percentiles of the corresponding Pareto-lognormal productivity distribution. The distributions in question are discretized to conduct numerical integration by using the trapezoid method.

[18] From the proof of Proposition 2 in Section A.1 in this Supplement, we have that, when there are no LBD effects (i.e., when $\zeta = 0$), $y_t(\theta^t)$ is given by the unique solution to

$$\left[1 + \rho^{t-1} \frac{1+\phi}{\theta_1 \gamma_1(\theta_1)} \right]^{-1} \theta_t^{1+\phi} = y_t^{\phi}.$$

Combining the above optimality conditions with the definitions of $e_1(\theta_1)$ and $e_2(\theta_1, \varepsilon_2)$ then leads to the formulas in the main text.

Regarding the deployed interpolation, for the results referring to the Pareto productivity distribution, we approximate the period-1 labor supply schedule $e_1(\theta_1)$ by means of a function $\hat{e}_1(\theta_1) + \varpi_1 \theta_1^{1/\phi}$, where $\hat{e}_1(\theta_1)$ is a simple polynomial of order 6. For the results referring to the Pareto-lognormal productivity distribution, instead, we approximate the period-1 labor supply schedule $e_1(\theta_1)$ by means of a function $\hat{e}_1(\theta_1) + \varpi_1 \bar{e}_1(\theta_1)$, where $\hat{e}_1(\theta_1)$ is a simple polynomial of order 8 and where $\bar{e}_1(\theta_1) = \left\{ \left[1 + \frac{1+\phi}{\theta_1 \gamma_1(\theta_1)} \right]^{-1} \theta_1 \right\}^{\frac{1}{\phi}}$ is the formula for the optimal period-1 effort under risk neutrality in the absence of LBD.

For the environments corresponding to Figures 2, 3, and 4 in the main body, and all the numerical results in Section B in the present supplement, we follow a procedure similar to the one described above, except for the following two changes when agents are risk averse (i.e., when $\eta > 0$). First, the policy $e_2(\theta_1, \varepsilon_2)$ is interpolated instead of being determined by an exact first-order condition given $e_1(\theta_1)$. Second, the policy $c_1(\theta_1)$ is also interpolated, instead of being set arbitrarily.[19] The interpolation coefficients for the policies $e_2(\theta_1, \varepsilon_2)$ and $c_1(\theta_1)$ are then optimized alongside the interpolation coefficients corresponding to the policy $e_1(\theta_1)$.

The promised utility level κ in the redistribution constraint is held fixed at the level that guarantees that, when the agents are risk neutral (that is, $\eta = 0$) and there are no LBD effects (i.e., $\zeta = 0$), the expected tax revenues under the optimal policies are equal to zero. Given the policies $e_1(\theta_1)$ and $e_2(\theta_1, \varepsilon_2)$, we then compute the policies $y_1(\theta_1)$ and $y_2(\theta_1, \theta_2)$ by letting $y_1(\theta_1) = \theta_1 e_1(\theta_1)$ and $y_2(\theta_1, \theta_2) = \theta_2 e_2 \left(\theta_1, \left(\frac{\theta_2}{\theta_1^{\rho} y_1(\theta_1)^{\zeta}} \right) \right)$. For any θ_1, we verify numerically that the function $y_2(\theta_1, \theta_2)$ is non-decreasing in θ_2, and that the functions $y_1(\theta_1)$ and $y_2(\theta_1, \theta_2)$ jointly satisfy all the integral monotonicity conditions of Section 4 in the main body. As explained in the main body, the above properties, along with the fact that the agents' utilities under the policies $y_1(\theta_1)$, $y_2(\theta_1, \theta_2)$, $c_1(\theta_1)$, and $c_2(\theta_1, \theta_2) = \hat{c}_2 \left(\theta_1, \left(\frac{\theta_2}{\theta_1^{\rho} y_1(\theta_1)^{\zeta}} \right) \right)$ satisfy the envelope necessary conditions for incentive compatibility, imply that the mechanism defined by the policies $y_1(\theta_1)$, $y_2(\theta_1, \theta_2)$, $c_1(\theta_1)$, and $c_2(\theta_1, \theta_2)$ so constructed is incentive compatible and constitutes a solution to the planner's original problem.

