

Taxation under Learning-by-Doing*

Miltiadis Makris
University of Kent

Alessandro Pavan
Northwestern University

November 2018

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 contribute to higher labor wedges under the optimal tax code. Next, we calibrate the model to US earnings data and show that simple taxes that are invariant to past incomes but age-dependent are approximately optimal and generate most of the welfare gains from the optimal reform. Under such simple codes, tax rates are lower for younger workers than for older ones. We also show 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 in the taxes paid by young workers, compared to the optimal reform that ignores learning-by-doing. Finally, we 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.

*For comments and useful suggestions, we thank the Editor and the referees, as well as Emmanuel Farhi, Mikhail Golosov, Marek Kapicka, Paul Klein, Alex Mennuni, Ludovic Renou, Stefanie Stantcheva, Arnau Valladares-Esteban, Pierre Boyer, 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, Edinburgh University, Northwestern University, the University of Western Ontario, the 2016 ESSET Conference, and the Fourth Taxation Theory Conference. Abdoulaye Ndiaye and Laura Doval provided excellent research assistance. The usual disclaimer applies. Email addresses: mmakris.econ@gmail.com (Makris); alepavan@northwestern.edu (Pavan).

1 Introduction

“On-the-job training or learning-by-doing appear to be at least as important as schooling in the formation of human capital” (Lucas, 1988).

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.¹ 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 as variations in earnings reflect variations in skills/productivities, these findings suggest that labor experience has a significant impact on human capital accumulation.² Learning-by-doing is also believed to be one of the key drivers of the negative relationship between firms’ unit production costs and their cumulative past production (see, e.g., Levitt et al. (2013)).³

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

¹For the general-equilibrium effects of learning-by-doing in macro models, see Becker (1964), Arrow (1972), and Lucas (1988) for seminal contributions, and Chang et al. (2002) and D’Alessandro et al. (2018) for more recent contributions.

²See 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.

³The literature documenting the impact of cumulative past production on unit-costs (especially early in the process) also includes Wright (1936), Benkard (2000), Thompson (2001), Thornton and Thompson (2001), and Thompson (2010, 2012). See also the discussion of the related literature below for more details about this strand of the literature.

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 affect the dynamics of wedges over the lifecycle. Third, as workers of different productivity expect different rents in future periods, the benefits of shifting the distribution of future productivity towards lower levels so as to economize on future rents typically depends on the workers' current productivity. Other things equal, LBD may thus also affect the progressivity of the wedges and of the associated optimal tax code. Clearly, the above effects interact with other forces that also play an important role in the determination of the level, dynamics, and progressivity of the wedges, such as the agents' risk aversion, the persistence of the agents' initial private information, the distribution from which the shocks to the agents' productivity are drawn, the elasticity of the agents' labor supply with respect to the (net of-taxes) wages, and the planner's preferences for redistribution. The contribution of the present paper is to develop a model that sheds light on the above interactions.

We consider a stylized, yet rich, economy in which the agents' working life is divided into two blocks, an earlier phase in which workers are young and learn on the job, and a second phase in which workers are older and take advantage of the impact of LBD on their productivity. To illustrate the novel effects brought in by LBD in the simplest possible way, 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 consider the more general case where agents may be risk averse (and have preferences for consumption smoothing) and where the planner may assign general non-linear Pareto weights to different agents' lifetime expected utility. We derive a general formula for the wedges under second-best allocations and show how the agents' risk aversion and the planner's preferences for redistribution interact with LBD in the determination of the labor wedges.

Equipped with the analytical results, we then calibrate a version of the model in which agents are risk averse and the planner assigns equal Pareto weights to all agents to match various moments of the US earnings distribution as reported in Huggett et al. (2011).⁴ The calibration also provides us with a parameter value for the intensity of LBD that is consistent with both the results in the micro literature (see, e.g., the meta-analysis in Best and Kleven (2013)) as well as the estimation results in the macro literature (see, e.g., Chang et al. (2002)). We show that reforming the existing US tax code by adopting the optimal one would yield an increase in expected lifetime utility equivalent to a 4% increase in consumption at each productivity history in the calibrated economy under the

⁴The approximation of the current US tax code is from Heathcote et. al. (2017).

existing US tax code. We also show that the utility that the agents derive from the second-best allocations, as well as the second-best earnings distributions, are very close to those that emerge under a simple history-independent tax code in which, in each period, taxes are invariant to past incomes and are given by a linear-power function of the form $T_s(y_s) = y_s - k_s y_s^{\tau_s} - b_s$, 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.⁵ Under this 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 less progressive for both young and old workers, are higher for young workers, and are higher at low income percentiles but lower at high income percentiles for old workers.

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 switching from the existing US tax code to the one implementing the second-best allocations can be generated through simple age-dependent linear taxes.⁶ 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 of reforming the US tax code (precisely, a 14% overestimation of the increase in consumption). We also show that, under LBD, wedges for young workers are higher than their counterparts without LBD. Importantly, that wedges are higher does not imply that taxes are also higher: in fact, marginal tax rates for young workers 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. We also 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.

We also conduct various comparative statics exercises that illustrate the effects on tax codes of variations in the levels of skill persistence and of the intensity of the LBD effects. Finally, we illustrate the role of the stochasticity of the LBD effects for the dynamics of tax rates over the lifecycle.

Outline. The rest of the manuscript is organized as follows. Below, we wrap up the Introduction with a discussion of the most pertinent literature. Section 2 introduces the model. Section 3 describes the first-best policies. Section 4 describes the second-best policies and derives various implications for the level, dynamics, and progressivity of the labor wedges. Section 5 calibrates the model to the US income distribution under the current US tax code, and contains all the quantitative results. Section 6 concludes. All proofs are either in the Appendix at the end of the document, or in the

⁵Similar functional forms have been considered in Benabou (2002) and Heathcote et al. (2017).

⁶See Kremer (2001), Weinzierl (2011), and Banks & Diamond (2011) for an earlier discussion of the benefits of age-dependent taxation, and Farhi and Werning (2013) and Golosov et al (2016) for recent developments.

Supplementary Material. The latter also contains a detailed description of the methods used to establish all numerical results, as well as additional comparative statics results of optimal tax codes with respect to the degree of the agents' risk aversion, the Frisch elasticity of the agents' labor supply, and the planner's preferences for redistribution.

Related Literature. The empirical literature on LBD is too vast to be succinctly summarized here. For an overview of the micro literature, we refer the reader to Thompson (2010, 2012). This literature draws on the labor economics and the industrial organization literature to document how LBD at both the individual and the firm level affects the dynamics of wages and firms' production costs in a variety of industries. While the intensity of LBD is found to vary across sectors, a common finding is that LBD is a function of both time and cumulative past output, and that its effects fade away after a certain number of periods (such sharp decline in the intensity of LBD is often referred to as "bounded learning" in the literature; see, e.g., Thomsson, 2010). The specification of the productivity process we consider in our quantitative analysis appears broadly consistent with these facts. We postulate that LBD is active only in the first twenty years of each agent's working life, that productivity in the second half of each agent's working life depends on the cumulative output generated in the first half, and that output generated in each of the first twenty years affects a worker's productivity in the second half with a weight that declines over time, capturing the idea that LBD is particularly strong in the first few years of employment.

While the works cited above document heterogeneity in the significance of LBD, a comprehensive analysis of how LBD varies across sectors is missing. The literature has proceeded more on a case-by-case basis by investigating the importance of LBD in a few specific markets for which suitable data are available. In the absence of a detailed analysis of how LBD varies across markets, we opted for a specification that treats the whole economy as a single homogenous market.

For an overview of the macro literature on LBD (both in growth and real business cycles models), we refer the reader to Chiang et al. (2002) and D'Alessandro et al. (2018). This literature finds that LBD is an important propagation mechanism of various macro shocks and an important determinant of economic growth. Using a rich DSGE model, Chiang et al. (2002) estimate a range of values for the elasticity of labor productivity with respect to past output. The numerical value we obtain in our calibration analysis is roughly in the middle of this range. This literature finds that the introduction of LBD effects improves significantly the ability of macro models to fit the dynamics of aggregate output, hours of work, inflation, and various other macro variables of interest.

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), Stantcheva (2017), Perrault (2017), and Makris and Pavan (2018b).⁷

Krause (2009) considers an economy in which productivity takes only two values and focuses on the effects of LBD on the "no-distortion-at-the-top" result.

⁷See also Stantcheva (2015).

Best and Kleven (2013) consider a two-period economy with risk-neutral agents, time-invariant private information, and deterministic LBD effects. That paper builds a comprehensive meta analysis of the existing micro literature on career effects and human capital accumulation to provide a range of plausible values for the intensity of LBD and other endogenous human capital and career effects. This range is broader than the one found in the macro literature, as reported in Chang et al (2002), with larger lower and upper bounds. Our calibrated value of the intensity of the LBD effects is within this range as well. In the theoretical part of the analysis, Best and Kleven (2013) restrict attention to tax codes where marginal tax rates are possibly age-dependent but invariant in past incomes. Contrary to our paper, they find that tax rates should decline with age. Our numerical analysis shows that conditioning tax rates on past incomes brings few welfare gains. On the other hand, our comparative statics exercises with respect to the stochasticity of the LBD effects show that the dynamics of taxes are reversed as one moves from an economy with deterministic to one with stochastic LBD effects (see the analysis in Subsubsection 5.5). Importantly, our calibration analysis reveals that the stochasticity of the LBD effects is not a mere theoretical curiosity, it is empirically relevant.⁸ When LBD has stochastic effects, young workers face risk in their consumption when old due to the volatility in their earnings. By charging higher taxes to the old, the planner reduces such volatility, which in turn increases welfare when agents are risk averse, as they are in our calibrated economy.

Kapicka (2006) considers the design of optimal history-independent tax codes in a dynamic economy where human capital is endogenous and evolves deterministically over time. The key finding is that, compared to an economy with exogenous human capital, steady-state tax rates are lower when the agents' productivity is endogenous. In contrast, in our economy, LBD has stochastic effects on the agents' productivity. We find that marginal tax rates under optimal tax codes are lower in our calibrated economy with LBD than in a fictitious economy that features the same productivity process but assumes the latter is exogenous. However, we show that this does not imply that labor wedges are also lower. Importantly, we study the dynamics of taxes over the lifecycle, whereas the focus of that paper is steady-state allocations.

Kapicka (2015a) and Kapicka (2015b) consider optimal taxation in an economy where workers' ability is constant over time and where agents' preferences are time-non-separable in their labor supply decisions. These papers show that optimal tax rates decline over the lifecycle, irrespective of whether labor supply decisions at different ages are substitutes (as in the Ben-Porath (1967) economy) or complements (as in the economy with LBD). We find the opposite dynamics. The reason why, in a Ben-Porath (1967) economy, tax rates decline over the lifecycle is that such dynamics induce workers to substitute labor with training earlier in their careers, which is desirable since it boosts the agents' productivity when old. The reason why in the LBD economy considered in Kapicka

⁸In addition to calibrating the intensity of the LBD effects, we calibrate the variance of the period-2 productivity shocks conditional on period-1 productivity and find that it is positive and plays a major role in the determination of the optimal tax code in the calibrated economy.

(2015b) tax rates decline over the lifecycle, despite labor supply decisions at different ages being complements, is that the anticipation of lower tax rates when old induces the young to work harder (this effect is referred to as the “anticipation effect” in Kapicka (2015b) and is qualitatively similar to the “elasticity effect” in Best and Kleven (2013)). The reason why we find the opposite dynamics is that, in our economy, LBD has a stochastic effect on the evolution of the agents’ productivity (the arguments are the same as in our discussion of Best and Kleven (2013)).⁹

Kapicka and Neira (2016) consider a two-period economy where productivity evolves stochastically over the workers’ lifecycle. However, contrary to our work, that paper assumes that the agents’ private information is constant over time. In addition to choosing their period-1 labor supply, agents invest in human capital. Such investment is unobservable while its effect on period-2 productivity is observable but stochastic. Formally, the above assumptions thus identify an economy with both adverse selection and moral hazard, but where both frictions are present only in the first period. That paper finds that tax rates should decline over the lifecycle. This is in contrast to our findings. The reason is that our economy is fundamentally different. It does not feature any moral hazard, the evolution of human capital originates in LBD instead of being the result of learning that takes place outside the workplace, and agents must be provided with incentives to reveal their private information in both periods.

Stantcheva (2017) considers a general multi-period economy where, in each period, agents invest in human capital in addition to providing their labor. As in our paper, agents receive private information over time and human capital has stochastic effects on the evolution of the agents’ productivity. Our work differs from that paper in two important dimensions. First, the evolution of human capital is a direct byproduct of labor supply. Second, in our economy the planner has only one instrument, whereas in that paper the planner can use both labor-income taxes and education and training subsidies to influence the agents’ joint determination of their labor choices and investments in human capital. Importantly, the fact that, in our model, investments in human capital cannot be controlled separately from labor supply decisions has important implications for the design of optimal tax codes and in particular for the progressivity of marginal tax rates.

In independent work, Perrault (2018) considers a special case of our economy in which period-2 productivity is invariant in period-1 productivity. As in our paper, he finds that whether marginal tax rates increase or decrease with age depends on the stochasticity of the LBD effects. Relative to that paper, our work shows that LBD also affects the progressivity of the optimal tax codes and that the latter crucially depends on the persistence of the productivity shocks. Given the importance that skill persistence has received in the new dynamic public finance literature (and the fact that skill persistence is present in the calibrated version of our economy), understanding the role of the interaction of skill persistence with LBD is important.

In Makris and Pavan (2018b), we consider a general multi-period economy in which the agents’

⁹Another difference between our economy and Kapicka (2015a) is that, in our economy, the progressivity of the tax code declines with age, whereas, in Kapicka (2015a), it is almost invariant over time.

private information evolves endogenously over time according to an arbitrary process. The two-period economy considered in the present paper is a special case of the general economy in the companion paper. The contribution of the two papers is, however, different. In Makris and Pavan (2018b), we show how one can use a recursive characterization of the second-best allocations to arrive at a general formula for the wedges that sheds light on the interaction between the various forces at play in various dynamic mechanism design problems with endogenous private information. In the present paper, instead, we focus on the specific predictions that LBD has for the design of optimal tax codes.

Related is also the work of Benabou (2002), Conesa and Krueger (2006), Heathcote et al. (2017), 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 an overview of recent developments,¹⁰ and to Golosov et al. (2006), Albanesi and Sleet (2006), Battaglini and Coate (2008), Kocherlakota (2010), Gorry and Oberfield (2012), Kapicka (2013), Farhi and Werning (2013), and Golosov et al. (2016) for applications to dynamic optimal taxation. In these papers, the agents' productivity evolves stochastically over the lifecycle, but is exogenous. One of the key findings of this literature is that the dynamics of the wedges crucially depends on the interaction between skill persistence and the agents' degree of risk aversion: wedges (weakly) decline over time under small degrees of risk aversion but increase for large degrees (see also Garrett and Pavan, 2015, for a discussion of the robustness of such findings). The key contribution of our paper relative to this literature is the investigation of the effects of the endogeneity of the type process on the dynamics of distortions.¹¹

We view the contribution of the present paper relative to the various literatures discussed above as twofold. First, we identify a novel mechanism by which the endogeneity of the agents' private information about their time-varying productivity affects the design of optimal tax codes. Second, we provide a flexible quantitative analysis that permits us to shed light on how LBD interacts with other empirically-relevant channels in shaping the design of optimal tax codes.

