

Taxation under Learning-by-Doing*

Miltiadis Makris
University of Southampton

Alessandro Pavan
Northwestern University

October 2017

Abstract

We study optimal labor-income taxation in a dynamic economy in which the agents' productivity is their own private information, is stochastic, and evolves endogenously over the lifecycle due to learning-by-doing. First, we identify novel distortions that originate in the endogeneity of the agents' private information and that contribute to higher labor wedges under the optimal tax code. Next, we calibrate the model to US earnings data and find that reforming the current US tax code while accounting for learning-by-doing calls for a smaller increase in marginal tax rates and for more progressivity, compared to the optimal reform that ignores learning-by-doing. We also find that ignoring learning-by-doing overestimates the benefits of the reform. At the same time, reforming the code while accounting for learning-by-doing brings significant welfare gains. Lastly, we show that simple age-dependent taxes that are invariant in past incomes are approximately optimal and generate most of the welfare gains from the optimal reform.

*For comments and useful suggestions, we thank Emmanuel Farhi, Mikhail Golosov, Marek Kapicka, Paul Klein, Alex Mennuni, Ludovic Renou, Stephanie Stantcheva, Arnau Valladares-Esteban, and seminar participants at the European Meetings of the Econometric Society, the Latin America and Caribbean Meetings of the Econometric Society, the World Congress of the Econometric Society, the second Shanghai Macroeconomics Workshop, CERGE-EI, Cergy-Pontoise, the Chicago FED Research Department, Northwestern University, the University of Western Ontario, and the 2016 ESSET Conference. Abdoulaye Ndiaye and Laura Doval provided excellent research assistance. The usual disclaimer applies. Email addresses: M.Makris@soton.ac.uk (Makris); alepavan@northwestern.edu (Pavan).

1 Introduction

Learning-by-doing refers to the positive effect of the time spent at work on the agents' productivity. One can think of it as human capital investment that is a side-product of the labor supply process. Learning-by-doing is believed to be a significant source of productivity growth. A vast literature in labor economics documents the effects of labor experience on wages. For example, Dustmann and Meghir (2005) find that, in the first two years of employment, wages grow at an average rate of 8.5% for the first year and 7.5% for the second. Insofar variations in earnings reflect variations in skills/productivities, these findings suggest that labor experience has a significant impact on human capital accumulation.¹

In this paper, we study the effects of learning-by-doing (hereafter, LBD) on optimal taxation. We consider a dynamic Mirrleesian economy in which the agents' productivity is their own private information, is stochastic, and evolves endogenously over the lifecycle as the result of LBD. We identify novel effects that contribute to higher labor wedges (distortions relative to the complete-information benchmark) and show that such effects have a significant impact on the level, progressivity, and dynamics of optimal tax codes.

In the presence of LBD, agents have incentives to work harder to boost their future productivity. Under complete information, this effect contributes positively to welfare. When the agents' productivity is their own private information, however, agents must receive rents to reveal their private information. Such rents represent welfare losses and call for downward distortions in labor supply. These rents are higher for highly productive agents. LBD, by shifting the productivity distribution in future periods towards higher levels, contributes to higher expected future rents and thus to higher expected future welfare losses.

The mechanism described above originates in the *endogeneity* of the agents' private information, which is a natural feature of economies with LBD. Importantly, this mechanism has significant implications for the labor wedges under optimal tax codes and thereby for the design of optimal taxes. First, it contributes to higher labor wedges. Second, as LBD is typically stronger for younger workers, the above mechanism may contribute to a decrease in wedges over the lifecycle. Third, as more productive workers expect, on average, higher rents in future periods than less productive ones, the benefits of shifting the distribution of future productivity towards lower levels are stronger for highly productive agents. Other things equal, LBD may thus contribute also to a higher progressivity of the wedges and of the associated optimal tax code.

We illustrate these effects by considering a stylized, yet rich, economy in which the agents' working life is divided in two blocks, an earlier phase in which workers are young and learn on the

¹See also Becker (1964), Arrow (1972), and Lucas (1988) for earlier contributions on the effects of learning-by-doing on workers' productivity, Willis (1986) and Altug and Miller (1998) for the effects of work experience on wage dynamics, Jacobson et al. (1993) for the effects of temporary job losses on future wages, and Topel (1991) for the effects of job tenure on wage profiles. See Blundell and MaCurdy (1999) and Farber (1999) for excellent overviews of the effects of work experience on wage dynamics.

job, and a second phase in which workers are older and take advantage of the impact of LBD on their productivity. We start by considering the textbook problem of a planner with extreme (Rawlsian) preferences for redistribution, facing a continuum of risk-neutral agents with quasilinear preferences over consumption and labor supply. We first derive the wedges using a direct approach, where we maximize welfare by optimizing directly over the agents' allocations (consumption and labor supply) under all the relevant incentive-compatibility constraints. We then show how the wedges under the optimal allocations can be interpreted in terms of the relevant elasticities by considering perturbations of the tax code in the spirit of Saez (2001)'s summary statistics approach, but accounting for the endogenous evolution of the agents' productivity. Next, we study the effects of the agents' risk aversion on the impact of LBD on labor wedges. Finally, we consider the case of a planner with no primitive preferences for redistribution (i.e., with an Utilitarian welfare objective) who nonetheless faces a continuum of risk-averse agents and hence recognizes the value of redistribution for social insurance.

Equipped with the analytical results, we then calibrate the model with risk-averse agents and Utilitarian preferences for redistribution to match various moments of the US labor income distribution as reported in Huggett et. al. (2011).² The calibration also provides us with a parameter value for the intensity of LBD that is consistent with the results of the meta-analysis in Best and Kleven (2013). We show that reforming the existing US tax code by adopting the optimal one would yield about a 4% increase in consumption at each productivity history. We also show that the tax code that implements the optimal allocations in the second best can be approximated by a simple code in which, in each period, taxes are invariant in past incomes and are given by a linear-power function of the form $T_s(y_s) = y_s - k_s y_s^\tau - b_s$, where y_s is period- s income, b_s is a lump-sum subsidy, and k_s and τ_s are age-dependent scalars that control for the progressivity of the code.³ Under such code, tax rates are mildly progressive for young workers and mildly regressive for old ones. Furthermore, compared to the existing US tax code, tax rates are higher and less progressive for young workers, whereas, for old workers, they are higher and less progressive at low income percentiles but lower and less progressive at high income percentiles. Finally, tax rates increase over the lifecycle from an average rate of about 38% for young workers to an average rate of about 43% for old workers. We also find that most of the welfare gains from reforming the tax code can be generated through simple age-dependent linear taxes.⁴ The marginal tax rates in the optimal linear code are equal to 38% for young workers and 46% for old ones.

To isolate the quantitative implications of LBD on the reform from all other effects, we consider a counterfactual economy that is identical to the calibrated one, except for the fact that productivity among old workers is exogenous while drawn from the same conditional distribution as in the economy with LBD. First, we show that ignoring LBD leads to a significant overestimation of the benefits

²The approximation of the current US tax code is from Heathcote et. al. (2016).

³Similar functional forms have been considered in Benabou (2002) and Heathcote et. al. (2016).

⁴See Weinzierl (2011) for an earlier documentation of the benefits of age-dependent taxes.

of reforming the US tax code (precisely, a 15% overestimation of the increase in consumption). We also show that wedges under LBD are higher than their counterparts without LBD. Importantly, that wedges are higher under LBD does not imply that taxes are also higher: in fact, marginal tax rates are lower with LBD. This is because wedges combine distortions in labor supply arising from current tax rates with distortions arising from the effect of variations in earnings in the current period on future tax bills. Finally, we show that, while tax rates are regressive without LBD, they are progressive with LBD. Together, the results indicate that LBD has pronounced effects on the structure of optimal tax codes.

Outline. The rest of the manuscript is organized as follows. Below, we wrap up the Introduction with a brief discussion of the most pertinent literature. Section 2 introduces the textbook model of a dynamic economy with Rawlsian preferences for redistribution and quasilinear preferences over consumption and labor decisions. Section 3 identifies the effects of LBD on the level, progressivity, and dynamics of wedges under optimal tax codes. Section 4 shows how optimal wedges can also be obtained through a perturbation approach in the spirit of Saez (2001)'s sufficient statistics, but adapted to the dynamic economy with endogenous types under examination. Section 5 introduces preferences for consumption smoothing and studies the effects of the agents' risk aversion on wedges under LBD. Section 6 considers an alternative economy in which the planner's objective is Utilitarian and the benefits of taxation originate in social insurance. Section 7 calibrates the model with Utilitarian preferences for redistribution and risk-averse agents to the US income distribution under the current US tax code, and contains all the quantitative results. Section 8 concludes. All proofs are either in the Appendix at the end of the document, or in the Supplementary Material.

Related literature. The closest body of work is the recent literature on optimal taxation with endogenous human capital, i.e., Krause (2009), Best and Kleven (2013), Kapicka (2006, 2015a,b), Kapicka and Neira (2016), and Stantcheva (2016).⁵ The key differences with respect to the first six papers is that we consider an economy in which LBD has a stochastic effect on the evolution of the agents' productivity and the agents receive time-varying private information over the lifecycle. The key difference with respect to Stantcheva (2016) is that, in our model, the evolution of human capital is a direct byproduct of labor supply, whereas, in Stantcheva (2016), it originates in external investments in training and/or education. Furthermore, in her model, the planner can use separate instruments to influence the agents' labor supply decisions and their investments in human capital (for example, the government can combine non-linear income taxes with direct subsidies for education and training expenses). In contrast, in our model, investments in human capital cannot be controlled separately from labor supply decisions. As a result, any manipulation of labor supply for redistributive reasons has direct effects on the evolution of the agents' productivity. Importantly, the above differences have significant qualitative and quantitative implications for the structure of optimal tax codes.

⁵See also Stantcheva (2015).

Related is also the work of Benabou (2002), Conesa and Krueger (2006), Heathcote et al. (2016), Kindermann and Krueger (2014), and Krueger and Ludwig, (2013). Following the Ramsey (1927) tradition, these papers characterize properties of optimal tax codes in an economy in which the planner has a restricted set of tax instruments. Our analysis reveals that simple tax schedules similar to those considered in this literature, but age-dependent, are approximately optimal.

Our paper is also related to the fast-growing literature on dynamic mechanism design. We refer the reader to Pavan, Segal, and Toikka (2014), Bergemann and Pavan (2015), and Pavan (2017) for a discussion of recent developments,⁶ and to Golosov et al. (2006), Albanesi and Sleet (2006), Kocherlakota (2010), Gorry and Oberfield (2012), Kapicka (2013), Farhi and Werning (2013), and Golosov et al. (2016) for applications to dynamic optimal taxation. The key contribution of the present paper relative to the dynamic mechanism design literature is the investigation of the effects of the endogeneity of the type process on the dynamics of distortions.⁷

2 The Economy

Agents, productivity, and information. The economy is populated by a unit-mass continuum of workers of heterogenous productivity. The lifecycle of each worker consists of two periods. We interpret the first period as the phase in which workers accumulate human capital through learning-by-doing, and the second period as the phase in which the workers take advantage of earlier investments in human capital. We capture this situation by letting productivity be exogenous in the first period but endogenous in the second.⁸

In each period $t = 1, 2$, each worker produces *income* $y_t \in Y_t = \mathbb{R}_+$ at a cost $\psi(y_t, \theta_t)$, where θ_t denotes the agent's *productivity* (equivalently, her skills) and is the agent's private information. The function $\psi(y_t, \theta_t)$ is thrice differentiable, increasing, and convex in y_t . Consistently with the rest of the literature⁹, we assume that ψ takes the familiar iso-elastic form

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

⁶See also Garrett and Pavan (2015) for a variational approach to the characterization of optimal contracts in a dynamic model with non-transferable utility, but with exogenous types.

⁷Pavan, Segal, and Toikka (2014) accommodate for endogenous types, but in an environment with transferable utility. Furthermore, that paper does not investigate the implications of such endogeneity for the dynamics of distortions under optimal contracts. The type process is endogenous in Bergemann and Valimaki (2010), and in Fershtman and Pavan (2016). These works, however, focus on experimentation in quasilinear settings. The questions asked and the nature of the results is fundamentally different from the ones in the present paper.

⁸As shown in the Supplementary Material, the above description can also be seen as a reduced-form representation of an economy in which agents work for an arbitrary number of periods. In this case, the lifecycle of each worker consists of two blocks. Productivity is constant in each of the two blocks, it is exogenous in the first block and endogenous in the second. The productivity in the second block is a stochastic weighted function of all labor supply decisions in the first block.

⁹See, among others, Kapicka (2013), Farhi and Werning (2013), and Best and Kleven (2013).

and then denote by

$$\psi_y(y_t, \theta_t) \equiv \partial\psi(y_t, \theta_t)/\partial y_t, \quad \psi_\theta(y_t, \theta_t) \equiv \partial\psi(y_t, \theta_t)/\partial\theta_t, \quad \text{and} \quad \psi_{y\theta}(y_t, \theta_t) \equiv \partial^2\psi(y_t, \theta_t)/\partial\theta_t\partial y_t$$

the partial and cross derivatives of the ψ function with respect to its arguments. Hence, under this specification, ϕ is the inverse Frish elasticity. Note that $\psi_\theta < 0$ and $\psi_{y\theta} < 0$. Higher types thus experience a lower disutility of effort and have a lower marginal cost for their labor supply.

Each worker's period-1 productivity is drawn independently across agents from a distribution F_1 that is absolutely-continuous over the entire real line with density f_1 strictly positive over the interval $\Theta_1 = (\underline{\theta}_1, \bar{\theta}_1)$. Each worker's period-2 productivity is endogenous and given by

$$\theta_2 = z_2(\theta_1, y_1, \varepsilon_2) = \theta_1^\rho y_1^\zeta \varepsilon_2 \tag{2}$$

with ε_2 drawn from some distribution G with support $E \subset \mathbb{R}_+$, independently across agents, and independently from all other random variables, where ρ and ζ are non-negative scalars. Given the above specification, for any $(\theta_1, y_1) \in \Theta_1 \times \mathbb{R}_+$, θ_2 is thus drawn from the conditional distribution

$$F_2(\theta_2|\theta_1, y_1) = G\left(\frac{\theta_2}{\theta_1^\rho y_1^\zeta}\right).$$

We assume that $\zeta \leq \phi/(1 + \phi)$ to ensure a well-defined solution to the first-period output schedule. We denote by

$$\Theta_2 = \{\theta_2 : \theta_2 = z_2(\theta_1, y_1, \varepsilon_2), (\theta_1, y_1, \varepsilon_2) \in \Theta_1 \times \mathbb{R}_+ \times E\}$$

the set of possible period-2 productivities, and by $\Theta = \Theta_1 \times \Theta_2$ the set of all possible productivity histories. The dependence of the agents' period-2 productivity on their period-1 income is what captures learning-by-doing (LBD). Because period-1 income is an increasing function of period-1 effort and period-1 productivity, the above representation is flexible enough to encompass both the case in which LBD comes from past effort, or labor supply, as well as the case in which it originates directly in past income/output. Note, further, that, under the specification in (2), the intensity of LBD is conveniently parametrized by the uni-dimensional parameter $\zeta \geq 0$; the case of no LBD corresponds to $\zeta = 0$, and higher values of ζ capture stronger LBD effects.

Also note that, under the specification in (2), for any $\theta \equiv (\theta_1, \theta_2) \in \Theta$, and any y_1 , the impulse response of θ_2 to θ_1 (that is, the marginal effect of a variation in θ_1 on θ_2 , holding fixed the shock $\varepsilon_2 = \theta_2/\theta_1^\rho y_1^\zeta$ that, together with θ_1 and y_1 , is responsible for θ_2) is given by

$$I_1^2(\theta, y_1) \equiv \left. \frac{\partial z_2(\theta_1, y_1, \varepsilon_2)}{\partial \theta_1} \right|_{\varepsilon_2 = \theta_2/\theta_1^\rho y_1^\zeta} = \rho \frac{\theta_2}{\theta_1}. \tag{3}$$

The advantage of the specification in (2) is that it implies that the only channel through which LBD affects wedges and optimal tax rates is by shifting the distribution of future productivity in a first-order-stochastic-dominance way. When, instead, impulse responses $I_1^2(\theta, y_1)$ also depend on period-1 income, there is a second channel through which LBD affects the wedges, namely through its

effect on the level of future rents, for given distribution of future productivity. This second channel is (a) similar to the one that operates through the accumulation of human capital in economies with exogenous private information, and (b) less robust than the first one in that it is sensitive to whether LBD affects highly productive workers more or less than less productive ones (equivalently, to the complementarity/substitutability between θ_1 and y_1 in the determination of θ_2). To highlight the novel effects due to the endogeneity of the agents' private information, and because the labor literature has not reached consensus on the degree of complementarity between past productivity and LBD in the determination of future productivity, hereafter we focus on the first channel. All the results, however, extend to economies in which impulse responses depend on y_1 , provided that $I_1^2(\theta, y_1)$ are either increasing or moderately decreasing in y_1 . Likewise, none of the analytical results hinges on the disutility of labor taking the iso-elastic form in (1). In the analysis below, we thus denote the conditional distribution of θ_2 by $F_2(\theta_2|\theta_1, y_1)$, the impulse response of θ_2 to θ_1 by $I_1^2(\theta, y_1)$, and the disutility of labor by $\psi(y_t, \theta_t)$, and highlight the dependence on the specific functional forms in (1) and (2) only when necessary.

Preferences. Denote by $c_t \in \mathbb{R}_+$ each agent's period- t consumption and let δ be the common discount factor.¹⁰ The *lifetime* utility of each agent is given by

$$U(\theta, y, c) \equiv \sum_t \delta^{t-1} (c_t - \psi(y_t, \theta_t)),$$

where $\theta \equiv (\theta_1, \theta_2)$, $y \equiv (y_1, y_2)$, and $c \equiv (c_1, c_2)$.

Planner's problem. The government's problem consists in designing an intertemporal tax code that maximizes an aggregator of the agents' lifetime utilities, subject to the constraint that the fiscal deficit be smaller than an exogenous level G . We will solve this problem by considering its dual in which the Government maximizes intertemporal tax revenues, subject to the constraint that the aggregator of the agents' lifetime utilities be greater than an exogenous threshold.

Formally, the dual problem can be stated as follows. Let $\chi : \Theta \rightarrow \mathbb{R}^{2T}$ denote an *allocation rule*, specifying, for each agent, the lifetime profile of income-consumption pairs $\chi(\theta) = (y_t(\theta^t), c_t(\theta^t))_{t=1,2}$, with $\theta^1 \equiv \theta_1$ and $\theta^2 \equiv \theta$. Then denote by $\lambda[\chi]$ the endogenous probability distribution over Θ that is obtained by combining the period-1 exogenous distribution F_1 with the endogenous period-2 distribution F_2 that one obtains when $y_1 = y_1(\theta_1)$. Further, let $\lambda[\chi]|\theta_1$ denote the endogenous distribution over Θ that obtains under the rule χ , when the agent's initial productivity is θ_1 . Finally, let

$$V_1(\theta_1) \equiv \mathbb{E}^{\lambda[\chi]|\theta_1} [U(\tilde{\theta}, \chi(\tilde{\theta}))] = \mathbb{E}^{\lambda[\chi]|\theta_1} \left[\sum_t \delta^{t-1} (c_t(\tilde{\theta}^t) - \psi(y_t(\tilde{\theta}^t), \tilde{\theta}_t)) \right]$$

denote the expected lifetime utility of each agent of initial productivity θ_1 , under the allocation rule χ . Importantly, note that the dependence on χ is both through the policies $c_t(\cdot)$ and $y_t(\cdot)$,

¹⁰In turn, δ is assumed to be equal to the inverse of the gross interest rate (see, for instance, Best and Kleven (2013), Kapicka (2013), Farhi and Werning (2013), Golosov et. al. (2016), and Stantcheva (2016) for similar assumptions).

and through the dependence of the period-2 distribution F_2 on period-1 income $y_1(\theta_1)$, with the dependence originating in LBD.

The planner's problem consists in maximizing expected tax revenues

$$R = \mathbb{E}^{\lambda[\chi]} \left[\sum_t \delta^{t-1} \left(y_t(\tilde{\theta}^t) - c_t(\tilde{\theta}^t) \right) \right]$$

subject to the constraint that

$$\min_{\theta_1} V_1(\theta_1) \geq \kappa$$

and the constraint that the rule χ be incentive compatible (that is, each agent finds it optimal to generate income over the life-cycle as specified by the policy $\chi(\cdot)$).

Remarks. Implicit in the above formulation are the following assumptions. (i) Consumption is a deterministic function of income, with the two linked by the tax code $\mathcal{T} \equiv (\mathcal{T}_1, \mathcal{T}_2)$ according to $c_t = y_t - \mathcal{T}_t(y^t)$ with $y^1 = y_1$ and $y^2 = y \equiv (y_1, y_2)$. Note that the above specification accommodates for the possibility that the taxes paid in period 2 depend on the income reported in the previous period. In equilibrium, an agent with lifetime productivity history $\theta = (\theta_1, \theta_2)$ thus consumes $c_1(\theta_1) = y_1(\theta_1) - \mathcal{T}_1(y_1(\theta_1))$, and $c_2(\theta) = y_2(\theta) - \mathcal{T}_2(y_1(\theta_1), y_2(\theta))$. (ii) The planner commits in advance to the intertemporal tax code \mathcal{T} . (iii) The planner is constrained to give each agent a lifetime expected utility greater than, or equal to, κ . This constraint amounts to an extreme preference for redistribution, and corresponds to a primal problem in which the Government designs the tax code so as to maximize the expected lifetime utility of those agents whose initial productivity is the lowest (i.e., $\theta_1 = \underline{\theta}_1$) subject to an exogenous fiscal budget constraint (Rawlsian problem). (iv) The agents are risk neutral when it comes to evaluating lotteries over intertemporal consumption streams.

The problem described above is a textbook problem in public economics, except for the endogeneity of the workers' private information (their period-2 productivity). We relax some of the assumptions in this problem in Sections 5 and 6 below.

3 Optimal Policies

To identify properties of optimal tax codes, we first characterize the income and consumption policies that solve the planner's dual problem. In the next section we then show how the tax schedules implementing such policies can also be obtained through a perturbation approach in the spirit of Saez (2001)'s sufficient statistics, but adapted to the dynamic problem with endogenous productivity under consideration.

3.1 First-Best Policies

Consider an economy in which each agent's productivity is verifiable (that is, agents do not possess private information). Let $\lambda[\chi]|\theta_1, y_1$ denote the endogenous distribution over Θ that obtains under the allocation rule χ , when the agent's initial productivity is θ_1 and period-1 output/income is y_1 .

The first-best policies (hereafter denoted with the superscript “*”) satisfy the following optimality conditions (see Lemma 1 in the Appendix)

$$\psi_y(y_1^*(\theta_1), \theta_1) = 1 + LD_1^{\chi^*}(\theta),$$

and

$$\psi_y(y_2^*(\theta), \theta_2) = 1,$$

where, given any policy $\chi = (y_t(\cdot), c_t(\cdot))_{t=1,2}$,

$$LD_1^{\chi}(\theta_1) \equiv \delta \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[y_2(\tilde{\theta}) - \psi(y_2(\tilde{\theta}), \tilde{\theta}_2) \right]. \quad (4)$$

In both periods, the first-best (FB) income policy equalizes each agent’s marginal disutility of labor with the marginal benefit of higher output. In the second period, the latter simply coincides with the extra resources that are made available when the agent works harder. In the first period, instead, the benefit of asking the agent for higher output also takes into account the effect that the latter has on the distribution of the period-2 productivity. Because the period-2 policies are set optimally, usual envelope arguments imply that, in a first-best world, the extra benefit of asking an agent of period-1 productivity θ_1 for higher period-1 output due to LBD is given by the function $LD_1^{\chi}(\theta)$ above. Importantly, note that this function is computed holding fixed the period-2 income policy, as specified by the allocation rule χ . The expectation in the formula for $LD_1^{\chi}(\theta)$ is thus with respect to the endogenous distribution over Θ under the rule χ , starting from period-1 productivity θ_1 and period-1 income $y_1 = y_1(\theta)$.

Because there are no preferences for intertemporal consumption smoothing in this simple baseline economy, given the optimal income policies, the optimal consumption policies are given by any combination of $c_1(\cdot)$ and $c_2(\cdot)$ that satisfy

$$c_1^*(\theta_1) + \delta \mathbb{E}^{\lambda[\chi^*]|\theta_1} [c_2^*(\theta)] = \kappa + \psi(y_1^*(\theta_1), \theta_1) + \delta \mathbb{E}^{\lambda[\chi^*]|\theta_1} [\psi(y_2^*(\theta), \theta_2)].$$

The first-best allocations can be sustained, for example, through non-linear income taxes of the form

$$\mathcal{T}_1(y_1^*(\theta_1), \theta_1) = y_1^*(\theta_1) - \psi(y_1^*(\theta_1), \theta_1) - K \text{ and } \mathcal{T}_2(y_2^*(\theta), \theta) = y_2(\theta) - \psi(y_2^*(\theta), \theta_2) - (\kappa - K) / \delta.$$

where K is a scalar.

3.2 Second-Best Policies

We now turn to the relevant case in which the agents’ productivities are their own private information. In this economy, the planner faces additional constraints to its ability to redistribute from more productive agents to less productive ones. In particular, incentive compatibility requires that highly productive workers be given informational rents necessary to dissuade them from mimicking the less productive workers. Let

$$V_2(\theta) \equiv c_2(\theta) - \psi(y_2(\theta), \theta_2)$$

denote the period-2 continuation utility of an agent with productivity history equal to θ . Period-2 incentive-compatibility requires that, for any θ_1 , $V_2(\theta_1, \cdot)$ be Lipschitz continuous and satisfy the familiar Mirrlees envelope formula from static optimal taxation

$$V_2(\theta_1, \theta_2) = V_2(\theta_1, \underline{\theta}_2) - \int_{\underline{\theta}_2}^{\theta_2} \psi_\theta(y_2(\theta_1, s), s) ds,$$

along with the requirement that the period-2 income schedule $y_2(\theta_1, \cdot)$ be nondecreasing in period-2 productivity θ_2 .