The interpolation of the policies $e_1(\theta_1)$, $e_2(\theta_1, \varepsilon_2)$, and $c_1(\theta_1)$ (as well as the verification of the validity of the first-order approach) is done by considering a uniform grid of N productivity shocks with the first and last points in the grid corresponding to the 0.05^{th} and 99.95^{th} percentiles of the corresponding productivity distribution. The distribution is discretized to conduct numerical integration by using the trapezoid method. The value for N is 150 for all results involving the Pareto-lognormal distribution and 75 for all results involving the Lognormal distribution.

For the verification of the validity of the first-order approach, we restrict attention to outcomes over a subset of the above grids. Namely, to the grid between the 0.1^{th} and 99.5^{th} percentiles for the results involving the Pareto-lognormal distribution, and between the 0.1^{th} and the 99.9^{th} percentiles for the results involving the Lognormal distribution. The reason for restricting attention to these ranges is that, outside these regions, we expect high numerical errors due to the deployed discretization of the distributions.

Regarding the deployed interpolation, we use simple polynomials of order 7 for $c(\theta_1)$. We

[19] Recall that, in these environments, agents have preferences for consumption smoothing, and hence the distribution of consumption over the two periods is part of the optimization.

approximate the period-1 labor supply schedule $e_1(\theta_1)$ by means of a function $\hat{e}_1(\theta_1) + \varpi_1 \bar{e}_1(\theta_1)$, where $\hat{e}_1(\theta_1)$ is a simple polynomial of order n_1 and $\bar{e}_1(\theta_1)$ is the formula for the optimal period-1 effort when agents are risk neutral, the planner has Utilitarian preferences for redistribution, and there are no LBD effects.[20]

We approximate the period-2 labor supply schedule $e_2(\theta_1, \varepsilon_2)$ by means of a function $\hat{e}_2(\theta_1, \varepsilon_2) + \varpi_2 \bar{e}_2(\theta_1, \varepsilon_2)$, where $\bar{e}_2(\theta_1, \varepsilon_2)$ is the known formula for the optimal period-2 labor supply when agents are risk neutral, the planner has Utilitarian preferences for redistribution, and there are no LBD effects,[21] and where $\hat{e}_2(\theta_1, \varepsilon_2)$ is a tensor product node-basis scheme with 4-degree simple polynomials for both dimensions of the $(\theta_1, \varepsilon_2)$ space. The coefficients ϖ_1 and ϖ_2, together with the coefficients of the polynomials for $\hat{e}_1(\theta_1)$, $c_1(\theta_1)$, and $\hat{e}_2(\theta_1, \varepsilon_2)$, are the control variables in the numerical optimization problem obtained from the original dual problem described above by letting the policy $\hat{c}_2(\theta_1, \varepsilon_2)$ be the one described above. For the results pertaining to the Pareto-lognormal distribution, we set to zero the coefficients of the polynomial for $\hat{e}_2(\theta_1, \varepsilon_2)$ that correspond to the terms $(1, 4)$, $(2, 3)$, $(2, 4)$, $(3, 3)$, $(3, 4)$, $(4, 2)$, $(4, 3)$, and $(4, 4)$ in the tensor product.[22] For the results pertaining to the Lognormal distribution, we set to zero the coefficients of the polynomial for $\hat{e}_2(\theta_1, \varepsilon)$ that correspond to the terms $(2, 4)$, $(3, 3)$, $(3, 4)$, $(4, 2)$, $(4, 3)$, and $(4, 4)$ in the tensor product. The value for n_1 is 9 for the results with the Pareto-lognormal distribution and 4 for the results with the Lognormal distribution.

