Taxation under Learning-by-Doing*

Miltiadis Makris
University of Kent

Alessandro Pavan
Northwestern University

October 2020

Abstract

We study optimal income taxation when workers' productivity is stochastic and evolves endogenously because of learning-by-doing. Learning-by-doing calls for higher wedges, and alters the relation between wedges and tax rates. In a calibrated model, we find that reforming the US tax code brings significant welfare gains and that a simple tax code invariant to past incomes is approximately optimal. We isolate the role of learning-by-doing by comparing the aforementioned tax code to its counterpart in an economy that is identical to the calibrated one except for the exogeneity of the productivity process. Ignoring learning-by-doing calls for fundamentally different proposals.

*For comments and useful suggestions, we thank the Editor, Harald Uhlig, various anonymous referees, 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, Bristol University, CERGE-EI, Cergy-Pontoise, the Chicago FED Research Department, Edinburgh University, Harvard University, the University of Kent, 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 the design of optimal tax codes. We consider a dynamic Mirrleesian economy in which the agents' productivity is their own private information, is stochastic, and evolves endogenously over the life cycle as the result of LBD. We show that, in the presence of LBD, labor wedges (i.e., the distortions relative to the complete-information benchmark) are higher than in the absence of LBD. Furthermore, the presence of LBD alters the relation between wedges and marginal tax rates under optimal tax codes, and has a significant impact on the level, the progressivity, and the dynamics of taxes over the life cycle.

In a model that is calibrated to US earnings data, we find that reforming the current US tax code brings significant welfare gains. We also find that most of the welfare gains from the optimal reform can be generated with a simple tax code where taxes are invariant to past incomes but age-dependent. The allocations under such simple code are close to the allocations under the fully-optimal code.

Compared to the current US tax code, such simple (yet quasi-optimal) code (a) features higher tax rates for young workers, whereas, for old workers, features higher tax rates at low income percentiles, but lower tax rates at high income percentiles, (b) is less progressive for young workers and is regressive, instead of progressive, for old workers, and (c) features a smaller differential between the average tax rate for old workers and the average tax rate for young workers.

[1] For 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.

[2] See 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.

[3] The 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.

To shed light on the role of LBD, we then compare such code to its counterpart in a counterfactual economy in which the workers' productivity follows the same process as in the calibrated economy but is exogenous.[4] We find that the value of reforming the current US tax code is 14% lower in the calibrated than in the counterfactual economy. Compared to the tax code in the counterfactual economy, the tax code in the calibrated economy (i) features lower tax rates for young workers, whereas, for old workers, features lower tax rates at low income percentiles but higher tax rates at high income percentiles, (ii) is progressive, instead of regressive for young workers, whereas, for old workers, is less regressive, and (iii) features a larger differential between the average tax rates for old and for young workers. Hence, reforming the current US tax code while accounting for the presence of LBD found in the data calls for very different proposals than the ones suggested by ignoring LBD.

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 productivity levels, contributes to higher expected future rents and thus to higher expected future welfare losses.

The mechanism described above is specific to economies in which the agents' productivity is their private information, is endogenous, and evolves stochastically over the life cycle, which are natural features of economies with LBD. The main contribution of the present paper is to illustrate the implications of such mechanism for the design of optimal tax codes. From previous work (e.g., Farhi and Werning (2013), and Golosov et al. (2016)), we know that uncertainty over future productivity contributes to an increasing intertemporal profile of wedges and tax rates over the life cycle. On the other hand, LBD, when deterministic, contributes to a decreasing intertemporal profile (e.g., Best and Kleven (2013)). The implications of LBD for the level, the progressivity, and the dynamics of taxes when the effects of LBD on the agents' productivity are stochastic are unknown. We develop a model that permits us to address this open question. Compared to the quasi-optimal tax code in an economy that is identical to the calibrated one except for the fact that LBD has deterministic effects on the workers' productivity, the quasi-optimal tax code in the economy with stochastic LBD (a) features lower tax rates for young workers at low income percentiles but higher tax rates at higher percentiles, whereas, for old workers, features higher tax rates at all percentiles, (b) is progressive, instead of regressive, for young workers and regressive, instead of progressive, for old workers, and (c) features an increasing, instead of decreasing, profile of taxes over the life cycle. Hence, reforming the current US tax code while accounting for the stochastic effects of LBD found in the data calls

[4] As in the calibrated economy, the welfare losses from using the quasi-optimal tax code in the counterfactual economy instead of the fully-optimal one are very small. Furthermore, the allocations under the quasi-optimal tax code are very close to those under the fully-optimal code. The same is true for the economies we consider for the various comparative statics exercises. Because quasi-optimal tax codes are particularly simple, and because most codes in the real world are history-independent, the discussion in the paper often focuses on such codes.

for very different proposals than those suggested by assuming deterministic LBD effects.

The shape of the optimal tax code depends on how LBD interacts with the agents' risk aversion, the persistence of the agents' initial productivity, the distribution from which the productivity shocks are drawn, the elasticity of the agents' labor supply with respect to the (net-of-taxes) wages, and the planner's preferences for redistribution. Another contribution of the present paper is to shed light on the above interactions when LBD has stochastic effects on the agents' productivity.

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 their productivity is determined by how much they worked when young. First, we identify the relevant channels through which LBD alters the relation between wedges and marginal tax rates under optimal allocations. Next, we consider 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. This benchmark permits us to illustrate in the simplest possible way the main mechanism through which LBD affects the labor wedges. We then turn to the more general case where the agents may be risk averse (with preferences for consumption smoothing) and where the planner may assign general non-linear Pareto weights to the agents' lifetime expected utilities. We derive a general formula for the second-best allocations that shows how the agents' risk aversion and the planner's preferences for redistribution interact with LBD in the determination of the labor wedges. We also illustrate how the distribution from which the productivity shocks are drawn interacts 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).[5] 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 (e.g., the meta-analysis in Best and Kleven (2013)) as well as the estimation results in the macro literature (e.g., Chang et al. (2002)). We show that reforming the existing US tax code by adopting the optimal one (i.e., the one implementing the second-best allocations) 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 existing US tax code. We also show that the utility that the agents derive under the optimal tax code, as well as the second-best allocations, are close to those that emerge under a simple 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 period-s lump-sum subsidy, and k_s and α_s are age-dependent parameters that control for the progressivity of the taxes.[6] Under this simple (yet quasi-optimal) code, tax rates are mildly progressive for young workers and mildly regressive for old ones and increase over the life cycle across all income percentiles with an average tax rate of 38% for young workers and of 43% for

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

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

old ones. We also find that most of the welfare gains from switching from the existing US tax code to the optimal one can be generated through simple age-dependent *linear* taxes.[7] The marginal tax rates in the optimal linear code are equal to 38% for young workers and 46% for older ones.