¹⁰See also Garrett and Pavan (2015) and Garrett, Pavan, and Toikka (2018) for a variational approach to the characterization of the dynamics of allocations under optimal contracts, in economies with exogenous types.

¹¹Pavan, 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 (2017). 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.

2 The Economy

Agents, productivity, and information. The economy is populated by a unit-mass continuum of workers with 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 by letting productivity be exogenous in the first period but endogenous in the second.¹²

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), which is her private information. The function $\psi(y_t, \theta_t)$ is thrice differentiable, increasing, and convex in y_t . Consistently with the rest of the literature¹³, we assume that ψ takes the familiar iso-elastic form

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

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 ψ . Hence, under this specification, ϕ is the inverse Frisch 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 $\Theta_1 = (\underline{\theta}_1, \bar{\theta}_1)$ and zero anywhere else. Each worker's period-2 productivity is endogenous and given by

$$\theta_2 = z_2(\theta_1, y_1, \varepsilon_2) = \theta_1^\rho y_1^\zeta \varepsilon_2 \quad (2)$$

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-best 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 (θ_1, θ_2) . The

¹²As 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.

¹³See, among others, Kapicka (2013), Farhi and Werning (2013), and Best and Kleven (2013).

dependence of the agents' period-2 productivity on their period-1 income is what captures learning-by-doing (LBD). When period-1 income is the product of period-1 effort and period-1 productivity (as it is in most of the literature on income taxation, following the seminal work of Mirrlees (1971)), 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 from 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

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

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 a 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,¹⁴ and (b) less robust than the first channel 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 complementarity/substitutability between past productivity and past labor supply 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. The *lifetime* utility of each agent is given by

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

where $y \equiv (y_1, y_2)$, $c \equiv (c_1, c_2)$, and $v : \mathbb{R} \rightarrow \mathbb{R}$ is an increasing, concave, twice differentiable function.

¹⁴See, for example, Kapicka (2015a,b) and Stantcheva (2017).

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

Formally, the dual problem can be stated as follows. Let $\chi = (y, c)$ denote an *allocation rule*, specifying, for each agent, the lifetime profile of income-consumption pairs $\chi(\theta) = (y_t(\theta^t), c_t(\theta^t))_{t=1,2}$ as a function of the agent's lifetime productivity history, 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

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

denote the expected lifetime utility of each agent of initial productivity θ_1 , under the allocation rule χ . Hereafter, we use tildes to denote random vectors. Importantly, note that the dependence on χ is both through the policies $c_t(\cdot)$ and $y_t(\cdot)$, 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 *dual problem* consists of maximizing expected discounted tax revenues¹⁵

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

subject to the constraint that

$$\int V_1(\theta_1) q(\theta_1) dF_1(\theta_1) \geq \kappa \tag{4}$$

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 $\chi(\cdot)$). The function $q : \Theta_1 \rightarrow \mathbb{R}_+$ in (4) describes the non-linear Pareto weights the planner assigns to the agents' expected lifetime utilities. Without loss of generality, the weights are normalized so that $\int q(\theta_1) dF_1(\theta_1) = 1$. Note that, when $q(\theta_1) = 1$ for all $\theta_1 \in \Theta_1$, $\int V_1(\theta_1) q(\theta_1) dF_1(\theta_1)$ reduces to the Utilitarian social welfare function, whereas the limiting case of $q(\theta_1) = 0$ for all $\theta_1 > \underline{\theta}_1 + \varepsilon$, and $q(\theta_1) = 1/[\varepsilon f_1(\theta_1)]$ for all $\theta_1 \in [\underline{\theta}_1, \underline{\theta}_1 + \varepsilon]$, with $\varepsilon > 0$, approximates, as $\varepsilon \rightarrow 0$, the redistribution environment of the Rawlsian social welfare function.

¹⁵Consistently with the rest of the literature, we assume that δ is 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 (2017)).

Implicit in the formulation of the dual problem above is the assumption that agents do not privately save (equivalently, their savings are controlled directly by the planner) and that the planner commits to the intertemporal tax code she chooses.

3 First-Best Policies

Suppose 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 $\chi = (y, c)$, when the agent's initial productivity is θ_1 and period-1 output/income is y_1 .

Proposition 1. *The first-best policies $\chi^* = (y^*, c^*)$ satisfy the following optimality conditions at all interior points, with $\lambda[\chi^*]$ -probability one.¹⁶*

$$\frac{\psi_y(y_1^*(\theta_1), \theta_1)}{v'(c_1^*(\theta_1))} = 1 + LD_1^{\chi^*}(\theta_1), \quad (5)$$

where

$$LD_1^{\chi}(\theta_1) \equiv \delta \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[y_2(\tilde{\theta}) - c_2(\tilde{\theta}) + \frac{v(c_2(\tilde{\theta})) - \psi(y_2(\tilde{\theta}), \tilde{\theta}_2)}{v'(c_2(\tilde{\theta}))} \right], \quad (6)$$

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

along with

$$v'(c_1^*(\theta_1)) = v'(c_2^*(\theta)), \quad (7)$$

$$v'(c_1^*(\theta_1))q(\theta_1) = v'(c_1^*(\theta'_1))q(\theta'_1) \quad \text{all } \theta_1, \theta'_1 \in \Theta_1, \quad (8)$$

and

$$\int V_1(\theta_1)q(\theta_1)dF_1(\theta_1) = \kappa. \quad (9)$$

The first two conditions describe the optimal output choices. In both periods, the first-best (FB) output policy equalizes each agent's marginal disutility of labor with the marginal benefit of higher output. The agents' disutility of labor is 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 function (tax revenues). Naturally, the monetary cost of compensating the agents for the extra disutility of labor is decreasing in the agents' marginal utility of consumption.

In the second period, the benefit of asking an agent for higher output simply coincides with the extra resources that are made available when the agent works harder. In the first period, instead, the benefit of asking for higher output also takes into account the effect that the output has on the distribution of the agent's period-2 productivity. Because the period-2 policies are set optimally, usual envelope arguments imply that, in a first-best world, the extra benefit, due to LBD, of asking an agent of period-1 productivity θ_1 for higher period-1 output is given by the function $LD_1^{\chi}(\theta)$ above.

¹⁶Hereafter we will always denote the FB policies with the superscript “*”.

Importantly, note that this function is computed holding fixed the period-2 income and consumption policies, as specified by the allocation rule $\chi = (y, c)$. The expectation in the formula for $LD_1^X(\theta)$ 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 $y_1 = y_1(\theta)$. Note that the term

$$\frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[y_2(\tilde{\theta}) - c_2(\tilde{\theta}) \right]$$

in the definition of $LD_1^X(\theta_1)$ is simply the change in expected period-2 tax revenues stemming from type θ_1 producing more in period one. The term

$$\frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[\frac{v(c_2(\tilde{\theta})) - \psi(y_2(\tilde{\theta}), \tilde{\theta}_2)}{v'(c_2(\tilde{\theta}))} \right],$$

instead, is the reduction in the period-1 compensation $c_1(\theta_1)$ the planner must provide to type θ_1 to hold her expected lifetime utility $V_1(\theta_1)$ constant, when asking her to produce more output in period one. This reduction is possible because, with LBD, the agent expects a higher continuation utility when working harder in period one.

The last three conditions describe the optimal choice of consumption. Optimality requires the equalization of the marginal utility of consumption between any two consecutive histories θ_1 and $\theta = (\theta_1, \theta_2)$, as well as the equalization of the marginal utility of consumption, scaled by the Pareto weights, between any pair of period-1 types, at a level consistent with a binding redistribution constraint.

4 Second-Best Policies

We now turn to the case in which the agents' productivities are their own private information. In this economy, the planner faces additional constraints to her 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

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

denote the period-2 continuation utility of an agent with productivity history equal to $\theta = (\theta_1, \theta_2)$. Period-2 incentive-compatibility requires that, for any $\theta_1 \in \Theta_1$, $V_2(\theta_1, \cdot)$ be Lipschitz continuous and satisfy the familiar Mirrlees envelope formula from static optimal taxation

$$V_2(\theta_1, \theta_2) = V_2(\theta_1, \underline{\theta}_2) - \int_{\underline{\theta}_2}^{\theta_2} \psi_{\theta}(y_2(\theta_1, s), s) ds, \quad (10)$$

along with the requirement that, for any $\theta_2, \hat{\theta}_2 \in \Theta_2$, the following period-2 integral monotonicity condition holds

$$\int_{\hat{\theta}_2}^{\theta_2} \psi_{\theta}(y_2(\theta_1, s), s) ds \leq \int_{\hat{\theta}_2}^{\theta_2} \psi_{\theta}(y_2(\theta_1, \hat{\theta}_2), s) ds.$$

The above period-2 integral monotonicity constraint is equivalent to 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 requires that each agent's expected lifetime utility $V_1(\theta_1)$ be Lipschitz continuous and satisfy an analogous envelope formula given by (see Pavan et al. (2014))

$$V_1(\theta_1) = V_1(\underline{\theta}_1) - \int_{\underline{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1(s), s) ds + \delta \mathbb{E}^{\lambda|s} \left[I_1^2(\tilde{\theta}, y_1(s)) \psi_\theta(y_2(s, \tilde{\theta}_2), \tilde{\theta}_2) \right] \right\} ds \quad (11)$$

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

$$\begin{aligned} & \int_{\hat{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1(s), s) + \delta \mathbb{E}^{\lambda|s, y_1(s)} \left[I_1^2(\tilde{\theta}, y_1(s)) \psi_\theta(y_2(s, \tilde{\theta}_2), \tilde{\theta}_2) \right] \right\} ds \\ & \leq \int_{\hat{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1(\hat{\theta}_1), s) + \delta \mathbb{E}^{\lambda|s, y_1(\hat{\theta}_1)} \left[I_1^2(\tilde{\theta}, y_1(\hat{\theta}_1)) \psi_\theta(y_2(\hat{\theta}_1, \tilde{\theta}_2), \tilde{\theta}_2) \right] \right\} ds. \end{aligned} \quad (12)$$

Hereafter, we follow the same first-order approach as in the rest of the literature by ignoring the monotonicity requirement on the policy $y_2(\theta_1, \cdot)$ and the integral monotonicity constraints in (12) and verify that these properties hold ex-post, once the solution to the planner's relaxed program described below is in hand.

The planner's problem in a second-best world is thus similar to its analog in a first-best world, except for the fact that, holding the labor supply choices fixed, the planner's ability to collect tax revenues is now constrained by the informational rents that must be left to the agents to induce them to reveal their private information. These rents must satisfy conditions (10) and (11). The presence of such constraints is what creates wedges in the second-best allocations, that is, marginal distortions vis-a-vis what is required by first-best efficiency.

Definition 1. The labor wedges are given by

$$W_1(\theta_1) \equiv 1 - \frac{\frac{\psi_y(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))}}{1 + LD_1^X(\theta_1)} \quad \text{and} \quad W_2(\theta) \equiv 1 - \frac{\psi_y(y_2(\theta), \theta_2)}{v'(c_2(\theta))}. \quad (13)$$

Recall that efficiency requires that the marginal disutility of generating extra period- t income be equalized to the marginal benefit, where, in the first period, the latter takes into account also the effect of 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 (which is equal to one) and the corresponding ratio at the proposed allocation. Importantly, in period-1, this 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 Subsubsection 4.1.1, the wedges W_t are related to marginal tax rates, that is, to 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 between the cost and the benefit of asking for higher output under the proposed policies.¹⁷ Hereafter, we will refer to \hat{W}_t as the *relative wedge*.

4.1 Rawlsian-Risk-Neutral Benchmark

As anticipated in the Introduction, to illustrate the key novel effects brought in by LBD in the simplest possible way, we start by considering a textbook economy in which the agents are risk neutral and the planner has extreme preferences for redistribution, in the sense that she assigns a positive Pareto weight only to those individuals with the lowest period-1 productivity. Using (11), it is easy to see that such agents are those for whom the expected lifetime utility is the lowest. Such extreme preferences for redistribution thus amount to a Rawlsian welfare function. With an abuse of notation, we then capture this case by replacing the redistribution constraint (4) with the constraint $V(\underline{\theta}_1) \geq \kappa$.¹⁸ The results in this benchmark economy are instrumental to the understanding of the findings in Subsection 4.2 where we return to the general case.

It is easy to see that, when the agents are risk neutral, the expected tax revenues are equal to

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

Using the IC constraint (11), and integrating by parts, we then have that the planner's objective can be expressed as

$$R = \mathbb{E}^{\lambda[x]} \left[\sum_t \delta^{t-1} \left(y_t(\tilde{\theta}^t) - \psi(y_t(\tilde{\theta}^t), \tilde{\theta}_t) + \frac{I_1^t(\tilde{\theta}, y_1(\tilde{\theta}_1))}{\gamma_1(\tilde{\theta}_1)} \psi_\theta(y_t(\tilde{\theta}^t), \tilde{\theta}_t) \right) - V_1(\underline{\theta}_1) \right], \quad (14)$$

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 . The second-best policies thus maximize (14) subject to the constraint that $V_1(\underline{\theta}_1) \geq \kappa$.

¹⁷For example, when $t = 1$, $\hat{W}_1(\theta_1) = W_1(\theta_1) / \left[\frac{\psi_y(y_1(\theta_1), \theta_1) / v'(c_1(\theta_1))}{1 + LD_1^x(\theta_1)} \right]$.

¹⁸Formally, this problem is not a special case of the problem stated above because of the difference in the redistribution constraint. However, it is easy to see that the solution to this problem is qualitatively similar to the solution to the general problem stated above in which the function q is such that $q(\theta_1) = 1/\varepsilon f_1(\theta_1)$ for all $\theta_1 \in [\underline{\theta}_1, \underline{\theta}_1 + \varepsilon]$ and $q(\theta_1) = 0$ for all $\theta_1 > \underline{\theta}_1 + \varepsilon$, with ε going to zero.

Note that the expectation of the “handicaps”

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

in the tax revenue formula in (14) coincides with the expectation of the information rents the planner must leave to the agents (over and above the utility $V_1(\underline{\theta}_1)$ left to type $\underline{\theta}_1$) to induce them to reveal their private information, where $y(\theta) \equiv (y_1(\theta_1), y_2(\theta_1, \theta_2))$.