C.2 Numerical results in Section 5 in main body

For the calibration of the benchmark economy with LBD, the calibrated parameters are set to minimize the sum of squared percentage deviations of the model-generated moments from the target moments. To ensure convergence, we impose bounds on the admissible values for the parameters in question, namely, $h_1 \geq 0$, $\zeta \in [0.2, 0.6]$, $\rho \in [0, 1]$, $\sigma \in [0, \sqrt{0.33}]$, and $\lambda \in [2.1, 6]$. The calibrated parameters that solve the above minimization problem (reported in Table 2 in the main body) are all interior to these ranges. To solve the above minimization problem, we use the constrained-optimization solver fmincon.m in MATLAB_R2014b. In calculating the moments (both in the case of LBD and in the counterfactual economy without LBD of Subsection 5.6 in the main text), we interpolate the workers' optimal decisions $y_1(\theta_1)$, $c_1(\theta_1)$, and $\hat{y}_2(\theta_1, \varepsilon_2) \equiv y_2(\theta_1, \theta_1^\rho y_1(\theta_1)^\zeta \varepsilon_2)/\theta_1^\rho y_1(\theta_1)^\zeta$ under the approximation of the US current tax code described in the main body. The interpolation is conducted using a 20-degree Chebyshev polynomial for $y_1(\theta_1)$, a 20-degree Chebyshev polynomial for $c_1(\theta_1)$, and a tensor product node-basis scheme with a 16-degree Chebyshev polynomial for the

[20]Precisely, $\bar{e}_1(\theta_1) = \frac{y_1(\theta_1)}{\theta_1}$, where $y_1(\theta_1)$ is given by the unique solution to

$$\theta_1^{1+\phi} = y_1^\phi$$

(see the formula for the FB output policy in Section 3 in the main body).

[21]Precisely, $\bar{e}_2(\theta_1, \varepsilon_2) = \frac{\hat{y}_2(\theta_1, \varepsilon_2)}{\theta_1^\rho \varepsilon_2}$, where $\hat{y}_2(\theta_1, \varepsilon_2)$ is given by the unique solution to

$$(\theta_1^\rho \varepsilon_2)^{1+\phi} = \hat{y}_2^\phi$$

(see the formula for the FB output policy in Section 3 in the main body).

[22]The term (i, j) of the tensor product in question is the term $\theta_1^{i-1} \varepsilon_2^{j-1}$.

θ_1 dimension and a 14-degree Chebyshev polynomial for the ε_2 dimension of the $(\theta_1, \varepsilon_2)$ space for the $\hat{y}_2(\theta_1, \varepsilon_2)$ policy. The period-2 consumption policy is determined residually from the other policies using the workers' intertemporal budget constraint for each productivity history, θ. The Chebyshev interpolation coefficients are set to minimize the sum of squared residuals in the workers' first-order conditions as described in Sub-subsection A.3.5 in the present Supplement. For the minimization in question, we used the routine lsqnonlin.m in MATLAB_R2014b.

Given the calibrated model (both in the case of LBD and in the counterfactual economy without LBD), we then derive the quasi-optimal tax schedule by solving the primal problem described in Section 5 in the main body using the optimization solver fminunc.m in MATLAB_R2014b. During the minimization procedure, given any tax schedule, we interpolate the optimal policies $y_1(\theta_1)$, $c_1(\theta_1)$, and $\hat{y}_2(\theta_1, \varepsilon_2) \equiv y_2(\theta_1, \theta_1^{\rho} y_1(\theta_1)^{\zeta} \varepsilon_2)/\theta_1^{\rho} y_1(\theta_1)^{\zeta}$ for the workers by approximating $y_1(\theta_1)$, $c_1(\theta_1)$, and $\hat{y}_2(\theta_1, \varepsilon_2)$ with Chebyshev polynomials and letting $\hat{c}_2(\theta_1, \varepsilon_2) \equiv c_2(\theta_1, \theta_1^{\rho} y_1(\theta_1)^{\zeta} \varepsilon_2)$ be determined residually from the other policies using the definition of the workers' intertemporal budget constraint. We choose the Chebyshev interpolation coefficients so as to minimize the sum of the squared residuals in the workers' optimality conditions using the routine lsqnonlin.m in MATLAB_R2014b. A similar procedure is followed to derive the other simple tax codes considered in Section 5 in the main body.