Period-1 incentive compatibility, instead, requires that each agent's expected lifetime utility $V_1(\theta_1)$ also be Lipschitz continuous in period-1 productivity and satisfy an analogous envelope formula given by (see Pavan et al. (2014))

$$V_1(\theta_1) = V_1(\underline{\theta}_1) - \int_{\underline{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1(s), s) ds + \delta \mathbb{E}^{\lambda|\chi|^s} \left[I_1^2(\tilde{\theta}, y_1(s)) \psi_\theta(y_2(\tilde{\theta}), \tilde{\theta}_2) \right] \right\} ds \quad (5)$$

along with the requirement that, for any pair $\theta_1, \hat{\theta}_1 \in \Theta_1$, the following integral-monotonicity condition holds

$$\begin{aligned} & \int_{\hat{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1(s), s) + \delta \mathbb{E}^{\lambda|\chi|^s, y_1(s)} \left[I_1^2(\tilde{\theta}, y_1(s)) \psi_\theta(y_2(s, \tilde{\theta}_2), \tilde{\theta}_2) \right] \right\} ds \\ & \leq \int_{\hat{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1(\hat{\theta}_1), s) + \delta \mathbb{E}^{\lambda|\chi|^s, y_1(\hat{\theta}_1)} \left[I_1^2(\tilde{\theta}, y_1(\hat{\theta}_1)) \psi_\theta(y_2(\hat{\theta}_1, \tilde{\theta}_2), \tilde{\theta}_2) \right] \right\} ds. \end{aligned} \quad (6)$$

Hereafter, we follow the same first-order approach as in the rest of the new dynamic public finance literature by ignoring the monotonicity requirement on $y_2(\theta_1, \cdot)$ and the integral monotonicity constraints in (6) and verify that they hold ex-post, once the solution to the planner's relaxed program described below is at hands.

Because the agents are risk neutral, expected tax revenues are equal to

$$R = \mathbb{E}^{\lambda|\chi|} \left[\sum_t \delta^{t-1} \left(y_t(\tilde{\theta}^t) - \psi(y_t(\tilde{\theta}^t), \tilde{\theta}_t) \right) - V_1(\tilde{\theta}_1) \right].$$

Using (5), and integrating by parts, we then have that, under asymmetric information, the planner's objective can be expressed as

$$R = \mathbb{E}^{\lambda|\chi|} \left[\sum_t \delta^{t-1} \left(y_t(\tilde{\theta}^t) - \psi(y_t(\tilde{\theta}^t), \tilde{\theta}_t) + \frac{I_1^t(\tilde{\theta}, y_1(\tilde{\theta}_1))}{\gamma_1(\tilde{\theta}_1)} \psi_\theta(y_t(\tilde{\theta}^t), \tilde{\theta}_t) \right) - V_1(\underline{\theta}_1) \right], \quad (7)$$

where $I_1^1(\theta, y_1(\theta_1)) \equiv 1$, all θ , and where

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

denotes the hazard rate of the period-1 exogenous productivity distribution F_1 . Because incentive compatibility requires $V_1(\theta_1)$ to be nondecreasing (see (5)), the second-best policies maximize (7) subject to the constraint that $V_1(\underline{\theta}_1) \geq \kappa$.

The Government's problem in a second-best world is thus similar to its problem in a first-best world, except for the fact that, holding the labor supply choices fixed, tax revenues are lower because of the informational rents that must be left to the agents to induce them to reveal their private information. Such rents are given by the expectation of the "handicaps"

$$h_1(\theta_1, y_1(\theta_1)) \equiv -\frac{1}{\gamma_1(\theta_1)}\psi_\theta(y_1(\theta_1), \theta_1) \text{ and } h_2(\theta, y(\theta)) \equiv -\frac{I_1^2(\theta, y_1(\theta_1))}{\gamma_1(\theta_1)}\psi_\theta(y_2(\theta), \theta_2)$$

in the above tax revenue formula, where $y(\theta) \equiv (y_1(\theta_1), y_2(\theta))$.

The second-best income policies are thus chosen to trade off the marginal effects of higher output on current and future surplus, as in a first-best world, with the marginal effects that higher output has on the agents' information rents, as captured by the expectation of current and future handicaps. Differentiating R with respect to $y_1(\theta_1)$ and $y_2(\theta)$, we have that the optimal policies must satisfy the following optimality conditions

$$\psi_y(y_2(\theta), \theta_2) - \frac{I_1^2(\theta, y_1(\theta_1))}{\gamma_1(\theta_1)}\psi_{\theta y}(y_2(\theta), \theta_2) = 1 \quad (8)$$

and

$$\psi_y(y_1(\theta_1), \theta_1) - \frac{1}{\gamma_1(\theta_1)}\psi_{\theta y}(y_1(\theta_1), \theta_1) - \delta \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[x]|\theta_1, y_1(\theta_1)} \left[\frac{I_1^2(\tilde{\theta}, y_1(\theta_1))}{\gamma_1(\theta_1)}\psi_\theta(y_2(\tilde{\theta}), \tilde{\theta}_2) \right] = 1 + LD_1^x(\theta_1), \quad (9)$$

where, for any period-2 policy $y_2(\cdot)$, the function $LD_1^x(\theta_1)$ is as in (4). The above conditions pin down the optimal output schedules. The left-hand side in each of these conditions is the marginal cost of asking the agent for higher output in period t , whereas the right-hand side is the marginal benefit.

Consider first (8). The marginal cost of asking for higher period-2 output from an agent of productivity history θ has two parts. The first one is the marginal adjustment $\psi_y(y_2(\theta), \theta_2)$ in the agent's consumption necessary to compensate him for the extra disutility of labor. This part is standard and is the same as in the first-best benchmark. The interesting part is the second one. Under asymmetric information, when the planner asks those agents with period-2 productivity history $\theta = (\theta_1, \theta_2)$ to marginally increase their period-2 income starting from $y_2(\theta)$, it then needs to increase by $\psi_{\theta y}(y_2(\theta), \theta_2)$ the consumption of all agents with period-2 productivity history (θ_1, θ'_2) , with $\theta'_2 > \theta_2$, to guarantee that these types do not mimick type θ_2 .

In period 1, the planner can then reduce the consumption of all agents with period-1 productivity equal to θ_1 by $\psi_{\theta y}(y_2(\theta), \theta_2)[1 - F_2(\theta_2|\theta_1, y_1(\theta_1))]$, for these agents now expect a higher period-2 compensation. So far, the adjustment comes with no extra rent for the agents. The problem is that the above adjustment also requires increasing the period-1 expected consumption of all agents with period-1 productivity above θ_1 . The increase in these latter agents' consumption is due to the fact that these agents, if they were to mimick type θ_1 in period 1, they would expect to receive the extra period-2 compensation $\psi_{\theta y}(y_2(\theta), \theta_2)$ with probability higher than $1 - F_2(\theta_2|\theta_1, y_1(\theta_1))$ due to

the positive serial correlation in the productivity process. To induce these workers to continue to produce the same period-1 income as prior to the policy change, the planner then needs to increase the consumption of all these workers by $I_1^2(\theta, y_1(\theta_1))\psi_{\theta y}(y_2(\theta), \theta_2)[1 - F_2(\theta_2|\theta_1, y_1(\theta_1))]$. Note that the impulse response $I_1^2(\theta, y_1(\theta_1))$ captures the extra probability of reaching a productivity level above θ_2 in period 2 assigned by the period-1 types just above θ_1 relative to the probability assigned by type θ_1 .

The marginal cost of such adjustment is naturally higher the higher the inverse hazard rate $1/\gamma_1(\theta_1) = [1 - F_1(\theta_1)]/f_1(\theta_1)$ of the period-1 productivity distribution and the higher the intertemporal informational linkage between period-1 and period-2 types, as captured by the impulse response $I_1^2(\theta, y_1(\theta_1))$ of θ_2 to θ_1 . The marginal cost of such increase in informational rents, as evaluated from the perspective of period one, is then given by the second term in the left-hand side of (8). Because this extra marginal cost is increasing in y_2 , at the second-best optimum, the labor supply of each worker of period-2 productivity history equal to θ is then distorted downwards relative to its first-best level.

Next, consider the optimal choice of period-1 income, as determined by (9). The benefits of asking for higher y_1 naturally take into account the effect of changing the distribution of period-2 productivity coming from LBD. These benefits are determined by the same function $LD_1^\chi(\theta)$ introduced above in the first-best case. However, for given (θ_1, y_1) , the value of the function $LD_1^\chi(\theta)$ is now different than under the first-best policies because the period-2 income policy $y_2(\cdot)$ specified by the rule χ is distorted downwards relative to its first-best counterpart, as explained above. As a result, the marginal benefits of LBD are reduced relative to the first-best. Furthermore, the marginal costs of asking for a higher period-1 output to those agents of period-1 productivity equal to θ_1 are higher than under symmetric information, as indicated by the second and the third terms in the left-hand side of (9). First, when the planner asks for a higher period-1 income to those agents of period-1 productivity equal to θ_1 , it then needs to increase the consumption of all workers of period-1 productivity above θ_1 . Again, such extra compensation is necessary to guarantee that such higher period-1 types do not mimic. As explained above, such higher types have a double advantage relative to an agent of period-1 productivity equal to θ_1 . First, they can generate the same period-1 income y_1 by working less, thus economizing on the disutility of labor. Second, because period-2 continuation utility $V_2(\theta_1, \cdot)$ is increasing in period-2 productivity θ_2 , such types also expect a higher period-2 continuation utility given that, on average, they expect to be more productive than type θ_1 also in the second period. To guarantee that such period-1 types do not mimic, when asking for a higher period-1 income to types θ_1 , the planner must then increase all former agents' consumption by

$$-\psi_{y\theta}(y_1(\theta_1), \theta_1) - \delta \frac{\partial}{\partial y_1} \frac{\partial}{\partial \theta_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[V_2(\tilde{\theta}) \right].$$

Now observe that¹¹

$$\begin{aligned} \frac{\partial}{\partial \theta_1} \mathbb{E}^{\lambda[x]|\theta_1, y_1(\theta_1)} \left[V_2(\tilde{\theta}) \right] &= \mathbb{E}^{\lambda[x]|\theta_1, y_1(\theta_1)} \left[I_1^2(\tilde{\theta}, y_1(\theta_1)) \frac{\partial V_2(\tilde{\theta})}{\partial \theta_2} \right] = \\ &= -\mathbb{E}^{\lambda[x]|\theta_1, y_1(\theta_1)} \left[I_1^2(\tilde{\theta}, y_1(\theta_1)) \psi_\theta(y_2(\tilde{\theta}), \tilde{\theta}_2) \right]. \end{aligned} \quad (10)$$

As explained above, the weight the planner assigns to increasing the rents of all agents with period-1 productivity above θ_1 , relative to the weight it assigns to asking type θ_1 for higher income, is equal to the inverse hazard rate $1/\gamma_1(\theta_1)$. It follows that the marginal cost of asking for a higher period-1 income from those workers of period-1 productivity equal to θ_1 due to asymmetric information is equal to the sum of the second and third terms in the left-hand side in (9). Note that the second term is the familiar one as in Mirrlees static analysis and coincides with the corresponding term in the optimality condition for period-2 output, except for the fact that the impulse response of θ_1 to itself, which is equal to $I_1^1(\theta, y_1(\theta_1)) = 1$, is typically higher than the impulse response $I_1^2(\theta, y_1(\theta_1))$ of θ_2 to θ_1 , due to the imperfect and declining persistence in the agents' productivity. The interesting novel effects due to LBD are captured by the third term in the left-hand side in (9), which is absent when period-2 productivity is exogenous.

To better understand the last term in the left-hand side in (9), note that, under LBD, asking those agents of period-1 productivity equal to θ_1 for a higher period-1 income affects the expectation of the period-2 handicaps $h_2(\theta, y(\theta))$ – equivalently, the expected rents the planner must provide to those workers of period-1 productivity above θ_1 – through two channels. The first one is through the change in the distribution of θ_2 , holding fixed the period-2 handicaps $h_2(\theta, y(\theta))$. The second channel is through the variation in the impulse responses I_1^2 holding the distribution of period-2 productivity constant (recall that these functions capture the comparative advantage of higher types stemming from the higher probability they assign to being highly productive also in the second period). This second channel is (a) absent under the specification in (2), (b) positive (thus contributing to higher expected rents) when period-1 productivity and period-1 income are complements in the determination of period-2 productivity (that is, when LBD benefits relatively more those workers of higher period-1 productivity), and (c) negative when period-1 productivity and period-1 income are substitutes in the determination of period-2 productivity.

As we show below, these novel effects have important implications for the level, the progressivity, and the dynamics of the labor wedges.

Definition 1. The labor wedges are given by

$$W_1(\theta_1) \equiv 1 - \frac{\psi_y(y_1(\theta_1), \theta_1)}{1 + LD_1^X(\theta_1)} \quad \text{and} \quad W_2(\theta) \equiv 1 - \psi_y(y_2(\theta), \theta_2).$$

Recall that efficiency requires that the marginal disutility of extra period- t income be equalized to the marginal benefit, where, in the first period, the latter takes into account also the effect of

¹¹The equalities in (10) follow from the fact that, given any Lipschitz continuous function $J(\theta_2)$, and any kernel $F_2(\theta_2|\theta_1)$, $\frac{\partial}{\partial \theta_1} \mathbb{E} \left[J(\tilde{\theta}_2)|\theta_1 \right] = \mathbb{E} \left[I_1^2(\tilde{\theta}) \frac{\partial J(\tilde{\theta}_2)}{\partial \theta_2} |\theta_1 \right]$.

higher period-1 income on the distribution of period-2 total surplus, as captured by the term $LD_1^X(\theta)$. The period- t wedge W_t is then defined as the discrepancy between the ratio of marginal cost and marginal benefit of higher period- t income at the efficient allocation and the corresponding ratio at the proposed allocation. Importantly, in period-1, such discrepancy is computed holding fixed the period-2 policies, so as to highlight the part of the inefficiency that pertains to the period-1 allocations. As we show in the next section, the wedges W_t are related to marginal tax rates; that is, the sensitivity of current taxes to current income, holding fixed past incomes and all future tax schedules (with the latter allowed to depend on the entire history of reported incomes).

As is customary in the taxation literature (see, among others, Diamond, 1998, and Saez, 2001), instead of studying the behavior of W_t , we will consider the monotone transformation

$$\hat{W}_t(\theta^t) \equiv \frac{W_t(\theta^t)}{1 - W_t(\theta^t)}$$

which measures the wedge relative to the ratio of the agent's marginal disutility of labor and social marginal benefit of higher output, under the proposed policies. Accordingly, we will refer to \hat{W}_t as to the *relative wedge*.¹²

Proposition 1. *With Rawlsian preferences for redistribution and risk-neutral agents, the relative wedges under optimal tax codes are given by (with $\lambda[\chi]$ -probability one)*

$$\hat{W}_1(\theta_1) = \hat{W}_1^{RRN}(\theta_1) + \Omega(\theta_1) \tag{11}$$

and $\hat{W}_2(\theta) = \hat{W}_2^{RRN}(\theta)$, where $\hat{W}_1^{RRN}(\theta_1)$ and $\hat{W}_2^{RRN}(\theta)$ are the relative wedges in the absence of LBD, and where the term

$$\Omega(\theta_1) \equiv -\delta \frac{\frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[\frac{I_1^2(\tilde{\theta}, y_1(\theta_1))}{\gamma_1(\theta_1)} \psi_\theta(y_2(\tilde{\theta}), \tilde{\theta}_2) \right]}{\psi_y(y_1(\theta_1), \theta_1)} \tag{12}$$

summarizes all the novel effects due to LBD.

The wedges $\hat{W}_t^{RRN}(\theta^t)$ are the well-known period- t wedges for a planner with Rawlsian preferences for redistribution facing risk-neutral agents (from which the acronym RRN), in the absence of LBD (see, e.g., Diamond, 1998, and Saez, 2001, for their static analogs). The novel effects due to LBD are captured by the term $\Omega(\theta_1)$. This term measures the (discounted) effect of LBD on expected future welfare losses due to the rents the planner has to leave to the agents under asymmetric information. Formally, Ω measures the variation in the expected period-2 rents due to a marginal variation in period-1 income.

¹²For example, when $t = 1$, $\hat{W}_1(\theta_1) = W_1(\theta_1) / \left[\frac{\psi_y(y_1(\theta_1), \theta_1)}{1 + LD_1^X(\theta_1)} \right]$.

3.3 Implications of LBD for Optimal Wedges

We now turn to the effects of LBD on the level, dynamics, and progressivity of the wedges under optimal tax codes. When the disutility of labor is iso-elastic, and period-2 productivity is given by (2),¹³

$$\hat{W}_t^{RRN}(\theta^t) = \rho^{t-1} \frac{1 + \phi}{\theta_1 \gamma_1(\theta_1)}, \quad (13)$$

and

$$\Omega(\theta_1) = \frac{\delta \rho}{\psi_y(y_1(\theta_1), \theta_1)} \hat{W}_1^{RRN}(\theta_1) \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda|\chi||\theta_1, y_1(\theta_1)} \left[\psi(y_2(\tilde{\theta}), \tilde{\theta}_2) \right]. \quad (14)$$

Under this specification, the wedges $\hat{W}_t^{RRN}(\theta^t)$ in the absence of LBD are independent of the agents' incomes y . Moreover, as anticipated above, the impulse responses $I_1^t(\theta, y_1(\theta_1))$ are invariant in period-1 income, in which case the novel effects due to LBD are summarized by the impact of y_1 on the expectation of period-2 disutility of labor $\psi(y_2(\theta), \theta_2)$.

First observe that, because y_2 is nondecreasing in period-2 productivity θ_2 , $\Omega(\theta_1) > 0$, meaning that LBD contributes to higher period-1 wedges. The reason is the one anticipated in the Introduction. By shifting the distribution of period-2 productivity towards levels that command lower period-2 rents, the government reduces the expected future welfare losses due to asymmetric information.

Second observe that, because the effects of LBD vanish in the last period, LBD also contributes to dynamics under which wedges decline over time.

Third, consider the effects of LBD on the progressivity of the wedges. Observe that $\hat{W}_t^{RRN}(\theta^t)$ are nonincreasing in productivity if, and only if, $\theta_1 \gamma_1(\theta_1)$ is nondecreasing in θ_1 , as typically assumed in the taxation literature. In the absence of LBD, the theory thus predicts wedges that are nonincreasing in earnings. Interesting, LBD can contribute to a higher progressivity of the period-1 wedges. This is the case if, and only if, $\Omega(\theta_1)$ is increasing. Because higher period-1 types expect, on average, larger rents in the future, the reduction in period-2 handicaps that obtains when the government shifts the period-2 distribution to the left, in a FOSD sense, is highest when applied to high period-1 types. Hence, Ω is typically increasing.

To illustrate, suppose period-1 productivity θ_1 and period-2 shocks ε_2 are drawn from a Pareto distribution (see, among others, Kapicka, 2013). In this case, $\theta_1 \gamma_1(\theta_1)$ is constant, implying that the wedges in the absence of LBD are constant across productivity levels. In this case, $\Omega(\theta_1)$ is strictly positive and increasing. LBD thus contributes to both a larger differential between period-1 and period-2 wedges and to a higher progressivity of the period-1 wedges. These effects are illustrated in Figure 1 below for the same Pareto distribution as in Kapicka (2013), and for income levels computed under the optimal policies (i.e. under the second-best rule χ). As the figure shows, stronger LBD effects (captured by a higher level of the parameter ζ in (2)) are responsible for higher period-1

¹³See Condition (30) in the Appendix.

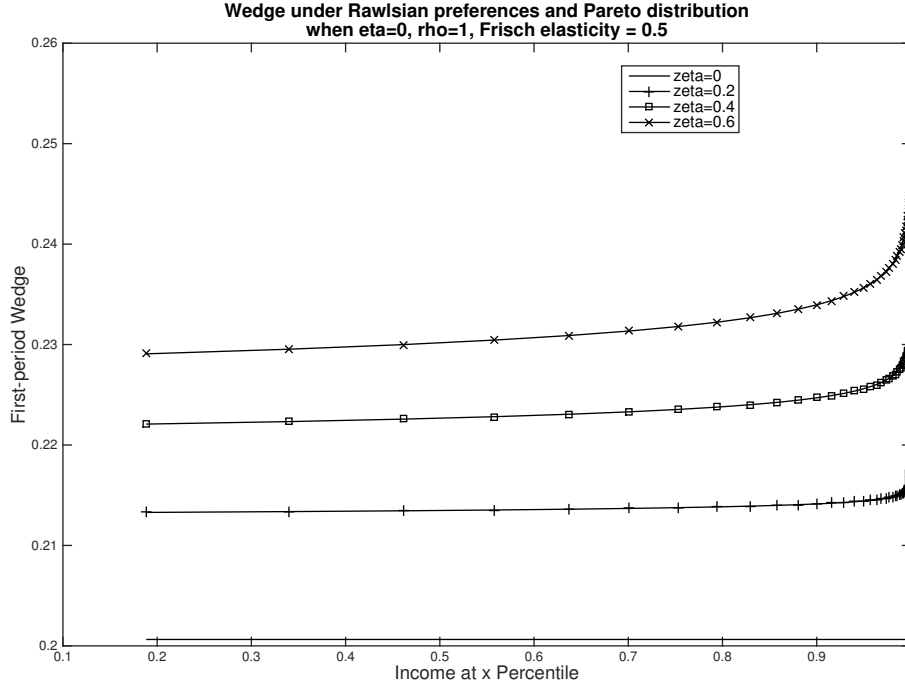


Figure 1: Period-1 wedges in the risk-neutral Rawlsian Pareto case

wedges and for more progressivity at all income percentiles, but in particular at high percentiles.¹⁴

This result is in fact robust to alternative distributions of the skill-shocks, as we show in the next proposition. For instance, Figure 2 below illustrates the first-period wedge function for a Pareto-lognormal skills distribution F_1 with Pareto-tail parameter equal to $\lambda = 5$, as in Diamond (1998).¹⁵ As the figure shows, LBD contributes to higher and more progressive wedges. However, contrary to the Pareto case depicted in Figure 1, the extra effects brought in by LBD are strong enough to turn the optimal period-1 wedge from regressive to progressive only at sufficiently high period-1 productivity levels.

We summarize the above results in Proposition 2 below. To do that, we first need to introduce some definitions and notation.

¹⁴The figure plots the non-normalized period-1 wedge function W_1 which is related to the normalized one by $W_1(\theta_1) = \frac{\hat{W}_1(\theta_1)}{1 + \hat{W}_1(\theta_1)}$. The parameter $\eta = 0$ in the figure indicates that the result is for the case of risk-neutral agents. Also note that the figure assumes $\phi = 2$, i.e., a Frisch elasticity of 0.5, as in Farhi and Werning (2013), Kapicka (2013), and Stantcheva (2016). Finally, the parameter $\rho = 1$ in the figure's caption indicates a skill-persistence parameter of 1 to facilitate the comparison to Farhi and Werning (2013), Kapicka (2013), Golosov et al. (2016), and Stantcheva (2016).

¹⁵A Pareto-lognormal distribution F_1 has support $(0, \infty)$, density $f_1(\theta_1) = \frac{\lambda}{\theta_1^{1+\lambda}} \exp(\lambda M + \lambda^2 \frac{\sigma^2}{2}) \Phi(\frac{\log(\theta_1) - M - \lambda \sigma^2}{\sigma})$, and cdf $F_1(\theta_1) = \Phi(\frac{\log(\theta_1) - M}{\sigma}) - \frac{1}{\theta_1^\lambda} \exp(\lambda M + \lambda^2 \frac{\sigma^2}{2}) \Phi(\frac{\log(\theta_1) - M - \lambda \sigma^2}{\sigma})$, where $\Phi(\cdot)$ is the c.d.f. of the standard Normal distribution. Such a distribution is similar to a Lognormal for small values of θ_1 but has a Pareto right tail. Its mean for $\lambda > 1$ is the product of means of $\text{Log} - N(M, \sigma)$ and $\text{Pareto}(1, \lambda)$ distributions, i.e., $\frac{\lambda}{\lambda-1} \exp(M + \frac{\sigma^2}{2})$.

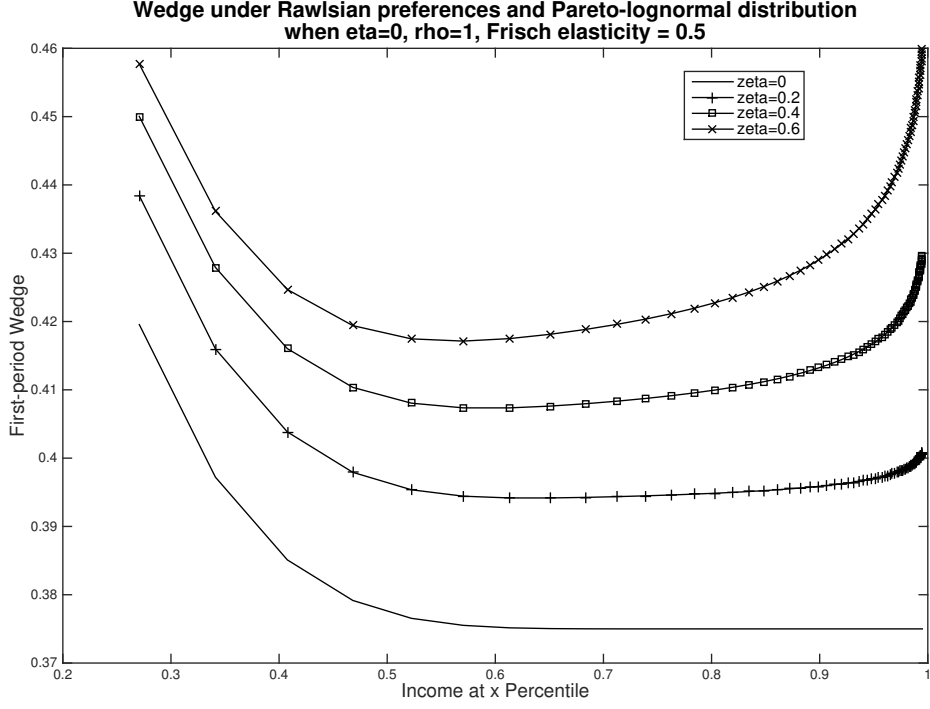


Figure 2: Period-1 wedges in the risk-neutral Rawlsian Pareto-lognormal case

Definition 2. The period-1 wedge is more progressive over the interval $(\theta'_1, \theta''_1) \subset \Theta_1$ in the presence of LBD than in its absence if, and only if, $\hat{W}_1(\cdot)$ is strictly steeper than $\hat{W}_1^{RRN}(\cdot)$ over (θ'_1, θ''_1) . The period-1 wedge under LBD, $\hat{W}_1(\cdot)$, is more progressive than the period-1 wedge in the absence of LBD, $\hat{W}_1^{RRN}(\cdot)$, if, and only if, $\hat{W}_1(\cdot)$ is weakly steeper than $\hat{W}_1^{RRN}(\cdot)$ over the entire support Θ_1 of the period-1 exogenous skill distribution, and strictly steeper over a subset $(\theta'_1, \theta''_1) \subset \Theta_1$.

Using the decomposition in (11), we have that the period-1 wedge is more progressive over the interval (θ'_1, θ''_1) in the presence of LBD than in its absence if, and only if, the function $\Omega(\theta_1)$ is strictly increasing over (θ'_1, θ''_1) . The proposition below identifies necessary and sufficient conditions for this to be the case over the entire support Θ_1 .