As anticipated above, to isolate the implications of LBD on the reform from all other effects, we compare the quasi-optimal tax code in the calibrated economy to its counterpart in a counterfactual economy that is identical to the calibrated one, except for the fact that workers' productivity follows the same process as in the calibrated economy but is exogenous.[8] We show that, in the economy with LBD, wedges for young workers are higher than in the counterfactual economy without LBD. Importantly, that wedges are higher does not imply that tax rates are also higher.[9] 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 via the change in the distribution from which future productivity is drawn. This novel "future tax bill effect", which is specific to economies with LBD, plays an important role when it comes to the differences discussed above between the tax code in the calibrated and in the counterfactual economy.

We also conduct various comparative statics exercises that illustrate the effects on the quasi-optimal tax code of the agents' skill persistence and of the intensity of the LBD effects. We show that, all other things equal, both higher levels of skill persistence and stronger LBD effects contribute to higher marginal tax rates (except for young workers at high percentiles), to less progressive taxes for young workers and to more progressive taxes for old workers.

Finally, we study the effects of the stochasticity of LBD on the level, the progressivity and the dynamics of taxes under quasi-optimal tax codes. When the effects of LBD on productivity are stochastic, young workers face risk in their consumption when old due to the volatility in their earnings. Such volatility reduces welfare when agents are risk averse, as they are in the calibrated economy. We show that, as the stochasticity in the agents' productivity increases, the tax rates on the old increase at all percentiles, whereas, for the young, they increase at high percentiles but decrease at low percentiles. Furthermore, the taxes on the young become more progressive whereas those on the old become more regressive. Interestingly, as mentioned above, as one moves from an economy with almost deterministic LBD effects to one with stochastic LBD effects as in the calibrated economy, both the progressivity of the taxes within each period and the dynamics of the taxes over the life cycle are reversed. As our calibration analysis reveals, the stochasticity of the LBD effects is empirically relevant.[10] The results thus bear important lessons for the design of optimal tax codes.

Outline. The rest of the manuscript is organized as follows. Below, we wrap up the Introduction

[7]See 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.

[8]The earning distributions in the counterfactual economy under the current US tax code are also very close to the empirical ones, as reported in Subsection 5.6.

[9]As mentioned above, marginal tax rates for young workers are in fact lower with LBD.

[10]In 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.

with a discussion of the most pertinent literature. Section 2 introduces the model. Section 3 describes the first-best policies. Section 4 relates wedges to tax rates, characterizes the second-best allocations, and derives various implications for the level, progressivity, and dynamics of the labor wedges. Section 5 calibrates the model to the US earnings 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 online Supplementary Material. The latter also contains a detailed description of the methods used to establish all numerical results and additional comparative statics exercises illustrating the effects on optimal wedges of variations in the degree of the agents' risk aversion, the Frisch elasticity of the agents' labor supply, and the planner's preferences for redistribution. Finally, it contains results showing how wedges and tax rates can also be obtained through a sufficient statistics approach in the spirit of Saez (2001), but adapted to the economy with LBD under consideration.

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 literatures 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., Thomspon, 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 (2018).[11]

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.

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 to 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, as mentioned above, our comparative statics exercises with respect to the stochasticity of the LBD effects show that the dynamics and progressivity of taxes are reversed as one moves from an economy with deterministic to one with stochastic LBD effects (see the analysis in Subsection 5.5).

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, for young workers, marginal tax rates are lower in our calibrated economy with LBD than in the counterfactual economy without it, whereas, for old workers, tax rates are lower in the calibrated economy only for low percentiles. Importantly, we show that this does not imply that labor wedges are also lower.

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 life cycle, 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 life cycle 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 (2015b) tax rates decline over the life cycle, 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

[11] See also Stantcheva (2015).

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) in the Introduction).[12]

Kapicka and Neira (2016) consider a two-period economy where productivity evolves stochastically. However, contrary to our work, 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 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 life cycle. 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, and agents must be provided with incentives to reveal their private information in both periods.

Stantcheva (2017) considers a multi-period economy where, in each period, agents invest in human capital in addition to providing their labor. Our work differs from that paper in a few important dimensions. In our model, the evolution of human capital comes from LBD, whereas in Stantcheva (2017) it comes from education, training, and other activities that are separate from labor supply. Because LBD is a direct byproduct of labor supply, in our model, the planner cannot affect investments in human capital without also affecting the agents' labor supply decisions. In contrast, in Stantcheva (2017), the planner can use instruments other than the marginal labor-income tax rates such as education and training subsidies to influence the agents' investments in human capital. Importantly, when, investments in human capital cannot be controlled separately from labor supply decisions, as is the case in economies with LBD, the structure of optimal tax codes, and in particular the progressivity of the taxes, is different than in economies where investments in human capital and labor supply decisions can be controlled separately.

In independent work, Perrault (2018) considers a special case of our economy in which period-2 productivity is invariant to 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 our calibrated economy), understanding the role of the interaction of skill persistence with LBD is important. Our work contributes towards this understanding.

In Makris and Pavan (2018), we consider a general multi-period economy in which the agents' 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 that paper. The contribution of the two papers is, however, different. In Makris and Pavan (2018), we show how

[12]Another 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.

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 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), Pavan (2017), and Bergemann and Välimäki (2019) for overviews of recent developments,[13] 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 life cycle, 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 version 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.[14]

We view the contribution of the present paper relative to the various strands of the literature discussed above as twofold. First, we identify a novel mechanism by which the endogeneity of the agents' private information about their stochastic and 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 such as the workers' skill persistence and the stochasticity of the agents' productivity in shaping the design of optimal tax codes.

2 Model

Agents, productivity, and information. The economy is populated by a unit-mass continuum of workers with heterogenous productivity. The life cycle of each worker consists of two periods. We interpret the first period as the phase in which workers accumulate human capital through LBD,

[13] See 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.