To derive the solution to the planner's relaxed (primal) program we use the following procedure. First, we derive the allocation that maximizes tax revenues among those yielding the workers the same ex-ante utility as under the quasi-optimal tax schedule. Refer to such an allocation as allocation A. If the tax revenues generated by allocation A are no more than 10^{-12}, we stop and (using duality) identify the allocation under the quasi-optimal tax schedule as the solution to the planner's relaxed program.[23] If, instead, the tax revenues under allocation A are more than 10^{-12} , we proceed as follows. We construct a new allocation by using the tax revenues under allocation A to increase consumption in an equiproportionate way across all histories, while keeping effort/output fixed at the same level as under allocation A. Refer to this new allocation as allocation B. By construction, allocation B yields zero tax revenues and an ex-ante lifetime utility to the workers which is higher than that under the quasi-optimal tax schedule. Next, we derive the allocation that maximizes tax revenues among those yielding the workers the same ex-ante utility as under the allocation B. Refer to such an allocation as allocation C. If the tax revenues under allocation C are less than 10^{-12}, we stop and (using duality) identify allocation B as the solution to the planner's relaxed program. If, on the other hand, the tax revenues under C exceed 10^{-12} , then we proceed by distributing uniformly across histories the tax revenues under allocation C while keeping effort/output fixed at the same level as under allocation C. The entire procedure described above is continued until the extra tax revenues do not exceed 10^{-12}.

At each round of the above procedure, the revenue-maximizing allocations are derived by approximating earnings and period-1 consumption with Chebyshev polynomials and by letting period-2 consumption be determined residually from the other policies using the envelope conditions for

[23] This case emerges when we study the importance of the stochasticity of the LBD effects, whereas it does not emerge when we conduct the robustness analysis with respect to the intensity of the LBD effects and the level of skill persistence.

incentive compatibility, the definition of the flow and intertemporal utility functions, and the redistribution constraint. The Chebyshev interpolation coefficients are set to minimize the sum of squared residuals in the planner's optimality conditions. For the minimization in question, we use routine lsqnonlin.m in MATLAB_R2014b.

In the interpolations mentioned above, we use n_1^C-degree Chebyshev interpolation for period-1 earnings and period-1 consumption, and a tensor product node-basis scheme with a n_2^C-degree Chebyshev polynomial for the θ_1 dimension and a 14-degree Chebyshev polynomial for the ε_2 dimension of the $(\theta_1, \varepsilon_2)$ space for the $\hat{y}_2(\theta_1, \varepsilon_2)$ policy. The values for n_1^C and n_2^C are, respectively, 20 and 16 for the allocations under quasi-optimal tax codes, and 25 and 45 for the allocations solving the planner's relaxed program.

Given the policies $y_1(\theta_1)$, $c_1(\theta_1)$, and $\hat{y}_2(\theta_1, \varepsilon_2)$, we then compute the policies $y_2(\theta_1, \theta_2)$ by letting $y_2(\theta_1, \theta_2) = \hat{y}_2\left(\theta_1, \left(\frac{\theta_2}{\theta_1^\rho y_1(\theta_1)^\zeta}\right)\right)\left[\theta_1^\rho y_1(\theta_1)^\zeta\right]$. For any θ_1, we then verify numerically that the function $y_2(\theta_1, \theta_2)$ so constructed is non-decreasing in θ_2, and that the functions $y_1(\theta_1)$ and $y_2(\theta_1, \theta_2)$ jointly satisfy all the integral monotonicity conditions of Section 4 in the main body. As explained in the main body, the above properties, along with the fact that the agents' utilities under the policies $y_1(\theta_1)$, $y_2(\theta_1, \theta_2)$, $c_1(\theta_1)$, and $c_2(\theta_1, \theta_2) = \hat{c}_2\left(\theta_1, \left(\frac{\theta_2}{\theta_1^\rho y_1(\theta_1)^\zeta}\right)\right)$ satisfy the envelope necessary conditions for incentive compatibility, imply that the mechanism defined by the policies $y_1(\theta_1)$, $y_2(\theta_1, \theta_2)$, $c_1(\theta_1)$, and $c_2(\theta_1, \theta_2)$ is incentive compatible and constitutes a solution to the planner's original problem.

For the calibration of the benchmark economy, the derivation of quasi-optimal and optimal policies, and the verification of the validity of the first-order approach (with and without LBD), we truncate the Pareto-lognormal distribution of the productivity shocks ε_t, $t = 1.2$, from below at the 1^{st} percentile, and we use a uniform grid of 399 productivity shocks with the last point in the grid corresponding to the 99.99^{th} percentile of the truncated distribution. We discretize the distribution to conduct numerical integration by using the trapezoid method. Given that the calibrated distribution is, in effect, very close to a Lognormal distribution, and the grid covers 99.99% of the distribution, we have chosen not to ignore the outcomes of any grid point.