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 the handicaps. Differentiating R with respect to $y_1(\theta_1)$ and $y_2(\theta)$, we have that the optimal income policies must satisfy the following optimality conditions

$$\psi_y(y_2(\theta), \theta_2) - \frac{I_1^2(\theta, y_1(\theta_1))}{\gamma_1(\theta_1)}\psi_{\theta y}(y_2(\theta), \theta_2) = 1 \quad (15)$$

and

$$\psi_y(y_1(\theta_1), \theta_1) - \frac{1}{\gamma_1(\theta_1)}\psi_{\theta y}(y_1(\theta_1), \theta_1) - \delta \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[x]|\theta_1, y_1(\theta_1)} \left[\frac{I_1^2(\tilde{\theta}, y_1(\theta_1))}{\gamma_1(\theta_1)}\psi_\theta(y_2(\tilde{\theta}), \tilde{\theta}_2) \right] = 1 + LD_1^X(\theta_1), \quad (16)$$

where, for any period-2 policy $y_2(\cdot)$, the function $LD_1^X(\theta_1)$ is as in (6). 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 (15). 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 $y_2(\theta)$, she 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 mimic 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 mimic type θ_1 in period 1, 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 $-\psi_{\theta y}(y_2(\theta), \theta_2) \partial[1 - F_2(\theta_2|\theta_1, y_1(\theta_1))]/\partial\theta_1$. Using the fact that the impulse response $I_1^2(\theta, y_1)$ of θ_2 to θ_1 is equal to the “extra” probability types just above θ_1 assign to reaching a productivity level above θ_2 in period 2 (normalized by the density $f_2(\theta_2|\theta_1, y_1)$), i.e.,

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

we have that the marginal cost of such an increase in informational rents, evaluated from the perspective of period one, is given by the second term in the left-hand side of (15). This marginal cost 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 . 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 θ is then distorted downwards relative to its first-best level.

Next, consider the optimal choice of period-1 income, as determined by (16). 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 any (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. Furthermore, the marginal costs of asking for a higher period-1 output to those agents of period-1 productivity θ_1 are higher than under symmetric information, as indicated by the second and third terms in the left-hand side of (16). First, when the planner asks for a higher period-1 income from those agents of period-1 productivity θ_1 , she then needs to increase the period-1 consumption of all workers with period-1 productivity above θ_1 . Again, this extra compensation is necessary to guarantee that such higher period-1 types do not mimic type θ_1 . These higher types have a double advantage relative to an agent of period-1 productivity 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 , these types also expect a higher period-2 continuation utility given that, on average, they expect to be more productive than θ_1 also in the second period. To guarantee these period-1 types do not mimic θ_1 , when asking for a higher period-1 income from types θ_1 , the planner must then increase the consumption of all period-1 types above θ_1 by

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

Now observe that¹⁹

$$\begin{aligned} \frac{\partial}{\partial \theta_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[V_2(\tilde{\theta}) \right] &= \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[I_1^2(\tilde{\theta}, y_1(\theta_1)) \frac{\partial V_2(\tilde{\theta})}{\partial \theta_2} \right] = \\ &= -\mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[I_1^2(\tilde{\theta}, y_1(\theta_1)) \psi_{\theta}(y_2(\tilde{\theta}), \tilde{\theta}_2) \right]. \end{aligned} \quad (17)$$

The weight the planner assigns to increasing the rents of all agents with period-1 productivity above θ_1 , relative to the weight she 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 θ_1 , due to asymmetric information, is equal to the sum of the second and third terms in the left-hand side of (16). 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 (note that the impulse response of θ_1 to itself is equal to $I_1^1(\theta, y_1(\theta_1)) = 1$). The interesting novel effects due to LBD are captured by the third term in the left-hand side of (16), which is absent when period-2 productivity is exogenous.

To better understand the last term in the left-hand side of (16), note that, under LBD, asking agents of period-1 productivity 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 workers of period-1 productivity above θ_1 – through two channels. The first is through the change in the distribution of θ_2 , holding fixed the period-2 handicaps $h_2(\theta, y(\theta))$. The second is through the variation in the impulse responses I_1^2 , holding the distribution of period-2 productivity constant. 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.

Such novel effects have important implications for the labor wedges. Let $\hat{W}_1^{NOLBD}(\theta_1)$ and $\hat{W}_2^{NOLBD}(\theta)$ be the relative wedges in the absence of LBD, and

$$\Omega(\theta_1) \equiv -\delta \frac{\frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[\frac{I_1^2(\tilde{\theta}, y_1(\theta_1))}{\gamma_1(\theta_1)} \psi_{\theta}(y_2(\tilde{\theta}), \tilde{\theta}_2) \right]}{\psi_y(y_1(\theta_1), \theta_1)}. \quad (18)$$

Using the optimality conditions (15) and (16), we have that the relative wedges under optimal tax codes are given by (with $\lambda[\chi]$ -probability one)

$$\hat{W}_1(\theta_1) = \hat{W}_1^{NOLBD}(\theta_1) + \Omega(\theta_1) \quad (19)$$

and $\hat{W}_2(\theta) = \hat{W}_2^{NOLBD}(\theta)$ where

$$\hat{W}_t^{NOLBD}(\theta) = -\frac{I_1^t(\theta^t, y_{t-1}(\theta^{t-1})) \psi_{\theta_y}(y_t(\theta^t), \theta_t)}{\gamma_1(\theta_1) \psi_y(y_t(\theta^t), \theta_t)}. \quad (20)$$

¹⁹The first equality in (17) follows from the fact that, given any Lipschitz continuous function $J(\theta_2)$, and any kernel $F_2(\theta_2|\theta_1)$, $\frac{\partial}{\partial \theta_1} \mathbb{E} \left[J(\tilde{\theta}_2) | \theta_1 \right] = \mathbb{E} \left[I_1^2(\tilde{\theta}) \frac{\partial J(\tilde{\theta}_2)}{\partial \theta_2} | \theta_1 \right]$. The second equality follows from the fact that IC requires that $\partial V_2(\theta_1, \theta_2) / \partial \theta_2 = -\psi_{y\theta}(y_2(\theta_1, \theta_2), \theta_2)$.

The terms $\hat{W}_t^{NOLBD}(\theta)$ are the period- t relative wedge when the process over θ is exogenous and equal to the one under χ . The term $\Omega(\theta_1)$, instead, summarizes all the novel effects due to LBD.²⁰ This new 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.

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),

$$\hat{W}_t^{NOLBD}(\theta^t) = \rho^{t-1} \frac{1 + \phi}{\theta_1 \gamma_1(\theta_1)}, \quad (21)$$

and

$$\Omega(\theta_1) = \frac{\delta \rho}{\psi_y(y_1(\theta_1), \theta_1)} \hat{W}_1^{NOLBD}(\theta_1) \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda|\chi||\theta_1, y_1(\theta_1)} \left[\psi(y_2(\tilde{\theta}), \tilde{\theta}_2) \right], \quad (22)$$

where the parameter ρ measures the exogenous persistence in the agents' productivity and the parameter ϕ measures the inverse Frisch elasticity.

Under this specification, the impulse responses $I_1^t(\theta, y_1)$ 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 $\psi(y_2(\theta), \theta_2)$. Moreover, $\psi(y_2(\tilde{\theta}), \tilde{\theta}_2)$ is nondecreasing in period-2 productivity θ_2 . As a result, $\Omega(\theta_1) > 0$, 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 planner reduces the agents' expected continuation rents.

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

Finally, consider the effects of LBD on the progressivity of the wedges. Observe that, under the assumed specification, $\hat{W}_t^{NOLBD}(\theta^t)$ are nonincreasing in productivity if, and only if, $\theta_1 \gamma_1(\theta_1)$ is nondecreasing in θ_1 . In the literature, this property is believed to hold at the upper tail of the distribution. In the absence of LBD, the literature then predicts wedges to be nonincreasing in productivity at the top. LBD can contribute positively or negatively to the progressivity of the period-1 wedges depending on whether Ω is increasing or decreasing in θ_1 . 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 all productivity levels. Under this specification, Ω 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 the left-hand panel of Figure 1 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,

²⁰The formal proof of the claim that relative wedges under optimal codes are consistent with the decomposition in (19) is in the Appendix (see the proof of Proposition 5).

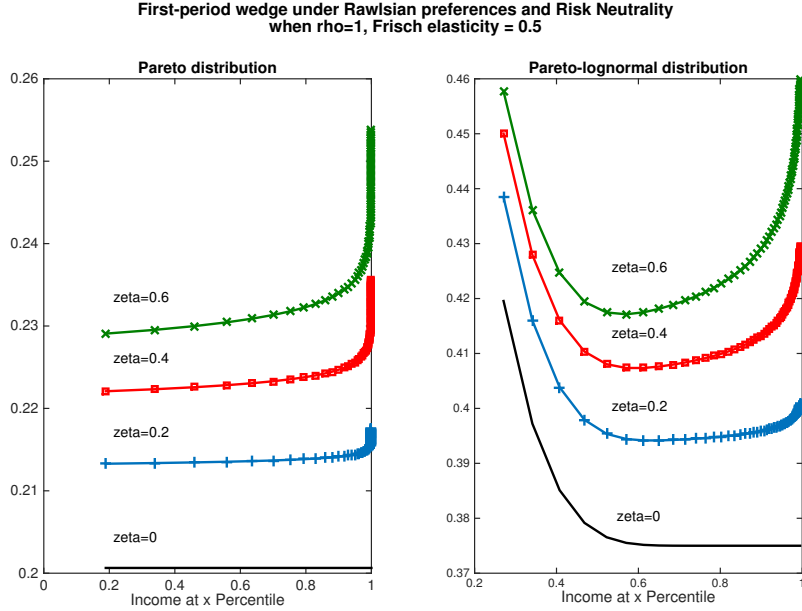


Figure 1: Period-1 wedges in the risk-neutral Rawlsian case

stronger LBD effects (captured by a higher level of the parameter ζ in (2)) are responsible for higher period-1 wedges and for more progressivity at all income percentiles, but in particular at the top.²¹

These properties extend to other distributions of the skill shocks. For instance, the right-hand panel of Figure 1 illustrates the period-1 wedge function for a Pareto-lognormal skills distribution F_1 with Pareto-tail parameter equal to $\xi = 5$, as in Diamond (1998). As the figure shows, LBD contributes to higher and more progressive wedges. However, contrary to the Pareto case depicted in the left-hand panel of 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, whose proof is in the online Supplementary Material. We first need to introduce some definitions and notation.

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 $\hat{W}_1^{NOLBD}(\cdot)$ over (θ'_1, θ''_1) . The period-1 wedge under LBD is more progressive than the period-1 wedge in the absence of LBD if, and only if, $\hat{W}_1(\cdot)$ is weakly steeper than $\hat{W}_1^{NOLBD}(\cdot)$ over the entire set Θ_1 , and strictly steeper over a subset $(\theta'_1, \theta''_1) \subset \Theta_1$.

²¹The 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 figure assumes $\phi = 2$, i.e., a Frisch elasticity of 0.5, as in Farhi and Werning (2013), Kapicka (2013), and Stantcheva (2017). Finally, the parameter $\rho = 1$ in the figure's caption refers to the exogenous skill persistence parameter. The assumption that $\rho = 1$ is made to facilitate the comparison to Farhi and Werning (2013), Kapicka (2013), Golosov et al. (2016), and Stantcheva (2017).

Using the decomposition in (19), 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 $\theta_1 \in \Theta_1$, $\hat{W}_1(\theta_1) > \hat{W}_1^{NOLBD}(\theta_1)$; (ii) For all $\theta = (\theta_1, \theta_2)$, $\hat{W}_1(\theta_1) - \hat{W}_2(\theta) > \hat{W}_1^{NOLBD}(\theta_1) - \hat{W}_2^{NOLBD}(\theta)$; (iii) Suppose F_1 is Pareto (in which case there exists $M \in \mathbb{R}_{++}$ such that $\theta_1 \gamma_1(\theta_1) = M$ for all θ_1). (a) In the absence of LBD, $\hat{W}_1^{NOLBD}(\theta_1) = (1 + \phi)/M$, 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 $\hat{W}_1^{NOLBD}(\theta_1)$ is nonincreasing; the solution to the relaxed program also solves the full program.*

Hence, when the disutility of labor is iso-elastic and period-2 productivity is given by the Cobb-Douglas specification in (2), LBD contributes to higher period-1 wedges across all productivity levels and to a higher differential between period-1 and period-2 wedges, across all histories $\theta = (\theta_1, \theta_2)$. Whether LBD also contributes to a higher progressivity of the period-1 wedges depends on the distribution from which the productivity shocks are drawn. When this distribution is Pareto, LBD contributes to more progressivity at all income percentiles. The proof of Proposition 2 in the Supplementary Material identifies a condition under which the same conclusion extends to other distributions.

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.

4.1.1 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.²² The purpose of this section is twofold. First, it helps us to relate 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 earnings distribution.

Let $\mathcal{T} = (\mathcal{T}_1, \mathcal{T}_2)$ be a generic tax code, with $\mathcal{T}_t(y^t)$ denoting the total period- t tax payment by an individual with period- t income history y^t . Given \mathcal{T} , for any y^t , 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 .

²²See also Golosov et al. (2014) and Kapicka (2015) for a variational approach in the context of a dynamic economy.

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 Supplementary Material)

$$W_1(\theta_1) = \frac{\tau_1(y_1(\theta_1)) + \delta \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1(\theta_1), y_2(\theta)) dF_2(\theta_2 | \theta_1, y_1(\theta_1))}{1 + \delta \frac{\partial}{\partial y_1} \int [y_2(\theta) - \psi(y_2(\theta), \theta_2)] dF_2(\theta_2 | \theta_1, y_1(\theta_1))} \quad (23)$$

and

$$W_2(\theta) = \tau_2(y_1(\theta_1), y_2(\theta)). \quad (24)$$

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

$$\hat{W}_1(\theta_1) = \frac{1}{\psi_{y_1}(y_1(\theta_1), \theta_1)} \left[\tau_1(y_1(\theta_1)) + \delta \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1(\theta_1), y_2(\theta)) dF_2(\theta_2 | \theta_1, y_1(\theta_1)) \right] \quad (25)$$

and

$$\hat{W}_2(\theta) = \frac{\tau_2(y_1(\theta_1), y_2(\theta))}{1 - \tau_2(y_1(\theta_1), y_2(\theta))}. \quad (26)$$

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 to period-1 incomes.

The representation in (25) and (26) 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 ones 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 planner's objective.

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

$$e_{\tau_2|y_1} \equiv \left. \frac{\partial \mathbb{E}[\tilde{\mathcal{T}}_2|y_1]}{\partial y_1} \right|_{\theta_1(y_1)=\text{constant}} \frac{y_1}{\mathbb{E}[\tilde{\mathcal{T}}_2|y_1]}$$

denote the elasticity of $\mathbb{E}[\tilde{\mathcal{T}}_2|y_1]$ with respect to period-1 income y_1 , holding $\theta_1(y_1)$ constant. 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:*

$$\frac{\tau_1(y_1)}{1 - \tau_1(y_1)} = \frac{1 - H_Y(y_1)}{y_1 \hat{h}_Y(y_1)} \frac{1}{\hat{E}_1(y_1)} \left[\frac{1}{1 + \delta e_{\tau_2|y_1} \frac{\mathbb{E}[\tilde{\mathcal{T}}_2|y_1]}{\tau_1(y_1) y_1}} \right]. \quad (27)$$

The formula in (27) generalizes the formula in Saez (2001) by accounting for the effect of a change in period-1 income on the expected period-2 tax payments. Importantly, the formula in (27) 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 a test requires 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 (27) has the advantage 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 complement each other.