[14] Pavan, 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, are fundamentally different from the ones in the present paper.

and the second period as the phase in which the workers take advantage of earlier investments in human capital. We capture this situation by letting productivity be exogenous in the first period but endogenous in the second.[15]

In each period $t = 1, 2$, each worker produces *income* $y_t \in Y_t = [0, \bar{y})$ at a cost $\psi(y_t, \theta_t)$, where $\bar{y} \leq \infty$, and where $\theta_t \in \Theta_t$ denotes the worker's period-t *productivity* (equivalently, his skills), and is the worker's private information. The function $\psi(y_t, \theta_t)$ is equi-Lipschitz continuous, thrice differentiable, increasing, and convex in y_t. We denote by

$$\psi_y(y_t, \theta_t) \equiv \frac{\partial \psi(y_t, \theta_t)}{\partial y_t}, \quad \psi_\theta(y_t, \theta_t) \equiv \frac{\partial \psi(y_t, \theta_t)}{\partial \theta_t}, \quad \psi_{y\theta}(y_t, \theta_t) \equiv \frac{\partial^2 \psi(y_t, \theta_t)}{\partial \theta_t \partial y_t}, \text{ and } \psi_{yy\theta}(y_t, \theta_t) \equiv \frac{\partial^3 \psi(y_t, \theta_t)}{\partial \theta_t \partial y_t^2}$$

the partial and cross derivatives of ψ, and assume that $\psi_\theta < 0$, $\psi_{y\theta} < 0$, and $\psi_{yy\theta} \leq 0$. Higher types thus experience a lower disutility of effort, and have a lower marginal cost for their labor supply. Furthermore, the rate by which their marginal cost increases with output is smaller than for lower types.

For the quantitative results, we further assume that ψ takes the familiar iso-elastic form[16]

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

and that $\bar{y}$ is finite but sufficiently large that, under both the first-best and the second-best policies, all workers produce income below $\bar{y}$. Under the above specification, ϕ is the inverse Frisch elasticity.

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 $\Theta_1 = (\underline{\theta}_1, \bar{\theta}_1)$, with $\underline{\theta}_1 > 0$, and zero anywhere else. Each worker's period-2 productivity is endogenous and given by

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

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

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

We assume that $\zeta \leq \phi/(1+\phi)$ to ensure a well-defined solution to the first-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-

[15] As 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 years. In this case, the life cycle 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.

[16] See, among others, Kapicka (2013), Farhi and Werning (2013), and Best and Kleven (2013).

2 productivities, and by $\Theta = \Theta_1 \times \Theta_2$ the set of all possible productivity histories (θ_1, θ_2). The dependence of the agents' period-2 productivity on their period-1 income is what captures 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. Under the specification in (2), the parameter $\rho \geq 0$ measures the exogenous persistence in the agents' productivity, whereas the parameter $\zeta \geq 0$ measures the intensity of LBD; the case of no LBD corresponds to $\zeta = 0$, and higher values of ζ capture stronger LBD effects.

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

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

The advantage of the specification in (2) is that the only channel through which LBD affects wedges and optimal tax rates is by shifting the distribution of period-2 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 similar to the one that operates through the accumulation of human capital in economies with exogenous private information.[17] To highlight the novel effects due to the endogeneity of the agents' private information, 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. 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 highlight the dependence of these functions on the specific functional forms in (2) when necessary.

Preferences. Let $c_t \in \mathbb{R}_+$ denote period-t consumption, $y \equiv (y_1, y_2)$ and $c \equiv (c_1, c_2)$ the income and consumption histories, respectively, and δ the common discount factor.[18] Each agent's *lifetime* utility is given by

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

where $v : \mathbb{R} \to \mathbb{R}$ is an increasing, concave, and twice differentiable function.

Planner's problem. The planner's problem consists of maximizing the weighted sum of the agents' expected lifetime utilities, subject to the constraint that the fiscal deficit be smaller than an

[17] See, for example, Kapicka (2015a,b) and Stantcheva (2017).

[18] Consistently with the rest of the literature, we assume that δ is also equal to the inverse of the gross return on savings (see, for instance, Best and Kleven (2013), Kapicka (2013), Farhi and Werning (2013), Golosov et. al. (2016), and Stantcheva (2017)).

exogenous level. We will solve this problem by considering its dual in which the planner maximizes expected intertemporal tax revenues, subject to the constraint that the weighted sum of the agents' expected lifetime utilities be greater than an exogenous threshold.

Formally, the dual problem can be stated as follows. Let $\chi \equiv (y(\cdot), c(\cdot))$ 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 $\theta^2 \equiv \theta = (\theta_1, \theta_2)$. 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, denote by

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

the lifetime utility that each agent of initial productivity θ_1 expects 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 $y(\cdot)$ and $c(\cdot)$, and through the dependence of the period-2 distribution F_2 on period-1 income, with the dependence originating in LBD.

The planner's *dual problem* consists of maximizing expected intertemporal tax revenues

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

subject to the constraint that

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

and the constraint that the rule χ be *incentive compatible* (that is, each agent finds it optimal to generate income over the life cycle according to the policy $y(\cdot)$ and consume according to the policy $c(\cdot)$, as specified by the rule χ). The function $q : \Theta_1 \to \mathbb{R}_+$ 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 limit 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 $\varepsilon \to 0$, the redistribution environment of the Rawlsian social welfare function.

Implicit in the formulation of the planner's problem are the assumptions that agents do not privately save (equivalently, their savings are controlled directly by the planner) and the planner commits to the intertemporal tax code she chooses. Furthermore, consistently with the rest of the literature, we also restrict attention to economies in which both the first-best and the second-best allocations are interior at all histories.

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 $\chi = (y(\cdot), c(\cdot))$ when the agent's initial productivity is θ_1 and period-1 output/income is y_1, where y_1 is an arbitrary income level, potentially different from $y_1(\theta_1)$.

Given any policy χ, let

$$LD_1^{\chi}(\theta_1) \equiv \delta \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[y_2(\tilde{\theta}) - c_2(\tilde{\theta}) + \frac{v(c_2(\tilde{\theta})) - \psi(y_2(\tilde{\theta}), \tilde{\theta}_2)}{v'(c_1(\tilde{\theta}_1))} \right]. \tag{5}$$

Proposition 1. *The first-best policies $\chi^* = (y^*(\cdot), c^*(\cdot))$ satisfy the following optimality conditions with $\lambda[\chi^*]$-probability one:*[19]

$$\frac{\psi_y(y_1^*(\theta_1), \theta_1)}{v'(c_1^*(\theta_1))} = 1 + LD_1^{\chi^*}(\theta_1), \tag{6}$$

$$\frac{\psi_y(y_2^*(\theta), \theta_2)}{v'(c_2^*(\theta))} = 1, \tag{7}$$

$$v'(c_1^*(\theta_1)) = v'(c_2^*(\theta)), \tag{8}$$

$$v'(c_1^*(\theta_1))q(\theta_1) = v'(c_1^*(\theta_1'))q(\theta_1') \quad all \quad \theta_1, \theta_1' \in \Theta_1, \tag{9}$$

and

$$\int V_1(\theta_1)q(\theta_1)dF_1(\theta_1) = \kappa. \tag{10}$$

Conditions (6) and (7) describe the optimal output choices. In both periods, the first-best (FB) output policy equalizes each agent's marginal disutility of generating extra output with the marginal benefit of higher output. To express the agents' disutility in the same units as in the planner's objective function (tax revenues), the disutility is weighted by the inverse marginal utility of consumption. Naturally, the monetary cost of compensating the agents for the extra disutility of higher output is decreasing in their 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 more. In the first period, instead, the benefit of asking for higher output also takes into account the effect that the latter 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 $LD_1^{\chi}(\theta)$ in (5). Importantly, note that this function is computed holding fixed the period-2 income and consumption policies, as specified by the allocation rule $\chi = (y(\cdot), c(\cdot))$. The expectation in the formula for $LD_1^{\chi}(\theta)$ is thus with respect to the endogenous distribution over Θ under the rule χ, starting from period-1

[19]Hereafter we will always denote the FB policies with the superscript "*".