Proposition 2. *Suppose the disutility of labor takes the iso-elastic form in (1) and the period-2 productivity is given by (2). The following are true: (i) For all $\theta_1 \in \Theta_1$, $\hat{W}_1(\theta_1) > \hat{W}_1^{RRN}(\theta_1)$; (ii) For all $\theta = (\theta_1, \theta_2)$, $\hat{W}_1(\theta_1) - \hat{W}_2(\theta) > \hat{W}_1^{RRN}(\theta_1) - \hat{W}_2^{RRN}(\theta)$; (iii) Suppose F_1 is Pareto (in which case there exists $\lambda \in \mathbb{R}_{++}$ such that $\theta_1 \gamma_1(\theta_1) = \lambda$ for all θ_1). (a) In the absence of LBD, $\hat{W}_1^{RRN}(\theta_1) = (1 + \phi)/\lambda$, for all θ_1 . (b) In the presence of LBD, $\hat{W}_1(\theta_1)$ is strictly increasing in θ_1 over the entire support $\Theta_1 = \mathbb{R}_+$. (iv) Assume $\hat{W}_1^{RRN}(\theta_1)$ is nonincreasing. The solution to the relaxed program also solves the full program.*

Hence, under the specification considered in the proposition, LBD contributes to higher period-1

wedges across all productivity levels. Furthermore, LBD contributes to a higher differential between period-1 and period-2 wedges, across all histories $\theta = (\theta_1, \theta_2)$. Finally, LBD may also contribute to a higher progressivity of the period-1 wedges. In the proposition, we illustrate such a possibility by considering the case of an economy in which period-1 productivity is drawn from a Pareto distribution. In the Appendix, we show that the result generalizes to arbitrary distributions provided that a certain condition holds.

Remark. When θ_1 and y_1 are complements in the determination of θ_2 , so that impulse responses $I_1^2(\theta, y_1)$ are increasing in y_1 , the effects of LBD on wedges documented in Proposition 2 are reinforced. When, instead, θ_1 and y_1 are substitutes, so that impulse responses $I_1^2(\theta, y_1)$ are decreasing in y_1 , the effects of LBD on period-1 wedges are smaller than under the specification in (2). However, provided the dependence of $I_1^2(\theta, y_1)$ on y_1 is small in absolute value, the results in Proposition 2 continue to hold.

4 Sufficient Statistics

We now show how optimal tax codes can also be obtained through local perturbations, in the spirit of Saez (2001), but adapted to the dynamic economy with LBD under consideration.¹⁶ The purpose of this section is twofold. First, it helps us relating the wedges to the tax code. Second, it permits us to express the formulas for the optimal tax code in terms of sufficient statistics of the empirical earning distribution.

Let $\mathcal{T} = (\mathcal{T}_1, \mathcal{T}_2)$ be a generic tax code, with $\mathcal{T}_t(y^t)$ denoting the period- t income tax schedule. Note that $\mathcal{T}_t(y^t)$ is the total period- t tax payment by any individual with period- t income history y^t . Given \mathcal{T} , for any y^t , then let $\tau_t(y^t) \equiv \partial \mathcal{T}_t(y^t) / \partial y_t$ denote the period- t marginal tax rate at history y^t .

Given the code \mathcal{T} , each individual optimally chooses how much income to generate in each period, taking into account the effects of LBD on the evolution of productivity. Consistently with the notation in the previous section, we denote by $y_1(\theta_1)$ and $y_2(\theta)$ the optimal income policies (the dependence of such policies on \mathcal{T} is dropped to ease the notation).

Given any tax code \mathcal{T} , the labor wedges at the allocations induced by \mathcal{T} , can be related to the tax code as follows (the details are in the Appendix)

$$W_1(\theta_1) = \frac{\tau_1(y_1(\theta_1)) + \delta \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1(\theta_1), y_2(\theta)) dF_2(\theta_2 | \theta_1, y_1(\theta_1))}{1 + \delta \frac{\partial}{\partial y_1} \int [y_2(\theta) - \psi(y_2(\theta), \theta_2)] dF_2(\theta_2 | \theta_1, y_1(\theta_1))} \quad (15)$$

and

$$W_2(\theta) = \tau_2(y_1(\theta_1), y_2(\theta)). \quad (16)$$

The relative wedges can also be related to the underlying tax code \mathcal{T} as follows

$$\hat{W}_1(\theta_1) = \frac{1}{\psi_{y_1}(y_1(\theta_1), \theta_1)} \left[\tau_1(y_1(\theta_1)) + \delta \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1(\theta_1), y_2(\theta)) dF_2(\theta_2 | \theta_1, y_1(\theta_1)) \right] \quad (17)$$

¹⁶See also Golosov et al. (2014) for a variational approach in the context of a dynamic economy.

and

$$\hat{W}_2(\theta) = \frac{\tau_2(y_1(\theta_1), y_2(\theta))}{1 - \tau_2(y_1(\theta_1), y_2(\theta))}. \quad (18)$$

Importantly, note that while the period-2 wedges coincide with the period-2 marginal tax rates, as in static economies, this is not true for the period-1 wedges. The latter combine the period-1 marginal tax rates with marginal variations in the expected period-2 tax bill and in the expected period-2 social surplus induced by a variation in the period-1 incomes. Note that, because of LBD, variations in y_1 affect the period-2 tax bill even when the period-2 tax schedules \mathcal{T}_2 are invariant in period-1 incomes.

The representation in (17) and (18) holds for any tax code. Below we show how wedges under the *optimal* tax code can be obtained through local perturbations that yield tax formulas in terms of sufficient statistics.

Consider first the period-1 tax schedules. The perturbations that lead to the optimal wedges are the familiar one from the optimal taxation literature, whereby the period-1 marginal tax rate is increased by $d\tau_1$ for all earnings in the bracket $[y_1, y_1 + dy_1)$, where y_1 is an income level generated by some type θ_1 under the tax code \mathcal{T} . The above perturbation comes with three effects on the government's objective.

First, all individuals with period-1 earnings (weakly) higher than $y_1 + dy_1$ pay higher taxes (for given earnings), in the amount of $d\tau_1 dy_1$. This is the familiar mechanical effect from the received literature.

Second, all individuals with period-1 earnings in the bracket $[y_1, y_1 + dy_1)$ reduce their period-1 earnings. This is the familiar behavioral effect also discussed at length in the received literature.

The interesting novel effect is the third one, which is specific to dynamic economies and which is affected by LBD. A change in the period-1 marginal tax rate, by triggering a change in the period-1 earnings of those individuals generating incomes in the bracket $[y_1, y_1 + dy_1)$ induces a variation in the period-2 tax revenues. Such variation combines the fact that the period-2 tax schedule $\mathcal{T}_2(y_1, y_2)$ may depend directly on period-1 incomes, 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 leads to a novel period-2 behavioral effect.

For the tax code $\mathcal{T} = (\mathcal{T}_1, \mathcal{T}_2)$ to be optimal, the sum of the above three effects must be zero. To illustrate the implications of this property, we need to introduce some notation. Let $H_Y(y_1)$ be the cumulative distribution of incomes generated by young workers, under the tax code $\mathcal{T} = (\mathcal{T}_1, \mathcal{T}_2)$. Next, consider a fictitious economy in which the original period-1 non-linear tax schedule \mathcal{T}_1 is replaced by the linear tax schedule $\hat{\mathcal{T}}_1$ with constant marginal tax rate $\tau_1 \equiv \tau_1(y_1)$, for some fixed earnings y_1 , and where the period-2 tax schedule \mathcal{T}_2 is the same as in the original economy. Let \hat{h}_Y be the density of the income distribution by young workers in the fictitious economy with tax code $\hat{\mathcal{T}} \equiv (\hat{\mathcal{T}}_1, \mathcal{T}_2)$. Denote by $\hat{y}_1(1 - \tau_1, \theta_1)$ the optimal period-1 income choice of an individual with

period-1 productivity θ_1 in the fictitious economy, and let

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

denote the elasticity of \hat{y}_1 with respect to the net-of-tax constant marginal wage rate $1 - \tau_1$ of those young workers of productivity $\theta_1(y_1)$, where $\theta_1(y_1)$ is the period-1 productivity of all agents whose period-1 income under the original tax code $\mathcal{T} = (\mathcal{T}_1, \mathcal{T}_2)$ is y_1 . Let $H_O(y_2|y_1)$ denote the conditional distribution of period-2 incomes of those workers generating income y_1 when young, under the original economy with tax code $\mathcal{T} = (\mathcal{T}_1, \mathcal{T}_2)$. Finally, let

$$\mathbb{E}[\tilde{\mathcal{T}}_2|y_1] \equiv \int \mathcal{T}_2(y_1, y_2) dH_O(y_2|y_1)$$

denote the average tax bill paid in period-2 by those workers whose period-1 income is y_1 , under the original tax code $\mathcal{T} = (\mathcal{T}_1, \mathcal{T}_2)$, and

$$e_{\mathcal{T}_2|y_1} \equiv \frac{\partial \mathbb{E}[\tilde{\mathcal{T}}_2|y_1]}{\partial y_1} \frac{y_1}{\mathbb{E}[\tilde{\mathcal{T}}_2|y_1]}$$

the elasticity of $\mathbb{E}[\tilde{\mathcal{T}}_2|y_1]$ with respect to period-1 income y_1 .¹⁷ We then have the following result:

Proposition 3. *Suppose the tax code $\mathcal{T} = (\mathcal{T}_1, \mathcal{T}_2)$ is optimal. The following property must hold for all y_1 in the support of the period-1 income distribution:*

$$\frac{\tau_1(y_1)}{1 - \tau_1(y_1)} = \frac{1 - H_Y(y_1)}{y_1 \hat{h}_Y(y_1)} \frac{1}{\hat{E}_1(y_1)} \left[\frac{1}{1 + \delta e_{\mathcal{T}_2|y_1} \frac{\mathbb{E}[\tilde{\mathcal{T}}_2|y_1]}{\tau_1(y_1) y_1}} \right]. \quad (19)$$

The formula in (19) generalizes the formula in Saez (2001) to a dynamic economy with LBD. In particular, the term in the square brackets in (19) adjusts the formula under no LBD to accommodate for the effect of a change in the period-1 income on the expected period-2 tax payments. Importantly, the formula in (19) can be used to test for the optimality of a given tax code \mathcal{T} . By relating the marginal tax rate of young workers to the average tax bill of older workers, the formula offers a concise *sufficient-statistic* test for the optimality of existing tax codes. Clearly, the viability of such test requires a panel data relating taxes paid by individuals early in their careers to the taxes paid by the same individuals later in their careers.

While the formula in (19) has the advantages of being easy to relate to observables, it does not permit one to see how LBD affects the level, progressivity, and dynamics of optimal taxes. These effects are best illustrated by the analysis in the previous section based on the allocation approach. The two approaches should thus be seen as complementary.

Next, consider the period-2 tax schedules. Let $\partial \tilde{H}_O(y_2|y_1)/\partial y_1$ denote the marginal variation of the conditional period-2 income distribution $H_O(y_2|y_1)$ with respect to y_1 , holding fixed the

¹⁷Such elasticity is computed holding $\theta_1(y_1)$ constant.

productivity of the period-1 agents at the level $\theta_1 = \theta_1(y_1)$, where recall that $\theta_1(y_1)$ is the period-1 productivity of those agents generating period-1 income equal to y_1 , under the original tax code $\mathcal{T} = (\mathcal{T}_1, \mathcal{T}_2)$. Next, let $\partial H_O(y_2|y_1)/\partial y_1$ denote the marginal variation of the conditional period-2 income distribution $H_O(y_2|y_1)$ with respect to y_1 , accounting for the variation in $\theta_1(y_1)$ (formally, the total derivative of $H_O(y_2|y_1)$ with respect to y_1 , taking into account also the dependence of $\theta_1(y_1)$ on y_1).

Consider the following reform of the tax code, which 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 with period-1 earnings 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

$$\delta \left(\frac{\partial \tilde{H}_O(y_2|y'_1)}{\partial y_1} - \frac{\partial H_O(y_2|y'_1)}{\partial y_1} \right) d\tau_2 dy_2$$

and (c) an income-contingent period-1 subsidy equal to $S(y'_1) = \delta[1 - H_O(y_2|y'_1)]d\tau_2 dy_2$ to all individuals generating period-1 income $y'_1 \in [y_1, y_1 + dy_1)$. Such perturbation is more sophisticated than the one leading to the formula for the optimal period-1 tax rates in (19). The role of parts (b) and (c) is to neutralize the impact of the variation in the period-2 marginal tax rate on period-1 earnings. They guarantee that the choice of period-1 income by any individual remains the same as prior to the reform. This, in turn, permits us to isolate the effects of the perturbation on period-2 tax revenues. In particular, the reform yields two effects. The first one is a static behavioral effect, originating from the fact that all individuals who, prior to the reform, would have generated period-1 earnings in the bracket $[y_1, y_1 + dy_1)$ and period-2 earnings in the bracket $[y_2, y_2 + dy_2)$ reduce their period-2 earnings.

The second effect is a mechanical effect *specific to dynamic economies* (with, or without, LBD). To understand this effect, note, first, that all individuals with period-1 earnings in the bracket $[y_1, y_1 + dy_1)$ and period-2 earnings above $y_2 + dy_2$ pay higher taxes in the second period in the amount of $d\tau_2 dy_2$, for given earnings in both periods. This means that any individual with period-1 income (prior to the reform) in the bracket $[y_1, y_1 + dy_1)$ expects to pay higher taxes when old. Furthermore, under the reform, all individuals generating period-1 earnings above $y_1 + dy_1$ pay higher taxes in period 1. Again, for the tax code to be optimal, the net effect of any such reform on the NPV of intertemporal tax revenues must be equal to zero. Now, paralleling the analysis for period-1 above, let $\hat{y}_2(1 - \tau_2, \theta_2)$ denote 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 τ_2 . Denote by $\hat{h}_O(y_2|y_1)$ the density of period-2 earnings among those workers generating period-1 earnings equal to y_1 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 $\hat{\mathcal{T}}_2(y_1, \cdot)$ with constant marginal tax rate $\tau_2 = \tau_2(y_1, y_2)$, for some fixed y_2 in the support of $H_O(y_2|y_1)$. Then let

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

denote the elasticity of \hat{y}_2 with respect to the net-of-tax constant wage rate $1 - \tau_2$ of those workers of period-2 productivity equal to $\theta_2(y_1, y_2)$, where $\theta_2(y_1, y_2)$ is the period-2 productivity of all agents who generate period-2 income y_2 after generating period-1 income y_1 , under the original tax code \mathcal{T} . We then have the following result:

Proposition 4. *Suppose the tax code $\mathcal{T} = (\mathcal{T}_1, \mathcal{T}_2)$ is optimal. The following property must hold for all income histories (y_1, y_2) in the support of the income distribution:*

$$\frac{\tau_2(y_1, y_2)}{1 - \tau_2(y_1, y_2)} = \left[\frac{\partial \tilde{H}_O(y_2|y_1)}{\partial y_1} - \frac{\partial H_O(y_2|y_1)}{\partial y_1} \right] \frac{1 - H_Y(y_1)}{h_Y(y_1)y_2\hat{h}_O(y_2|y_1)} \frac{1}{\hat{E}_2(y_1, y_2)}. \quad (20)$$

The formula in (20) generalizes the formula in Saez (2001) to a dynamic economy with LBD and serially correlated types. Once again, such formula complements the one derived in the previous section by relating tax rates to the empirical income distribution in both periods. In the Supplementary Material, we also verify that the wedges under the allocations induced by the optimal tax code derived through the perturbation approach in this section coincide with those derived through the allocation approach in the previous section.

5 Risk Aversion

We now turn to the effects of risk aversion on the impact of LBD on wedges. To this purpose, assume each agent's *lifetime* utility is now given by

$$U(\theta, y, c) \equiv \sum_t \delta^{t-1} (v(c_t) - \psi(y_t, \theta_t))$$

with $v : \mathbb{R} \rightarrow \mathbb{R}$ increasing, strictly concave, twice differentiable, and satisfying the Inada condition $\lim_{c \rightarrow 0} v'(c) = \infty$.¹⁸

When the agents are risk averse, the first-best allocation rule $\chi^* = (y^*, c^*)$ satisfies the following optimality conditions, at all interior points, with $\lambda[\chi^*]$ -probability one (see Lemma 1 in the Appendix):

$$1 + LD_1^{\chi^*}(\theta_1) = \frac{\psi_y(y_1^*(\theta_1), \theta_1)}{v'(c_1^*(\theta_1))}, \quad (21)$$

$$1 = \frac{\psi_y(y_2^*(\theta), \theta_2)}{v'(c_2^*(\theta))},$$

and

$$v'(c_1^*(\theta_1)) = v'(c_2^*(\theta)), \quad (22)$$

¹⁸To highlight the effects specific to LBD, and consistently with most of the literature, we assume the agent cannot privately save.

where, as in the baseline model,

$$LD_1^X(\theta_1) \equiv \delta \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda|x||\theta_1, y_1(\theta_1)} \left[y_2(\tilde{\theta}) - c_2(\tilde{\theta}) + \frac{v(c_2(\tilde{\theta})) - \psi(y_2(\tilde{\theta}), \tilde{\theta}_2)}{v'(c_2(\tilde{\theta}))} \right] \quad (23)$$

captures the effect of a marginal change in period-1 output on expected continuation surplus. Note that, in each period, the agents' disutility of labor, as well as the period-2 continuation utility $v(c_2) - \psi(y_2, \theta_2)$, are weighted by the agents' inverse marginal utility of consumption, so as to express the agents' utility in the same units as in the planner's objective (tax revenues). Importantly, as in the baseline model, the function $LD_1^X(\theta_1)$ is computed holding fixed the period-2 policies $c_2(\cdot)$ and $y_2(\cdot)$. Finally observe that the function $LD_1^X(\theta_1)$, as well as the above optimality conditions, reduce to their counterparts in the baseline model when the agents are risk neutral, that is, when $v(c) = c$, all c .

The first two conditions describe the optimal output choices. They parallel the corresponding conditions in the risk-neutral case, adapted to account for the fact that, under risk aversion, the monetary cost of compensating the agents for the extra disutility of labor is decreasing in the agents' marginal utility of consumption.

The third condition describes the optimal choice of consumption. While, under risk neutrality, the dynamics of consumption are indeterminate, under risk aversion, optimality requires the equalization of the marginal utility of consumption between any two consecutive histories θ_1 and $\theta = (\theta_1, \theta_2)$.

Now consider the second-best policies. Because utility is separable in consumption, the characterization of incentive compatibility is unaffected by the introduction of risk aversion. The second-best policies, and thereby the wedges, however, are affected by the introduction of risk aversion, as we show next.

First observe that, under risk aversion, the definition of the wedges must be adjusted to account for the fact that the cost of compensating the agents for the extra disutility of labor is now inversely proportional to their marginal utility of consumption.¹⁹

Definition 3. Under risk aversion, the labor wedges are given by

$$W_1(\theta_1) \equiv 1 - \frac{\frac{\psi_y(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))}}{[1 + LD_1^X(\theta_1)]} \quad \text{and} \quad W_2(\theta) \equiv 1 - \frac{\psi_y(y_2(\theta), \theta_2)}{v'(c_2(\theta))}. \quad (24)$$

The following result extends the characterization of the labor wedges under the optimal tax code to an economy with risk-averse agents.

¹⁹Under risk aversion, there are also "savings" wedges, that is, distortions in intertemporal consumption smoothing. They are given by

$$W_1^S(\theta^t) \equiv 1 - \frac{v'(c_1(\theta_1))}{\int v'(c_2(\theta)) dF_2(\theta_2|\theta_1, y_1(\theta_1))}$$

Note that the definition uses the assumption that the rate of return on savings (net of any linear capital tax) is equal to $1/\delta$. This wedge has been studied extensively in the received New Dynamic Public Finance literature. Furthermore, LBD impacts these wedges only indirectly. Given the agents' labor supply, the optimal saving wedges are determined as in the economy without LBD. Hereafter, we thus focus on labor wedges.

Proposition 5. *With Rawlsian preferences for redistribution and risk-averse agents, the relative wedges under the optimal tax code are given by (with $\lambda[\chi]$ -probability one)*

$$\hat{W}_1(\theta_1) = \hat{W}_1^{RRA}(\theta_1) + RA(\theta_1)\Omega(\theta_1),$$

and $\hat{W}_2(\theta) = \hat{W}_2^{RRA}(\theta)$, where $\hat{W}_1^{RRA}(\theta_1)$ and $\hat{W}_2^{RRA}(\theta)$ are the wedges in the absence of LBD, $\Omega(\theta_1)$ is the same function capturing the effects of LBD as in the benchmark with risk-neutral agents (as defined in (12)), and

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

is a correction term due to the agents' risk aversion.

What is interesting about the result in the proposition is how risk aversion amplifies the novel effects due to LBD. As in the risk-neutral case, when the planner increases by one unit the expected lifetime utility of a worker of period-1 productivity θ_1 , she also needs to increase the expected lifetime utility of all workers of higher period-1 productivity by the same amount. The correction term $RA(\theta_1)$ represents the extra cost to the planner, in consumption terms (equivalently, in tax revenues terms) of providing such additional utility, stemming from the fact that types above θ_1 have a lower marginal utility of consumption than type θ_1 . To see this, recall that the term Ω in the period-1 wedges captures the marginal variation in the expected net present value of future consumption triggered by a marginal variation in period-1 output, due to LBD. The term $\Omega(\theta_1)v'(c_1(\theta_1))$ translates such increase in utility terms, whereas the term

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

is the total cost (in consumption units) of increasing the utility of all workers with period-1 productivity above θ_1 , accounting for the heterogeneity in such agents' marginal utility of consumption.

We now illustrate how risk aversion interacts with LBD in shaping the level, dynamics, and progressivity of the wedges. For this purpose, consider the same economy as in the previous section, but assume the agents' preferences over consumption are CRRA with coefficient of relative-risk aversion η . To facilitate the comparison with the calibrated economy examined in the next section, all the figures below assume that the period-1 productivity θ_1 and the period-2 shock ε_2 are drawn from a Pareto-lognormal distribution.²⁰

Figure 3 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 $\zeta = 0.6$. As the figure illustrates, in the absence of LBD, risk aversion contributes to higher period-1 wedges. The reason is 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

²⁰The Appendix contains details about the numerical computations behind the figures in this and the following sections.

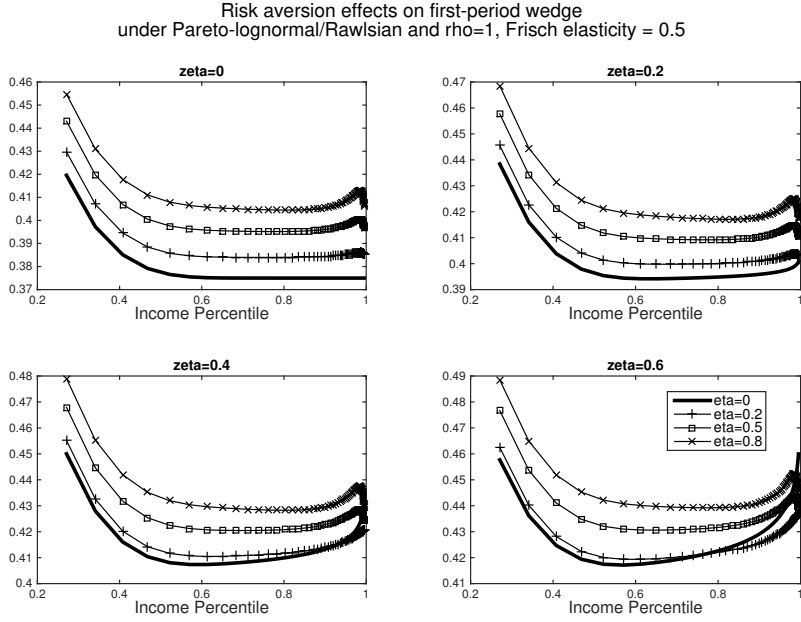


Figure 3: Period-1 wedges in the risk-averse Rawlsian Pareto-lognormal case

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 effect of risk aversion on the period-1 wedges is more convoluted. While, at lower income percentiles, risk aversion continues to contribute to higher wedges, this is not necessarily the case at higher percentiles.

To gain more insights about the interaction of risk aversion with LBD in the determination of the optimal wedges, Figures 4, 5, and 6 illustrate the effects of risk aversion on the various components of the period-1 wedges, as identified in Proposition 5. In particular, Figure 4 illustrates the impact of risk aversion on Ω . To understand the figure, recall that Ω measures the variation in the expected NPV of future information rents triggered by a variation in period-1 labor supply, due to LBD. 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 NPV of period-2 rents. The benefit of shifting the distribution of period-2 productivity towards lower levels to reduce the expected NPV of period-2 rents are thus diminished by the presence of risk aversion. Higher degrees of risk aversion thus contribute to lower levels of Ω . Furthermore, because highly productive agents have a lower marginal utility of consumption than less productive ones, the reduction in period-2 labor supply is

Risk aversion impact on Omega
under Pareto-lognormal/Rawlsian and rho=1, Frisch elasticity = 0.5

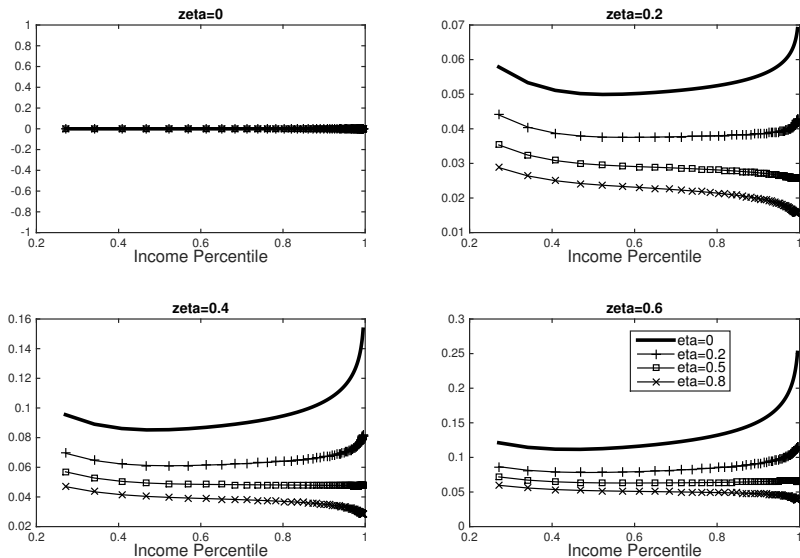


Figure 4: The LBD term Ω in the Rawlsian Pareto-lognormal case

most pronounced at the top of the period-2 distribution. Because productivity is serially correlated, in turn this implies that the reduction in Ω is stronger for high income percentiles. Risk aversion thus also contributes to a reduction in the progressivity of Ω .