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 productivity of the period-1 agents at the level $\theta_1 = \theta_1(y_1)$. 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 $d\tau_2$ in 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 $y'_1 \in [y_1, y_1 + dy_1)$. This perturbation is more sophisticated than the one leading to the formula for the optimal period-1 tax rates in (27). 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 by an

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 net present value of intertemporal tax revenues must be equal to zero. Now, paralleling the analysis for period one, 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 with period-2 productivity $\theta_2(y_1, y_2)$, 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:*

$$\frac{\tau_2(y_1, y_2)}{1 - \tau_2(y_1, y_2)} = \left[\frac{\partial \tilde{H}_O(y_2|y_1)}{\partial y_1} - \frac{\partial H_O(y_2|y_1)}{\partial y_1} \right] \frac{1 - H_Y(y_1)}{h_Y(y_1)y_2\hat{h}_O(y_2|y_1)} \frac{1}{\hat{E}_2(y_1, y_2)}. \quad (28)$$

Once again, the formula in (28) complements the one derived in the previous section by relating marginal tax rates to the empirical income distribution in both periods. In the online Supplementary Material, we also verify that the formulas for 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.

4.2 General Case

We now return to the general case where the agents may be risk averse with preferences for consumption smoothing, and where the planner may assign different non-linear Pareto weights to different period-1 types according to the general function $q(\theta_1)$ described above.

Proposition 5. *The relative wedges under the optimal tax code are given by (with $\lambda[\chi]$ -probability one)*

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

and $\hat{W}_2(\theta) = \hat{W}_2^{NOLBD}(\theta)$. The terms $\hat{W}_1^{NOLBD}(\theta_1)$ and $\hat{W}_2^{NOLBD}(\theta)$ are the relative wedges in the absence of LBD, $\Omega(\theta_1)$ is the function defined in (18) capturing the effects of LBD in the benchmark economy with risk-neutral agents and Rawlsian preferences for redistribution,

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

is a correction term due to risk aversion, and

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

is a correction term reflecting the benefit the planner assigns to increasing the expected lifetime utility of those agents with period-1 productivity above θ_1 .

What is interesting about this result is how risk aversion and the planner's preferences for redistribution (as captured by the non-linear Pareto weights $q(\theta_1)$) interact with 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 to ensure incentive-compatibility. The correction term $RA(\theta_1)$ represents the extra cost to the planner, in consumption terms (equivalently, in tax revenue terms), of providing this 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 captures the marginal variation in expected discounted future consumption triggered by a marginal change in period-1 output, with the variation originating in the endogeneity of the period-2 productivity distribution. The term $\Omega(\theta_1)v'(c_1(\theta_1))$ translates such variation into utility terms. The term

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

thus represents the average cost (in consumption units) of increasing the utility of all workers with period-1 productivity above θ_1 by $v'(c_1(\theta_1))\Omega(\theta_1)$, accounting for the heterogeneity in such agents' marginal utility of consumption.

Next, consider the term $D(\theta_1)$. This extra correction term controls for the higher Pareto weights the planner assigns to all agents whose initial productivity is above θ_1 , relative to the benchmark with Rawlsian preferences for redistribution. Other things equal, this term naturally contributes to lower wedges. When the planner assigns strictly positive Pareto weights to each agent whose period-1 productivity exceeds θ_1 (as in the case of a planner with Utilitarian preferences for redistribution considered in the next section), increasing the lifetime utility of each type $\theta'_1 > \theta_1$ by $v'(c_1(\theta_1))\Omega(\theta_1)$, as required by incentive compatibility, comes with the benefit of relaxing the redistribution constraint

(4). The benefit for the planner in revenue terms is equal to²³

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

In the Supplement to this article, we discuss in detail how the correction term $RA - D$ interacts with the uncorrected LBD term Ω in the determination of the period-1 wedges. We show that the term $RA - D$ can be increasing in the agents' degree of risk aversion. This is because higher degrees of risk aversion imply, other things equal, higher costs to compensate the agents for higher effort and thereby higher rents to dissuade them from mimicking other types. To reduce such rents, the planner then asks the agents to provide lower period-2 output. In turn, this means that the benefit of shifting the period-2 distribution towards lower productivity levels so as to economize on period-2 rents are lower the higher the agents' risk aversion. As a result, the uncorrected LBD term Ω tends to be decreasing in the agents' risk aversion. Whether higher degrees of risk aversion contribute to higher or lower period-1 wedges then depends on whether the effects on the correction term $RA - D$ dominate over those on the uncorrected term Ω . In the online Supplementary Material, we also discuss how the terms $RA - D$ and Ω are affected by variations in (a) the Frisch elasticity of the agents' labor supply, as captured by the term $1/\phi$ in the agent's disutility of labor, and (b) the planner's preferences for redistribution, as captured by the function $q(\cdot)$.

We conclude this section by highlighting that the progressivity of the wedges depends critically on the shape of the skills distribution.²⁴ To illustrate this point, Figures 2 and 3 depict the period-1 wedges under Utilitarian preferences for redistribution for four different levels of the coefficient η of relative risk aversion (namely for $\eta = 0$, $\eta = 0.2$, $\eta = 0.5$, and $\eta = 0.8$) and for four different levels of the LBD intensity (namely, for $\zeta = 0$, $\zeta = 0.2$, $\zeta = 0.4$, and $\zeta = 0.6$). Figure 2 corresponds to the case where the period-1 productivity θ_1 and the period-2 shocks ε_2 are drawn from the Pareto-lognormal distribution introduced above. Figure 3, instead, corresponds to the case where θ_1 and ε_2 are drawn from the Lognormal distribution in Farhi and Werning (2013).²⁵ When the agents are risk averse, wedges are progressive across all income percentiles in the Pareto-lognormal case, but regressive at top percentiles in the Lognormal case. Furthermore, in the Pareto-Lognormal case, an increase in the intensity of LBD increases the progressivity of the period-1 wedges, whereas the opposite is true

²³Observe that $v'(c_1(\theta_1))\Omega(\theta_1) \int_{\theta_1}^{\bar{\theta}_1} q(s) \frac{dF_1(s)}{1-F_1(\theta_1)}$ is the weighted-average benefit of increasing by $v'(c_1(\theta_1))\Omega(\theta_1)$ the expected lifetime utility of each period-1 type $\theta'_1 > \theta_1$, where the average accounts for the different Pareto weights assigned by the planner to types above θ_1 . The term $\int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} dF_1(s)$, instead, is the shadow value, in revenue terms, of increasing all agents' expected lifetime utility uniformly. To understand this last term, observe that, if the planner were to increase by one unit the expected lifetime utility of all period-1 types, then incentive compatibility would be preserved and the cost to the planner in terms of ex-ante revenues would be equal to $\int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} dF_1(s)$. The product of the two terms $\left[v'(c_1(\theta_1))\Omega(\theta_1) \int_{\theta_1}^{\bar{\theta}_1} q(s) \frac{dF_1(s)}{1-F_1(\theta_1)} \right]$ and $\int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} dF_1(s)$ thus represents the weighted-average benefit, in revenue terms, of increasing by $v'(c_1(\theta_1))\Omega(\theta_1)$ the expected lifetime utility of all period-1 types above θ_1 .

²⁴For a related discussion, see also Golosov et al. (2016).

²⁵Specifically, the productivity shock is drawn from a Lognormal distribution with mean one and variance parameter $\sqrt{0.0095}$. The latter is the middle value of the three values considered in Farhi and Werning (2013).

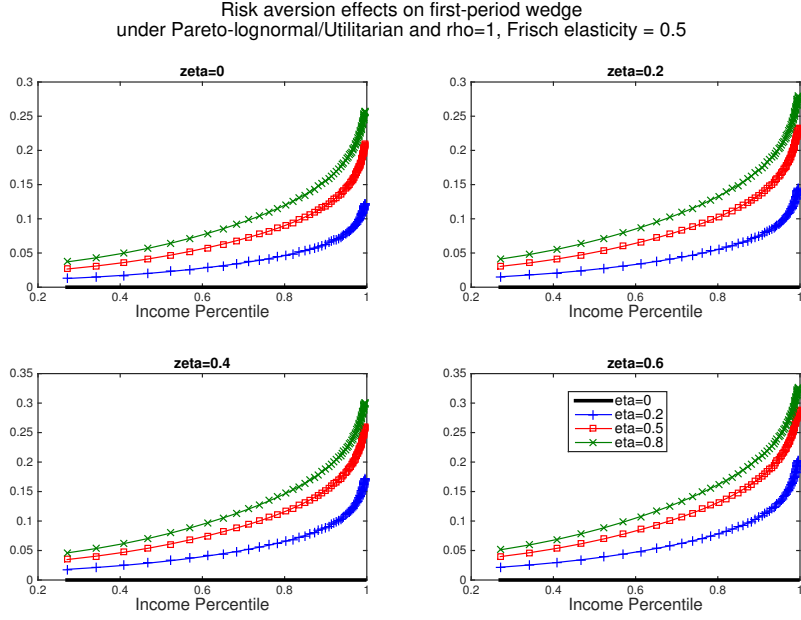


Figure 2: Period-1 wedges in the Utilitarian Pareto-lognormal case

in the Lognormal case, as can be seen from Figure 4. These observations will be important for the interpretation of the quantitative results in the next section.²⁶

5 Quantitative Analysis

We now turn to the quantitative implications of our analysis.²⁷ In Subsection 5.1, we calibrate the model to match various moments of the US earnings distribution under the existing US tax code. In Subsection 5.2, we derive the optimal wedges for the calibrated economy. In Subsection 5.3, we show that most of the welfare gain from the optimal reform of the existing tax code can be generated through simple (history-independent) taxes. In Subsection 5.4, we conduct some comparative statics of optimal tax codes with respect to alternative specifications of the intensity of the LBD effects and the exogenous persistence of the agents' productivity. In Subsection 5.5, we illustrate the importance of the stochasticity of the LBD effects for the structure of optimal taxes. Finally, in Subsection 5.6, we isolate the role of LBD by conducting a counterfactual analysis where we compare the optimal tax code in the calibrated economy with LBD to the optimal tax code in an economy that features the same productivity process as in the calibrated economy but assumes the latter is exogenous.

²⁶As is well known, under a Utilitarian welfare objective, in the absence of LBD, irrespective of the shock distribution, wedges are identically equal to zero when $\eta = 0$, i.e., when agents are risk neutral. The same remains true with LBD.

²⁷Details on the computations can be found in the online Supplementary Material.

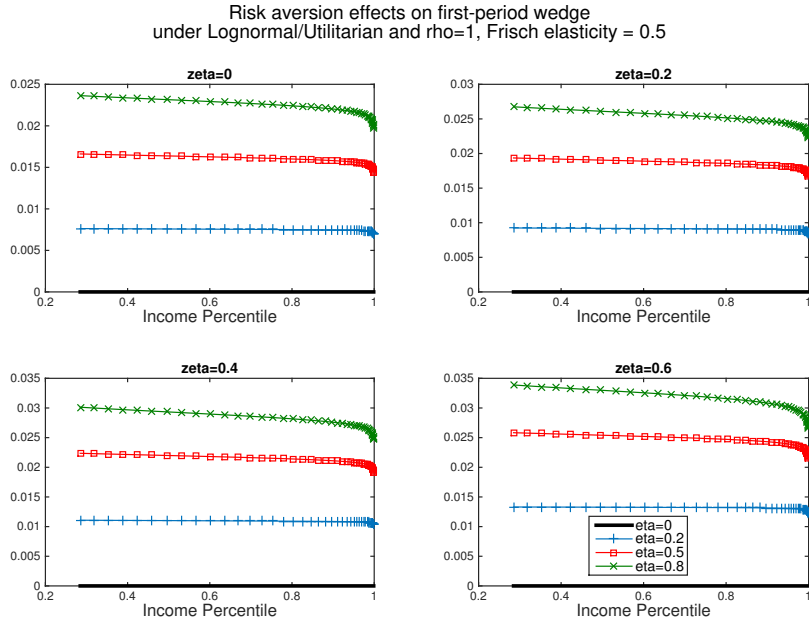


Figure 3: Period-1 wedges in the Utilitarian Lognormal case

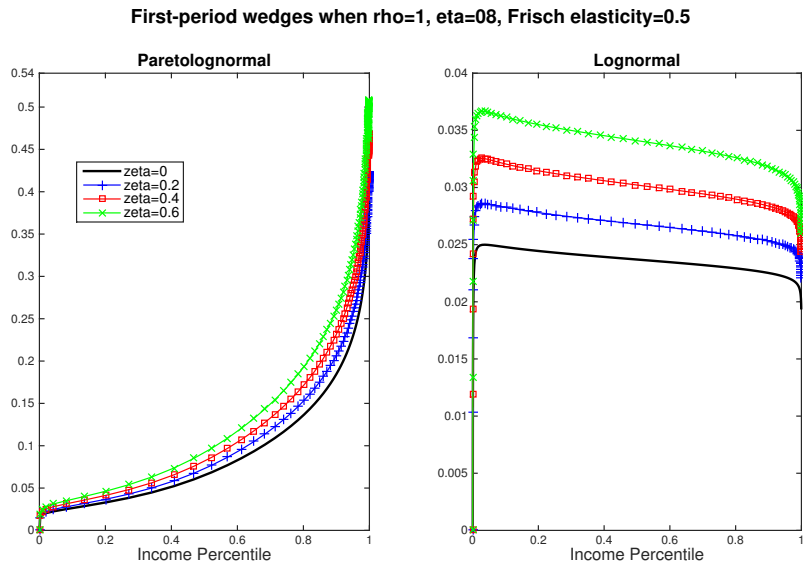


Figure 4: Period-1 wedges in the Utilitarian case under Pareto-lognormal and Log-normal distributions

5.1 Calibration

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, 2017). 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 effects of $(y_s)_{s=1}^{20}$ on θ_2 are summarized by a weighted average of the income levels in the first twenty years, with the weights $\hat{\beta}_s / \sum_{r=1}^{20} \hat{\beta}_r$ $s = 1, \dots, 20$, declining over time, and with each $\hat{\beta}_s$ being a positive scalar. This specification is thus consistent with the empirical evidence that LBD in earlier periods has more pronounced effects on wages and productivity than in later periods, and that the effects of LBD on future productivity eventually fade away. See, for instance, Dustmann and Meghir (2005), Levitt et al. (2013), and Thompson (2012).