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

$$\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1,y_1(\theta_1)}\left[\frac{v(c_2(\tilde{\theta})) - \psi(y_2(\tilde{\theta}),\tilde{\theta}_2)}{v'(c_1(\tilde{\theta}))}\right],$$

instead, is the reduction in the period-1 compensation the planner must provide to type θ_1 to hold the agent's expected lifetime utility $V_1(\theta_1)$ constant, when asking the agent 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 consumption policy. Optimality requires the equalization of the marginal utility of consumption between any two consecutive histories θ_1 and $\theta = (\theta_1, \theta_2)$, and the equalization of the marginal utility of consumption, scaled by the Pareto weights q, 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 (IC) requires that highly productive workers be given informational rents to dissuade them from mimicking the less productive ones. Let

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

denote the period-2 continuation utility of an agent with productivity history $\theta = (\theta_1, \theta_2)$. Period-2 incentive-compatibility is equivalent to the requirements 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

$$V_2(\theta_1, \theta_2) = V_2(\theta_1, \underline{\theta}_2) - \int_{\underline{\theta}_2}^{\theta_2} \psi_\theta(y_2(\theta_1, s), s)ds, \tag{11}$$

and 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 is equivalent to the requirements that each agent's expected lifetime utility $V_1(\theta_1)$ be Lipschitz continuous and satisfy an envelope formula analogous to that in (11) and given by

$$V_1(\theta_1) = V_1(\underline{\theta}_1) - \int_{\underline{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1(s), s)ds + \delta \mathbb{E}^{\lambda[\chi]|s} \left[I_1^2(\tilde{\theta}, y_1(s)) \psi_\theta(y_2(s, \tilde{\theta}_2), \tilde{\theta}_2) \right] \right\} ds, \tag{12}$$

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

$$\begin{aligned} &\int_{\hat{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1(s), s) + \delta \mathbb{E}^{\lambda[\chi]|s, y_1(s)} \left[I_1^2(\tilde{\theta}, y_1(s)) \psi_\theta(y_2(s, \tilde{\theta}_2), \tilde{\theta}_2) \right] \right\} ds \\ \leq \; &\int_{\hat{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1(\hat{\theta}_1), s) + \delta \mathbb{E}^{\lambda[\chi]|s, y_1(\hat{\theta}_1)} \left[I_1^2(\tilde{\theta}, y_1(\hat{\theta}_1)) \psi_\theta(y_2(\hat{\theta}_1, \tilde{\theta}_2), \tilde{\theta}_2) \right] \right\} ds. \end{aligned} \tag{13}$$

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 (13). We verify that the omitted monotonicity conditions hold under the solution to the planner's relaxed program described below.

The planner's relaxed problem is similar to the one in Section 3, except for the fact that the agents' utilities must now satisfy constraints (11) and (12) above. Such constraints describe the rents the agents must receive to reveal their private information. The presence of such constraints 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 under the policies $\chi = (y(\cdot), c(\cdot))$ are given by*

$$W_1(\theta_1) \equiv 1 + LD_1^\chi(\theta_1) - \frac{\psi_y(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))} \qquad \textit{and} \qquad W_2(\theta) \equiv 1 - \frac{\psi_y(y_2(\theta), \theta_2)}{v'(c_2(\theta))}. \tag{14}$$

Recall that efficiency requires that the marginal cost of generating extra period-t income, in consumption terms, 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 expected sum of the planner's and the workers' payoffs, as captured by the function $LD_1^\chi(\theta)$ defined in (5). The period-t wedge W_t is the discrepancy between the marginal benefit and the marginal cost of higher period-t income. 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.

The wedges are related to the tax code *implementing* the policies χ by the following relationships.[20] Let $\mathcal{T} = (\mathcal{T}_t(\cdot))$ be a generic tax code, with $\mathcal{T}_t(y^t)$ denoting the total period-t tax payment

[20] A (differentiable and equi-Lipschitz continuous) tax code $\mathcal{T}$ implements the policies $\chi = (y(\cdot), c(\cdot))$ if, and only if,

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. Given any income policy $y(\cdot)$ induced by the tax code $\mathcal{T}$, for any t and θ^t, let $\hat{\mathcal{T}}_t(\theta^t) = \mathcal{T}_t(y^t(\theta^t))$ denote the period-t total tax paid by an individual at the productivity history θ^t and $\hat{\tau}_t(\theta^t) = \tau_t(y^t(\theta^t))$ the corresponding marginal tax rate.

Proposition 2. *The tax code $\mathcal{T} = (\mathcal{T}_t(\cdot))$ implements the policies $\chi = (y(\cdot), c(\cdot))$ only* if, with $\lambda[\chi]$-probability one,

$$W_1(\theta_1) = \hat{\tau}_1(\theta_1) + \delta \mathbb{E}^{\lambda[\chi]|\theta_1} \left[\frac{\partial \mathcal{T}_2(y_1(\theta_1), y_2(\tilde{\theta}))}{\partial y_1} \frac{v'(c_2(\tilde{\theta}))}{v'(c_1(\theta_1))} \right] + \delta \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[\hat{\mathcal{T}}_2(\tilde{\theta}) \right] \tag{15}$$

and $W_2(\theta) = \hat{\tau}_2(\theta)$.

The derivative in the second term in the right-hand side of (15) is the derivative of the period-2 tax bill with respect to the period-1 income y_1, holding the period-2 policies $(y_2(\cdot), c_2(\cdot))$ constant. In contrast, the derivative in the third term in the right-hand side of (15) is the derivative of the expected period-2 tax bill that one obtains by differentiating the distribution over θ_2, holding fixed the period-2 equilibrium tax function $\hat{\mathcal{T}}_2(\cdot)$.