Next, consider the effects of risk aversion on the correction term RA , which are illustrated in Figure 5. As discussed above, this term controls for the effects of the heterogeneity in the 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. Recall that the correction term RA consists of two parts. The first part, captured by the term $v'(c_1(\theta_1))$, 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 the distribution of period-2 productivity towards levels that command lower period-2 rents. The second part, captured by the term

$$\int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} \frac{dF_1(s)}{1 - F_1(\theta_1)}, \quad (25)$$

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 . 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. 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. Under the parameters' specification in the figure, higher degrees of risk aversion contribute to a higher incentives cost and hence to a higher level of the correction term, RA . As Figure 5 shows, this is true across all income

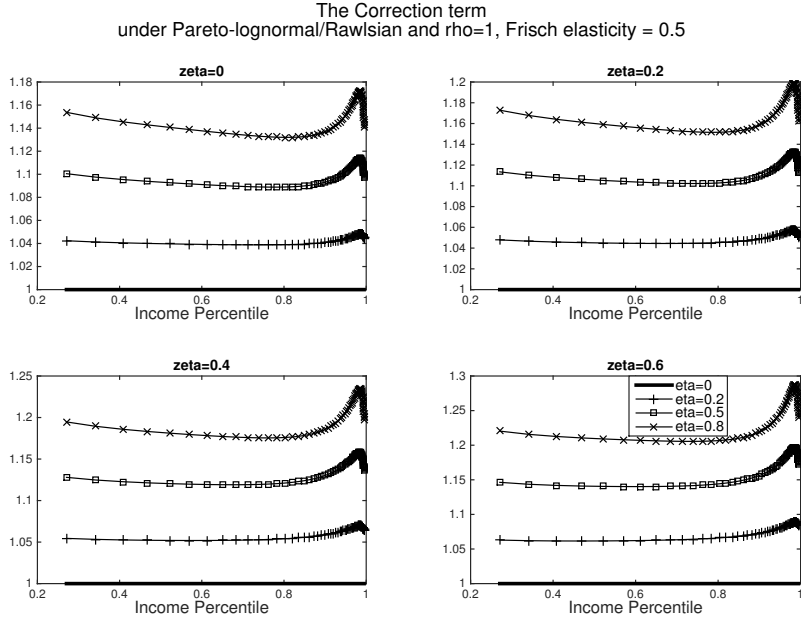


Figure 5: The RA correction term in the Rawlsian Pareto-lognormal case

percentiles and irrespective of the intensity of LBD. The figure also shows that the correction is nonmonotone in the income percentile or, equivalently, in the agents' period-1 productivity. In particular, the effect of an increase in θ_1 on the part of the correction term that controls for the average inverse marginal utility of consumption among the most productive agents, as captured by the term (25) in RA , dominates over the effect of an increase in θ_1 on $v'(c_1(\theta_1))$ for intermediate levels of the period-1 productivity, while it is dominated by the effect of an increase in θ_1 on $v'(c_1(\theta_1))$ for low and high values of θ_1 . As a result, the contribution of the correction term to the progressivity of the period-1 wedges depends on the income percentile.

Finally, Figure 6 depicts the product of the correction term RA with the LBD term Ω . Recall that this product summarizes the net effect of LBD on period-1 wedges, in the presence of risk aversion. As the figure shows, the effect of risk aversion on Ω prevails over the effect of risk aversion on RA , making the term $RA \cdot \Omega$ decrease with risk aversion across all income percentiles. While the overall effect of LBD on wedges remains positive (that is, LBD contributes to higher wedges), the impact of LBD on the level and progressivity of the period-1 wedges is lower under high degrees of risk aversion. This property explains the non-monotone effect of risk aversion on the level and progressivity of the period-1 wedges documented in Figure 3.

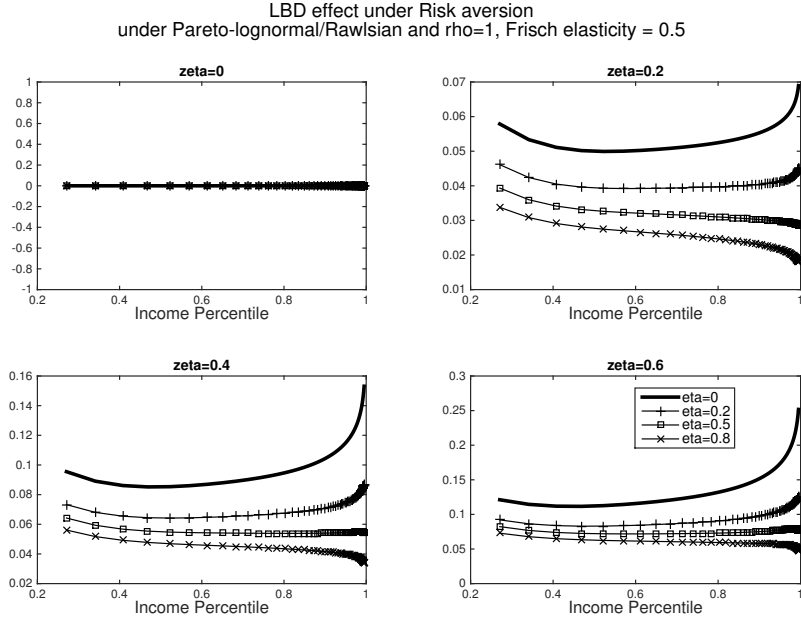


Figure 6: The risk-adjusted LBD term $RA \cdot \Omega$ in the Rawlsian Pareto-lognormal case

6 Utilitarian Preferences for Redistribution

We now turn to economies in which the planner's aversion to inequality is less extreme than in the Rawlsian case considered so far. We consider the opposite extreme case in which the planner assigns equal Pareto weights to each agent's lifetime utility (i.e., has Utilitarian preferences for redistribution). It is well known that such benchmark requires the agents to be risk averse, for otherwise, wedges are identically equal to zero, irrespective of LBD. The analysis in this section is also instrumental to the understanding of the results in the next section where we calibrate the model in this section to US earnings data.

Consider an economy that is identical to the one in the previous section, except for the fact that the planner's redistribution constraint

$$\int V_1(\theta_1) dF_1(\theta_1) \geq \kappa \quad (26)$$

now reflects the equal Pareto weights assigned to the various agents. In particular, the characterization of the FB policies and thereby of the LD^x terms is unaffected by the change in the planner's preferences for redistribution. The second-best policies, instead, are affected, as described in the following proposition.

Proposition 6. *With Utilitarian preferences for redistribution and risk-averse agents, the relative wedges under the optimal tax code are given by (with $\lambda[\chi]$ -probability one)*

$$\hat{W}_1(\theta_1) = \hat{W}_1^{URA}(\theta_1) + [RA(\theta_1) - D(\theta_1)] \Omega(\theta_1),$$

and $\hat{W}_2(\theta) = \hat{W}_2^{URA}(\theta)$. The terms $\hat{W}_1^{URA}(\theta_1)$ and $\hat{W}_2^{URA}(\theta)$ are the relative wedges in the absence of LBD, $\Omega(\theta_1)$ is the same function capturing the effects of LBD as in the benchmark with risk-neutral agents and Rawlsian preferences for redistribution, $RA(\theta_1)$ is the same correction term due to risk aversion as in the economy with Rawlsian preferences for redistribution and risk-averse agents, and

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

is a novel correction term reflecting the higher Pareto weights the planner assigns to all agents with productivity above $\underline{\theta}_1$ under Utilitarian preferences for redistribution.

As in the Rawlsian case, the term $RA(\theta_1)$ scales the LBD term $\Omega(\theta_1)$ to account for the extra costs to the planner of asking for higher period-1 output to type θ_1 , due to the heterogeneity in the agents' marginal utility of consumption. As shown above, this term contributes to an amplification of the LBD term Ω_1 . The novel term here is $D(\theta_1)$. This term captures the higher Pareto weights the planner assigns to all agents whose initial productivity is above $\underline{\theta}_1$. Other things equal, this term naturally contributes to lower wedges. When the planner assigns a Pareto weight equal to one to the utility of each agent whose period-1 productivity exceeds θ_1 , then increasing such agents' lifetime utility by $v'(c_1(\theta_1))\Omega(\theta_1)$, as required by incentive compatibility²¹, comes with the benefit of relaxing the redistribution constraint (26). The benefit for the planner in revenue terms is then equal to²²

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

Equipped with the decomposition in Proposition 6, we now turn to the effects of variations in the degree of risk aversion on the contribution of LBD to period-1 wedges. Figure 7 illustrates how the period-1 wedges under Utilitarian preferences for redistribution vary with the degree of risk aversion and with the intensity of LBD. As anticipated above, wedges are identically equal to zero when $\eta = 0$, i.e., when agents are risk neutral, and increase in the degree of risk aversion, across all intensities of LBD. Higher degrees of risk aversion also contribute to a higher progressivity of the optimal wedges, reflecting the fact that the cost of compensating the agents for higher effort increases with the agents' period-1 productivity due to the lower marginal utility of consumption of highly productive workers.

Figures 8, 9, and 10 focus on the components of the period-1 wedges that are specific to LBD. In particular, Figure 8 illustrates the effect of risk aversion on the Ω term. As in the Rawlsian case, higher degrees of risk aversion contribute to lower period-2 output and hence to a lower benefit of

²¹Recall that $\Omega(\theta_1)$ is the increase in the expected NPV of consumption the planner must provide to type θ_1 when the planner asks the latter agent to work harder in period one, due to LBD. The term $v'(\theta_1)\Omega(\theta_1)$ is the extra utility associated with such increase in consumption.

²²To understand this term, observe that the shadow value, in revenue terms, of providing one extra util in period 1 to all agents, so as to maintain incentive compatibility, is equal to $\int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} dF_1(s)$. This benefit is, obviously, absent under Rawlsian preferences for redistribution, for, in this case, the planner does not attach any positive weight to the utility of any period-1 type above $\underline{\theta}_1$.

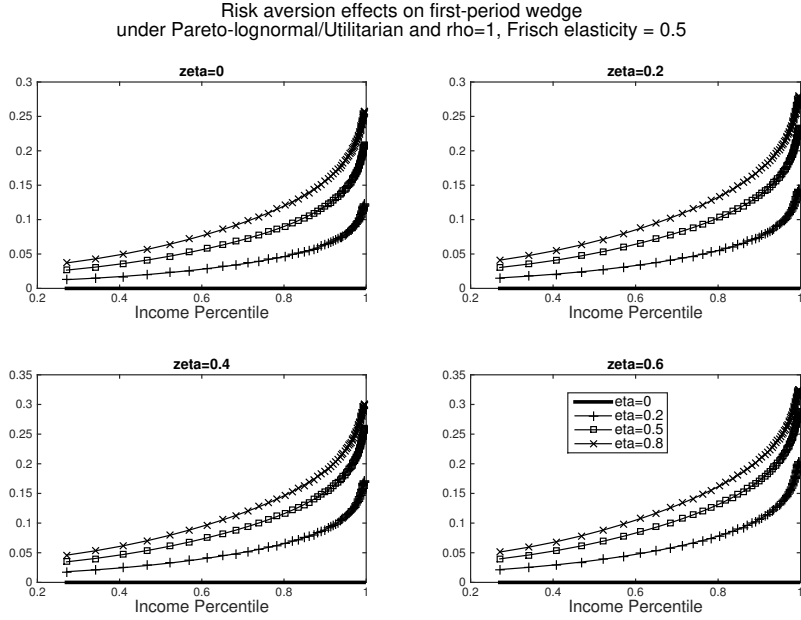


Figure 7: Period-1 wedges in the Utilitarian Pareto-lognormal case

shifting the distribution of period-2 productivity towards levels that command lower information rents. As a result, Ω decreases with the degree of risk aversion, at all income percentiles, and across all intensities of LBD. Higher degrees of risk aversion also contribute to a smaller progressivity of Ω , for the same reasons as in the Rawlsian case.

Figure 9 illustrates the effects of risk aversion on the correction term $RA - D$. Two observations are worth emphasizing. First, relative to the Rawlsian case, the correction term is smaller, reflecting the lower value the planner assigns to redistribution. Second, the progressivity of the correction term is larger than in the Rawlsian case. Again, this property is a direct consequence of the smaller costs the planner assigns to the compensation of the highly productive agents.

Finally, Figure 10 illustrates the net contribution $[RA - D]\Omega$ of LBD to the period-1 wedges. The effect of risk aversion on the corrected term $[RA - D]\Omega$ appears more ambiguous than in the Rawlsian case.

We conclude this section by highlighting that the progressivity of the wedges depends critically on the shape of the skills distribution.²³ To illustrate this point, Figure 11 depicts the wedges under Utilitarian preferences for redistribution and various degrees of risk aversion and LBD, but for the Lognormal skills distribution in Farhi and Werning (2013).²⁴ As the figure illustrates, wedges are now regressive for top percentiles across all degrees of risk aversion. This observation will be important

²³For a related discussion, see also Golosov et al. (2016).

²⁴Specifically, the productivity shock is a Lognormal random variable with mean one and variance parameter $\sqrt{0.0095}$. The latter is the middle value out of the three values considered in Farhi and Werning (2013).

Risk aversion impact on Omega
under Pareto-lognormal/Utilitarian and rho=1, Frisch elasticity = 0.5

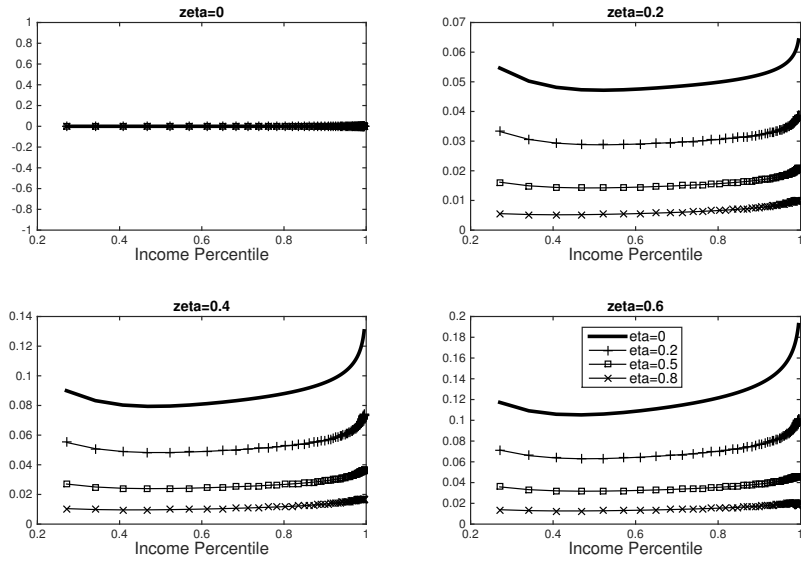


Figure 8: The LBD term Ω in the Utilitarian Pareto-lognormal case

The Correction term
under Pareto-lognormal/Utilitarian and rho=1, Frisch elasticity = 0.5

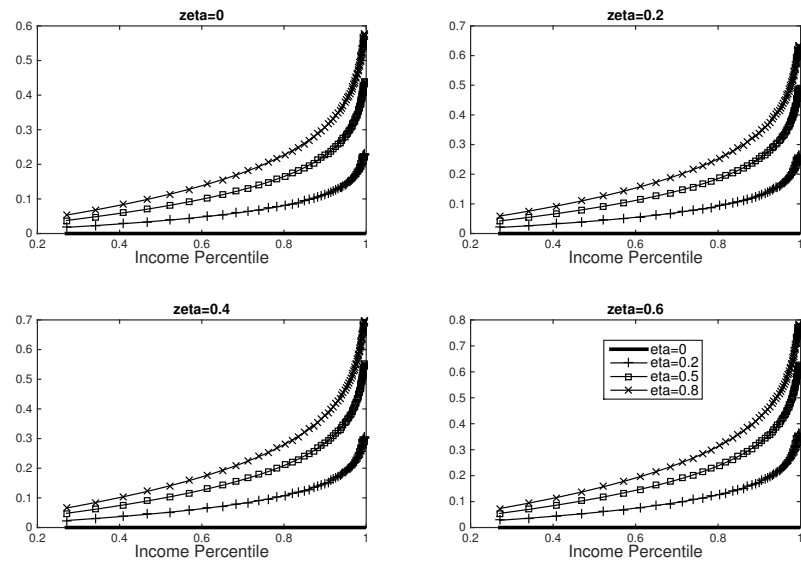


Figure 9: The $RA - D$ correction term in the Utilitarian Pareto-lognormal case

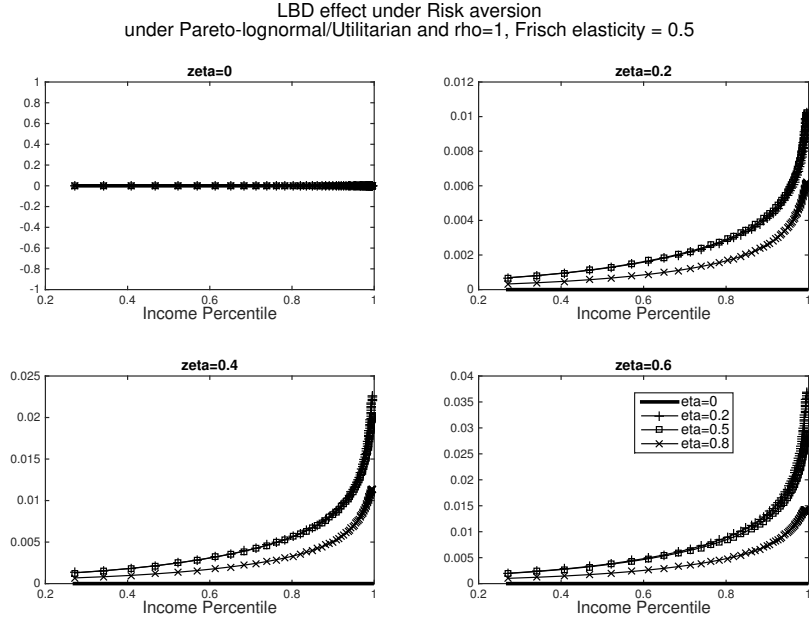


Figure 10: Risk-adjusted LBD term $[RA - D] \cdot \Omega$ in the Utilitarian Pareto-lognormal case

for the interpretation of the quantitative results in the next section.

7 Quantitative Analysis

We now turn to the quantitative implications of our analysis.²⁵ First, we calibrate the model to capture various moments of the US earnings distribution under the existing US tax code. We then derive the optimal wedges for the calibrated economy and conduct a few counterfactual exercises aimed at illustrating the quantitative effects of LBD on the level, progressivity, and dynamics of the wedges and taxes under optimal codes. Finally, we derive the optimal tax code when taxes are restricted to be either linear or linear-power functions of reported incomes and compare it to the existing US tax code.

As in most of the dynamic public finance literature, we assume that agents work for 40 years (see, among others, Farhi and Werning, 2013, Golosov et al., 2016, and Stantcheva, 2016). Consistently with the analysis in the previous sections, however, we assume that productivity changes only in the middle of each agent's working life (as in Best and Kleven, 2013, and Kapicka and Neira, 2016). We allow productivity in the second half to depend on labor supply in each of the periods in the first half. We assume that the effect of $(y_s)_{s=1}^{20}$ on θ_2 is summarized by a weighted average of the income levels in the first twenty years, with the weights $\hat{\beta}_s / \sum_{s=1}^{20} \hat{\beta}_s$ declining over time, $s = 1, \dots, 20$, and with each $\hat{\beta}_s$ being a positive scalar. This specification is thus consistent with empirical evidence

²⁵Details on the computations can be found in the Appendix.

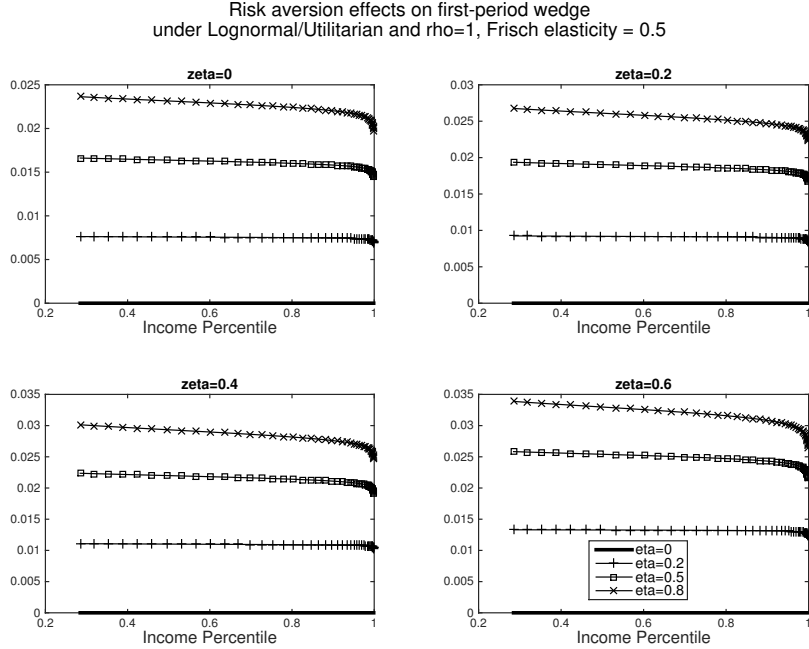


Figure 11: Period-1 wedges in the Utilitarian Lognormal case

documenting that LBD in earlier periods has more pronounced effects on wages and productivity than in later periods. See, for instance, Dustmann and Meghir (2005).

Interestingly, when the weights decay geometrically, that is, when the vector $(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_{20})$ is proportional to $(1, \beta, \dots, \beta^{19})$, where β is the annual discount factor, and the latter equals the inverse of the gross interest rate (as typically assumed in most of the literature), then consumption and earnings under both the existing US tax code and the optimal one are constant over each of the two 20-years blocks. That is, consumption and earnings are equal to (c_1, y_1) in each of the first 20 years, and then equal to (c_2, y_2) in each of the subsequent 20 years. Moreover, the allocations in this economy coincide with the corresponding ones in the two-period model of the previous sections, after setting $\delta = \beta^{20}$. This equivalence permits us to retain insights from the analysis in the previous sections, while also permitting us to draw comparisons with the existing literature.²⁶

7.1 Calibration

To calibrate the model, we first fix a few parameters to the levels typically assumed in the literature (see Table 1).²⁷ As in the rest of the literature, we assume that the disutility of labor takes the

²⁶See the Supplement to this article for a formal proof of the equivalence between the two economies.

²⁷The acronyms in the table should be interpreted as follows. BK stands for Best and Kleven (2013); FW stands for Farhi and Werning (2013); GTT for Golosov et al. (2016); K for Kapicka (2013); KN for Kapicka and Neira (2016); S for Stantcheva (2016). FW and S assume that $\beta = 0.95$, GTT that $\beta = 0.98$, K and KN that $\beta = 0.96$, and BK that $\beta = 1$. Our choice of $\beta = 0.9615$ is consistent with the assumptions in almost all these papers that the annual

Definition	Symbol	Value	As in
CRRA parameter	η	1	FW, K, GTT, S, KN
Frisch elasticity of labor	$1/\phi$	0.5	FW, GTT, S, BK
Annual interest rate	r	4%	KN
Annual discount factor	β	$1/(1+r)$	FW, K, GTT, S, BK
Working years per period	–	20	BK, KN
Cutoff year	–	21	BK

Table 1: Exogenous parameters

iso-elastic form in (1) and that the utility of consumption takes the log specification $v(c) = \ln(c)$. Next, we postulate that period-1 productivity is given by

$$\theta_1 = h_1 \varepsilon_1$$

where h_1 is a positive scalar whereas θ_2 is given by

$$\theta_2 = h_2 \theta_1^\rho \left(\frac{\sum_{s=1}^{20} \beta^{s-1} y_s}{\sum_{s=1}^{20} \beta^{s-1}} \right)^\zeta \varepsilon_2,$$

as in the previous sections, except for the fact that y_1 is replaced by

$$\bar{y}(\theta) \equiv \frac{\sum_{s=1}^{20} \beta^{s-1} y_s(\theta)}{\sum_{s=1}^{20} \beta^{s-1}}.$$

We also assume that the productivity shocks in the two periods are i.i.d. random draws from a Pareto-lognormal distribution with parameters (μ, σ^2, λ) . The parameter λ governs the Pareto right tail of the distribution, whereas the parameters (μ, σ^2) govern the Lognormal left tail of the distribution. We truncate the distribution at the 1% percentile and set μ so that the mean of the truncated distribution is equal to one.

To calibrate the parameters $(h_1, \rho, \zeta, \sigma, \lambda)$, we use the following estimation

$$T(y) = y - e^{\tau_0} y^{1-0.181}$$

of the existing US income tax code from Heathcote et. al. (2016), with the parameter $\tau_0 = -0.1005$ set so that the total tax revenues are normalized to zero.²⁸ Given the above tax code, in the benchmark economy, workers maximize their expected lifetime utility by choosing income and consumption in each period, taking as given the exogenous net interest rate, and accounting for the effects of LBD

discount factor is equal to the inverse of the annual gross interest rate, while being somewhere in the middle of the aforementioned choices for the annual discount factor, excluding the one in BK.

²⁸Golosov et. al. (2016) use a similar estimation, but from the 2014 version of the Heathcote et. al. paper. Namely, they assume that $T(y) = y - e^{\tau_0} y^{1-0.151}$.

Symbol	Value	Target Moment	Data	Absolute Percentage Deviation
ρ	0.4505	mean earnings ratio	0.868	0.0015%
ζ	0.2175	Var. log-earnings young	0.335	1%
h_1	0.4795	Var. log-earnings old	0.435	0.009%
σ	0.5573	Gini earnings young	0.3175	1.7%
λ	5.9907	mean-to-median earnings young	1.335	1.25%

Table 2: Calibrated Parameters

on the evolution of their productivity.²⁹ The parameters $(h_1, \rho, \zeta, \sigma, \lambda)$ are calibrated by minimizing the sum of the squared deviations of five simulated moments of the earnings distribution under the above tax code from their corresponding moments in the data, as reported in Huggett et. al. (2011), with each deviation expressed as a percentage of the target moment.

The first target moment is the ratio between the mean earnings of young workers and the mean earnings of old workers, as in Kapicka and Neira (2016).³⁰ The remaining four target moments are (a) the variance of log-earnings for young workers (years 1-20), (b) the variance of log-earnings for old workers (years 21-40), (c) the Gini coefficient for the earning distribution of young workers, and (d) the mean-to-median ratio in the earning distribution of young workers. These moments are computed by taking the average of the corresponding annual moments in Figure 1 in Huggett et. al. (2011) over the first 20 years and over the second 20 years in the workers' lifecycle, using year 21 as the cutoff age. Table 2 reports the calibrated parameters, the target moments, and the absolute percentage deviations of the model-generated moments from the target moments. We note that the calibrated value for the LBD parameter ζ is consistent with the estimated range $[0.2, 0.6]$ in the metadata analysis of Best and Kleven (2013).