Interestingly, 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 is 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-year blocks. Moreover, the allocations in this economy coincide with the corresponding ones in the two-period model of the previous sections, after setting $\delta = \beta^{20}$. 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. 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.²⁸

To calibrate the model, we first fix a few parameters to the levels typically assumed in the literature (see Table 1).²⁹ As in the rest of the literature, we assume that the disutility of labor takes the 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 and where ε_1 is a random variable described below. Period-2 productivity, instead, is given by

$$\theta_2 = \theta_1^\rho \left(\frac{\sum_{s=1}^{20} \beta^{s-1} y_s}{\sum_{s=1}^{20} \beta^{s-1}} \right)^\zeta \varepsilon_2, \quad (29)$$

²⁸See the online Supplementary Material for a formal proof of the equivalence between the two economies.

²⁹The 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 (2017). 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 assumption in almost all these papers that the annual discount factor is equal to the inverse of the annual gross interest rate, while being somewhere in the middle of the range of values in the above papers, with the exception of BK.

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

where ε_2 is a random variable. The specification in (29) is the same as in (2), 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 assume the productivity shocks ε_1 and ε_2 are i.i.d. draws from a Pareto-lognormal distribution with parameters (μ, σ^2, ξ) . The parameter ξ governs the Pareto right tail of the distribution, whereas the parameters (μ, σ^2) govern the moments of the distribution for the given tail parameter ξ . We truncate the distribution at the 1st percentile and set μ so that the mean of the truncated distribution is equal to one.³⁰

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

$$T(y) = y - e^{\tau_0} y^{1-0.181} \quad (30)$$

of the existing US income tax code from Heathcote et. al. (2017), with the parameter $\tau_0 = -0.1005$ set so that the total tax revenues are normalized to zero.³¹ Given the above tax code, workers maximize their expected lifetime utility by choosing income and consumption in each period, taking as given the exogenous net interest rate on their savings, and accounting for the effects of LBD on the evolution of their productivity.³² The parameters $(h_1, \rho, \zeta, \sigma, \xi)$ 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). 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

³⁰A Pareto-lognormal distribution G has support $(0, \infty)$, density $g(\varepsilon) = \frac{\xi}{\varepsilon^{1+\xi}} \exp(\xi\mu + \xi^2 \frac{\sigma^2}{2}) \Phi(\frac{\log(\varepsilon) - \mu - \xi\sigma^2}{\sigma})$, and cdf $G(\varepsilon) = \Phi(\frac{\log(\varepsilon) - \mu}{\sigma}) - \frac{1}{\varepsilon^\xi} \exp(\xi\mu + \xi^2 \frac{\sigma^2}{2}) \Phi(\frac{\log(\varepsilon) - \mu - \xi\sigma^2}{\sigma})$, 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 ε but has a Pareto right tail.

³¹Golosov et. al. (2016) use a similar estimation, but from the 2014 version of the Heathcote et. al. paper. Namely, they assume that $T(y) = y - e^{\tau_0} y^{1-0.151}$.

³²The annual interest rate can be thought of as net of any (exogenous) linear capital tax rate.

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%
ξ	5.9907	mean-to-median earnings young	1.335	1.25%

Table 2: Calibrated Parameters

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.³³ Table 2 reports the calibrated parameters, the target moments, and the absolute percentage deviations of the model-generated moments from the target moments.³⁴ As anticipated in the Introduction, the calibrated value for the LBD parameter ζ is consistent with both the estimated range [0.2, 0.6] in the metadata analysis of Best and Kleven (2013), and the estimated range [0.004, 0.353] of Chang et al. (2002).

5.2 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.³⁵ We do so by solving the primal to the planner's problem as described in Section 2. That is, we identify the allocations that maximize the agents' ex-ante expected utility subject to the constraint that expected tax revenues be non-negative. 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

³³Contrary to our paper, Kapicka and Neira (2016) separate the data in Huggett et. al. (2011) by using year 20 instead of year 21 as the cutoff year. However, the mean earnings reported in Panel A of Figure 1 in Huggett et. al. (2011) for year 20 and year 21 are virtually identical, so the distinction is not quantitatively relevant.

³⁴As a robustness check, we verified that the earning distribution under the existing US tax code in the calibrated economy is broadly consistent with the empirical distribution also in terms of the following two non-targeted moments: (a) the Gini coefficient of the old workers' unconditional earning distribution, and (b) the mean-to-median ratio in the unconditional earnings distribution of old workers. Specifically, the absolute percentage deviations of the aforementioned two model-generated moments from their empirical counterparts are equal to 9.16% and 5.74%, respectively.

³⁵The assumption that the planner has Utilitarian preferences for redistribution eases the comparison with the pertinent quantitative literature. The same preferences are assumed in the numerical analysis in Farhi and Werning (2013), Kapicka (2013), Golosov et al (2016), Kapicka and Neira (2016), and Stantcheva (2017). 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.

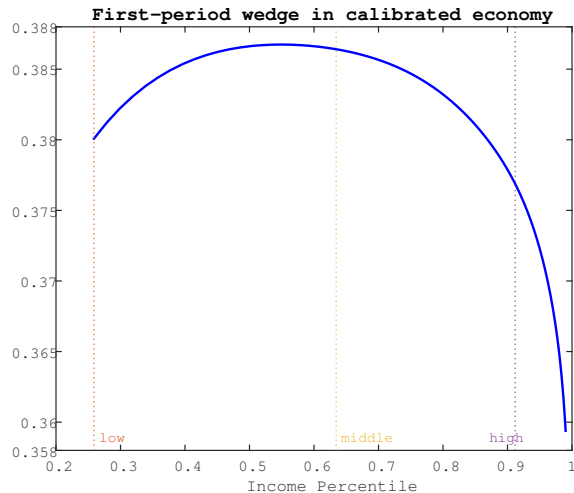


Figure 5: Period-1 optimal wedge. The vertical lines indicate the period-1 income percentiles corresponding to low, middle and high earnings.

relaxed program satisfies all the remaining incentive-compatibility constraints. The verification is done by checking that (i) period-2 earnings are nondecreasing in period-2 shock/productivity, for any given level of period-1 productivity, and (ii) all the integral monotonicity conditions in (12) are satisfied.

We first report that reforming the existing tax code by adopting the optimal one would yield an increase in the agents' ex-ante expected utility equal to that generated by a 4.0409% equiproportionate increase in consumption at each history, starting from the allocations under the existing US tax code, while keeping output choices constant.³⁶ This appears to be a significant improvement, although it is lower than the figure that one would obtain by assuming no LBD effects (see below).

We now 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 5 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 99.9th 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 a distribution, wedges have an inverted U-shape (see Golosov et al. (2016) as well as the discussion at the end of the previous section).

³⁶This number is calculated as follows. Observe that, because $v(c) = \ln(c)$, if consumption at each history increases by $x\%$, then lifetime utility increases by $\Delta V = [1 + \delta] \log(1 + x)$, where $\delta = \beta^{20}$. The welfare gains brought about by changing the code from the current one to the optimal one are $\Delta V = 0.057693$. Therefore, when translated in consumption terms, it is as if the reform yields an equi-proportionate increase in consumption at every history equal to $(e^{\Delta V})^{\frac{1}{1+\delta}} - 1 = 0.040409$, that is 4.0409%.

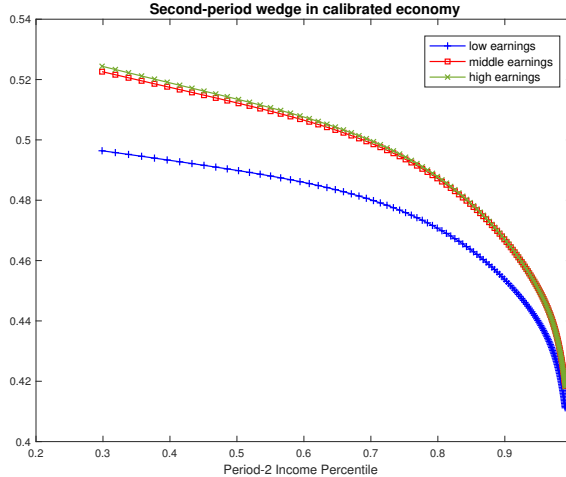


Figure 6: Period-2 wedges as a function of period-2 earnings percentiles for low, middle and high period-1 earnings.

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 mean earnings. The low (alternatively, high) earnings correspond to earnings approximately equal to half (alternatively, twice) the mean of the period-1 earnings.

Figure 6 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. Period-2 wedges also have an inverse-U shape. However, contrary to their period-1 counterparts, period-2 wedges decrease after the 30th percentile. Period-1 wedges thus exhibit progressivity over a range of income percentiles for which the period-2 wedges are regressive (e.g., between the 30th and 60th percentile). Also note that, at any percentile of the period-2 distribution, period-2 wedges are increasing in period-1 incomes, across the selected histories described above.

Next, we discuss the dynamics of the wedges. Figure 7 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 period-2 earnings $y_2(\theta)$ coincides with the conditional period-2 mean earnings given the corresponding period-1 income percentile (i.e., for each period-1 income percentile $x \in [0, 1]$, $y_2(\theta) = \mathbb{E}[y_2(\theta_1, \tilde{\theta}_2) | \theta_1, y_1(\theta_1)]$, with $F_1(\theta_1) = x$), and (b) the conditional period-2 income percentile coincides with the period-1 percentile (equivalently, the 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. To reduce the volatility of the period-2 consumption choices, the planner distorts downwards the agents' period-2 income choices more than it does with their period-1 income choices. This effect is not specific to economies with LBD (See for example Farhi and Werning, 2013, and Golosov et al., 2016). What is interesting is that this effect remains dominant even in the

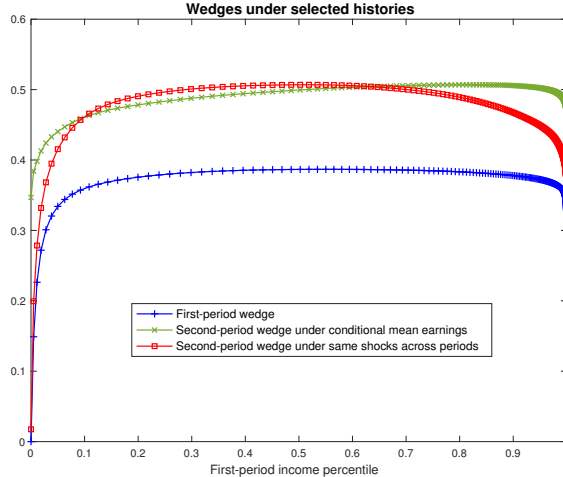


Figure 7: Optimal wedges for selected histories

	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

presence of LBD.

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. For any such history, wedges increase with age, reflecting again the pattern documented in the previous figure.

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 8 (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.

The property that wedges tend to increase over the lifecycle is consistent with the results in Farhi and Werning (2013) and Golosov et al. (2016). The reason why, in our calibrated economy, wedges increase over the lifecycle, is that agents are highly risk averse (the coefficient of relative risk aversion is $\eta = 1$) and are exposed to significant risk (the variance of the period-2 productivity shock is 0.35).

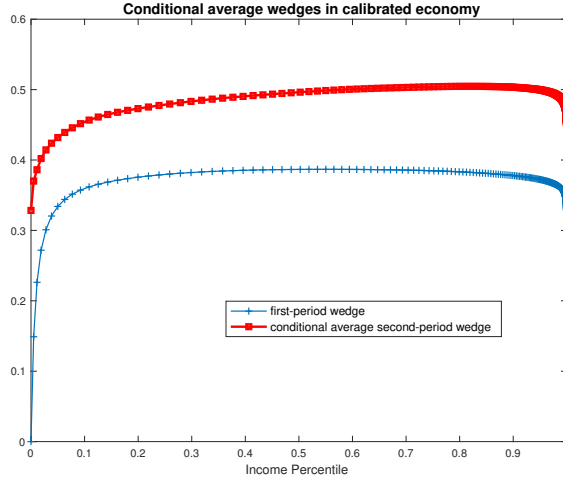


Figure 8: Period-1 wedges and conditional period-2 wedges as a function of period-1 income percentile.

It is known that, in such circumstances, in the absence of LBD, wedges increase over the lifecycle. Our results show that, when the intensity of the LBD effects are of the magnitude of the calibrated economy, the same dynamics obtain in the presence of LBD, but the level of the wedges is higher.³⁷ This property is consistent with the results in the previous sections that LBD tends to contribute to higher wedges.³⁸

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.³⁹

5.3 Simple (History-Independent) Taxes

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

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

³⁷In each period, the average wedges in our calibrated economy are higher than the corresponding averages in the first and in the second halves of each agent's working life in the Lognormal simulations in the aforementioned two papers. Golosov et al. (2016), however, report the wedges only for top earners.

³⁸See also the discussion in Subsection 5.5 about the role of the stochasticity of the LBD effects for the dynamics of optimal taxes.

³⁹Note 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).

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 current US tax code, as estimated in Heathcote et. al. (2017), corresponds to the case $B = 0$, $\tau_{0,1} = \tau_{0,2}$ 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 expected lifetime utility of workers, subject to the constraint that total tax revenues be at least zero (recall that average tax revenues are equal to zero in the benchmark economy, i.e., under the tax code in Heathcote et. al. (2017) that best approximates the current US tax code). We refer to the solution to this problem as the quasi-optimal tax code.

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 (the one implementing the second-best allocations associated with the fully optimal wedges discussed above). In particular, while adopting the fully optimal tax code yields an increase in expected lifetime utility equal to an equiproportionate increase in consumption of 4.0409% at all histories, starting from the allocation under the current US tax code, adopting the quasi-optimal tax code yields an increase in expected lifetime utility equal to a 3.8859% equiproportionate increase in consumption. The loss from simple taxes is thus only 0.155% in consumption terms.⁴⁰ Therefore, the quasi-optimal tax code is approximately optimal. Importantly, as with the optimal wedges, marginal tax rates increase over the lifecycle, contrary to what is found in Best and Kleven (2013) and Kapicka (2015a,b). As anticipated in the Introduction, the difference originates in the stochasticity of the LBD effects – see also the discussion in Subsection 5.5.

Importantly, 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.⁴¹ Adopting the optimal linear tax code yields an increase in expected lifetime utility equal to a 3.8842% equi-proportionate increase in consumption at all histories and periods, starting from the allocations under the current US tax code, which is only 0.1567% less than the increase under the fully optimal tax code. These

⁴⁰The 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, it is as if the reform yields an equiproportionate increase in consumption at every period and history equal to $(e^{\Delta V})^{\frac{1}{1+\delta}} - 1 = 0.038859$, that is 3.8859%.

⁴¹The welfare gains from replacing the current tax code with the optimal linear one are $\Delta V = 0.055499$. When translated into consumption terms, the optimal linear tax code thus yields an equiproportionate increase in consumption at all histories, starting from the allocations under the current US tax code, equal to $x = (e^{\Delta V})^{\frac{1}{1+\beta^{20}}} - 1 = 0.038842$, that is, 3.8842%.