Wedges are thus related to both present and future taxes. The dependence of future taxes on current incomes is both direct, through the dependence of future tax schedules on past incomes (the second term in the right-hand side of (15)), and indirect, through the dependence of the distribution of future productivity on current income (the third term in the right-hand side of (15)). Note that the second term in the right-hand side of (15) is not specific to economies with LBD; it is a natural feature of dynamic economies in which taxes are history-dependent. The third term, instead, is specific to economies with LBD. As we show in Section 5, the second-best allocations in the calibrated economy can be approximated with history-independent tax codes.[21] However, in the presence of LBD, even if taxes are history-independent, wedges and marginal tax rates do not coincide, due to the third term in the right-hand side of (15).

To understand the result in Proposition 2, note that, faced with the tax code $\mathcal{T} = (\mathcal{T}_t(\cdot))$, the period-1 income that a worker of productivity θ_1 chooses is given by the optimality condition (see the proof of Proposition 2 in the Appendix)

$$\begin{aligned} \frac{\psi_y(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))} = 1 - \hat{\tau}_1(\theta_1) - \delta \mathbb{E}^{\lambda[\chi]|\theta_1} \left[\frac{\partial \mathcal{T}_2(y_1(\theta_1), y_2(\tilde{\theta}))}{\partial y_1} \frac{v'(c_2(\tilde{\theta}))}{v'(c_1(\theta_1))} \right] \\ + \delta \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[\frac{V_2(\tilde{\theta})}{v'(c_1(\theta_1))} \right] \end{aligned} \tag{16}$$

The left-hand side of (16) is the marginal disutility of extra output. The first three terms in the right-hand side are the marginal effects of higher period-1 income on the utility of current and future

faced with the proposed tax code and with a (net of any linear capital-tax) interest rate on savings equal to $1/\delta - 1$, agents have incentives to generate the same income and consumption choices as specified by the policies χ.

[21]This result is consistent with what found in previous work focusing on economies without LBD; recall the discussion in the Introduction.

consumption. The last term is the marginal change in the worker's continuation utility due to the change in the period-2 productivity distribution. All these effects are in period-1 consumption terms.

In a first-best world, instead, where taxes can be made income-independent, the planner induces the worker to choose output according to the condition

$$\tfrac{\psi_y(y_1(\theta_1),\theta_1)}{v'(c_1(\theta_1))} = 1 + \delta \tfrac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[\tfrac{V_2(\tilde{\theta})}{v'(c_1(\theta_1))} \right] + \delta \tfrac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[y_2(\tilde{\theta}) - c_2(\tilde{\theta}) \right]. \tag{17}$$

The difference between the above two conditions is the period-1 wedge, $W_1(\theta_1)$.

From the above optimality conditions, it is easy to see that the first two terms in (15) capture the reduction in the utility the agent derives from his intertemporal consumption due to the dependence of taxes (in both periods) on period-1 income. The term

$$\frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[\hat{\mathcal{T}}_2(\tilde{\theta}) \right] = \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[y_2(\tilde{\theta}) - c_2(\tilde{\theta}) \right], \tag{18}$$

instead, is the extra benefit (to the planner) in terms of extra resources available in the second period which arises from shifting the period-2 productivity distribution.[22] This effect is not internalized by the worker and hence contributes to the period-1 wedge. As mentioned above, this effect is specific to economies with LBD and plays a fundamental role in the difference between wedges and tax rates under optimal tax codes, as we show in Section 5.

As is customary in the taxation literature (see, among others, Diamond, 1998, and Saez, 2001), hereafter, instead of focusing on the wedges W_t, which depend on the units of measure, we will focus on the wedges normalized by the marginal costs (in units of consumption)

$$\hat{W}_t(\theta^t) \equiv W_t(\theta^t) / \left(\frac{\psi_y(y_t(\theta^t), \theta_t)}{v'(c_t(\theta^t))} \right)$$

so as to obtain an absolute (i.e., percentage) measure of the marginal distortions. We will refer to $\hat{W}_t$ as the *relative wedges*.

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 (12), 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

[22]The equality in (18) follows from the fact that, for any θ, $y_2(\theta) - c_2(\theta) = \mathcal{T}(y_1(\theta_1), y_2(\theta)) - (y_1(\theta_1) - \mathcal{T}_1(\theta_1) - c_1(\theta_1))/\delta$, where $(y_1(\theta_1) - \mathcal{T}_1(\theta_1) - c_1(\theta_1))/\delta$ is the gross return on the period-1 savings. Because the latter does not depend on θ_2, the equality in (18) holds. See the Appendix for the details.

$V(\underline{\theta}_1) \geq \kappa$.[23] 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

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

Using the IC constraint (12), and integrating by parts, the above can be expressed as

$$R = \mathbb{E}^{\lambda[\chi]}\left[\sum_t \delta^{t-1}\left(y_t(\tilde{\theta}^t) - \psi(y_t(\tilde{\theta}^t), \tilde{\theta}_t) + \frac{I_1^t(\tilde{\theta}, y_1(\tilde{\theta}_1))}{\gamma_1(\tilde{\theta}_1)}\psi_\theta(y_t(\tilde{\theta}^t), \tilde{\theta}_t)\right) - V_1(\underline{\theta}_1)\right], \tag{19}$$

where $I_1^1(\theta, y_1(\theta_1)) \equiv 1$, for 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 (19) subject to the constraint that $V_1(\underline{\theta}_1) \geq \kappa$.

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 (19) coincides with the expectation of the information rents the planner must leave to the agents (over and above the utility $V_1(\underline{\theta}_1)$ given to the lowest period-1 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 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

$$\psi_y(y_2(\theta), \theta_2) - \frac{I_1^2(\theta, y_1(\theta_1))}{\gamma_1(\theta_1)}\psi_{\theta y}(y_2(\theta), \theta_2) = 1 \tag{20}$$

and

$$\psi_y(y_1(\theta_1), \theta_1) - \frac{1}{\gamma_1(\theta_1)}\psi_{\theta y}(y_1(\theta_1), \theta_1) - \delta\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)}\left[\frac{I_1^2(\tilde{\theta}, y_1(\theta_1))}{\gamma_1(\theta_1)}\psi_\theta(y_2(\tilde{\theta}), \tilde{\theta}_2)\right] = 1 + LD_1^\chi(\theta_1), \tag{21}$$

[23]Formally, 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 $\theta_1 > \underline{\theta}_1 + \varepsilon$, provided ε is sufficiently small.