7.2 Optimal wedges

Given the above calibration, we then characterize the optimal wedges for the calibrated economy, assuming that the planner has Utilitarian preferences for redistribution.³¹ We let the parameter κ in the planner's dual problem be equal to the average lifetime utility in the benchmark econ-

²⁹The annual interest rate can be thought of as net of any (exogenous) linear capital tax rate.

³⁰Contrary to our paper, Kapicka and Neira (2016) separate the data in Heathcote et. al. (2016) by using year 20 instead on year 21 as the cutoff year. However, the mean earnings reported in Panel A of Figure 1 in Heathcote et. al. (2016) for year 20 and year 21 are virtually identical, so the distinction is not quantitatively relevant.

³¹Assuming the planner has Utilitarian preferences for redistribution eases the comparison with the pertinent quantitative literature. In fact, the same preferences for redistribution are assumed in the numerical analysis in Farhi and Werning (2013), Kapicka (2013), Golosov et al (2016), Stantcheva (2016), and Kapicka and Neira (2016). Best and Kleven (2013), instead, assume non-linear Pareto weights, because workers in their model are risk neutral, implying that, under linear Pareto weights, all wedges would be identically equal to zero.

omy, and notice that, under the above calibration, $\kappa = -1.645$. Consistently with the analysis in the previous sections, we replace the agents' incentive constraints with the local first-order (envelope) conditions and then verify that the solution to the relaxed program satisfies all the remaining incentive-compatibility constraints. The latter is done by verifying that (i) second-period earnings are increasing in second-period shock/productivity, for any given level of first-period productivity, and (ii) all the integral monotonicity conditions in (6) are satisfied.

We first report that reforming the existing tax code by adopting the optimal one would yield a 4.0348% increase in consumption in each period and at each productivity level.³² This appears a significant improvement, although it is lower than the figure that one would obtain by assuming no LBD effects (see below).

We then turn to the optimal wedges in the calibrated economy. Period-1 wedges have an inverse-U shape with respect to the period-1 income percentile, which in turn tracks period-1 productivity. Figure 12 zooms into the distribution of the period-1 wedge, focusing on the top 75% of the distribution. The inverse-U shape of the wedges appears to follow from the fact that the calibrated Pareto-lognormal distribution has a Pareto right tail only asymptotically, i.e., for productivity levels exceeding the 99.9-percentile. That is, for the percentiles reported in the various figures in this section, it is as if the shock distribution is Lognormal. Under such distribution, wedges have an inverted U-shape (see Golosov et al. as well as the discussion at the end of the previous section).

The figure also highlights three selected income percentiles corresponding to low, middle, and high earnings. The middle earnings correspond to earnings approximately equal to the period-1 unconditional mean earnings. The low (alternatively, high) earnings correspond to earnings approximately equal to half (alternatively, twice) the mean of the period-1 earnings.

Figure 13 in turn shows the period-2 wedge as a function of the period-2 earnings percentile, for each period-1 productivity shock corresponding to the low, middle, and high earnings defined above. The figure also reports the average period-2 wedge as a function of the period-2 income percentile for a weighted average of period-1 productivity shocks (with weights equal to the respective likelihoods). As with Figure 12, we also highlight three selected period-2 percentiles corresponding to low, median, and high earnings. The middle earnings correspond to period-2 earnings approximately equal to the unconditional mean earnings in the second period, when period-1 earnings are the middle earnings defined above. The low (alternatively, high) period-2 earnings correspond to earnings approximately equal to half (alternatively, twice) the mean earnings in the second period, when period-1 earnings are equal to the low (alternatively, high) levels defined above. Period-2 wedges also have an inverse-U shape. However, contrary to their period-1 counterparts, such period-2 wedges decrease after the

³²This number is calculated as follows. Holding the agents' expected lifetime utility constant, reforming the tax code yields an extra total discounted tax revenue of $\Delta R = 0.0287$. Such extra revenue can be rebated to the agents in a way that permits a constant proportional increase of $x\%$ in consumption in each period and each history. The number reported in the main text is then obtained by dividing ΔR by the net present value of total consumption in the benchmark economy, which is equal to 0.7121.

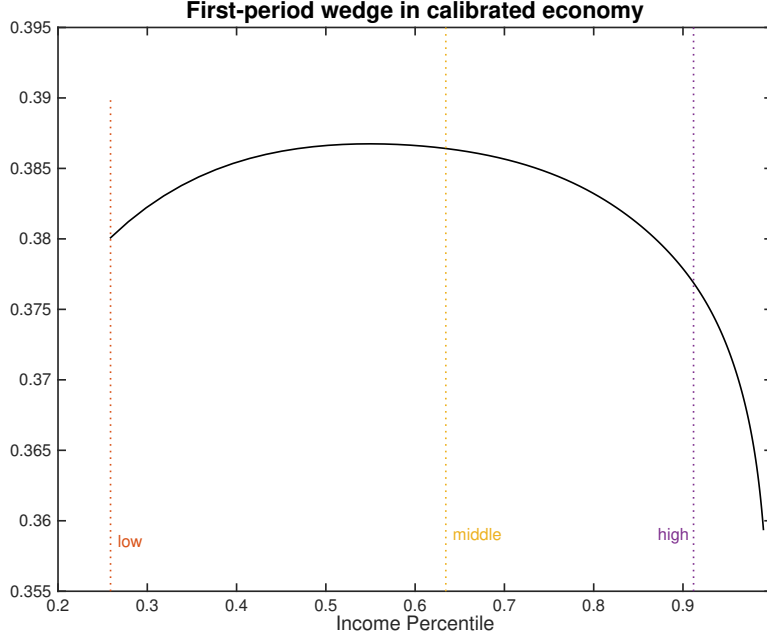


Figure 12: Period-1 optimal wedge. The vertical lines indicate the period-1 income percentiles corresponding to low, middle and high earnings.

30% percentile. Period-1 wedges exhibit progressivity over a range of (conditional) income percentiles for which period-2 wedges are regressive (e.g. between the 30% and 60% percentile). Also note that, at any percentile, period-2 wedges are increasing in period-1 incomes, across the selected histories described above.

Next, we discuss the dynamics of the wedges. Figure 14 shows the period-1 wedge as a function of the period-1 income percentile (solid line). It also shows the corresponding period-2 wedge at histories at which (a) the second-period earnings $y_2(\theta)$ coincides with the period-2 mean earnings given the corresponding period-1 income percentile, i.e., $\mathbb{E}[y_2(\theta_1, \theta_2) | \theta_1, y_1(\theta_1)]$ (dash line), and (b) the productivity shocks experienced in the two periods coincide, i.e., $\varepsilon_1 = \varepsilon_2$. For these histories, wedges increase over the lifecycle, due to the high degree of risk aversion (recall the discussion in the previous section).

Table 3 reports the wedges for the three particular histories of productivity shocks mentioned above. The middle earnings history corresponds to productivity shocks that yield earnings approximately equal to the unconditional mean earnings *in each* period. The low (alternatively, high) earnings history corresponds to productivity shocks that yield earnings approximately equal to half (alternatively, twice) the mean earnings in each period. In each period, wedges are inverted-U-shaped in productivity, confirming the result in the previous figures. Furthermore, for any such history, wedges increase with age, reflecting again the pattern documented in the previous figure.

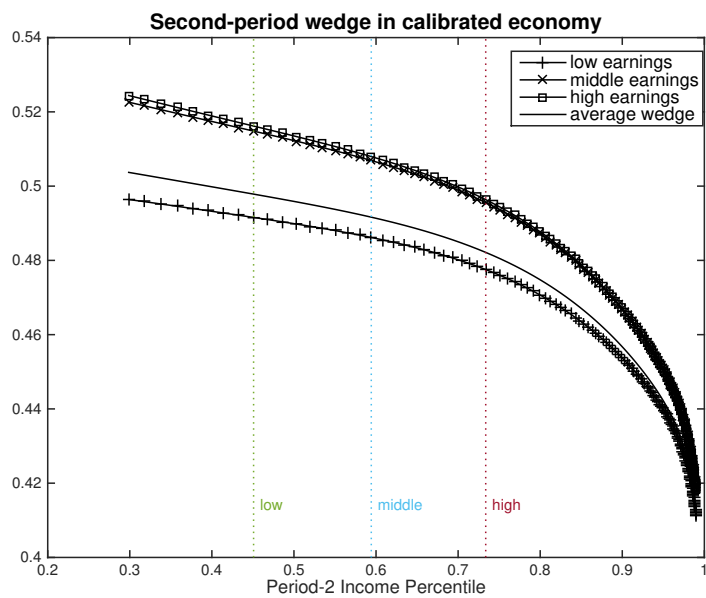


Figure 13: Period-2 wedges as a function of the period-2 earnings percentiles for (a) low, middle and high period-1 earnings, and (b) the weighted average of period-1 productivities. The vertical lines indicate the conditional income percentiles for the low, middle and high conditional period-2 earnings.

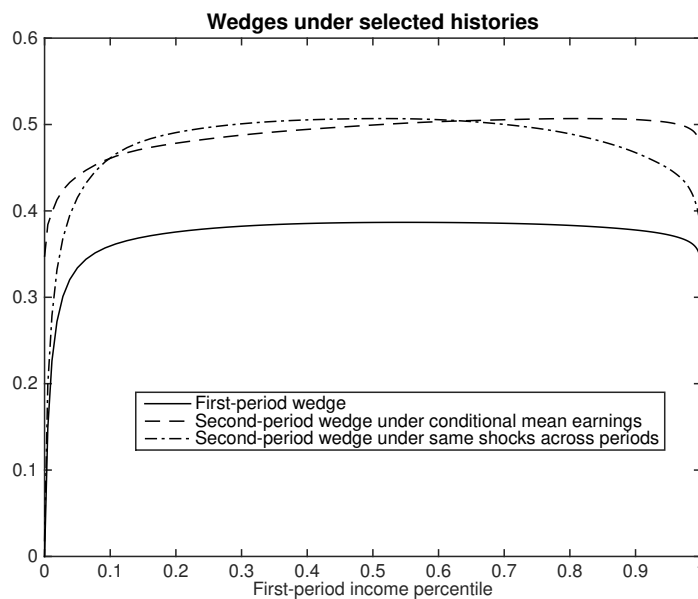


Figure 14: Optimal wedges for selected histories

	First-period wedge	Second-period wedge
Low earnings	0.3801	0.4916
Middle earnings	0.3864	0.5069
High earnings	0.3769	0.4964

Table 3: Optimal wedges for selected histories

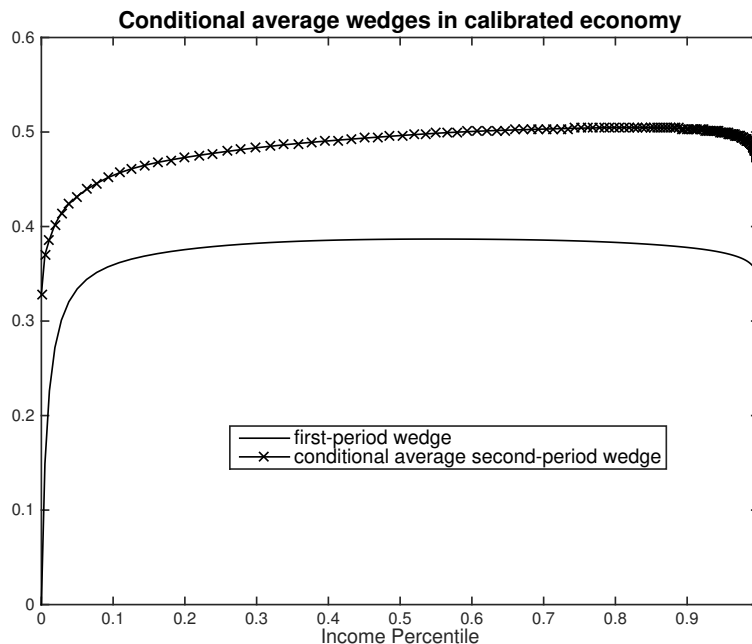


Figure 15: Period-1 wedges and conditional period-2 wedges as a function of period-1 income percentile.

The pattern that emerges from all these figures is that wedges tend to increase over the lifecycle. While there exist histories for which the period-2 wedge is lower than the corresponding period-1 wedge, the conditional average period-2 wedge is higher than the corresponding period-1 wedge at any given period-1 income percentile, as can be seen from Figure 15 (that is, $\mathbb{E}[W_2(\theta_1, \theta_2) | \theta_1, y_1(\theta_1)] > W_1(\theta_1)$). The average period-1 wedge (across all histories) is equal to 0.3733, whereas the average period-2 wedge (also across all histories) is equal to 0.4854.

These results echo those in Farhi and Werning (2013) and Golosov et al. (2016), but are in contrast to those in Best and Kleven (2013). Interestingly, the average wedges in our calibrated economy are higher than the corresponding averages in the first and in the second halves of a working life in the Lognormal simulations in Farhi and Werning (2013) and Golosov et al (2016), which is consistent with the findings in the previous sections.³³

³³Golosov et al. (2016), however, report the wedges only for top earners.

Next, consider the progressivity of the wedges. As anticipated above, the inverted-U shape largely comes from the high variance of the calibrated Pareto-Lognormal distribution of the productivity shocks which makes the latter de facto very similar to a Lognormal distribution, with $[1 - G(\varepsilon)]/\varepsilon g(\varepsilon)$ decreasing in ε and approaching λ only asymptotically. As discussed after Proposition 2 and at the end of the previous section, under such a distribution, wedges should not be expected to be progressive over the entire range of income percentiles (Note that the shape of the wedges for high earnings percentiles in the last two figures is similar to the one in Figure 5 in Golosov et al. (2016)).

A second reason why first-period wedges are regressive at high income percentiles is the high degree of risk aversion ($\eta = 1$) assumed in the simulations. As indicated in the previous section, a high degree of risk aversion implies high compensation costs in the second period and thereby low second-period information rents. In turn, this effect alleviates the contribution of LBD to the progressivity of the period-1 wedges. See Figure 10 for the effects of risk aversion on the level and progressivity of the component of the wedge affected by LBD.

Finally, note that the moderate degree of skill persistence ($\rho = 0.4505$) and the moderate level of LBD ($\zeta = 0.2175$) in the calibrated economy also contribute to an alleviation of the impact of LBD on the progressivity of the period-1 wedges. Recall that the lower these values the smaller the effects of current types and incomes on the distribution of future rents. As a result, the lower these values, the smaller the benefits from distorting the labor supply of high earners to economize on future rents.

7.3 Simple Taxes

We now turn to the question of whether simple taxes yield most of the welfare gains from reforming the current tax code. Consider the following class of age-dependent taxes

$$T_1(y_1) = -B + y_1 - e^{\tau_{0,1}} y_1^{1-\tau_1}$$

and

$$T_2(y_2) = y_2 - e^{\tau_{0,2}} y_2^{1-\tau_2}.$$

Observe that the marginal income tax in period t is increasing in τ_t . Also note that the special case of linear taxes corresponds to $\tau_1 = \tau_2 = 0$, in which case the constant marginal tax rates are equal to $1 - e^{\tau_{0,1}}$ and $1 - e^{\tau_{0,2}}$ for young and old workers, respectively. The age-independent tax code that approximates the US current tax code, as estimated in Heathcote et. al. (2016), corresponds to the case $B = 0$ and $\tau_1 = \tau_2 = 0.181$.

To derive the optimal tax code within this class, we solve for the values of $B, \tau_{0,1}, \tau_{0,2}, \tau_1, \tau_2$ that maximize the average lifetime utility of each worker, subject to the constraint that total tax revenues be at least zero (recall that average tax revenues were normalized to zero in the benchmark economy, i.e., under the current tax code). We refer to the solution to this problem as the quasi-optimal tax

code.³⁴

The quasi-optimal income tax code is given by $B = 0.2603$, $\tau_{0,1} = -0.4769$, $\tau_{0,2} = -0.6231$, $\tau_1 = 0.0055$, and $\tau_2 = -0.0186$. The code is mildly progressive for young workers and mildly regressive for the old ones. Interestingly, it yields almost all of the welfare gains of reforming the existing tax code by adopting the fully optimal unconstrained code reported above. In particular, while adopting the fully optimal code yields an increase in consumption of 4.0348% at all histories and periods, adopting the quasi-optimal tax code yields an increase in consumption of 3.8859%. The loss from simple taxes is thus only 0.1489% in consumption.³⁵ Therefore, the quasi-optimal tax code is approximately optimal.

As anticipated in the Introduction, virtually all of the welfare gains from adopting the quasi-optimal income tax code can also be generated by adopting a code where taxes are restricted to be linear:³⁶ Adopting the optimal linear tax code yields an equi-proportionate increase in consumption at all histories and periods equal to 3.8842%, which is only 0.1506% less than the increase under the fully optimal code. These results thus indicate that simple linear, but age-dependent, taxes are approximately optimal. The marginal tax rates in the optimal linear tax code are equal to 38% and 46% for young and old workers, respectively.³⁷ Consistently with the results reported above for the optimal wedges, marginal tax rates increase over the lifecycle.

In Figure 16, we plot together the marginal income tax rates as functions of (unconditional) income percentiles, when the income tax code is (a) the quasi-optimal one (solid line), (b) the optimal linear one (dash line) and (c) the estimated current US one. The quasi-optimal and linear tax rates are increasing with age, and higher than the ones in the current US tax code for young workers. For old workers, the quasi-optimal and linear tax rates are higher than the ones in the current US tax code at low income percentiles, but lower at higher percentiles. Importantly, the tax rates in the linear code are very close to those in the quasi-optimal code.

Finally, suppose the planner is constrained to use age-independent taxes (that is, $\tau_{0,1} = \tau_{0,2}$ and $\tau_1 = \tau_2$). The optimal age-independent tax code is given by $B = 0.2624$, $\tau_0 = -0.5312$ and $\tau = 0.0022$. Not surprisingly, the optimal age-independent code is close to an average of the two schedules of the quasi-optimal tax code. The welfare gains in consumption terms from replacing the current code with the optimal age-independent one are equal to a 3.7988% equi-proportionate increase in consumption at all histories and periods. Therefore, the loss from using this code rather

³⁴Note that the problem here is the primal of the dual problem considered above to derive the fully optimal wedges.

³⁵Observe that, because $v(c) = \ln(c)$, if consumption at each period and history increases by $x\%$, then lifetime utility increases by $\Delta V = [1 + \delta] \log(1 + x)$. The welfare gains brought about by changing the code from the current one to the quasi-optimal one are $\Delta V = 0.055521$. Therefore, when translated in consumption terms, the reform yields an equi-proportionate increase in consumption at every period and history equal to $(e^{\Delta V})^{\frac{1}{1+\delta}} - 1 = 0.038859$.

³⁶The welfare gains from replacing the current tax code with the optimal linear one are $\Delta V = 0.055499$. When translated in consumption terms, the reform yields an equi-proportionate increase in consumption at all histories and periods equal to $x = (e^{\Delta V})^{\frac{1}{1+\beta^{20}}} - 1 = 0.038842$, that is, 3.8842%.

³⁷The optimal linear code is given by $B = 0.2599$, $1 - e^{\tau_{0,1}} = 0.3806$, and $1 - e^{\tau_{0,2}} = 0.4595$.

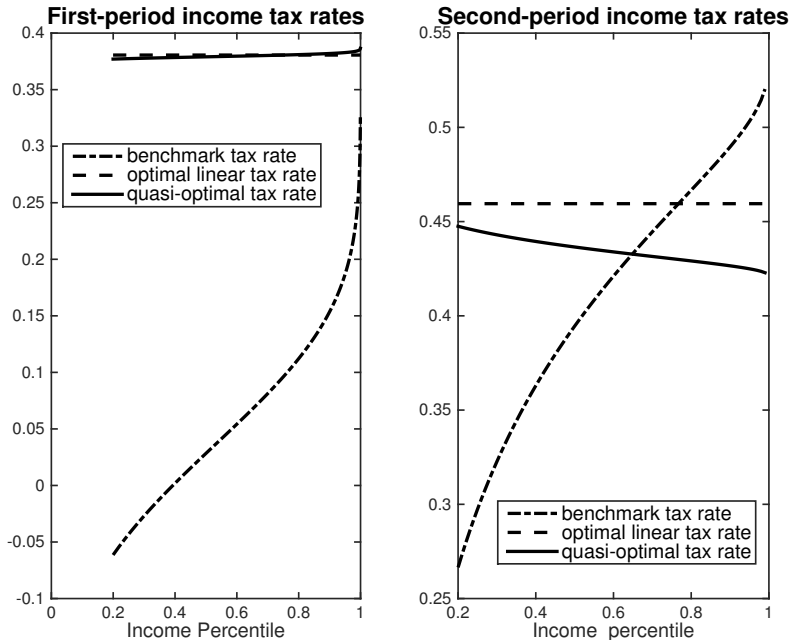


Figure 16: Tax rates as functions of income

than the quasi-optimal one is 0.087% in consumption.³⁸

If the planner is further constrained to use age-independent linear taxes, the optimal code is given by $B = 0.2631$ and an $\tau_0 = -0.5318$, which implies a constant tax rate of 41.25%. The welfare gains in consumption terms from replacing the current code with the optimal age-independent linear code are equal to a 3.7987% equi-proportionate increase in consumption at all histories and periods. Thus, the loss from using this code rather than the quasi-optimal one is 0.0871% in consumption terms.

7.4 Importance of LBD

We now turn to the quantitative effects of LBD on wedges and taxes in the calibrated economy. For this purpose, suppose that period-2 productivity were exogenous and given by $\theta_2 = h_2 \theta_1^{\hat{\rho}} \varepsilon_2$, where h_2 and $\hat{\rho}$ are positive scalars. All the parameters of the model are the same as in the previous subsections, except for ζ and ρ which are replaced by h_2 and $\hat{\rho}$. Now let the values of h_2 and $\hat{\rho}$ be determined so as to minimize the sum of the squared residuals $\left(\theta_1^\rho y_1(\theta_1)^\zeta - h_2 \theta_1^{\hat{\rho}}\right)^2$ across the two models (in percentage term), with $y_1(\theta_1)$ denoting the period-1 incomes in the economy with LBD, under the existing US tax code. The values of h_2 and $\hat{\rho}$ that minimize the squared residuals are

³⁸Note that if we were to further constrain the planner to using a tax code in which $B = 0$, with age-independent taxes, the optimal tax code in this class would feature $\tau = 0.3846$. Such code is significantly more progressive than the current US tax code, as estimated in Heathcote et. al. (2016).

Symbol	Value	Target Moment	Data	Absolute Percentage Deviation
$\hat{\rho}$	0.6981	mean earnings ratio	0.868	0.064%
h_2	0.9866	Var. log-earnings young	0.335	0.43%
h_1	0.4795	Var. log-earnings old	0.435	2.6%
σ	0.5573	Gini earnings young	0.3175	2.4%
λ	5.9907	mean-to-median earnings young	1.335	1.5%

Table 4: Calibrated Parameters

$\hat{\rho}=0.6981$ and $h_2 = 0.9866$. Under these values, the maximum absolute percentage squared residual is 2.3%. The distance of the earnings distribution under such parameters configuration from the one in the data is also very small as one can see from Table 4.

Therefore, the above two parameter values, along with the other ones in Table 4, represent a suitable calibration of this alternative economy with no LBD. Note that, by construction, the distribution of θ_1 in this alternative economy is identical to the one in the economy with LBD. Likewise, the conditional distribution of θ_2 for each θ_1 is also the same in the two economies. The only difference across the two economies is the endogeneity of the period-2 productivity. Importantly, such endogeneity has significant implications for (a) the structure of the optimal wedges, (b) the value of reforming the tax code, and (c) the structure of simple taxes approximating the optimal code. Because the productivity distributions are the same in the two economies, such differences would also be present if the analyst could measure productivity directly. The comparison across these two economies thus permits us to isolate the quantitative effects of LBD.

First, consider the value of reforming the tax code. In this alternative economy, moving from the existing tax code to the optimal one yields a 4.65% increase in consumption at all histories and periods. Note that this figure is 15% larger than the one in the economy with LBD. Ignoring LBD, thus leads to a significant overestimation of the benefits of reforming the current tax code.

Next, consider the wedges. In this alternative economy, the period-1 optimal wedges are distinctively lower than in the corresponding economy with LBD, for all income percentiles. Moreover, it is more regressive over various segments of the income distribution. This result, which is consistent with the discussion in the previous section, is illustrated in Figure 17.

The comparison is sharper if one looks at the allocations under the quasi-optimal tax codes (the linear-power codes defined above). This comparison is also motivated by the fact that, with or without LBD, the period-1 and period-2 earnings distributions generated by the quasi-optimal code are very close to the corresponding distributions under the fully optimal unconstrained tax code, as measured, for the five moments in Tables 2 and 4, by the absolute percentage difference between the moments in the economy under the quasi-optimal code and the moments in the economy under the

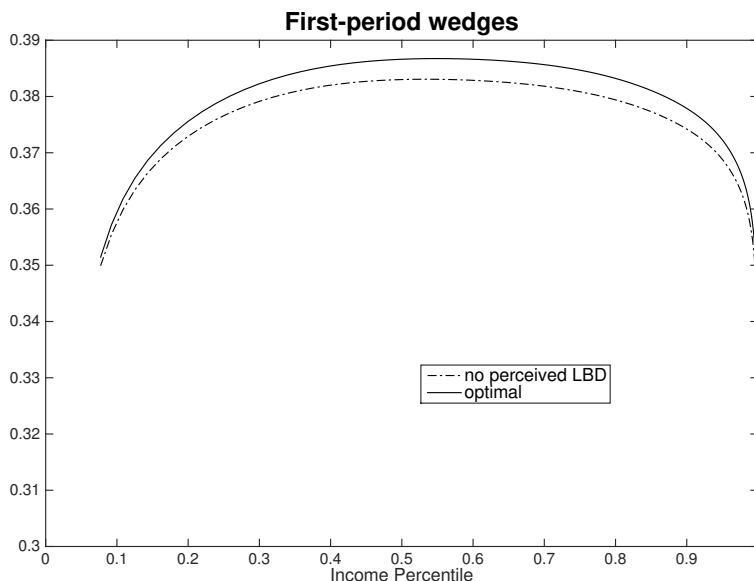


Figure 17: First-period wedges with and without LBD

fully optimal tax code.³⁹ That is, the quasi-optimal codes can be viewed as approximately the same as the fully optimal ones.

Figure 18 plots the period-1 wedges in the two economies under the quasi-optimal tax codes for each of the two economies. Period-1 wedges are higher with LBD across all income percentiles.