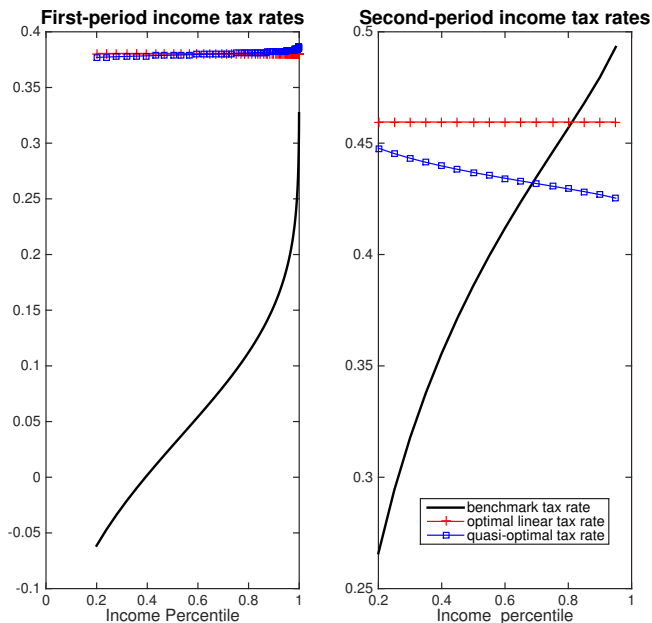


Figure 9: Tax rates as functions of income

results thus indicate that simple linear, but age-dependent, taxes are approximately optimal.⁴² The marginal tax rates in the optimal linear tax code are equal to 38% and 46% for young and old workers, respectively.⁴³

In Figure 9, we plot together the marginal income tax rates as functions of (unconditional) income percentiles, under (a) the existing US tax code (solid line), (b) the quasi optimal tax code (squared line), and (c) the optimal linear tax code (crossed line). The tax rates under the quasi-optimal and linear tax codes are increasing with age, and higher than the ones in the current US tax code for young workers. For old workers, the 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 optimal linear tax code are very close to those in the quasi-optimal tax code.

Next, suppose the planner is constrained to using 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 tax 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 US tax code with the optimal age-independent one are equal to a 3.7988% equiproportionate increase in consumption at all histories. Therefore, the gains from using the quasi-optimal tax code instead of the optimal age-independent one are equal to a 0.087% equiproportionate increase in

⁴²The optimal linear tax code is given by $B = 0.2599$, $1 - e^{\tau_{0,1}} = 0.3806$, and $1 - e^{\tau_{0,2}} = 0.4595$.

⁴³That optimal linear age-dependent taxes may generate most of the welfare gains from reforming the existing US tax code is consistent with the findings in Farhi and Werning (2013) and Stantcheva (2017).

consumption at all histories, starting from the allocations under the current US tax code. Note that if the planner, in addition to being constrained to use age-independent taxes was also constrained to use a tax code without lump-sum transfers (formally captured here by the restriction to $B = 0$), the optimal tax code in this class would feature $\tau = 0.3846$. Such a code is significantly more progressive than the current US tax code, as estimated in Heathcote et. al. (2017).

Finally, suppose the planner is constrained to using a code with age-independent and linear taxes. The optimal code in this class is given by $B = 0.2631$ and $\tau_0 = -0.5318$, which implies a constant marginal tax rate of 41.25%. The welfare gains in consumption terms from replacing the current US tax code with the optimal age-independent linear code are equal to a 3.7987% equiproportionate increase in consumption at all histories, starting from the allocations under the current US tax code. Thus, the benefits from using the quasi-optimal tax code instead of the optimal age-independent linear one are equal to a 0.0871% equiproportionate increase in consumption at all histories, starting from the allocations under the current US tax code.

The overall lesson we draw from the above results is that most of the welfare gains from reforming the current US tax code can be generated with simple taxes.

5.4 Comparative Statics

Next we provide comparative statics of the quasi-optimal taxes with respect to the intensity of the LBD effects and skill persistence. We focus here on quasi-optimal taxes because in all the cases considered below, we find that the welfare losses from reforming the existing US tax code by adopting the quasi-optimal taxes instead of the fully optimal ones (i.e., those inducing the second-best allocations) are very small (precisely, of the same order as the corresponding losses in the calibrated economy reported above).⁴⁴ We thus expect the comparative statics results below to also capture the response of the fully optimal tax codes to the corresponding changes in the parameters of interest.

5.4.1 Intensity of LBD effects

Figure 10 illustrates the impact of stronger LBD effects on marginal tax rates. First, consider income percentiles below the very top ones. For those agents at such percentiles, an increase in the intensity of the LBD effects leads to higher tax rates in both periods. The increase in such agents' period-1 tax rates appears to follow from the fact that, when LBD effects are stronger, the benefits of distorting downward the period-1 incomes to economize on future costs of incentives are higher (this is the mechanism discussed in the previous sections). The increase in period-2 tax rates, instead, follows from optimal consumption smoothing (see Condition (55) in the proof of Proposition 5 in the Appendix). Period-1 consumption drops as the result of the increase in period-1 taxes. Consumption

⁴⁴In all cases considered below, adopting the fully optimal tax code instead of the quasi-optimal one yields welfare gains equivalent to an equi-proportionate increase in consumption starting from the allocations in the economy under the current US tax code of at most 0.2%.

smoothing then calls for lower consumption also in period 2, which is induced through higher period-2 tax rates. Next, consider agents at top income percentiles. For such agents, period-1 tax rates decrease whereas period-2 tax rates increase as the intensity of the LBD effects increases. The reason seems to be that, for these agents, the efficiency cost of making them less productive in period 2 is very high when LBD effects are strong. Indeed, under the specification in (2), for any level of ζ , the effects of y_1 on θ_2 are stronger the larger θ_1 is. When LBD effects grow large, the planner then optimally reduces the marginal tax rates for the most productive period-1 types to make such types highly productive also in period 2. Making such types work harder in period 1, however, implies an increase in the expected period-2 rents that the planner must grant to such agents. To contain such rents, the planner then increases the period-2 marginal tax rates at the top of the period-2 income distribution, for these rates are the most relevant ones for the top period-1 types. An implication of the above effects is that higher levels of LBD then come with lower progressivity of taxes in period one, but higher progressivity in period two.

We also note that, for the levels of LBD intensity considered in Figure 10, reforming the current US tax code by adopting the optimal one generates welfare gains of the same order as in the calibrated economy.⁴⁵ Interestingly, the welfare gains are increasing in the intensity of LBD. Finally, we note that, as in the calibrated economy, in each of the economies covered by Figure 10, most of the welfare gains from reforming the current US tax code can also be generated through simple age-dependent linear taxes.⁴⁶

5.4.2 Skill Persistence

Next, consider variations in skill persistence. As Figure 11 shows, an increase in skill persistence leads to higher marginal tax rates in both periods. The increase in period-2 tax rates follows from the fact that higher persistence implies higher impulse responses of period-2 types to period-1 types, and hence higher period-2 handicaps (equivalently, higher expected period-2 rents). Because period-2 handicaps/rents are increasing in period-2 incomes, to contain the increase in expected period-2 rents, the planner then optimally increases period-2 tax rates. The increase in period-1 tax rates, instead, follows from two mechanisms that operate in the same direction. First, an increase in period-2 handicaps/rents increases the benefit of distorting downwards the period-1 incomes so as

⁴⁵Namely, these gains are equivalent to those brought by an equi-proportionate increase in consumption of $x\%$ starting from the allocations under the current US tax code, where x is equal to 4.02, 4.04, 4.06, 4.1, 4.16, and 4.22, respectively for the economy with $\zeta = 0.15$, $\zeta = 0.217$, $\zeta = 0.25$, $\zeta = 0.3$, $\zeta = 0.35$, and $\zeta = 0.4$.

⁴⁶Precisely, adopting the optimal linear tax code yields welfare gains equivalent to an equi-proportionate increase in consumption of $x\%$ starting from the allocations under the current US tax code, where x is equal to 3.82, 3.89, 3.92, 3.99, 4.07, and 4.15, respectively for the case where $\zeta = 0.15$, $\zeta = 0.217$, $\zeta = 0.25$, $\zeta = 0.3$, $\zeta = 0.35$, and $\zeta = 0.4$. Therefore, for each of the above economies, the welfare gains of adopting the fully optimal tax code instead of the optimal linear one are equivalent to an equi-proportionate increase in consumption of $y\%$ starting from the allocations under the current US tax code, where y is equal to 0.2, 0.15, 0.14, 0.11, 0.09, and 0.07, respectively.

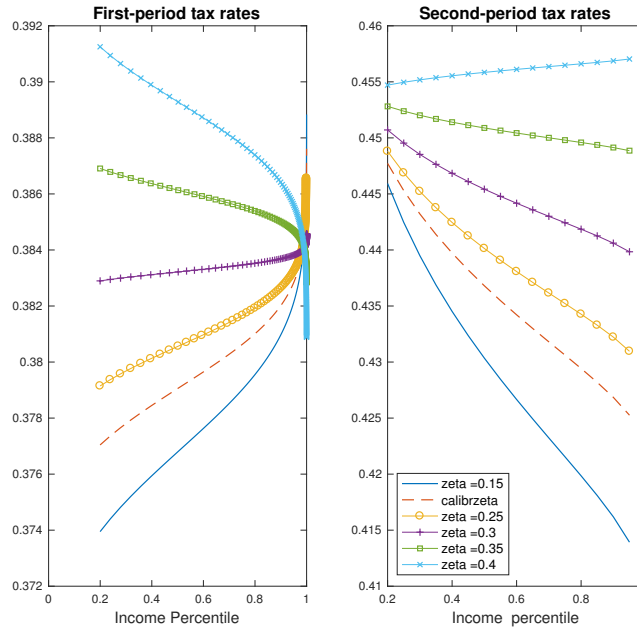


Figure 10: Effects of LBD intensity on quasi-optimal taxes

to shift the distribution of period-2 productivity towards levels commanding lower rents (this is the mechanism discussed throughout the entire paper). Second, because period-2 tax rates are higher, period-2 consumption is lower, which in turn calls for a reduction in period-1 consumption. This is obtained through an increase in period-1 tax rates (this is the same consumption smoothing channel discussed above for the comparative statics with respect to the intensity of the LBD effects).

The property that more persistence reduces the progressivity of the period-1 taxes seems to originate in the distribution of the productivity shocks in the calibrated economy. Under this distribution, the relative value of distorting the labor supply of the most productive period-1 agents, as captured by the term $\frac{1-F_1(\theta_1)}{\theta_1 f_1(\theta_1)}$, converges to a small constant as θ_1 grows large. Hence, at the very top, the benefits of increasing marginal tax rates (in either period) to contain the effects that higher persistence has on impulse responses (and hence on rents) are small compared to the corresponding benefits at lower percentiles. The planner thus increases marginal tax rates at top period-1 percentiles relatively less than at lower percentiles. As a result, the progressivity of the period-1 taxes is smaller when persistence is higher.

For any level of persistence considered in Figure 11, reforming the current US tax code by adopting the optimal one generates welfare gains of the same order as in the calibrated economy.⁴⁷ Interestingly, the higher the degree of skill persistence, the larger these welfare gains are. Moreover,

⁴⁷Namely, these gains are equivalent to those brought by an equi-proportionate increase in consumption of $x\%$ starting from the allocations under the current US tax code, where x is equal to 3.6, 3.74, 3.89, 4.04, 4.19, 4.34 and 4.49, respectively for the economy with $\rho = 0.3$, $\rho = 0.35$, $\rho = 0.4$, $\rho = 0.4505$, $\rho = 0.5$, $\rho = 0.55$, and $\rho = 0.6$.

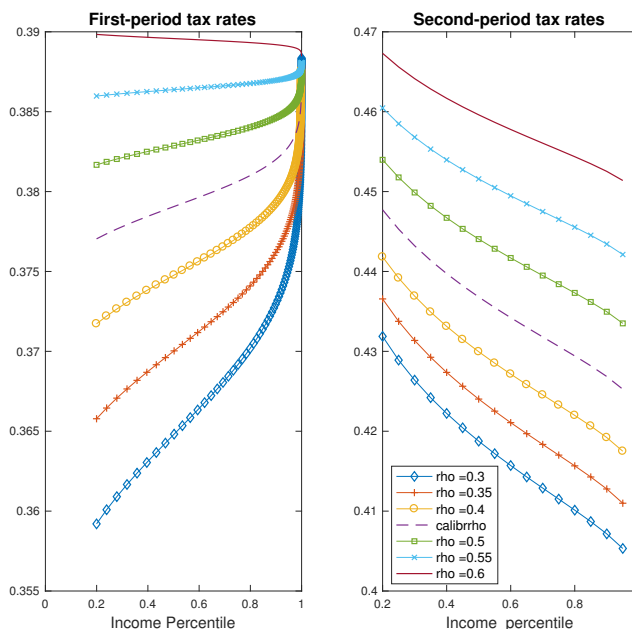


Figure 11: Effects of skill persistence on quasi-optimal taxes

as in the calibrated economy, most of these welfare gains can also be generated through simple age-dependent linear taxes.⁴⁸

5.5 Importance of the stochasticity of the LBD effects

Next, consider the role played by the stochasticity of the productivity process. Figure 12 shows how the optimal period-1 tax rates vary when one reduces the variance of the period-2 shocks ε_2 holding the Pareto tail parameter constant.

As the figure shows, a reduction in stochasticity leads to a reduction in period-2 tax rates accompanied by an increase in period-1 tax rates. The reason is the following. A reduction in stochasticity implies a reduction in income and consumption risk, which in turn calls for an increase in period-2 labor supply. This mechanism is the same as in Farhi and Werning (2013) and Golosov et al (2016). The reduction in period-2 taxes in turn implies an increase in the expected period-2 handicaps/rents, which in turn implies a higher benefit of distorting downwards period-1 incomes to economize on

⁴⁸To be precise, adopting the optimal linear tax code yields welfare gains equivalent to an equi-proportionate increase in consumption of $x\%$ starting from the allocations under the current US tax code, where x is equal to 3.4, 3.56, 3.72, 3.89, 4.04, 4.21, and 4.37, respectively for the case where $\rho = 0.3$, $\rho = 0.35$, $\rho = 0.4$, $\rho = 0.4505$, $\rho = 0.5$, $\rho = 0.55$, and $\rho = 0.6$. Therefore, for each of the above economies, the welfare gains of adopting the fully optimal tax code instead of the optimal linear one are equivalent to an equi-proportionate increase in consumption of $y\%$ starting from the allocations under the current US tax code, where y is equal to 0.2, 0.18, 0.17, 0.15, 0.15, 0.13, and 0.12, respectively.