where the function $LD_1^{\chi}(\theta_1)$ is as in (5). Note that, when the agents are risk neutral, $LD_1^{\chi}(\theta_1)$ does not depend on the agents' consumption choices. The above conditions thus pin down the optimal output schedules, independently of the consumption 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 the optimality condition (20). The marginal cost of asking for higher period-2 output from an agent of productivity history $\theta = (\theta_1, \theta_2)$ 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 $\theta_2' > \theta_2$. Such adjustment is necessary to guarantee that these latter types do not mimic type θ_2.[24] Because the above adjustment implies an increase by $-\psi_{\theta y}(y_2(\theta), \theta_2)[1 - F_2(\theta_2|\theta_1, y_1(\theta_1))]$ in the lifetime utility expected by those agents of period-1 productivity equal to θ_1, the planner can then reduce by the same amount the consumption of such agents to keep their expected lifetime utility constant. So far, the adjustment comes with no extra intertemporal rent for the agents. The problem is that the above adjustment also requires increasing by $-\psi_{\theta y}(y_2(\theta), \theta_2)\partial[1 - F_2(\theta_2|\theta_1, y_1(\theta_1))]/\partial\theta_1$ the consumption of all agents with period-1 productivity just above θ_1 while keeping their income constant. The increase in the rents of these latter agents is necessary because these latter agents, being more productive, if they were to produce the same period-1 earnings of those agents of period-1 productivity equal to θ_1, they would expect to receive the extra period-2 rent $-\psi_{\theta y}(y_2(\theta), \theta_2)$ with probability

$$1 - F_2(\theta_2|\theta_1 + \varepsilon, y_1(\theta_1)) > 1 - F_2(\theta_2|\theta_1, y_1(\theta_1)),$$

with ε small, due to the fact that, because of skill persistence, these agents are more likely to have a period-2 productivity exceeding θ_2 than those agents of period-1 productivity equal to θ_1. Furthermore, once the planner increases by $-\psi_{\theta y}(y_2(\theta), \theta_2)\partial[1-F_2(\theta_2|\theta_1, y_1(\theta_1))]/\partial\theta_1$ the consumption/rent of those agents of period-1 productivity equal to $\theta_1+\varepsilon$, she also needs to increase by the same amount

[24]To see this, consider an agent with productivity history $(\theta_1, \theta_2 + \varepsilon)$, with ε small. When the planner asks those agents of period-2 productivity history equal to $\theta = (\theta_1, \theta_2)$ to increase their labor income by one unit starting from $y_2(\theta)$, she needs to increase their consumption by $\psi_y(y_2(\theta), \theta_2)$. Furthermore, to discourage those agents of period-2 productivity history equal to $(\theta_1, \theta_2 + \varepsilon)$ from mimicking the former agents, she needs to increase the latter agents' consumption by $-\psi_{\theta y}(y_2(\theta), \theta_2)$, while keeping their income constant. The increase in the rents of these latter types is necessary because these latter agents, being more productive, if they were to produce the same period-2 earnings of those agents of period-2 productivity equal to θ_2, they would incur a smaller disutility of labor, with the cost-saving being equal to $-\psi_{\theta y}(y_2(\theta), \theta_2)$. Once the planner increases by $-\psi_{\theta y}(y_2(\theta), \theta_2)$ the consumption/rent of those agents of period-2 productivity history equal to $(\theta_1, \theta_2+\varepsilon)$, she then needs to increase by the same amount the consumption/rent of all agents of period-2 productivity history equal to (θ_1, θ_2'), with $\theta_2' > \theta_2 + \varepsilon$, for otherwise such agents would prefer to mimic their downward adjacent types, as formally implied by Condition (11).

the rent of all agents of period-1 productivity above $\theta_1 + \varepsilon$, for, otherwise, such agents would prefer to mimic their downward adjacent types.

Next, observe that the impulse response of θ_2 to θ_1 is equal to

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

that is, it is equal to the "extra" probability $\partial[1 - F_2(\theta_2|\theta_1, y_1(\theta_1))]/\partial\theta_1$ that a type just above θ_1 assigns to reaching a productivity level above θ_2 in period 2 relative to the probability assigned by type θ_1, normalized by the density $f_2(\theta_2|\theta_1, y_1(\theta_1))$.

Combining the above observations, we thus have that the marginal cost to the planner (in terms of extra intertemporal rents) of increasing the period-2 income of those agents of period-2 productivity history equal to $\theta = (\theta_1, \theta_2)$ above $y_2(\theta)$ is equal to

$$-\psi_{\theta y}(y_2(\theta), \theta_2) I_1^2(\theta, y_1(\theta_1)) f_2(\theta_2|\theta_1, y_1(\theta_1))[1 - F_1(\theta_1)].$$

The optimality condition in (20) then equalizes the (net) marginal benefit $1 - \psi_y(y_2(\theta), \theta_2)$ of asking for a higher period-2 income at history $\theta = (\theta_1, \theta_2)$ with the above marginal cost, while accounting for the fact that the benefit occurs with probability density $f_1(\theta_1) f_2(\theta_2|\theta_1, y_1(\theta_1))$. Relative to the marginal benefit, the above 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. Furthermore, because this extra marginal cost is increasing in y_2, at the second-best optimum, the earnings of each worker of period-2 productivity history θ are distorted downwards relative to their first-best level.[25]

Next, consider the optimal choice of period-1 income, as determined by (21). The marginal benefit of asking for higher period-1 output naturally takes into account the effect of changing the distribution of period-2 productivity due to LBD. This extra marginal benefit is determined by the same function $LD_1^\chi(\theta_1)$ as in the first-best case. However, for any (θ_1, y_1), the value of the function $LD_1^\chi(\theta_1)$ is different than under the first-best policies because the second-best period-2 income policy $y_2(\cdot)$ is distorted downwards relative to its first-best counterpart, as explained above.

The marginal cost of asking for higher period-1 output has three parts, corresponding to the three terms in the left-hand side of (21). The first term, $\psi_y(y_1(\theta_1), \theta_1)$, is the increase in the consumption the planner has to provide to those agents of period-1 productivity equal to θ_1 when she asks them to increase their income above $y_1(\theta_1)$. This term is the same as under symmetric information. The second and third terms in the left-hand side of (21) are extra costs due to the fact that, under asymmetric information, when the planner asks those agents of period-1 productivity equal to θ_1 to increase their income, she then needs to increase the consumption/rent of all agents of period-1

[25]Recall that $\psi_{\theta yy}(y_t, \theta_t) \leq 0$.

productivity above θ_1, while keeping these latter agents' income constant. The reasons are analogous to the ones discussed above for period-2 output.