Importantly, that wedges are higher with LBD does not imply that taxes are also higher, as one can see from Figure 19. The figure plots the marginal tax rates in the two economies as a function of the (unconditional) income percentiles. While the period-1 quasi-optimal tax code is progressive with LBD, it is regressive without LBD. In period two, both codes are regressive, but the regressivity is higher without LBD. It is also worth noticing that the magnitude of the tax rates is significantly lower when accounting for LBD. The intuition for these results is the following. With LBD, the welfare implications of discouraging agents from working are higher, due to the benefits of inducing a more favorable productivity distribution in later periods. In reforming the tax code, the benefits from raising the marginal tax rates vis-a-vis the levels in the current US tax code are thus lower when accounting for LBD. As a result, taxes are lower with LBD. However, other-things-equal, distortions are higher with LBD. Furthermore, under LBD, the planner also gains on the margin from shifting the distribution of future productivity towards lower levels to economize on the welfare costs of the distortions in the second half of the agents' working life (recall the discussion in the previous sections). As a result, wedges are higher with LBD despite taxes being lower.

³⁹Specifically, these differences are, in the order the moments are stated in Tables 2 and 4, (2.91%, 3.62, 2.19%, 2.15%, 0.90%) for the case of LBD, and (1.75%, 2.13%, 1.77%, 0.27%, 0.48) for the case of no LBD.

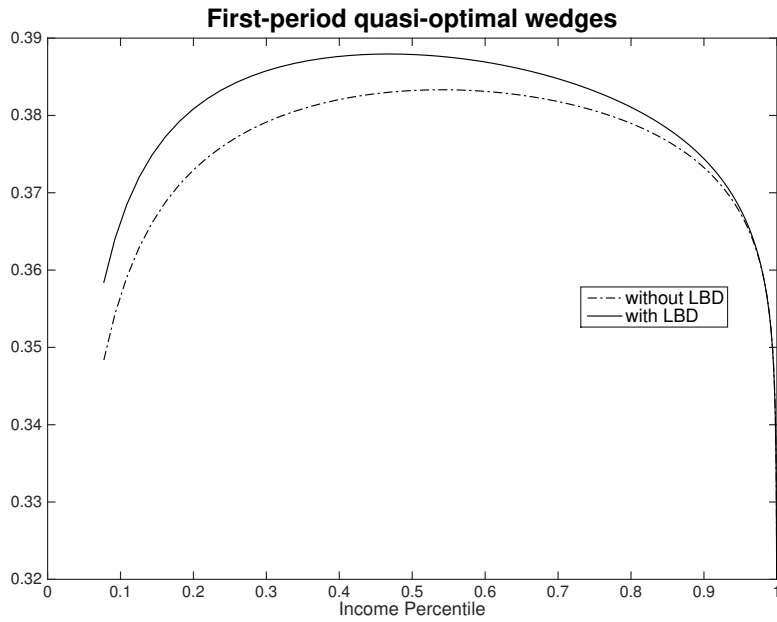


Figure 18: Quasi-optimal first-period wedges with and without LBD

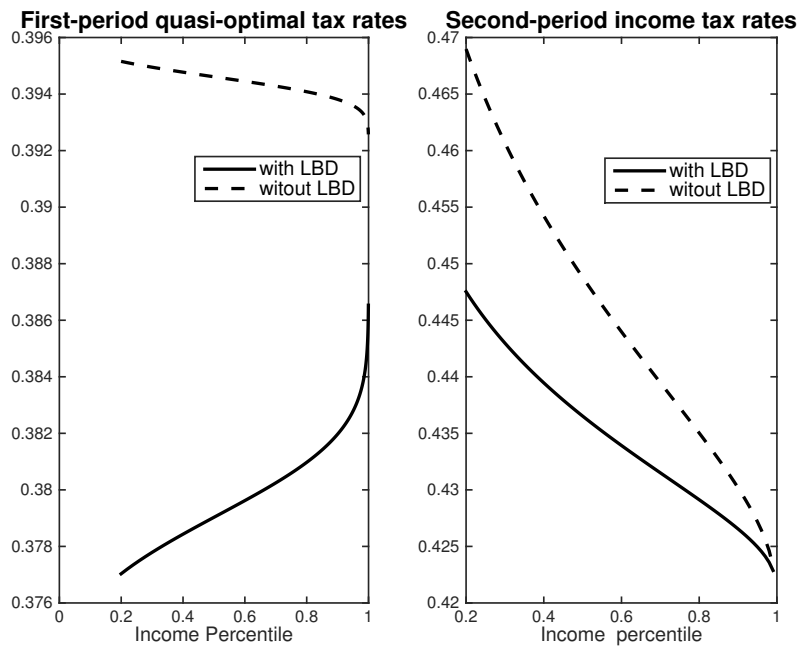


Figure 19: Quasi-optimal income tax rates with and without LBD

The above results also warn against possible difficulties in extrapolating properties of tax rates from optimal wedges in dynamic economies.⁴⁰ Wedges are higher with LBD than without it (except for low income percentiles), whereas the opposite is true for tax rates. Furthermore, while period-1 wedges tend to be regressive for high income percentiles irrespective of LBD, period-1 tax rates are progressive with LBD but not in its absence. These differences originate from the confounding effects on wedges of variations in labour supply and in consumption across time and income percentiles.

8 Conclusions

We study optimal taxation in a dynamic economy in which the workers' productivity evolves endogenously over time as the result of (on the job) learning-by-doing. We show that learning-by-doing contributes to higher wedges (i.e., to higher distortions relative to the first-best benchmark) but calls for lower marginal tax rates, compared to what predicted by models that assume productivity evolves exogenously over the lifecycle. Furthermore, learning-by-doing contributes to a higher progressivity of the optimal tax code for younger workers and to a lower regressivity for older ones. We also find that the benefits of reforming the existing US tax code are significant but lower than what predicted by ignoring learning-by-doing. Finally, we show that simple taxes that are invariant in past incomes but age-dependent are approximately optimal.

We believe these insights can guide the reform of existing tax codes and, more broadly, the restructuring of social insurance and redistributive programs in various economies. In future work, it would be interesting to extend the analysis to accommodate for hidden savings and/or limited commitment on the planner's side.

References

- [1] Albanesi, S. and C. Sleet (2006): "Dynamic Optimal Taxation with Private Information," *Review of Economic Studies*, 73(1): 1-30.
- [2] Arrow, K., (1962): "The Economic Implications of Learning by Doing", *Review of Economic Studies*, 29(3), 155-173.

⁴⁰We note here that the counterpart of (15) for the case of risk-averse agents is

$$W_1(\theta_1) = \frac{\tau_1(y_1(\theta_1)) + \delta \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1(\theta_1), y_2(\theta_1, \theta_2)) dF_2(\theta_2 | \theta_1, y_1(\theta_1))}{1 + LD_1^X(\theta_1)} + \frac{\delta \mathbb{E}^{\lambda[x] | \theta_1} \left[\frac{\partial}{\partial y_1} \mathcal{T}_2(y_1, (\theta_1) y_2(\bar{\theta})) \frac{v'(c_2(\bar{\theta}))}{v'(c_1(\theta_1))} \right]}{1 + LD_1^X(\theta_1)}.$$

Therefore, differences between the wedges and the tax rates can be attributed to the impact of higher earnings on (a) the expected surplus, taking into account the marginal utility of consumption in the second period, (b) the expected future tax bill for given earnings over time, and (c) the expected future tax bill for given probability distribution of period-2 productivity, adjusted to take into account the different marginal utilities of consumption over the two periods, for the given history of productivities. The first two effects are specific to LBD, whereas the third effect is present even in a dynamic economy without LBD.

- [3] Altugw, S., and R. Miller, (1998): “The Effect of Work Experience on Female Wages and Labour Supply”, *Review of Economic Studies*, 65(1), 45-85.
- [4] Becker, G. (1964): “Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education”, University of Chicago Press.
- [5] Benabou, R., (2002): “Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?”, *Econometrica*, 70(2), 481–517.
- [6] Bergemann, D. and A. Pavan (2015): “Introduction to JET Symposium on Dynamic Contracts and Mechanism Design,” *Journal of Economic Theory*, Volume 159, 679-701.
- [7] Best, M.C., and H.J. Kleven, (2013): “Optimal Income Taxation with Career Effects of Work Effort,” mimeo, LSE.
- [8] Blundell, R. and T. MaCurdy (1999): “Labor Supply: A Review of Alternative Approaches,” Chap. 27, pp1559–1695 in Ashenfelter, Orley, and Card, David (eds), *Handbook of Labor Economics*, Vol. 3. North Holland, Elsevier.
- [9] Conesa, J.C. and D. Krueger, (2006): “On the Optimal Progressivity of the Income Tax Code”, *Journal of Monetary Economics*, 53(7):1425–1450.
- [10] Diamond, P.A., (1998): “Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates,” *American Economic Review*, 88(1), 83-95.
- [11] Dustmann, C., and K. Meghir, (2005): “Wages, Experience, and Seniority,” *Review of Economic Studies*, 72, 77-108.
- [12] Farber, H. S., (1999), “Mobility and Stability: The Dynamics of Job Change in Labor Markets,” Chap. 37 in Ashenfelter, Orley, & Card, David (eds), *Handbook of Labor Economics*, Vol.3.
- [13] Farhi, E., and I. Werning, (2013): “Insurance and Taxation over the Life Cycle,” *Review of Economic Studies*, 80, 596-635.
- [14] Fershtman, D., and A. Pavan (2016): “Matching Auctions,” mimeo Northwestern University.
- [15] Garrett, D., and A. Pavan, (2016): “Dynamic Managerial Compensation: a Variational Approach,” *Journal of Economic Theory* 159, 775–818.
- [16] Golosov, M., A. Tsyvinski, and N. Werquin (2014): “A Variational Approach to the Analysis of Tax Systems,” NBER Working Paper No. 20780.
- [17] Golosov, M., M. Troshkin, and A. Tsyvinski, (2016): “Redistribution and Social Insurance,” *American Economic Review*, 106(2): 359-86.

- [18] Golosov, M., A. Tsyvinski, and I. Werning, (2006): “New Dynamic Public Finance: A User’s Guide”, NBER Macroeconomic Annual 2006, MIT Press.
- [19] Gorry, A. and E. Oberfield, (2012): “Optimal Taxation Over the Life Cycle”, *Review of Economic Dynamics*, 15(4):551–572.
- [20] Huggett, M., G. Ventura, and A. Yaron, (2011): “Sources of Lifetime Inequality”, *American Economic Review*, 101(7), 2923-2954.
- [21] Jacobson L., R. LaLonde, and D. Sullivan, (1993): “Earnings Losses of Displaced Workers,” *American Economic Review*, 83(4), 685-709.
- [22] Kapicka, M., (2006): “Optimal Income Taxation with Human Capital Accumulation and Limited Record Keeping”, *Review of Economic Dynamics*, 9, 612–639.
- [23] Kapicka, M., (2013): “Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First Order Approach,” *Review of Economic Studies*, 80(3), 1027-1054.
- [24] Kapicka, M., (2015a): “Optimal Mirrleesian Taxation in a Ben-Porath Economy,” *American Economic Journal: Macroeconomics*, Vol. 7, no. 2, 219-48.
- [25] Kapicka, M., (2015b): “Pareto Efficient Income Taxation with Learning by Doing”, mimeo, CERGE-EI.
- [26] Kapicka, M. and J. Neira (2016): “Optimal Taxation with Risky Human Capital,” mimeo, CERGE-EI.
- [27] Kindermann, F. and D. Krueger (2014), “High Marginal Tax Rates on the Top 1%? Lessons from a Life Cycle Model with Idiosyncratic Income Risk,” NBER Working Paper 20601.
- [28] Kocherlakota, N. R., (2010): *The New Dynamic Public Finance*. Princeton University Press, USA.
- [29] Krause, A., (2009): “Optimal Nonlinear Income Taxation with Learning-by-doing”, *Journal of Public Economics*, 93(9-10), 1098-1110.
- [30] Krueger, D. and A. Ludwig, (2013): “Optimal Progressive Labor Income Taxation and Education Subsidies When Education Decisions and Intergenerational Transfers Are Endogenous”, *American Economic Review*, 103(3):496–501.
- [31] Lucas, R.E., (1988): “On the Mechanics of Economic Development” *Journal of Monetary Economics*, 22(1), 33-42.
- [32] Mirrlees, J., (1971): “An Exploration in the Theory of Optimal Income Taxation,” *Review of Economic Studies*, 38, 175-208.

- [33] Pavan, A., S. Segal, and J. Toikka, (2014): “Dynamic Mechanism Design: A Myersonian Approach,” *Econometrica*, 82(2), 601-653.
- [34] Pavan, A., (2017): “Dynamic Mechanism Design: Robustness and Endogenous Types,” *Advances in Economics and Econometrics*, Proceeding of the 2015 World Congress of the Econometric Society, forthcoming.
- [35] Ramsey, F.P., (1927): “A Contribution to the Theory of Taxation”, *Economic Journal*, 37(145):47–61.
- [36] Rogerson, W.P., (1985): “Repeated Moral Hazard,” *Econometrica*, 53(1), 69-76.
- [37] Saez, E., (2001): “Using Elasticities to Derive Optimal Income Tax Rates,” *Review of Economic Studies*, 68, 205-229.
- [38] Stantcheva, S., (2015): “Learning and (or) Doing: Human Capital Investments and Optimal Taxation” NBER Working Paper 21381.
- [39] Stantcheva, S., (2016): “Optimal Taxation and Human Capital Policies over the Life Cycle,” *Journal of Political Economy*, forthcoming.
- [40] Topel, R., (1991): “Specific Capital, Mobility and Wages: Wages Rise with Seniority,” *Journal of Political Economy*, 99(t), 145-176.
- [41] Weinzierl, M., (2011): “The Surprising Power of Age-Dependent Taxes,” *Review of Economic Studies*, 78(4): 1490-1518.
- [42] Willis, R., (1986): “Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions,” in *Handbook of Labor Economics*, ed. by O. Ashenfelter and R. Layard, New York, Elsevier Science Publisher, Vol. 1, 525-602.

9 Appendix

9.1 Proofs

Lemma 1. *In each of the three economies examined in the paper (Rawlsian preferences for redistribution with risk neutral agents; Rawlsian preferences with risk-averse agents; Utilitarian preferences with risk averse agents), the first-best allocation rule $\chi^* = (y^*, c^*)$ satisfies the following optimality conditions (at all interior points with $\lambda[\chi^*]$ -probability one)*

$$1 + LD_1^{\chi^*}(\theta_1) = \frac{\psi_y(y_1^*(\theta_1), \theta_1)}{v'(c_1^*(\theta_1))},$$

$$1 = \frac{\psi_y(y_2^*(\theta), \theta_2)}{v'(c_2^*(\theta))},$$

$$v'(c_1^*(\theta_1)) = v'(c_2^*(\theta)),$$

where

$$LD_1^X(\theta_1) \equiv \delta \frac{\partial}{\partial y_1} \int \left\{ y_2(\theta) - c_2(\theta) + \frac{v(c_2(\theta)) - \psi(y_2(\theta), \theta_2)}{v'(c_1(\theta_1))} \right\} dF_2(\theta_2 | \theta_1, y_1(\theta_1)).$$

Proof of Lemma 1. Under full information, the optimal allocation rule $\chi^* = (y^*, c^*)$ maximizes expected tax revenue

$$R = \mathbb{E}^{\lambda[\chi]} \left[\sum_t \delta^{t-1} \left(y_t(\tilde{\theta}^t) - c_t(\tilde{\theta}^t) \right) \right]$$

subject to the redistribution constraint

$$(1-r)V_1(\theta_1) + r \int V_1(s) dF_1(s) \geq \kappa, \quad \text{all } \theta_1,$$

where $r = 0$ in case of Rawlsian preferences for redistribution, and $r = 1$ in case of Utilitarian preferences for redistribution, with

$$V_1(\theta_1) = \mathbb{E}^{\lambda[\chi]|\theta_1} \left[\sum_t \delta^{t-1} \left(v(c_t(\tilde{\theta}^t)) - \psi(y_t(\tilde{\theta}^t), \tilde{\theta}_t) \right) \right]$$

denoting type θ_1 's expected lifetime utility.

Letting $C \equiv v^{-1}$, and noting that the redistribution constraint binds at the optimum, we can rewrite the planner's first-best (FB) problem as

$$\begin{aligned} \max_{y_1(\cdot), y_2(\cdot), c_2(\cdot), V_1(\cdot)} \int & \left\{ y_1(\theta_1) - C \left(V_1(\theta_1) + \psi(y_1(\theta_1), \theta_1) - \delta \int [v(c_2(\theta)) - \psi(y_2(\theta), \theta_2)] dF_2(\theta_2 | \theta_1, y_1(\theta_1)) \right) \right. \\ & \left. + \delta \int [y_2(\theta) - c_2(\theta)] dF_2(\theta_2 | \theta_1, y_1(\theta_1)) \right\} dF_1(\theta_1) \end{aligned}$$

subject to

$$(1-r)V_1(\theta_1) + r \int V_1(s) dF_1(s) - \kappa = 0, \quad \text{all } \theta_1. \quad (27)$$

Consider first the case in which $r = 1$. Let π be the multiplier of the redistribution constraint (27), which is an integral constraint. At the optimum, the following necessary conditions must hold with $\lambda[\chi^*]$ -probability one:

$$1 - \frac{\psi_y(y_1^*(\theta_1), \theta_1) - \delta \frac{\partial}{\partial y_1} \int \{v(c_2(\theta)) - \psi(y_2(\theta), \theta_2)\} dF_2(\theta_2 | \theta_1, y_1(\theta_1))}{v'(c_1^*(\theta_1))} + \delta \frac{\partial}{\partial y_1} \int \{y_2(\theta) - c_2(\theta)\} dF_2(\theta_2 | \theta_1, y_1(\theta_1)) = 0$$

$$-\frac{\delta \psi_y(y_2^*(\theta), \theta_2)}{v'(c_1^*(\theta_1))} + \delta = 0$$

$$\frac{\delta v'(c_2^*(\theta))}{v'(c_1^*(\theta_1))} - \delta = 0$$

$$-\frac{1}{v'(c_1^*(\theta_1))} + \pi = 0 \quad (28)$$

along with

$$\int V_1(\theta_1) dF_1(\theta_1) - \kappa = 0 \quad (29)$$

and

$$c_1(\theta_1) = C \left(V_1(\theta_1) + \psi(y_1(\theta_1), \theta_1) - \delta \int [v(c_2(\theta)) - \psi(y_2(\theta), \theta_2)] dF_2(\theta_2 | \theta_1, y_1(\theta_1)) \right).$$

Rearranging, and using the definition of $LD_1^{FB;\chi}(\theta_1)$, we obtain the conditions in the lemma.

Next, consider the case in which $r = 0$. Let $\pi(\theta_1)$ be the multiplier associated with the constraint $V_1(\theta_1) - \kappa \geq 0$. The optimality conditions are then the same as in the case $r = 1$ except for conditions (28) and (29) which must be replaced with

$$-\frac{f_1(\theta_1)}{v'(c_1^*(\theta_1))} + \pi(\theta_1) = 0$$

and $V_1(\theta_1) = \kappa$. Q.E.D.

Proof of Proposition 1. Using (8) and (9), we have that wedges under the second-best allocations are given by

$$W_2(\theta) \equiv 1 - \psi_y(y_2(\theta), \theta_2) = -\frac{I_1^2(\theta, y_1(\theta_1))}{\gamma_1(\theta_1)} \psi_{\theta y}(y_2(\theta), \theta_2)$$

and

$$W_1(\theta_1) \equiv 1 - \frac{\psi_y(y_1(\theta_1), \theta_1)}{1 + LD_1^X(\theta_1)} = -\frac{1}{\gamma_1(\theta_1)} \frac{\psi_{\theta y}(y_1(\theta_1), \theta_1)}{1 + LD_1^X(\theta_1)} - \frac{\delta}{1 + LD_1^X(\theta_1)} \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi] | \theta_1, y_1(\theta_1)} \left[\frac{I_1^2(\tilde{\theta}, y_1(\theta_1))}{\gamma_1(\theta_1)} \psi_{\theta}(y_2(\tilde{\theta}), \tilde{\theta}_2) \right].$$

Furthermore, by definition, $1 - W_2(\theta) = \psi_y(y_2(\theta), \theta_2)$ and $1 - W_1(\theta_1) = \psi_y(y_1(\theta_1), \theta_1) / [1 + LD_1^X(\theta_1)]$.

We conclude that the *relative* wedges under the second-best allocations are given by

$$\hat{W}_2(\theta) \equiv \frac{W_2(\theta)}{1 - W_2(\theta)} = -\frac{I_1^2(\theta, y_1(\theta_1))}{\gamma_1(\theta_1)} \frac{\psi_{\theta y}(y_2(\theta), \theta_2)}{\psi_y(y_2(\theta), \theta_2)}$$

and

$$\hat{W}_1(\theta_1) \equiv \frac{W_1(\theta_1)}{1 - W_1(\theta_1)} = -\frac{1}{\gamma_1(\theta_1)} \frac{\psi_{\theta y}(y_1(\theta_1), \theta_1)}{\psi_y(y_1(\theta_1), \theta_1)} - \delta \frac{\frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi] | \theta_1, y_1(\theta_1)} \left[\frac{I_1^2(\tilde{\theta}, y_1(\theta_1))}{\gamma_1(\theta_1)} \psi_{\theta}(y_2(\tilde{\theta}), \tilde{\theta}_2) \right]}{\psi_y(y_1(\theta_1), \theta_1)}.$$

It is immediate to see that, in the absence of LBD, that is, when the process over θ is exogenous and equal to the one under χ ,

$$\hat{W}_t^{RRN}(\theta) = -\frac{I_1^t(\theta^t, y_{t-1}(\theta^{t-1}))}{\gamma_1(\theta_1)} \frac{\psi_{\theta y}(y_t(\theta^t), \theta_t)}{\psi_y(y_t(\theta^t), \theta_t)} \quad (30)$$

Using the fact that $I_1^1(\theta_1) = 1$, we then obtain the result in the proposition. Q.E.D.

Proof of Proposition 2. Let $y_1(\theta_1)$ be the unique solution to the following equation

$$\left[1 + \hat{W}_1^{RN}(\theta_1)\right]^{-1} \theta_1^{1+\phi} + \delta \zeta \bar{\varepsilon}(\phi) \theta_1^{\frac{(1+\phi)^2}{\phi}} \left[\frac{1 + \rho \hat{W}_1^{RRN}(\theta_1)}{1 + \hat{W}_1^{RRN}(\theta_1)} \right] \left[1 + \rho \hat{W}_1^{RN}(\theta_1)\right]^{-\frac{1+\phi}{\phi}} y_1^{\frac{\zeta(1+\phi)-\phi}{\phi}} - y_1^\phi = 0 \quad (31)$$

with $\bar{\varepsilon}(\phi) \equiv \mathbb{E} \left[\varepsilon_2^{\frac{1+\phi}{\phi}} \right]$. Observe that the assumption that $\zeta \leq \phi/(1+\phi)$ implies that the left-hand-side of (31) is strictly decreasing in y_1 . In turn, this implies that the unique solution $y_1(\theta_1)$ to (31) is nondecreasing in θ_1 whenever $\hat{W}_1^{RRN}(\theta_1)$ is nonincreasing.

Then, let

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

and

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

The proof proceeds in four steps. Step 1 shows that the period-1 wedge is given by

$$\hat{W}_1(\theta_1) = \hat{W}_1^{RRN}(\theta_1) \left\{ 1 + \left[\frac{1 + \rho \hat{W}_1^{RRN}(\theta_1)}{1 + \hat{W}_1^{RRN}(\theta_1)} \right]^{-1} \frac{\frac{\rho}{\phi} \theta_1^{\frac{1+\phi}{\phi}} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}{\frac{1}{\delta \zeta \bar{\varepsilon}(\phi)} \left[1 + \rho \hat{W}_1^{RRN}(\theta_1) \right]^{\frac{1}{\phi}} + \theta_1^{\frac{1+\phi}{\phi}} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}} \right\} \quad (32)$$

with $y_1(\theta_1)$ defined by the unique solution to equation (31). 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, $\hat{W}_1(\theta_1) - \hat{W}_2(\theta) > \hat{W}_1^{RRN}(\theta_1) - \hat{W}_2^{RRN}(\theta)$ 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 if $J(\theta_1) \geq 0$, 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 $\hat{W}_1^{RRN}(\theta_1)$ is nonincreasing, the policies (y_1, y_2) that solve the relax program satisfy the integral monotonicity conditions in (6), which implies that the solution to the relaxed program also solves the full program.

Step 1. Recall from (14) that the effects of LBD on the period-1 wedge are summarized by the term

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

Next use (9), 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

$$1 + LD_1^\chi(\theta_1) - \delta \rho \hat{W}_1^{RRN}(\theta_1) \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda|\chi||\theta_1, y_1(\theta_1)} \left[\psi(y_2(\tilde{\theta}), \tilde{\theta}_2) \right] = \psi_y(y_1(\theta_1), \theta_1) \left[1 + \hat{W}_1^{RRN}(\theta_1) \right] \quad (33)$$

where recall that

$$LD_1^X(\theta_1) = \delta \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda|X|\theta_1, y_1(\theta_1)} \left[y_2(\tilde{\theta}) - \psi(y_2(\tilde{\theta}), \tilde{\theta}_2) \right]. \quad (34)$$

Now, use (8) to see that $y_2(\theta)$ must solve the first-order-condition

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

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

$$y_2(\theta) = (1 + \phi) \left[1 + \rho \hat{W}_1^{RRN}(\theta_1) \right] \psi(y_2(\theta), \theta_2). \quad (35)$$

Replacing (35) into (34), we have that

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

and hence that

$$\begin{aligned} LD_1^X(\theta_1) - \delta \rho \hat{W}_1^{RRN}(\theta_1) \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda|X|\theta_1, y_1(\theta_1)} \left[\psi(y_2(\tilde{\theta}), \tilde{\theta}_2) \right] \\ = \delta \phi \left[1 + \rho \hat{W}_1^{RRN}(\theta_1) \right] \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda|X|\theta_1, y_1(\theta_1)} \left[\psi(y_2(\tilde{\theta}), \tilde{\theta}_2) \right]. \end{aligned} \quad (36)$$

Using (33), in turn we have that

$$\frac{1}{\psi_y(y_1(\theta_1), \theta_1)} = \frac{1 + \hat{W}_1^{RRN}(\theta_1)}{1 + \delta \phi \left[1 + \rho \hat{W}_1^{RRN}(\theta_1) \right] \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda|X|\theta_1, y_1(\theta_1)} \left[\psi(y_2(\tilde{\theta}), \tilde{\theta}_2) \right]} \quad (37)$$

Replacing (37) into (14), we have that

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

or, equivalently,

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

where we used the shortcut notation

$$\Lambda^X(\theta_1, y_1(\theta_1)) \equiv \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda|X|\theta_1, y_1(\theta_1)} \left[\psi(y_2(\tilde{\theta}), \tilde{\theta}_2) \right]. \quad (38)$$

Next, use (35) to observe that, when ψ is isoelastic,

$$y_2(\theta) = \theta_2^{\frac{1+\phi}{\phi}} \left[1 + \rho \hat{W}_1^{RRN}(\theta_1) \right]^{-\frac{1}{\phi}} \quad (39)$$

and hence

$$\psi(y_2(\theta), \theta_2) = \frac{1}{1 + \phi} \left[1 + \rho \hat{W}_1^{RRN}(\theta_1) \right]^{-\frac{1+\phi}{\phi}} \cdot \theta_2^{\frac{1+\phi}{\phi}}$$

It follows that

$$\Lambda^X(\theta_1, y_1(\theta_1)) = \frac{1}{1+\phi} \left[1 + \rho \hat{W}_1^{RRN}(\theta_1) \right]^{-\frac{1+\phi}{\phi}} \frac{\partial}{\partial y_1} \mathbb{E} \left[\theta_2^{\frac{1+\phi}{\phi}} | \theta_1, y_1(\theta_1) \right].$$

Using

$$\bar{\varepsilon}(\phi) = \mathbb{E} \left[\tilde{\varepsilon}_2^{\frac{1+\phi}{\phi}} \right]$$

we have that⁴¹

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

This implies that

$$\Lambda^X(\theta_1, y_1(\theta_1)) = \left\{ \frac{1}{1+\phi} \left[1 + \rho \hat{W}_1^{RRN}(\theta_1) \right]^{-\frac{1+\phi}{\phi}} \frac{\zeta(1+\phi)}{\phi} \bar{\varepsilon}(\phi) \theta_1^{\frac{1+\phi}{\phi}} \right\} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}. \quad (40)$$

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

$$\Omega_1(\theta_1) = \left[\frac{1 + \rho \hat{W}_1^{RRN}(\theta_1)}{1 + \hat{W}_1^{RRN}(\theta_1)} \right]^{-1} \frac{\frac{\rho}{\phi} \hat{W}_1^{RRN}(\theta_1) \theta_1^{\frac{1+\phi}{\phi}} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}{\frac{1}{\delta \zeta \bar{\varepsilon}(\phi)} \left[1 + \rho \hat{W}_1^{RRN}(\theta_1) \right]^{\frac{1}{\phi}} + \theta_1^{\frac{1+\phi}{\phi}} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}. \quad (41)$$

Replacing (41) into (11) for $t = 1$ permits us to establish the formula for $\hat{W}_1(\theta_1)$ in (32).