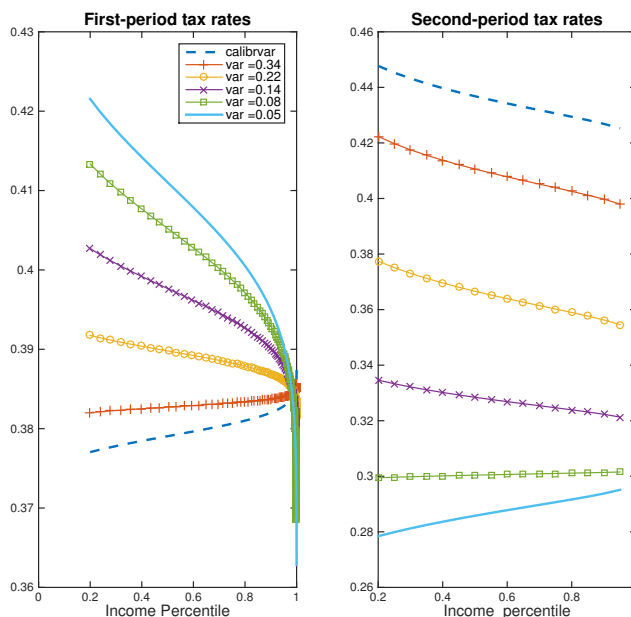


Figure 12: Effects of reductions in stochasticity of LBD on quasi-optimal taxes

expected period-2 rents. As a result, a reduction in stochasticity implies an increase in period-1 tax rates. Interestingly, when the variance of the period-2 shocks vanishes, the dynamics of taxes is reversed compared to the one identified in the calibrated economy: taxes are higher for young workers than for old ones. This property is consistent with the findings in Best and Kleven (2013) and Kapicka (2015) who document a declining pattern of taxes over the lifecycle. We also find that the welfare gains of reforming the current US tax code by adopting the optimal one are lower the lower the stochasticity is. For very small stochasticity levels, these gains are less than half of those in the calibrated economy.⁴⁹ These findings underscore the importance for tax design to account for the stochastic effects of LBD on the evolution of the workers' productivity, and hence the focus of this paper.

5.6 Isolating the role of LBD: Counterfactual Analysis

We conclude this section by discussing the role that LBD plays for wedges and taxes in the calibrated economy. For this purpose, we conduct the following counterfactual analysis. Suppose that period-2 productivity was 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

⁴⁹In particular, when the variance of the shocks is close to zero, the welfare gains from reforming the current US tax code are equivalent to an equi-proportionate increase in consumption starting from the allocations under the US tax code of at most 2%.

Symbol	Value	Target Moment	Data	Absolute Percentage Deviation
$\hat{\rho}$	0.6722	mean earnings ratio	0.868	3.54%
h_2	0.9966	Var. log-earnings young	0.335	1.09%
h_1	0.4795	Var. log-earnings old	0.435	0.97%
σ	0.5573	Gini earnings young	0.3175	2.66%
ξ	5.9907	mean-to-median earnings young	1.335	1.57%

Table 4: Calibrated Parameters

replaced by $\zeta = 0$ and $\hat{\rho}$, with the new set of parameters now including also h_2 .

Let the values of h_2 and $\hat{\rho}$ be determined so as to minimize the sum of the squared percentage residuals $\left(\frac{\theta_1^\rho y_1(\theta_1)^\zeta - h_2 \theta_1^{\hat{\rho}}}{\theta_1^\rho y_1(\theta_1)^\zeta}\right)^2$ across the two models, 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 sum of squared percentage residuals are $\hat{\rho} = 0.6722$ and $h_2 = 0.9966$. Under these values, the maximum absolute residual (as a percentage of $\theta_1^\rho y_1(\theta_1)^\zeta$), is 0.18%. The distance of the earnings distribution under this parameter configuration from the one in the data is also very small, as one can see from Table 4.

Therefore, the parameter values in Table 4 represent a suitable calibration of this alternative economy with no LBD effects. 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 (approximately) 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 between 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, reforming the current US tax code by adopting the optimal one yields welfare gains that are equivalent to those brought in by a 4.6178% equi-proportionate increase in consumption at all histories starting from the allocations under the existing tax code. Note that this figure is 14.27% larger than the corresponding 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. This result, which is consistent with the discussion in the previous subsections, is illustrated in Figure 13.

Importantly, that wedges are higher with LBD does not imply that taxes are also higher, as

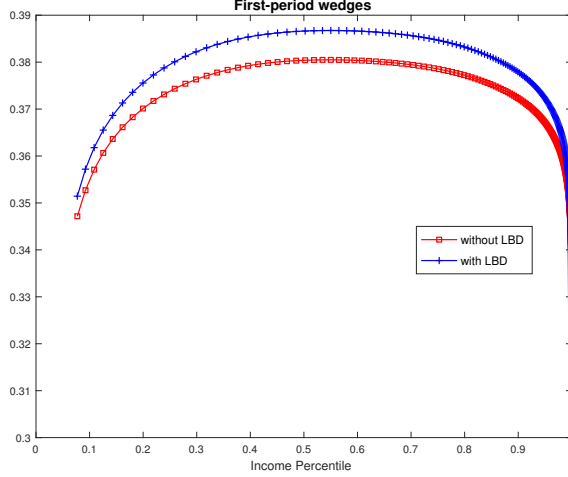


Figure 13: First-period wedges with and without LBD

one can see from Figure 14. The figure plots the marginal tax rates in the quasi-optimal tax codes for each of the two economies as a function of the (unconditional) income percentiles.⁵⁰ 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 with LBD. These findings can be explained using the comparative statics results in Subsections 5.4.1 and 5.4.2. Consider period-1 tax rates. Recall that lower LBD effects command lower and more progressive marginal tax rates, whereas higher persistence commands higher and more regressive marginal tax rates. As Figure 14 reveals, the effects due to higher persistence prevail under the quantitative specifications of the two economies under consideration. The combination of the effects that arise from the joint reduction in the intensity of LBD and the increase in persistence is also responsible for the differences in the period-2 marginal tax between the two economies.

The above results also warn against possible difficulties in extrapolating properties of tax rates from optimal wedges in dynamic economies.⁵¹ Period-1 wedges are higher with LBD than without

⁵⁰As in the economy with LBD, the quasi-optimal tax code in the counterfactual economy without LBD yields most of the welfare gains from reforming the current US tax code. Precisely, adopting the fully optimal tax code instead of the quasi-optimal one yields welfare gains equivalent to a 0.1919% equi-proportionate increase in consumption at all histories, starting from the allocations under the current US tax code.

⁵¹We note here that the counterpart of (23) for the case of risk-averse agents is

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

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 a given probability distribution of period-2 productivity, adjusted to take into account the different marginal utilities of consumption over the two

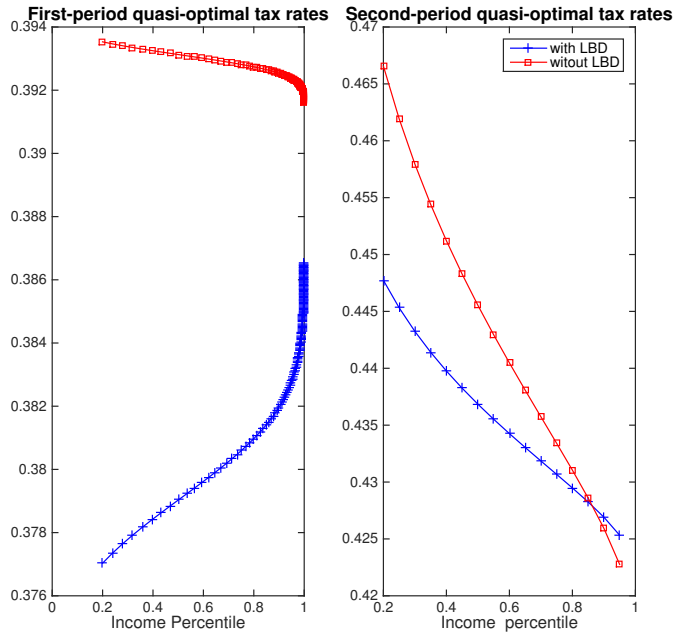


Figure 14: Quasi-optimal income tax rates with and without LBD

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 labor supply and in consumption across time and income percentiles.

6 Conclusions

This paper studies optimal taxation in a dynamic economy in which the workers' productivity evolves endogenously over the lifecycle as the result of (on-the-job) learning-by-doing. We show that, for younger workers, 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 is 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 is predicted by ignoring learning-by-doing. Finally, we show that simple taxes that are invariant to past incomes but age-dependent are approximately optimal.

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.

These taxes are higher but less progressive for younger workers than those under the current US tax code.

We believe these insights can guide the debate regarding the reform of existing tax codes and the alleviation of inequality in developed economies. In future work, it would be interesting to extend the analysis to accommodate for hidden savings, retirement, limited commitment on the planner's side, and the possibility that the intensity of LBD is sector-specific with agents choosing occupation in addition to their labor supply.⁵²

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.
- [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] Banks, J. and P. Diamond, (2011): "The Base for Direct Taxation," in "Dimensions of Tax Design: The Mirrlees Review," Mirrlees, J., A. Stuart, T. Besley, R. Blundell, R. Bond, S. Chote, R. Gammie, M. Johnson, P. Myles, and J. Poterba (eds), Oxford: Oxford University Press.
- [5] Battaglini, M., and S. Coate, (2008): "Pareto Efficient Income Taxation with Stochastic Abilities," *Journal of Public Economics* 92 (3-4), 844-68.
- [6] Becker, G., (1964): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press.
- [7] Ben-Porath, Y., (1967): "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy* 75(4), pp. 352-365.
- [8] 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.
- [9] Benkard, L., (2000): "Learning and Forgetting: The Dynamics of Aircraft Production," *American Economic Review*, 90(4), 1043-1054.

⁵²See also Ndiaye (2018) for an analysis of the benefits of restructuring social security and tax codes when retirement is endogenous.

- [10] Bergemann, D. and A. Pavan, (2015): “Introduction to JET Symposium on Dynamic Contracts and Mechanism Design,” *Journal of Economic Theory*, 159, 679-701.
- [11] Best, M.C. and H.J. Kleven, (2013): “Optimal Income Taxation with Career Effects of Work Effort,” mimeo, LSE.
- [12] Blundell, R. and T. MaCurdy, (1999): “Labor Supply: A Review of Alternative Approaches,” Chap. 27, pp. 1559–1695 in Ashenfelter, Orley, and Card, David (eds), *Handbook of Labor Economics*, Vol. 3. North Holland, Elsevier.
- [13] Chang, Y., J. Gomes, and F. Schorfheide (2002): “Learning-by-Doing as a Propagation Mechanism,” *American Economic Review*, 92(5), 1498-1520.
- [14] Conesa, J.C. and D. Krueger, (2006): “On the Optimal Progressivity of the Income Tax Code,” *Journal of Monetary Economics*, 53(7), 1425-1450.
- [15] 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.
- [16] D’Alessandro A., G. Fella and L. Melosi (2018): “Fiscal Stimulus with Learning-By-Doing,” mimeo Chicago Federal Reserve.
- [17] Dustmann, C. and K. Meghir, (2005): “Wages, Experience, and Seniority,” *Review of Economic Studies*, 72, 77-108.
- [18] 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.
- [19] Farhi, E. and I. Werning, (2013): “Insurance and Taxation over the Life Cycle,” *Review of Economic Studies*, 80, 596-635.
- [20] Fershtman, D. and A. Pavan, (2017): “Matching Auctions,” mimeo, Northwestern University.
- [21] Garrett, D. and A. Pavan, (2015): “Dynamic Managerial Compensation: a Variational Approach,” *Journal of Economic Theory*, 159, 775-818.
- [22] Garrett, D., A. Pavan, and J. Toikka, (2018): “Robust Predictions of Dynamic Screening,” mimeo, MIT, TSE, Northwestern.
- [23] Golosov, M., A. Tsyvinski, and N. Werquin, (2014): “A Variational Approach to the Analysis of Tax Systems,” NBER Working Paper No. 20780.
- [24] Golosov, M., M. Troshkin, and A. Tsyvinski, (2016): “Redistribution and Social Insurance,” *American Economic Review*, 106(2), 359-86.

- [25] Golosov, M., A. Tsyvinski, and I. Werning, (2006): “New Dynamic Public Finance: A User’s Guide”, NBER Macroeconomic Annual 2006, MIT Press.
- [26] Gorry, A. and E. Oberfield, (2012): “Optimal Taxation Over the Life Cycle,” *Review of Economic Dynamics*, 15(4), 551–572.
- [27] Heathcote, J., K. Storesletten, and G. Violante, (2017): “Optimal Tax Progressivity: An Analytical Framework,” *Quarterly Journal of Economics*, 132(4), 1693-1754
- [28] Huggett, M., G. Ventura, and A. Yaron, (2011): “Sources of Lifetime Inequality,” *American Economic Review*, 101(7), 2923-2954.
- [29] Jacobson L., R. LaLonde, and D. Sullivan, (1993): “Earnings Losses of Displaced Workers,” *American Economic Review*, 83(4), 685-709.
- [30] Kapicka, M., (2006): “Optimal Income Taxation with Human Capital Accumulation and Limited Record Keeping”, *Review of Economic Dynamics*, 9, 612–639.
- [31] 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.
- [32] Kapicka, M., (2015a): “Optimal Mirrleesian Taxation in a Ben-Porath Economy,” *American Economic Journal: Macroeconomics*, 7(2), 219-48.
- [33] Kapicka, M., (2015b): “Pareto Efficient Income Taxation with Learning by Doing,” mimeo, CERGE-EI.
- [34] Kapicka, M. and J. Neira, (2016): “Optimal Taxation with Risky Human Capital,” mimeo, CERGE-EI.
- [35] 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.
- [36] Kocherlakota, N. R., (2010): *The New Dynamic Public Finance*. Princeton University Press.
- [37] Krause, A., (2009): “Optimal Nonlinear Income Taxation with Learning-by-doing,” *Journal of Public Economics*, 93(9-10), 1098-1110.
- [38] Kremer, M., (2001): “Should Taxes be Independent of Age?” mimeo, Harvard University.
- [39] 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.
- [40] Levitt, S.D., J. List, and C. Syverson, (2013): “Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant,” *Journal of Political Economy*, 121(4), 643-681.