In particular, consider those agents of period-1 productivity equal to $\theta_1 + \varepsilon$. When the planner increases marginally the period-1 income of the θ_1-agents, starting from $y_1(\theta_1)$, and adjusts the latter agents' consumption by $\psi_y(y_1(\theta_1), \theta_1)$, the $(\theta_1 + \varepsilon)$-agents, if they were to mimic the θ_1-agents, would benefit in two ways. First, they would incur a lower disutility of labor, with the cost-saving being equal to $-\psi_{y\theta}(y_1(\theta_1), \theta_1)$. Second, because of LBD, the θ_1-agents, when they increase their period-1 earnings above $y_1(\theta_1)$, they also increase their expected continuation utility by $\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)}\left[V_2(\tilde{\theta})\right]$. Because the $(\theta_1+\varepsilon)$-agents are more likely to be highly productive in the second period than the θ_1-agents, they expect a higher continuation utility than the θ_1-agents when generating the same period-1 earnings as the latter agents. The extra continuation utility expected by the $(\theta_1 + \varepsilon)$-agents is equal to[26]

$$\frac{\partial}{\partial \theta_1}\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)}\left[V_2(\tilde{\theta})\right] = \frac{\partial}{\partial y_1}\frac{\partial}{\partial \theta_1}\mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)}\left[V_2(\tilde{\theta})\right] = \frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)}\left[I_1^2(\tilde{\theta}, y_1(\theta_1))\frac{\partial V_2(\tilde{\theta})}{\partial \theta_2}\right] = \frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)}\left[-I_1^2(\tilde{\theta}, y_1(\theta_1))\psi_\theta(y_2(\tilde{\theta}), \tilde{\theta}_2)\right]. \tag{22}$$

To guarantee that the $(\theta_1 + \varepsilon)$-agents do not mimic the θ_1-agents, the planner must then increase the consumption/rent of the $(\theta_1 + \varepsilon)$-agents by

$$-\psi_{\theta y}(y_1(\theta_1), \theta_1) + \delta\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)}\left[-I_1^2(\tilde{\theta}, y_1(\theta_1))\psi_\theta(y_2(\tilde{\theta}), \tilde{\theta}_2)\right] \tag{23}$$

when asking the θ_1-agents to increase their period-1 earnings above $y_1(\theta_1)$ (the formula in (23) is for the limit case in which ε vanishes). Furthermore, she must also increase by the same amount the consumption/rent of all agents of period-1 productivity above $\theta_1 + \varepsilon$ to discourage these agents from mimicking their downward adjacent types.

The weight the planner assigns to increasing the rents of those agents of period-1 productivity above θ_1, relative to the weight she assigns to asking the θ_1-agents for higher period-1 output, is equal to the inverse hazard rate $1/\gamma_1(\theta_1) = [1 - F_1(\theta_1)]/f_1(\theta_1)$ of the period-1 productivity distribution. It follows that, relative to the marginal benefit, the marginal cost of asking for higher period-1 output from those agents of period-1 productivity equal to θ_1, due to asymmetric information, is equal to the sum of the second and third terms in the left-hand side of (21). The second term is the familiar one from 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 (21), which is equal to zero when period-2 productivity is exogenous.

[26]The first equality in (22) is self-evident. The second equality follows from the fact that, given any Lipschitz continuous function $J(\theta_2)$, $\frac{\partial}{\partial \theta_1}\mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)}\left[J(\tilde{\theta}_2)\right] = \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)}\left[I_1^2(\tilde{\theta}, y_1(\theta_1))\frac{\partial J(\tilde{\theta}_2)}{\partial \theta_2}\right]$, where the first derivative is obtained by differentiating the measure $F_2(\cdot|\theta_1, y_1(\theta_1))$, holding y_1 fixed at $y_1(\theta_1)$. The third equality follows from the fact that IC requires that $\partial V_2(\theta_1, \theta_2)/\partial \theta_2 = -\psi_\theta(y_2(\theta_1, \theta_2), \theta_2)$.

To better understand the last term in the left-hand side of (21), note that, under LBD, when the planner asks the θ_1-agents to increase their period-1 income above $y_1(\theta_1)$, she affects those agents' expected period-2 rents – equivalently, the expected period-2 handicaps $h_2(\theta, y(\theta))$ – through two channels. The first is through the change in the distribution of θ_2, holding fixed the period-2 handicaps h_2. 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. Using the optimality conditions (20) and (21), we have that, with LBD, the relative wedges under optimal tax codes are given by (with $\lambda[\chi]$-probability one)

$$\hat{W}_1(\theta_1) = \hat{W}_1^{NOLBD}(\theta_1) + \Omega(\theta_1) \tag{24}$$

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

$$\hat{W}_t^{NOLBD}(\theta) \equiv -\frac{I_1^t(\theta^t, y_{t-1}(\theta^{t-1}))}{\gamma_1(\theta_1)} \frac{\psi_{\theta y}(y_t(\theta^t), \theta_t)}{\psi_y(y_t(\theta^t), \theta_t)}. \tag{25}$$

are the same functions describing the period-t relative wedges in the absence of LBD[27] and where the function

$$\Omega(\theta_1) \equiv -\delta \frac{\frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[\frac{I_1^2(\tilde{\theta}, y_1(\theta_1))}{\gamma_1(\theta_1)} \psi_\theta(y_2(\tilde{\theta}), \tilde{\theta}_2) \right]}{\psi_y(y_1(\theta_1), \theta_1)} \tag{26}$$

summarizes the novel effects due to the change in the period-2 distribution, originating in LBD.[28] This new term measures the (discounted) marginal 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 extra marginal rents, due to LBD, that the planner must leave to those agents with period-1 productivity above θ_1 when she increases the period-1 income of the θ_1-agents above $y_1(\theta_1)$.

We turn to the effects of LBD on the level, dynamics, and progressivity of the optimal relative wedges. When the disutility of labor is iso-elastic, and period-2 productivity is given by (2), we have that

$$\hat{W}_t^{NOLBD}(\theta^t) = \rho^{t-1} \frac{1+\phi}{\theta_1 \gamma_1(\theta_1)}, \tag{27}$$

[27] That is, the functions in (25) are the same functions describing the wedges under optimal tax codes in a fictitious economy in which the productivity process over Θ is exogenous and coincides with the one induced by the second-best allocations in the economy with LBD. Note, however, that the policies $y(\cdot)$ in $\hat{W}_t^{NOLBD}(\theta)$ are the second-best policies for the economy with LBD and not those for the economy without it.