We conclude this step by showing that $y_1(\theta_1)$ is implicitly given by equation (31). This follows from combining (33), (36) and (38) with (40).

Step 2. Differentiating $\Omega_1(\theta_1)$ in (41), and simplifying the derivative using the fact that, at the

⁴¹Observe that, given any Lipschitz continuous function $J(\theta_2)$, and any kernel $F_2(\theta_2 | \theta_1, y_1)$, $\frac{\partial}{\partial y_1} \mathbb{E} \left[J(\tilde{\theta}_2) | \theta_1, y_1 \right] = \mathbb{E} \left[\left(-\frac{\frac{\partial F_2(\tilde{\theta}_2 | \theta_1, y_1)}{\partial y_1}}{f_2(\tilde{\theta}_2 | \theta_1, y_1)} \right) \frac{\partial J(\tilde{\theta}_2)}{\partial \theta_2} | \theta_1, y_1 \right]$.

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

$$\begin{aligned}
J(\theta_1) \equiv & \left[\frac{\epsilon^{\hat{W}_1^{RRN}}(\theta_1) \hat{W}_1^{RRN}(\theta_1)}{\theta_1} \frac{(1-\rho)}{\left[1 + \rho \hat{W}_1^{RRN}(\theta_1)\right]^2} \frac{\frac{\rho}{\phi} \hat{W}_1^{RRN}(\theta_1) \theta_1^{\frac{1+\phi}{\phi}} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}{\frac{1}{\delta\zeta\bar{\epsilon}(\phi)} \left[1 + \rho \hat{W}_1^{RRN}(\theta_1)\right]^{\frac{1}{\phi}} + \theta_1^{\frac{1+\phi}{\phi}} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}} \right] \times \\
& \left[\frac{\left[\frac{1+\rho \hat{W}_1^{RRN}(\theta_1)}{1+\hat{W}_1^{RRN}(\theta_1)} \right]^{-1} \frac{\hat{W}_1^{RRN}(\theta_1)}{\theta_1} \frac{\rho}{\phi} \theta_1^{\frac{1+\phi}{\phi}} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}{\left[\frac{1}{\delta\zeta\bar{\epsilon}(\phi)} \left[1 + \rho \hat{W}_1^{RRN}(\theta_1)\right]^{\frac{1}{\phi}} + \theta_1^{\frac{1+\phi}{\phi}} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}} \right]^2} \right]^{-1} + \\
& \epsilon^{\hat{W}_1^{RRN}}(\theta_1) \left\{ \frac{1}{\delta\zeta\bar{\epsilon}(\phi)} \left[1 + \rho \hat{W}_1^{RRN}(\theta_1)\right]^{\frac{1}{\phi}} \left[1 - \frac{\rho \hat{W}_1^{RRN}(\theta_1)}{\phi(1 + \rho \hat{W}_1^{RRN}(\theta_1))}\right] + \theta_1^{\frac{1+\phi}{\phi}} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}} \right\} + \\
& \left[\frac{1+\phi}{\phi} + \frac{\zeta(1+\phi)-\phi}{\phi} e^{y_1(\theta_1)} \right] \frac{1}{\delta\zeta\bar{\epsilon}(\phi)} \left[1 + \rho \hat{W}_1^{RRN}(\theta_1)\right]^{\frac{1}{\phi}}
\end{aligned}$$

is non-negative.

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

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

and equation (31) reduces to

$$P_1(\theta_1) + P_2(\theta_1) y_1^{\frac{\zeta(1+\phi)-\phi}{\phi}} - y_1^\phi = 0 \quad (42)$$

where

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

and

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

Furthermore, in this case,

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

Now use (42) to obtain that

$$\frac{dy_1(\theta_1)}{d\theta_1} = - \frac{\frac{dP_1(\theta_1)}{d\theta_1} + \frac{dP_2(\theta_1)}{d\theta_1} y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}{\frac{\zeta(1+\phi)-\phi}{\phi} P_2(\theta_1) y_1(\theta_1)^{\frac{\zeta(1+\phi)-2\phi}{\phi}} - \phi y_1(\theta_1)^{\phi-1}}. \quad (43)$$

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

$$\frac{dy_1(\theta_1)}{d\theta_1} = - \frac{\frac{dP_1(\theta_1)}{d\theta_1} y_1(\theta_1) + \frac{dP_2(\theta_1)}{d\theta_1} y_1(\theta_1)^{\frac{\zeta(1+\phi)}{\phi}}}{\frac{\zeta(1+\phi)-\phi}{\phi} P_2(\theta_1) y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}} - \phi y_1(\theta_1)^\phi}. \quad (44)$$

Replacing

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

into (44), 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

$$\frac{dy_1(\theta_1)}{d\theta_1} = -\frac{\frac{dP_1(\theta_1)}{d\theta_1}y_1(\theta_1) + \frac{dP_2(\theta_1)}{d\theta_1}y_1(\theta_1)^{\frac{\zeta(1+\phi)}{\phi}}}{\frac{(\zeta-\phi)(1+\phi)}{\phi}P_2(\theta_1)y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}} - \phi P_1(\theta_1)}. \quad (45)$$

Now note that

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

Replacing these functions into (45), and letting $n(\theta_1) \equiv \delta\zeta\bar{\varepsilon}(\phi)\theta_1^{\frac{(1+\phi)}{\phi}} \left[1 + \rho\frac{1+\phi}{\lambda} \right]^{-\frac{1}{\phi}}$, we then have that

$$\epsilon^{y_1}(\theta_1) = -\frac{1+\phi + \frac{(1+\phi)^2}{\phi}n(\theta_1)y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}{\frac{(\zeta-\phi)(1+\phi)}{\phi}n(\theta_1)y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}} - \phi}. \quad (46)$$

It follows that

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

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

$$1 + \left(\frac{\phi}{1+\phi} - \zeta \right) \left[\frac{1+\phi + \frac{(1+\phi)^2}{\phi}n(\theta_1)y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}}}{\frac{(\zeta-\phi)(1+\phi)}{\phi}n(\theta_1)y_1(\theta_1)^{\frac{\zeta(1+\phi)-\phi}{\phi}} - \phi} \right] > 0. \quad (47)$$

Now fix θ_1 and observe that the left-hand side of (47) is nondecreasing in $y_1(\theta_1)$. A sufficient condition for $J(\theta_1) > 0$ is thus that the inequality in (47) holds when $y_1(\theta_1) = 0$. It is easy to see that, when $y_1(\theta_1) = 0$, the left-hand side of (47) reduces to

$$\zeta \frac{1+\phi}{\phi}$$

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 (39) to observe that, for any θ_1 , $y_2(\theta_1, \theta_2)$ is nondecreasing in θ_2 . Next note that (6) is equivalent to

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

with $\hat{\theta}_1 < \theta_1$. Define now 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 noting 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

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

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 (39), we have that

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

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) $\left[1 + \rho \hat{W}_1^{RRN}(\theta_1)\right]^{-\frac{1}{\phi}} y_1(\theta_1)^{\frac{\zeta}{\phi}}$ 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 (31), is nondecreasing whenever $\hat{W}_1^{RRN}(\theta_1)$ is nonincreasing. Q.E.D.

Proof of relationship between tax codes and wedges in (15), (16), (17) and (18).

Faced with the code \mathcal{T} , the problem of a worker with period-1 productivity θ_1 consists in choosing a period-1 income y_1 and a contingent period-2 income schedule $\bar{y}_2(\theta_2; y_1)$ so as to maximize⁴²

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

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

$$\begin{aligned} 1 - \tau_1(y_1) &= \Gamma(y_1, \theta_1) \\ &\equiv \psi_y(y_1, \theta_1) + \delta \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1, \bar{y}_2(\theta_2; y_1)) dF_2(\theta_2|\theta_1, y_1) \\ &\quad - \delta \frac{\partial}{\partial y_1} \int [\bar{y}_2(\theta_2; y_1) - \psi(\bar{y}_2(\theta_2; y_1), \theta_2)] dF_2(\theta_2|\theta_1, y_1) \end{aligned} \quad (48)$$

and

$$1 - \tau_2(y_1, \bar{y}_2(\theta_2; y_1)) = \psi_y(\bar{y}_2(\theta_2; y_1), \theta_2). \quad (49)$$

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

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

in (48) 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 $y_2(\theta) = \bar{y}_2(\theta_2; y_1(\theta_1))$, where the dependence of such policies on the tax code \mathcal{T} is dropped to ease the notation.⁴³

Using the above FOCs, we then have that

$$W_1(\theta_1) \equiv 1 - \frac{\psi_y(y_1(\theta_1), \theta_1)}{1 + LD_1^X(\theta_1)} = \frac{\tau_1(y_1(\theta_1)) + \delta \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1(\theta_1), y_2(\theta)) dF_2(\theta_2|\theta_1, y_1(\theta_1))}{1 + \delta \frac{\partial}{\partial y_1} \int [y_2(\theta) - \psi(y_2(\theta), \theta_2)] dF_2(\theta_2|\theta_1, y_1(\theta_1))} \quad (50)$$

and

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

⁴²That 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 $\bar{y}_2(\theta_2; y_1)$.

⁴³When we find it useful to highlight such dependence, we will do it by denoting the optimal income policies by $y_1(\theta_1; \mathcal{T})$, and $y_2(\theta; \mathcal{T}) = \bar{y}_2(\theta_2; y_1(\theta_1; \mathcal{T}), \mathcal{T})$.

where we used the fact that

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

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 (17) and (18). Q.E.D.

Proof of Proposition 3. 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), in the amount of $d\tau_1 dy_1$. Assuming dy_1 is small, this mechanical effect is equal to

$$d\tau_1 dy_1 [1 - H_Y(y_1)] \quad (51)$$

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 $\frac{\partial \hat{y}_1(1 - \tau_1(y'_1), \theta_1(y'_1))}{\partial(1 - \tau_1)} d\tau_1$, where $\theta_1(y'_1)$ is the period-1 productivity of all agents whose period-1 income under the original tax code \mathcal{T} is y'_1 .⁴⁴ Using again the fact that dy_1 is small, we have that the total reduction in first-period tax revenues from such individuals is equal to

$$- d\tau_1 \frac{y_1}{1 - \tau_1(y_1)} \hat{E}_1(y_1) \tau_1(y_1) \hat{h}_Y(y_1) dy_1, \quad (52)$$

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 the non-linear period-1 tax schedule \mathcal{T}_1 is replaced by the linear tax schedule $\hat{\mathcal{T}}_1$ with constant marginal tax rate equal to $\tau_1(y_1)$, and (b)

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

Third, a change in the period-1 marginal tax rate, by triggering a change in the period-1 earnings of those individuals generating income $y'_1 \in [y_1, y_1 + dy_1)$, also induces a variation in the period-2 tax revenue. Such 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

$$- d\tau \frac{y_1}{1 - \tau_1(y_1)} \hat{E}_1(y_1) \left[\delta \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1, y_2) dH_O(y_2 | y_1) \right] \hat{h}_Y(y_1) dy_1 \quad (53)$$

⁴⁴The function $\theta_1(y'_1)$ is implicitly defined by the FOC $1 - \tau_1(y'_1) = \Gamma(y'_1, \theta_1)$.

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 (53) combines the direct effect of a change in period-2 taxes for given distribution of period-2 incomes with the indirect effect due to a variation in the distribution of period-2 income for 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 $\theta_1(y_1)$.⁴⁵

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, Condition (19) in the proposition holds. Q.E.D.

Proof of Proposition 4. Consider the reform of the tax schedule described in the main text. Recall that such 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 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

$$\delta \left(\frac{\partial \tilde{H}_O(y_2|y'_1)}{\partial y_1} - \frac{\partial H_O(y_2|y_1)}{\partial y_1} \right) d\tau_2 dy_2$$

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 individual with period-1 income $y'_1 \in [y_1, y_1 + dy_1)$.

Such 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 $-\frac{\partial \hat{y}_2(1-\tau_2(y'_1, y'_2), \theta_2(y'_1, y'_2))}{\partial(1-\tau_2)} d\tau_2$, where $\hat{y}_2(1-\tau_2, \theta_2)$ 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 , such 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

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

⁴⁵Formally, 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.

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

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

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

$$- \delta \left[d\tau_2 \frac{y_2}{1 - \tau_2(y_1, y_2)} \hat{E}_2(y_1, y_2) \tau_2(y_1, y_2) dy_2 dy_1 \hat{h}_O(y_2|y_1) h_Y(y_1) \right]. \quad (54)$$

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) in the 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) equal to $y'_1 \in [y_1, y_1 + dy_1)$ expects a net increase in his lifetime taxes equal to

$$\begin{aligned} & \delta [1 - F_2(\theta_2(y'_1, y_2) | \theta_1(y'_1), y'_1)] d\tau_2 dy_2 + \delta d\tau_2 dy_2 \int_{y_1}^{y'_1} \left(\frac{\partial \tilde{H}_O(y_2|s)}{\partial y_1} - \frac{\partial H_O(y_2|s)}{\partial y_1} \right) ds \\ & - \delta [1 - H_O(y_2|y'_1)] d\tau_2 dy_2. \end{aligned} \quad (55)$$

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.⁴⁶ 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 in the amount of $dy_1 d\tau_2 dy_2 \delta \left[\frac{\partial \tilde{H}_O(y_2|y_1)}{\partial y_1} - \frac{\partial H_O(y_2|y_1)}{\partial y_1} \right]$. 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

$$dy_1 d\tau_2 dy_2 \delta \left[\frac{\partial \tilde{H}_O(y_2|y_1)}{\partial y_1} - \frac{\partial H_O(y_2|y_1)}{\partial y_1} \right] [1 - H_Y(y_1)], \quad (56)$$

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

Under any optimal tax code, the sum of the above behavioral and mechanical effects on the NPV of intertemporal tax revenues must be equal to zero.⁴⁷ For this to be the case, it must be that, for

⁴⁶This is a direct consequence of the fact that the derivative of the term in (55) with respect to y'_1 is zero. To see the latter, 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) $\partial \tilde{H}_O(y_2|y'_1)/\partial y_1$ 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 (55) must be computed holding θ_1 fixed at $\theta_1(y'_1)$.

⁴⁷Note that, for $dy_1 \rightarrow 0$, for any $y'_1 \in [y_1, y_1 + dy_1)$, $\delta d\tau_2 dy_2 \int_{y_1}^{y'_1} \left[\frac{\partial \tilde{H}_O(y_2|s)}{\partial y_1} - \frac{\partial H_O(y_2|s)}{\partial y_1} \right] ds + \delta d\tau_2 dy_2 [1 - H_O(y_2|y'_1)] - S(y'_1) \rightarrow 0$ at the optimum.

any (y_1, y_2) in the support of the induced income distribution, Condition (20) in the proposition holds. Q.E.D.

Proof of Propositions 5 and 6. The two propositions are proved together given that many of the steps are in common. The proof is in two steps. Step 1 characterizes the first-order conditions for the second-best allocations. Step 2 uses the first-order conditions to show that wedges under the second-best allocations satisfy the properties in the propositions.

Step 1. The government's problem is the same as in the proof of Lemma 1, augmented by the local IC conditions, as summarized by the envelope formulas

$$\frac{\partial V_2(\theta)}{\partial \theta_2} = -\psi_\theta(y_2(\theta), \theta_2), \text{ all } \theta_1 \in \Theta_1, \text{ almost all } \theta_2 \in \text{Supp}[F_2(\cdot | \theta_1, y_1(\theta_1))],$$

and

$$\frac{\partial V_1(\theta_1)}{\partial \theta_1} = -\psi_\theta(y_1(\theta_1), \theta_1) - \delta \mathbb{E}^{\lambda[x] | \theta_1} \left[I_1^2(\tilde{\theta}, y_1(\theta_1)) \psi_\theta(y_2(\tilde{\theta}), \tilde{\theta}_2) \right], \text{ almost all } \theta_1 \in \Theta_1. \quad (57)$$

Let $r \in \{0, 1\}$, with $r = 0$ in case of Rawlsian preferences for redistribution, and $r = 1$ in case of Utilitarian preferences for redistribution. Equipped with this notation, we have that the planner's problem can be conveniently reformulated as follows:

$$\begin{aligned} \max_{y_1(\cdot), V_1(\cdot), \Pi_2(\cdot), Z_2(\cdot)} \int \{ & y_1(\theta_1) - C(V_1(\theta_1) + \psi(y_1(\theta_1), \theta_1) - \delta \Pi_2(\theta_1)) \\ & + \delta Q_2(\theta_1, y_1(\theta_1), \Pi_2(\theta_1), Z_2(\theta_1)) \} dF_1(\theta_1) \end{aligned}$$

subject to

$$(1 - r)V_1(\underline{\theta}_1) + r \int V_1(s) dF_1(s) - \kappa = 0, \quad (58)$$

and

$$\frac{\partial V_1(\theta_1)}{\partial \theta_1} = -\psi_\theta(y_1(\theta_1), \theta_1) + \delta Z_2(\theta_1), \quad (59)$$

where

$$Q_2(\theta_1, y_1(\theta_1), \Pi_2(\theta_1), Z_2(\theta_1)) \equiv \max_{y_2(\theta_1, \cdot), V_2(\theta_1, \cdot)} \int \{ y_2(\theta) - C(V_2(\theta) + \psi(y_2(\theta), \theta_2)) \} dF_2(\theta_2 | \theta_1, y_1(\theta_1))$$

subject to

$$\Pi_2(\theta_1) = \int V_2(\theta) dF_2(\theta_2 | \theta_1, y_1(\theta_1)), \quad (60)$$

$$Z_2(\theta_1) = - \int I_1^2(\theta, y_1(\theta_1)) \psi_\theta(y_2(\theta), \theta_2) f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2, \quad (61)$$

and

$$\frac{\partial V_2(\theta)}{\partial \theta_2} = -\psi_\theta(y_2(\theta), \theta_2). \quad (62)$$

Note that, with asymmetric information, in case of Rawlsian preferences for redistribution ($r = 0$), the redistribution constraint $V(\theta_1) \geq \kappa$ all θ_1 is binding only for $\theta_1 = \underline{\theta}_1$, which explains the constraint in (58).⁴⁸

⁴⁸To see this, it suffices to use (57) to see that incentive compatibility requires that $V_1(\theta_1)$ be nondecreasing.

The above optimization problem thus consists of two interdependent optimal control problems, one for each period. We start with the case of non-moving supports. We solve first the problem for Utilitarian preferences for redistribution ($r = 1$) and then then the one for Rawlsian preferences for redistribution ($r = 0$). We then analyze the case of moving supports.

Non-moving supports.

Using the property that

$$\begin{aligned} & - \int I_1^2(\theta, y_1(\theta_1)) \psi_\theta(y_2(\theta), \theta_2) f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2 \\ & = \int I_1^2(\theta, y_1(\theta_1)) \frac{\partial V_2(\theta)}{\partial \theta_2} f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2 = \int V_2(\theta) \frac{\partial}{\partial \theta_1} f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2, \end{aligned} \quad (63)$$

we have that the integral constraint (61) can be conveniently rewritten as

$$Z_2(\theta_1) = \int V_2(\theta) \frac{\partial}{\partial \theta_1} f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2. \quad (64)$$

Case $r = 1$ (i.e., Utilitarian preferences for redistribution).

As usual, we proceed backwards, by solving first the period-2 problem defining the value function $Q_2(\theta_1, y_1(\theta_1), \Pi_2(\theta_1), Z_2(\theta_1))$. This is a standard optimal control problem with two integral constraints, (60) and (64). The control variable is $y_2(\theta_1, \cdot)$, the state variable is $V_2(\theta_1, \cdot)$, and the law of motion for the state variable is given by (62).

Let $\pi_2(\theta_1)$ and $\xi_2(\theta_1)$ be the multipliers of the two integral constraints (60) and (64) and $\mu_2(\theta)$ the costate variable for the law of motion of $V_2(\theta_1, \theta_2)$.

Along with (60), (62), and (64), the following necessary optimality conditions must hold for almost all $\theta_2 \in \text{Supp}[F_2(\cdot | \theta_1, y_1(\theta_1))]$:

$$1 - \frac{\psi_y(y_2(\theta), \theta_2)}{v'(c_2(\theta))} - \mu_2(\theta) \frac{\psi_{\theta_y}(y_2(\theta), \theta_2)}{f_2(\theta_2 | \theta_2, y_1(\theta_1))} = 0, \quad (65)$$

$$\frac{\partial \mu_2(\theta)}{\partial \theta_2} = f_2(\theta_2 | \theta_1, y_1(\theta_1)) \cdot \left\{ \frac{1}{v'(c_2(\theta))} + \pi_2(\theta_1) + \xi_2(\theta_1) \frac{\partial f_2(\theta_2 | \theta_2, y_1(\theta_1)) / \partial \theta_1}{f_2(\theta_2 | \theta_2, y_1(\theta_1))} \right\}, \quad (66)$$

along with the boundary conditions

$$\mu_2(\theta_1, \underline{\theta}_2) = 0, \quad (67)$$

$$\mu_2(\theta_1, \bar{\theta}_2) = 0, \quad (68)$$

where

$$c_2(\theta) = C(V_2(\theta) + \psi(y_2(\theta), \theta_2)).$$

Next, consider the choice of the period-1 policies. Let $\mu_1(\theta_1)$ be the costate variable associated with the constraint (59) and π_1 the multiplier associated with the redistribution constraint

$$\int V_1(\theta_1) dF_1(\theta_1) = \kappa. \quad (69)$$

In addition to (59) and (69), the following optimality conditions must hold:

$$1 - \frac{\psi_y(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))} + \delta \int \{y_2(\theta) - c_2(\theta) - \pi_2(\theta_1)V_2(\theta)\} \frac{\partial}{\partial y_1} f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2 \quad (70)$$

$$+ \delta \xi_2(\theta_1) \frac{\partial}{\partial y_1} \int I_1^2(\theta, y_1(\theta_1)) \psi_\theta(y_2(\theta), \theta_2) f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2 - \mu_1(\theta_1) \frac{\psi_{\theta y}(y_1(\theta_1), \theta_1)}{f_1(\theta_1)} = 0,$$

$$\frac{\partial \mu_1(\theta_1)}{\partial \theta_1} = f_1(\theta_1) \cdot \left\{ \frac{1}{v'(c_1(\theta_1))} + \pi_1 \right\}, \quad (71)$$

$$\frac{1}{v'(c_1(\theta_1))} + \pi_2(\theta_1) = 0, \quad (72)$$

$$\mu_1(\theta_1) + \xi_2(\theta_1) f_1(\theta_1) = 0, \quad (73)$$

along with the boundary conditions

$$\mu_1(\underline{\theta}_1) = 0, \quad (74)$$

$$\mu_1(\bar{\theta}_1) = 0, \quad (75)$$

where

$$c_1(\theta_1) = C(V_1(\theta_1) + \psi(y_1(\theta_1), \theta_1) - \delta \Pi_2(\theta_1)).$$

Note that, in computing the FOCs with respect to $\Pi_2(\theta_1)$, $Z_2(\theta_1)$ and $y_1(\theta_1)$, we have used the following properties: $\frac{\partial Q_2}{\partial \Pi_2} = \pi_2(\theta_1)$, $\frac{\partial Q_2}{\partial Z_2} = \xi_2(\theta_1)$, and (63) along with the fact that

$$\begin{aligned} \frac{\partial Q_2}{\partial y_1} &= \int \{y_2(\theta) - c_2(\theta)\} \frac{\partial}{\partial y_1} f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2 - \pi_2(\theta_1) \int V_2(\theta) \frac{\partial}{\partial y_1} f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2 \\ &\quad + \xi_2(\theta_1) \frac{\partial}{\partial y_1} \int I_1^2(\theta, y_1(\theta_1)) \psi_\theta(y_2(\theta), \theta_2) f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2. \end{aligned}$$

Now use (71) along with the boundary conditions (74) and (75) to obtain that

$$0 = \int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{\partial \mu_1(\theta_1)}{\partial \theta_1} d\theta_1 = \int_{\underline{\theta}_1}^{\bar{\theta}_1} f_1(\theta_1) \cdot \left\{ \frac{1}{v'(c_1(\theta_1))} + \pi_1 \right\} d\theta_1,$$

which implies that

$$\pi_1 = - \int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(\theta_1))} dF_1(\theta_1). \quad (76)$$

Replacing the value of π_1 into (71) we have that

$$\mu_1(\theta_1) = - \int_{\theta_1}^{\bar{\theta}_1} f_1(s) \cdot \left\{ \frac{1}{v'(c_1(s))} - \int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} dF_1(s) \right\} ds,$$

from which we obtain that

$$\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} = \frac{1 - F_1(\theta_1)}{f_1(\theta_1)} \left[\int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} \frac{dF_1(s)}{1 - F_1(\theta_1)} - \int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} dF_1(s) \right]. \quad (77)$$

Next, use (72) and (73) to rewrite the FOC for $y_1(\theta_1)$ as follows

$$1 - \frac{\psi_y(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))} + \delta \int \left\{ y_2(\theta) - c_2(\theta) - \frac{V_2(\theta)}{v'(c_1(\theta_1))} \right\} \frac{\partial}{\partial y_1} f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2$$

$$+ \left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \delta \frac{\partial}{\partial y_1} \int I_1^2(\theta, y_1(\theta_1)) \psi_\theta(y_2(\theta), \theta_2) f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2 + \left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \psi_{\theta y}(y_1(\theta_1), \theta_1) = 0, \quad (78)$$

with $-\mu_1(\theta_1)/f_1(\theta_1)$ given by (77).