- [41] Lucas, R.E., (1988): “On the Mechanics of Economic Development,” *Journal of Monetary Economics*, 22(1), 33-42.
- [42] Mirrlees, J., (1971): “An Exploration in the Theory of Optimal Income Taxation,” *Review of Economic Studies*, 38, 175-208.
- [43] Ndiaye, A., (2018): “Flexible Retirement and Optimal Taxation,” mimeo, Northwestern University.
- [44] Pavan, A., S. Segal, and J. Toikka, (2014): “Dynamic Mechanism Design: A Myersonian Approach,” *Econometrica*, 82(2), 601-653.
- [45] 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.
- [46] Perrault, L., (2018): “Optimal Taxation with Work Experience as a Risky Investment,” mimeo, Georgia State University.
- [47] Ramsey, F.P., (1927): “A Contribution to the Theory of Taxation,” *Economic Journal*, 37(145), 47-61.
- [48] Rogerson, W.P., (1985): “Repeated Moral Hazard,” *Econometrica*, 53(1), 69-76.
- [49] Saez, E., (2001): “Using Elasticities to Derive Optimal Income Tax Rates,” *Review of Economic Studies*, 68, 205-229.
- [50] Stantcheva, S., (2015): “Learning and (or) Doing: Human Capital Investments and Optimal Taxation,” NBER Working Paper 21381.
- [51] Stantcheva, S., (2017): “Optimal Taxation and Human Capital Policies over the Life Cycle,” *Journal of Political Economy*, 125(6), 1931–1990.
- [52] Thompson, P., (2001): “How Much Did the Liberty Shipbuilders Learn? New Evidence for an Old Case Study,” *Journal of Political Economy*, 109(1), 103-137.
- [53] Thompson, P., (2012): “The Relationship Between Unit Cost and Cumulative Quantity and the Evidence for Organizational Learning-by-Doing,” *Journal of Economic Perspectives*, 26(3), 203-224.
- [54] Thornton, R. and P. Thompson, (2001): “Learning from Experience and Learning from Others: An Exploration of Learning and Spillovers in Wartime Shipbuilding,” *American Economic Review*, 91(5), 1350-1368.

- [55] Topel, R., (1991): “Specific Capital, Mobility and Wages: Wages Rise with Seniority,” *Journal of Political Economy*, 99(1), 145-176.
- [56] Weinzierl, M., (2011): “The Surprising Power of Age-Dependent Taxes,” *Review of Economic Studies*, 78(4), 1490-1518.
- [57] 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.
- [58] Wright, T.P., (1936): “Factors Affecting the Cost of Airplanes,” *Journal of the Aeronautical Sciences*, 3(4), 122-128.

7 Appendix

Proof of Proposition 1. Under full information, the optimal allocation rule $\chi^* = (y^*, c^*)$ maximizes expected tax revenue $R = \mathbb{E}^{\lambda[\chi]} \left[\sum_t \delta^{t-1} (y_t(\tilde{\theta}^t) - c_t(\tilde{\theta}^t)) \right]$ subject to the redistribution constraint $\int V_1(\theta_1)q(\theta_1)dF_1(\theta_1) \geq \kappa$, with

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

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

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

subject to

$$\int V_1(\theta_1)q(\theta_1)dF_1(\theta_1) - \kappa = 0. \quad (31)$$

Let π be the multiplier of the redistribution constraint (31), which is an integral constraint. At the optimum, the following necessary conditions must hold with $\lambda[\chi^*]$ -probability one:

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

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

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

$$- \frac{1}{v'(c_1^*(\theta_1))} + \pi q(\theta_1) = 0, \quad (32)$$

along with

$$\int V_1(\theta_1) q(\theta_1) dF_1(\theta_1) - \kappa = 0 \quad (33)$$

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 $LD_1^X(\theta_1)$, we obtain the conditions in the proposition. Q.E.D.

Proof of Proposition 5. 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 proposition.

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

$$\frac{\partial V_1(\theta_1)}{\partial \theta_1} = -\psi_\theta(y_1(\theta_1), \theta_1) - \delta \mathbb{E}^{\lambda^{[x]|\theta_1}} \left[I_1^2(\tilde{\theta}, y_1(\theta_1)) \psi_\theta(y_2(\tilde{\theta}), \tilde{\theta}_2) \right], \text{ almost all } \theta_1 \in \Theta_1 \quad (34)$$

and

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

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

$$\frac{\partial V_1(\theta_1)}{\partial \theta_1} = -\psi_\theta(y_1(\theta_1), \theta_1) + \delta Z_2(\theta_1), \quad (35)$$

and

$$\int V_1(\theta_1) q(\theta_1) dF_1(\theta_1) - \kappa = 0, \quad (36)$$

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

$$\Pi_2(\theta_1) = \int V_2(\theta) dF_2(\theta_2 | \theta_1, y_1(\theta_1)), \quad (37)$$

$$Z_2(\theta_1) = - \int I_1^2(\theta, y_1(\theta_1)) \psi_\theta(y_2(\theta), \theta_2) dF_2(\theta_2 | \theta_1, y_1(\theta_1)), \quad (38)$$

and

$$\frac{\partial V_2(\theta)}{\partial \theta_2} = -\psi_\theta(y_2(\theta), \theta_2). \quad (39)$$

The above optimization problem thus consists of two interdependent optimal control problems, one for each period.

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, (37) and (38). 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 (39).

Let $\pi_2(\theta_1)$ and $\xi_2(\theta_1)$ be the multipliers of the two integral constraints (37) and (38) and $\mu_2(\theta)$ the costate variable for the law of motion of $V_2(\theta_1, \theta_2)$.

Along with (37), (38), and (39), the following necessary optimality conditions must hold for almost all $\theta_2 \in \text{Supp}[F_2(\cdot | \theta_1, y_1(\theta_1))]$:

$$1 - \frac{\psi_y(y_2(\theta), \theta_2)}{v'(c_2(\theta))} - \mu_2(\theta) \frac{\psi_{\theta y}(y_2(\theta), \theta_2)}{f_2(\theta_2 | \theta_2, y_1(\theta_1))} + \xi_2(\theta_1) I_1^2(\theta, y_1(\theta_1)) \psi_{\theta y}(y_2(\theta), \theta_2) = 0, \quad (40)$$

$$\frac{\partial \mu_2(\theta)}{\partial \theta_2} = f_2(\theta_2 | \theta_1, y_1(\theta_1)) \cdot \left\{ \frac{1}{v'(c_2(\theta))} + \pi_2(\theta_1) \right\}, \quad (41)$$

along with the boundary conditions

$$\mu_2(\theta_1, \underline{\theta}_2) = 0, \quad (42)$$

$$\mu_2(\theta_1, \bar{\theta}_2) = 0, \quad (43)$$

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 (35) and π_1 the multiplier associated with the constraint (36). In addition to (35) and (36), the following optimality conditions must hold:

$$1 - \frac{\psi_y(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))} + \delta \frac{\partial}{\partial y_1} \int [y_2(\theta) - c_2(\theta) - \pi_2(\theta_1) V_2(\theta)] dF_2(\theta_2 | \theta_1, y_1(\theta_1)) \quad (44)$$

$$+ \delta \xi_2(\theta_1) \frac{\partial}{\partial y_1} \int I_1^2(\theta, y_1(\theta_1)) \psi_\theta(y_2(\theta), \theta_2) dF_2(\theta_2 | \theta_1, y_1(\theta_1)) - \mu_1(\theta_1) \frac{\psi_{\theta y}(y_1(\theta_1), \theta_1)}{f_1(\theta_1)} = 0,$$

$$\frac{\partial \mu_1(\theta_1)}{\partial \theta_1} = f_1(\theta_1) \cdot \left\{ \frac{1}{v'(c_1(\theta_1))} - \pi_1 q(\theta_1) \right\}, \quad (45)$$

$$\frac{1}{v'(c_1(\theta_1))} + \pi_2(\theta_1) = 0, \quad (46)$$

$$\mu_1(\theta_1) + \xi_2(\theta_1) f_1(\theta_1) = 0, \quad (47)$$

along with the boundary conditions

$$\mu_1(\underline{\theta}_1) = 0, \quad (48)$$

$$\mu_1(\bar{\theta}_1) = 0, \quad (49)$$

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 $y_1(\theta_1)$, $\Pi_2(\theta_1)$, and $Z_2(\theta_1)$, we have used the following properties:

$$\begin{aligned} \frac{\partial Q_2}{\partial y_1} &= \frac{\partial}{\partial y_1} \int [y_2(\theta) - c_2(\theta)] dF_2(\theta_2 | \theta_1, y_1(\theta_1)) - \pi_2(\theta_1) \frac{\partial}{\partial y_1} \int V_2(\theta) dF_2(\theta_2 | \theta_1, y_1(\theta_1)) \\ &\quad + \xi_2(\theta_1) \frac{\partial}{\partial y_1} \int I_1^2(\theta, y_1(\theta_1)) \psi_\theta(y_2(\theta), \theta_2) dF_2(\theta_2 | \theta_1, y_1(\theta_1)), \end{aligned}$$

$$\frac{\partial Q_2}{\partial \Pi_2} = \pi_2(\theta_1), \text{ and } \frac{\partial Q_2}{\partial Z_2} = \xi_2(\theta_1).$$

Now use (45) along with the boundary conditions (48) and (49) to obtain that

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

which implies that

$$\pi_1 = \int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(\theta_1))} dF_1(\theta_1), \quad (50)$$

where we used the fact that $\int_{\theta_1}^{\bar{\theta}_1} q(\theta_1) dF_1(\theta_1) = 1$. Furthermore, using (45) and (49) again, we have that

$$\mu_1(\theta_1) = - \int_{\theta_1}^{\bar{\theta}_1} f_1(s) \cdot \left\{ \frac{1}{v'(c_1(s))} - \pi_1 q(s) \right\} ds,$$

from which we obtain that

$$\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} = \frac{1 - F_1(\theta_1)}{f_1(\theta_1)} \left[\int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} \frac{dF_1(s)}{1 - F_1(\theta_1)} - \pi_1 \int_{\theta_1}^{\bar{\theta}_1} q(s) \frac{dF_1(s)}{1 - F_1(\theta_1)} \right]. \quad (51)$$

Next, use (46) and (47) to rewrite the FOC for $y_1(\theta_1)$ as follows

$$\begin{aligned} &1 - \frac{\psi_{y_1}(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))} + \delta \frac{\partial}{\partial y_1} \int \left[y_2(\theta) - c_2(\theta) + \frac{V_2(\theta)}{v'(c_1(\theta_1))} \right] dF_2(\theta_2 | \theta_1, y_1(\theta_1)) \\ &+ \left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \delta \frac{\partial}{\partial y_1} \int I_1^2(\theta, y_1(\theta_1)) \psi_\theta(y_2(\theta), \theta_2) dF_2(\theta_2 | \theta_1, y_1(\theta_1)) + \left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \psi_{\theta y_1}(y_1(\theta_1), \theta_1) = 0, \end{aligned} \quad (52)$$

with $-\mu_1(\theta_1)/f_1(\theta_1)$ given by (51).

Note that, together, Conditions (41) and (46), imply that

$$\frac{\partial \mu_2(\theta)}{\partial \theta_2} = f_2(\theta_2 | \theta_1, y_1(\theta_1)) \cdot \left\{ \frac{1}{v'(c_2(\theta))} - \frac{1}{v'(c_1(\theta_1))} \right\}. \quad (53)$$

Combining (53) with the boundary conditions (42) and (43), we have that

$$0 = \int_{\theta_2}^{\bar{\theta}_2} \frac{\partial \mu_2(\theta_1, s)}{\partial \theta_2} ds = \int_{\theta_2}^{\bar{\theta}_2} \frac{1}{v'(c_2(\theta_1, s))} dF_2(s | \theta_1, y_1(\theta_1)) - \frac{1}{v'(c_1(\theta_1))}. \quad (54)$$

Next note that (54) yields the familiar Rogerson inverse-Euler condition

$$\frac{1}{v'(c_1(\theta_1))} = \int_{\theta_2}^{\bar{\theta}_2} \frac{1}{v'(c_2(\theta_1, s))} dF_2(s | \theta_1, y_1(\theta_1)). \quad (55)$$

Combining (53) and (43) with (55), we have that

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

from which we obtain that

$$\begin{aligned} \frac{-\mu_2(\theta_1, \theta_2)}{f_2(\theta_2|\theta_1, y_1(\theta_1))} &= \frac{1-F_2(\theta_2|\theta_1, y_1(\theta_1))}{f_2(\theta_2|\theta_1, y_1(\theta_1))} \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{1}{v'(c_2(\theta_1, s))} \frac{dF_2(s|\theta_1, y_1(\theta_1))}{1-F_2(\theta_2|\theta_1, y_1(\theta_1))} \\ &\quad - \frac{1-F_2(\theta_2|\theta_1, y_1(\theta_1))}{f_2(\theta_2|\theta_1, y_1(\theta_1))} \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{1}{v'(c_2(\theta_1, s))} dF_2(s|\theta_1, y_1(\theta_1)). \end{aligned} \quad (56)$$

Step 2. We now show how the above optimality conditions permit us to arrive at the expressions for the wedges in the proposition.

Consider first the period-2 wedges. Using the FOC for period-2 output (40), we have that the period-2 wedges, under the second-best allocations, are given by

$$W_2(\theta) \equiv 1 - \frac{\psi_y(y_2(\theta), \theta_2)}{v'(c_2(\theta))} = -\psi_{\theta y}(y_2(\theta), \theta_2) \left(\frac{-\mu_2(\theta)}{f_2(\theta_2|\theta_2, y_1(\theta_1))} - \frac{\mu_1(\theta_1)}{f_1(\theta_1)} I_1^2(\theta, y_1(\theta_1)) \right).$$

Hence, the period-2 relative wedges are given by

$$\hat{W}_2(\theta) \equiv \frac{W_2(\theta)}{1 - W_2(\theta)} = \frac{-v'(c_2(\theta))}{\psi_y(y_2(\theta), \theta_2)} \psi_{\theta y}(y_2(\theta), \theta_2) \left(\frac{-\mu_2(\theta)}{f_2(\theta_2|\theta_2, y_1(\theta_1))} - \frac{\mu_1(\theta_1)}{f_1(\theta_1)} I_1^2(\theta, y_1(\theta_1)) \right)$$

with $-\mu_2(\theta)/f_2(\theta_2|\theta_2, y_1(\theta_1))$ given by (56) and $-\mu_1(\theta_1)/f_1(\theta_1)$ given by (51).

Next, consider the period-1 wedges. Using the FOC for period-1 output (52) and the definition of the $LD_1^X(\theta_1)$ function in (6), we have that

$$\begin{aligned} W_1(\theta_1) &\equiv 1 - \frac{\frac{\psi_y(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))}}{[1+LD^X(\theta_1)]} \\ &= - \frac{\left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \delta \frac{\partial}{\partial y_1} \int I_1^2(\theta, y_1(\theta_1)) \psi_{\theta}(y_2(\theta), \theta_2) dF_2(\theta_2|\theta_1, y_1(\theta_1)) + \left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \psi_{\theta y}(y_1(\theta_1), \theta_1)}{1+LD^X(\theta_1)}. \end{aligned}$$

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 (51).

Using the definition of $\Omega(\theta_1)$ from (18) we thus have that

$$\hat{W}_1(\theta_1) = \left[\left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \gamma_1(\theta_1) v'(c_1(\theta_1)) \right] \Omega(\theta_1) + \left(\frac{\mu_1(\theta_1)}{f_1(\theta_1)} \right) \frac{\psi_{\theta y}(y_t(\theta^t), \theta_t)}{\psi_y(y_t(\theta^t), \theta_t)} v'(c_1(\theta_1)).$$

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

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

Using (50) and (51), 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 planner's preferences for redistribution is equal to

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

Q.E.D.