Note that, together, Conditions (66), (72), and (73) imply that

$$\frac{\partial \mu_2(\theta)}{\partial \theta_2} = f_2(\theta_2 | \theta_1, y_1(\theta_1)) \cdot \left\{ \frac{1}{v'(c_2(\theta))} - \frac{1}{v'(c_1(\theta_1))} - \frac{\mu_1(\theta_1)}{f_1(\theta_1)} \frac{\partial f_2(\theta_2 | \theta_1, y_1(\theta_1))/\partial \theta_1}{f_2(\theta_2 | \theta_1, y_1(\theta_1))} \right\}. \quad (79)$$

Combining (79) with the boundary conditions (67) and (68), we have that

$$0 = \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{\partial \mu_2(\theta_1, s)}{\partial \theta_2} ds = \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{1}{v'(c_2(\theta_1, s))} dF_2(s | \theta_1, y_1(\theta_1)) - \frac{1}{v'(c_1(\theta_1))}, \quad (80)$$

where we used the fact that

$$\int_{\underline{\theta}_2}^{\bar{\theta}_2} \partial f_2(\theta_2 | \theta_1, y_1(\theta_1)) / \partial \theta_1 = 0.$$

Next note that (80) yields the familiar Rogerson inverse-Euler condition

$$\frac{1}{v'(c_1(\theta_1))} = \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{1}{v'(c_2(\theta_1, s))} dF_2(s | \theta_1, y_1(\theta_1)). \quad (81)$$

Combining (79) with (81), we have that

$$\mu_2(\theta_1, \theta_2) = - \int_{\underline{\theta}_2}^{\bar{\theta}_2} f_2(s | \theta_1, y_1(\theta_1)) \left\{ \frac{1}{v'(c_2(\theta_1, s))} - \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{1}{v'(c_2(\theta_1, s))} f_2(s | \theta_1, y_1(\theta_1)) ds \right\} ds$$

$$+ \frac{\mu_1(\theta_1)}{f_1(\theta_1)} \int_{\underline{\theta}_2}^{\bar{\theta}_2} [\partial f_2(s | \theta_1, y_1(\theta_1)) / \partial \theta_1] ds.$$

Using the fact that

$$\frac{\int_{\underline{\theta}_2}^{\bar{\theta}_2} [\partial f_2(s | \theta_2, y_1(\theta_1)) / \partial \theta_1] ds}{f_2(\theta_2 | \theta_1, y_1(\theta_1))} = \frac{\partial [1 - F_2(\theta_2 | \theta_1, y_1(\theta_1))] / \partial \theta_1}{f_2(\theta_2 | \theta_1, y_1(\theta_1))} = I_1^2((\theta_1, \theta_2), y_1(\theta_1))$$

we have that

$$\frac{-\mu_2(\theta_1, \theta_2)}{f_2(\theta_2 | \theta_1, y_1(\theta_1))} = \frac{1 - F_2(\theta_2 | \theta_1, y_1(\theta_1))}{f_2(\theta_2 | \theta_1, y_1(\theta_1))} \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{1}{v'(c_2(\theta_1, s))} \frac{dF_2(s | \theta_1, y_1(\theta_1))}{1 - F_2(\theta_2 | \theta_1, y_1(\theta_1))}$$

$$- \frac{1 - F_2(\theta_2 | \theta_1, y_1(\theta_1))}{f_2(\theta_2 | \theta_1, y_1(\theta_1))} \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{1}{v'(c_2(\theta_1, s))} f_2(s | \theta_1, y_1(\theta_1)) ds - \frac{\mu_1(\theta_1)}{f_1(\theta_1)} I_1^2((\theta_1, \theta_2), y_1(\theta_1)). \quad (82)$$

Case $r = 0$ (i.e., Rawlsian preferences for redistribution).

Here, the redistribution constraint (69) is absent and replaced by the constraint $V_1(\underline{\theta}_1) = \kappa$, in the optimal control problem yielding the period-1 policies. As a result, the analysis parallels the one

for the case $r = 1$ with the following two differences. First, the multiplier π_1 is now irrelevant and, thereby, set to zero. Second, the only relevant first-period transversality condition is now $\mu_1(\bar{\theta}_1) = 0$.

Setting $\pi_1 = 0$ in (71) and using the transversality condition $\mu_1(\bar{\theta}_1) = 0$ we thus obtain that

$$0 = \mu_1(\underline{\theta}_1) + \int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{\partial \mu_1(\theta_1)}{\partial \theta_1} d\theta_1 = \mu_1(\underline{\theta}_1) + \int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(\theta_1))} dF_1(\theta_1),$$

which implies that

$$\mu_1(\underline{\theta}_1) = - \int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(\theta_1))} dF_1(\theta_1).$$

Using this expression, together with (71), we have that

$$\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} = \frac{1 - F_1(\theta_1)}{f_1(\theta_1)} \left[\int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} \frac{dF_1(s)}{1 - F_1(\theta_1)} \right], \quad (83)$$

which replaces Condition (77) in the characterization of the necessary optimality conditions for the optimal policies. All other conditions are the same as in the Utilitarian case ($r = 1$).

Moving supports.

To accommodate for moving supports, let $\mathbf{1}_{\underline{\theta}_2}$ be the indicator function taking value 1 if, and only if, $\underline{\theta}_2$ depends on $(\theta_1, y_1(\theta_1))$. Given that $\underline{\theta}_2 = z_2(\theta_1, y_1, \varepsilon_2) = \theta_1^\rho y_1^\zeta \varepsilon_2$, this is the case if, and only if, $\underline{\varepsilon} > 0$. Similarly, let $\mathbf{1}_{\bar{\theta}_t}$ be the indicator function taking value 1 if, and only if, $\bar{\theta}_2$ depends on $(\theta_1, y_1(\theta_1))$. In our setup, this is the case if, and only if, $\bar{\varepsilon} < +\infty$. Given that $\bar{\theta}_2 = z_2(\theta_1, y_1, \varepsilon_2) = \theta_1^\rho y_1^\zeta \varepsilon_2$, observe that the functions that map $(\theta_1, y_1(\theta_1))$ into $\underline{\theta}_2$ and $\bar{\theta}_2$ respectively are differentiable.

With this notation, and using the fact that $\bar{\theta}_2 = \theta_1^\rho y_1^\zeta \bar{\varepsilon}$, $\underline{\theta}_2 = \theta_1^\rho y_1^\zeta \underline{\varepsilon}$ and $I_1^2(\theta_2, \theta_1, y_1(\theta_1)) = \rho \frac{\theta_2}{\theta_1}$, we have that the integral constraint (61) can now be rewritten as

$$\begin{aligned} Z_2(\theta_1) &= - \int I_1^2(\theta, y_1(\theta_1)) \psi_\theta(y_2(\theta), \theta_2) f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2 = \int I_1^2(\theta, y_1(\theta_1)) \frac{\partial V_2(\theta)}{\partial \theta_2} f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2 \\ &= V_2(\bar{\theta}_2) I_1^2(\bar{\theta}_2, \theta_1, y_1(\theta_1)) f_2(\bar{\theta}_2 | \theta_1, y_1(\theta_1)) \mathbf{1}_{\bar{\theta}_t} - V_2(\underline{\theta}_2) I_1^2(\underline{\theta}_2, \theta_1, y_1(\theta_1)) f_2(\underline{\theta}_2 | \theta_1, y_1(\theta_1)) \mathbf{1}_{\underline{\theta}_t} \\ &\quad + \int V_2(\theta) \frac{\partial}{\partial \theta_1} f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2 \end{aligned} \quad (84)$$

where, for the last equality, we used the property that, for any differentiable function $h(\theta_2)$,

$$\int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{\partial h(\theta_2)}{\partial \theta_2} I_1^2(\theta, y_1) f_2(\theta_2 | \theta_1, y_1) d\theta_2 = \frac{\partial}{\partial \theta_1} \int_{\underline{\theta}_2}^{\bar{\theta}_2} h(\theta_2) f_2(\theta_2 | \theta_1, y_1) d\theta_2.$$

The last equality then follows from this property along with Leibniz rule.

Now let $\varsigma(\theta)$ be the function defined by the differential equation

$$\frac{\partial \varsigma(\theta)}{\partial \theta_2} = V_2(\theta) \frac{\partial}{\partial \theta_1} f_2(\theta_2 | \theta_1, y_1(\theta_1))$$

with boundary condition

$$\varsigma(\theta_1, \underline{\theta}_2) = 0.$$

The constraint (84) can then be conveniently rewritten as

$$\begin{aligned} \varsigma(\theta_1, \bar{\theta}_2) &= Z_2(\theta_1) - V_2(\bar{\theta}_2) I_1^2(\bar{\theta}_2, \theta_1, y_1(\theta_1)) f_2(\bar{\theta}_2 | \theta_1, y_1(\theta_1)) \mathbf{1}_{\bar{\theta}_t} \\ &\quad + V_2(\underline{\theta}_2) I_1^2(\underline{\theta}_2, \theta_1, y_1(\theta_1)) f_2(\underline{\theta}_2 | \theta_1, y_1(\theta_1)) \mathbf{1}_{\underline{\theta}_t}. \end{aligned}$$

Denoting with $\widehat{\xi}(\theta)$ the costate variable for the law of motion of the auxiliary state variable $\varsigma(\theta)$ defined above, we then have that, at the optimum, $\widehat{\xi}(\theta)$ is independent of θ_2 .⁴⁹ Then let $\xi_2(\theta_1) \equiv -\widehat{\xi}(\theta_1, \theta_2)$. Equipped with this notation, we then have that, at the optimum, Conditions (66), (72), and (73), and thereby (79), must continue to hold. The boundary conditions (67) and (68), instead, should be replaced by

$$\mu_2(\theta_1, \underline{\theta}_2) = -\xi_2(\theta_1) I_1^2(\underline{\theta}_2, \theta_1, y_1(\theta_1)) f_2(\underline{\theta}_2 | \theta_1, y_1(\theta_1)) \mathbf{1}_{\underline{\theta}_t} \quad (85)$$

and

$$\mu_2(\theta_1, \bar{\theta}_2) = -\xi_2(\theta_1) I_1^2(\bar{\theta}_2, \theta_1, y_1(\theta_1)) f_2(\bar{\theta}_2 | \theta_1, y_1(\theta_1)) \mathbf{1}_{\bar{\theta}_t}. \quad (86)$$

Combining (73) and (79) with these boundary conditions, and using the fact that, for any given y_1 ,

$$\begin{aligned} I_1^2(\bar{\theta}_2, \theta_1, y_1) f_2(\bar{\theta}_2 | \theta_1, y_1) \mathbf{1}_{\bar{\theta}_t} - I_1^2(\underline{\theta}_2, \theta_1, y_1) f_2(\underline{\theta}_2 | \theta_1, y_1) \mathbf{1}_{\underline{\theta}_t} + \int_{\underline{\theta}_2}^{\bar{\theta}_2} \partial f_2(\theta_2 | \theta_1, y_1) / \partial \theta_1 = \\ \frac{\partial}{\partial \theta_1} \int_{\underline{\theta}_2}^{\bar{\theta}_2} f_2(\theta_2 | \theta_1, y_1) = 0, \end{aligned}$$

we obtain once again the Rogerson-inverse-Euler condition (81).

Conditions (73), (81) and (79), together with the boundary condition (85), and the fact that, for any y_1 ,

$$\begin{aligned} -I_1^2(\underline{\theta}_2, \theta_1, y_1) f_2(\underline{\theta}_2 | \theta_1, y_1) \mathbf{1}_{\underline{\theta}_t} + \int_{\underline{\theta}_2}^{\bar{\theta}_2} \partial f_2(\theta_2 | \theta_1, y_1) / \partial \theta_1 = \\ \frac{\partial}{\partial \theta_1} \int_{\underline{\theta}_2}^{\bar{\theta}_2} f_2(\theta_2 | \theta_1, y_1) = \frac{\partial}{\partial \theta_1} F_2(\theta_2 | \theta_1, y_1), \end{aligned}$$

in turn imply that

$$\begin{aligned} \mu_2(\theta_1, \theta_2) &= - \int_{\underline{\theta}_2}^{\bar{\theta}_2} f_2(s | \theta_1, y_1(\theta_1)) \left\{ \frac{1}{v'(c_2(\theta_1, s))} - \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{1}{v'(c_2(\theta_1, s))} f_2(s | \theta_1, y_1(\theta_1)) ds \right\} ds \\ &\quad - \frac{\mu_1(\theta_1)}{f_1(\theta_1)} \frac{\partial}{\partial \theta_1} F_2(\theta_2 | \theta_1, y_1). \end{aligned}$$

Using the fact that

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

we obtain again Condition (82).

We thus conclude that the optimality conditions for the moving-support case are the same as for the non-moving-support case.

⁴⁹To see this, recall that the maximum principle requires that the derivative of $\widehat{\xi}(\theta)$ with respect to θ_2 be equal to the negative of the derivative of the Hamiltonian with respect to the auxiliary variable $\varsigma(\theta)$, which is zero.

Step 2. we now show how the above optimality conditions permits us to arrive at the expressions for the wedges in the two propositions.

Consider first the period-2 wedges. Combining Definition 3 with (65), we have that the period-2 wedges, under the second-best allocations, are given by

$$W_2(\theta) \equiv 1 - \frac{\psi_y(y_2(\theta), \theta_2)}{v'(c_2(\theta))} = \psi_{\theta y}(y_2(\theta), \theta_2) \left(\frac{\mu_2(\theta)}{f_2(\theta_2 | \theta_2, y_1(\theta_1))} \right).$$

Hence, the period-2 relative wedges are given by

$$\hat{W}_2(\theta) \equiv \frac{W_2(\theta)}{1 - W_2(\theta)} = \frac{-v'(c_2(\theta))}{\psi_y(y_2(\theta), \theta_2)} \psi_{\theta y}(y_2(\theta), \theta_2) \left(\frac{-\mu_2(\theta)}{f_2(\theta_2 | \theta_2, y_1(\theta_1))} \right)$$

with $-\mu_2(\theta)/f_2(\theta_2 | \theta_2, y_1(\theta_1))$ given by (82).

Next, consider the period-1 wedges. Definition 3, together with (78) and the definition of the $LD_1^X(\theta_1)$ function in (23), imply that

$$\begin{aligned} W_1(\theta_1) &\equiv 1 - \frac{\frac{\psi_y(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))}}{[1 + LD_1^{FB;X}(\theta_1)]} \\ &= - \frac{\left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \delta \frac{\partial}{\partial y_1} \int I_1^2(\theta, y_1(\theta_1)) \psi_{\theta}(y_2(\theta), \theta_2) f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2 + \left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \psi_{\theta y}(y_1(\theta_1), \theta_1)}{1 + LD_1^{FB;X}(\theta_1)}. \end{aligned}$$

It follows that the period-1 relative wedges under the second-best allocations are given by

$$\begin{aligned} \hat{W}_1(\theta_1) &\equiv \frac{W_1(\theta_1)}{1 - W_1(\theta_1)} \\ &= - \frac{\left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \delta \frac{\partial}{\partial y_1} \int I_1^2(\theta, y_1(\theta_1)) \psi_{\theta}(y_2(\theta), \theta_2) f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2 + \left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \psi_{\theta y}(y_1(\theta_1), \theta_1)}{\frac{\psi_y(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))}} \end{aligned}$$

with $-\mu_1(\theta_1)/f_1(\theta_1)$ given by (77) in case of Utilitarian preferences for redistribution ($r = 1$) and by (83) in case of Rawlsian preferences for redistribution ($r = 0$).

Using the definition of $\Omega(\theta_1)$ from Proposition 1 and the definition of W_t^{RRN} , $t = 1, 2$, from (30) we thus have that

$$\hat{W}_1(\theta_1) = \left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \gamma_1(\theta_1) v'(c_1(\theta_1)) \left\{ \Omega(\theta_1) + \hat{W}_1^{RRN}(\theta_1) \right\}$$

and

$$\hat{W}_2(\theta) = \left(\frac{-\mu_2(\theta)}{f_2(\theta_2 | \theta_2, y_1(\theta_1))} \right) \frac{\gamma_1(\theta_1) v'(c_2(\theta))}{I_1^2(\theta, y_1(\theta_1))} \hat{W}_2^{RRN}(\theta)$$

It is also easy to see that the only term affected by the presence of LBD is

$$\left[\left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \gamma_1(\theta_1) v'(c_1(\theta_1)) \right] \Omega(\theta_1).$$

Case $r = 0$ (Rawlsian preferences for redistribution). Using (83), and the definition of $\gamma_1(\theta_1)$, we have that the correction in the period-1 relative wedge due to risk aversion is equal to

$$\left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \gamma_1(\theta_1) v'(c_1(\theta_1)) = v'(c_1(\theta_1)) \int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} \frac{dF_1(s)}{1 - F_1(\theta_1)} \equiv RA(\theta_1),$$

and that, in the absence of LBD,

$$\hat{W}_1^{RA}(\theta_1) = RA(\theta_1)\hat{W}_1^{RRN}(\theta_1)$$

and

$$\hat{W}_2^{RA}(\theta) = \hat{W}_2(\theta) = \left(\frac{-\mu_2(\theta)}{f_2(\theta_2 | \theta_2, y_1(\theta_1))} \right) \frac{\gamma_1(\theta_1)v'(c_2(\theta))}{I_1^2(\theta, y_1(\theta_1))} \hat{W}_2^{RRN}(\theta)$$

where $-\mu_2(\theta)/f_2(\theta_2 | \theta_2, y_1(\theta_1))$ is given by (82), and where the term $-\mu_1(\theta_1)/f_1(\theta_1)$ in (82) is given by (83).

Case $r = 1$ (*Utilitarian preferences for redistribution*). Using (77), and the definition of $\gamma_1(\theta_1)$, we have that the correction in the period-1 relative wedge due to the combination of risk aversion and the lower aversion to inequality is equal to

$$\begin{aligned} \left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \gamma_1(\theta_1)v'(c_1(\theta_1)) &= v'(c_1(\theta_1)) \left[\int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} \frac{dF_1(s)}{1-F_1(\theta_1)} - \int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} dF_1(s) \right] \\ &= RA(\theta_1) - D(\theta_1), \end{aligned}$$

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

Furthermore, in the absence of LBD,

$$\hat{W}_1^{URA}(\theta_1) = [RA(\theta_1) - D(\theta_1)]\hat{W}_1^{RRN}(\theta_1)$$

and

$$\hat{W}_2^{URA}(\theta) = \hat{W}_2(\theta) = \left(\frac{-\mu_2(\theta)}{f_2(\theta_2 | \theta_2, y_1(\theta_1))} \right) \frac{\gamma_1(\theta_1)v'(c_2(\theta))}{I_1^2(\theta, y_1(\theta_1))} \hat{W}_2^{RRN}(\theta)$$

where $-\mu_2(\theta)/f_2(\theta_2 | \theta_2, y_1(\theta_1))$ is given by (82), and where the term $-\mu_1(\theta_1)/f_1(\theta_1)$ in (82) is given by (77). Q.E.D.

9.2 Computational Appendix

9.2.1 Numerical results in Sections 5 and 6

For the numerical results reported in Sections 5 and 6 we have solved numerically the planner's relaxed program as described in Sections 3.2, 5, and 6. We solved this program by (i) interpolating the optimal policies $e_1(\theta_1) \equiv y_1(\theta_1)/\theta_1$, $c_1(\theta_1)$, and $e_2(\theta_1, \varepsilon_2) \equiv y_2(\theta_1, \theta_1^\rho y_1(\theta_1)^{\zeta\varepsilon})/\theta_1^\rho y_1(\theta_1)^{\zeta\varepsilon}$ (the interpolating functions are described below), (ii) setting the interpolation coefficients as control variables, and (iii) solving the ensued numerical optimization problem by using the gradient-free optimization solver `fminsearch` in MATLAB_R2014b. The second-period consumption policy $c_2(\theta_1, \theta_2)$ has then been obtained as a residual of the other policies by using the envelope incentive-compatibility conditions and the definition of the flow utility functions, for each productivity history, θ .

We verified numerically that the earnings and consumption policies that solve the relaxed program satisfy all the remaining incentive-compatibility constraints. The latter verification was done by verifying that (i) the second-period earnings policies are increasing in second-period shock/productivity,

for any given level of first-period productivity, and (ii) all the integral monotonicity conditions in (6) are satisfied.

To construct the relaxed optimization problem (and verify its validity) we have used a uniform grid of N productivity shocks with the first and last points in the grid corresponding to the 0.05% and 99.95% percentiles of the deployed distribution. The distribution has been discretized to conduct numerical integration by using the trapezoid method. The value for N is 150 for all results involving the ParetoLognormal distribution, and 75 for all results involving the Lognormal distribution. We have restricted attention to outcomes over a subset of the above grids. Namely, to the grid between the 0.01% and 99.5% percentiles for the results involving the ParetoLognormal distribution, and between the 0.01% and the 99.9% percentiles for the results involving the Lognormal distribution. The reason for restricting attention to such ranges is that, outside these regions, we expect high numerical error due to the deployed discretization of the distributions.

Regarding the deployed interpolation, we used simple polynomials of order 7 for $c(\theta_1)$. We approximated the period-1 labor supply schedule $e_1(\theta_1)$ by means of a function $\hat{e}_1(\theta_1) + \varpi_1 e_1^{RN}(\theta_1)$, where $\hat{e}_1(\theta_1)$ is a simple polynomial of order n_1 and where $e_1^{RN}(\theta_1)$ is the known formula for the optimal period-1 effort under risk neutrality. We approximated the period-2 labor supply schedule $e_2(\theta_1, \varepsilon_2)$ by means of a function $\hat{e}_2(\theta_1, \varepsilon_2) + \varpi_2 e_2^{RN}(\theta_1, \varepsilon_2)$, where $e_2^{RN}(\theta_1, \varepsilon_2)$ is the known formula for the optimal period-2 labor supply under risk neutrality (as a function of period-1 productivity and period-2 shock), and where $\hat{e}_2(\theta_1, \varepsilon_2)$ is a tensor product node-basis scheme with 4-degree and 4-degree simple polynomials for the respective dimensions of the $(\theta_1, \varepsilon_2)$ space. The coefficients ϖ_1 and ϖ_2 , together with the coefficients of the simple polynomials deployed for the interpolation of $e_1(\theta_1)$, $c_1(\theta_1)$, and $e_2(\theta_1, \varepsilon_2)$, were the control variables in the numerical optimization problem we solved, after setting to zero the coefficients of the polynomial $\hat{e}_2(\theta_1, \varepsilon_2)$ that correspond to the following terms of the tensor product: (2, 4), (3, 3), (3, 4), (4, 2), (4, 3), (4, 4), and, for the results involving the ParetoLognormal distribution, (1, 4) and (2, 3).⁵⁰ The value for n_1 is 8 for the results for the ParetoLognormal distribution with Rawlsian preferences for redistribution, 9 for the results for the ParetoLognormal distribution with Utilitarian preferences for redistribution, and 4 for the results for the Lognormal distribution with Utilitarian preferences for redistribution.

9.2.2 Numerical results in Section 7

For the calibration of the benchmark economy with LBD, the required parameters were set to minimize the sum of squared percentage deviations of the model-generated moments from the target moments by using the constrained-optimization solver `fmincon` in `MATLAB_R2014b`. To ensure convergence we imposed bounds on the admissible values for the parameters in question; namely, $h_1 \geq 0$, $\zeta \in [0.2, 0.6]$, $\rho \in [0, 1]$, $\sigma \in [0, \sqrt{0.33}]$, $\lambda \in [2.1, 6]$. The derived solution for all calibrated parameters (reported in Table 2), however, turned out to be interior.

⁵⁰The term (i, j) of the tensor product in question is the term $\theta_1^{i-1} e_2^{j-1}$.

In calculating the moments, we interpolated workers' optimal decisions (for given tax code) for $y_1(\theta_1)$, $c_1(\theta_1)$ and $\hat{y}_2(\theta_1, \varepsilon_2) \equiv y_2(\theta_1, \theta_1^\rho y_1(\theta_1)^\zeta \varepsilon) / \theta_1^\rho y_1(\theta_1)^\zeta$ by using 20-degree Chebyshev polynomials for $y_1(\theta_1)$ and $c_1(\theta_1)$, and a tensor product node-basis scheme with 16-degree and 14-degree Chebyshev polynomials for the respective dimensions of the $(\theta_1, \varepsilon_2)$ space for $\hat{y}_2(\theta_1, \varepsilon_2)$.

The second-period consumption policy was then determined as a residual of the other policies by using the definition of the workers' intertemporal budget constraint for each productivity history, θ . The Chebyshev interpolation coefficients were set to minimize the sum of squared residuals in the workers' optimality conditions. For the minimization in question, we used the routine `lsqnonlin.m` in MATLAB_R2014b.

Given the calibrated model (with or without LBD), we interpolated the pseudo-optimal and optimal policies for the relaxed program by approximating earnings and first-period consumption with Chebyshev polynomials. The second-period consumption policy was then obtained as a residual of the other policies using the envelope incentive-compatibility condition and the definition of the flow utility function, for each productivity history θ . The Chebyshev interpolation coefficients were set to minimize the sum of squared residuals in the optimality conditions from the planner's corresponding problem. For the minimization in question, we have used the routine `lsqnonlin.m` in MATLAB_R2014b. We have used n_1^C -degree Chebyshev interpolation for period-1 earnings and consumption, and a tensor product node-basis scheme with n_2^C -degree and 14-degree Chebyshev polynomials, for the respective dimensions of the $(\theta_1, \varepsilon_2)$ space, for the auxiliary earnings $\hat{y}_2(\theta_1, \varepsilon_2)$. The values for n_1^C and n_2^C are, respectively, 20 and 16 for pseudo-optimal policies, and 25 and 45 for optimal policies.

For the calibration of the benchmark economy and the derivation of pseudo-optimal and optimal policies (with or without LBD), we truncated the ParetoLognormal distribution of productivity shocks ε from below at the 1% percentile, and used a uniform grid of 399 productivity shocks with the last point in the grid corresponding to the 99.99% percentile of the truncated distribution. In addition, we have discretized 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 the 99.99% of the distribution, we have chosen to not ignore the outcomes of any grid point.