[28] The formal proof of the claim that relative wedges under optimal codes are consistent with the decomposition in (24) is in the Appendix (see the proof of Proposition 4).

and

$$\Omega(\theta_1) = \frac{\delta\rho}{\psi_y(y_1(\theta_1), \theta_1)} \hat{W}_1^{NOLBD}(\theta_1) \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[\psi(y_2(\tilde{\theta}), \tilde{\theta}_2) \right]. \tag{28}$$

Under this specification, the impulse responses $I_1^t(\theta, y_1)$ are invariant to period-1 income, and the novel effects due to LBD are summarized by the impact of y_1 on the expectation of the period-2 disutility of labor $\psi(y_2(\theta), \theta_2)$. Moreover, $\psi(y_2(\theta), \theta_2)$ is nondecreasing in period-2 productivity θ_2. As a result, $\Omega(\theta_1) > 0$, meaning that LBD contributes to higher period-1 relative wedges. The reason is that by making the θ_1-agents less productive in period 2, the planner reduces the θ_1-agents' expected continuation utility. This permits the planner to reduce the rent she must leave to all agents with period-1 productivity above θ_1 (see the discussion above for the details of the mechanism).

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

Finally, consider the effects of LBD on the progressivity of the relative 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 relative wedges to be nonincreasing in productivity, at the top. LBD can contribute positively or negatively to the progressivity of the period-1 relative wedges depending on whether Ω is increasing or decreasing in θ_1. To illustrate, suppose that the period-1 productivity θ_1 and the 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 relative 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 relative wedges and to a higher progressivity of the period-1 relative 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, stronger LBD effects (captured by a higher level of the parameter ζ in (2)) are responsible for higher period-1 relative wedges and for more progressivity at all income percentiles, but in particular at the top.[29]

These properties extend to other distributions of the skill shocks. For instance, the right-hand panel of Figure 1 illustrates the period-1 relative wedge as a function of the period-1 income percentile for a Pareto-lognormal skills distribution F_1 with Pareto-tail parameter equal to $\xi = 5$, and the Pareto right tail applying to income percentiles over the 85^{th}, as in Diamond (1998). As the figure shows, LBD contributes to higher and more progressive relative 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

[29]The figure plots the relative period-1 wedge $\hat{W}_1$ as a function of the period-1 income percentile. 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).

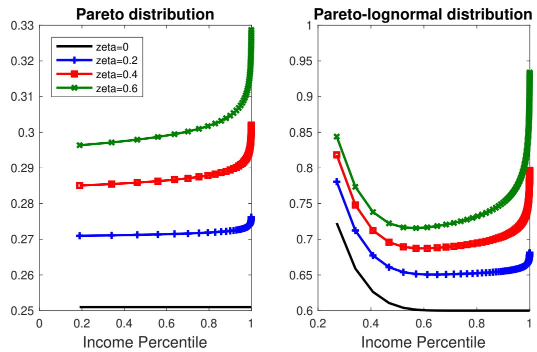

Figure 1: Period-1 relative wedges in Rawlsian-risk-neutral case ($\rho = 1$; $\phi = 2$)

enough to turn the optimal period-1 relative wedge from regressive to progressive only at sufficiently high income percentiles.

We summarize the above results in Proposition 3 below, whose proof is in the online Supplementary Material. We first need to introduce some definitions and notation.

Definition 2. *The period-1 relative 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, $\partial \hat{W}_1(\theta_1)/\partial\theta_1 > \partial \hat{W}_1^{NOLBD}(\theta_1)/\partial\theta_1$ for $\theta_1 \in (\theta_1', \theta_1'')$. The period-1 relative wedge under LBD is more progressive than the period-1 relative wedge in the absence of LBD if, and only if, $\partial \hat{W}_1(\theta_1)/\partial\theta_1 \geq \partial \hat{W}_1^{NOLBD}(\theta_1)/\partial\theta_1$ for all θ_1, with the inequality strict for a subset $(\theta_1', \theta_1'') \subset \Theta_1$.*

Using the decomposition in (24), we have that, when $\theta_1\gamma_1(\theta_1)$ is nondecreasing in θ_1 and the disutility of labor takes the iso-elastic form in (1), the period-1 relative 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 3. *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) if $\hat{W}_1^{NOLBD}(\theta_1)$ is nonincreasing, then the solution to the relaxed program also solves the full program.*

When 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), in the absence of LBD, $\hat{W}_1^{NOLBD}(\theta_1) = (1+\phi)/M$ for all θ_1, whereas, in the presence of LBD, $\partial \hat{W}_1(\theta_1)/\partial\theta_1 > 0$ for all $\theta_1 \in \Theta_1 = \mathbb{R}_+$.

Hence, when the disutility of labor is iso-elastic and period-2 productivity is given by the specification in (2), LBD contributes to higher period-1 relative wedges across all productivity levels and to a higher differential between period-1 and period-2 relative wedges, across all histories $\theta = (\theta_1, \theta_2)$. Whether LBD contributes to a higher progressivity of the period-1 relative 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 3 in the Supplementary Material identifies a condition under which the same conclusion extends to other distributions.[30]

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 earlier.

Proposition 4. *The relative wedges under the optimal tax code are given by (with $\lambda[\chi]$-probability one)*

$$\hat{W}_1(\theta_1) = \hat{W}_1^{NOLBD}(\theta_1) + [RA(\theta_1) - D(\theta_1)]\,\Omega(\theta_1), \tag{29}$$

and $\hat{W}_2(\theta) = \hat{W}_2^{NOLBD}(\theta)$. The functions $\hat{W}_1^{NOLBD}(\theta_1)$ and $\hat{W}_2^{NOLBD}(\theta)$ are the same functions describing the relative wedges in the absence of LBD introduced in (25), $\Omega(\theta_1)$ i*s the function capturing the novel effects of LBD in the benchmark economy with risk-neutral agents and Rawlsian preferences for redistribution as defined in (26),*

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

is a correction term due to risk aversion, and

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

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 the result in Proposition 4 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 in the determination of the relative wedges under optimal tax codes.

Recall from the discussion in the Rawlsian-risk-neutral case that, when the planner asks type θ_1 to increase her period-1 income above $y_1(\theta_1)$, she then needs to increase by

$$-\psi_{\theta y}(y_1(\theta_1), \theta_1) + \delta \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[-I_1^2(\tilde{\theta}, y_1(\theta_1)) \psi_\theta(y_2(\tilde{\theta}), \tilde{\theta}_2) \right] \tag{30}$$

[30]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 relative wedges documented in Proposition 3 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 relative 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 3 continue to hold.

the consumption/rent of the $(\theta_1+\varepsilon)$-agents (Again, the formula in (30) is for ε vanishing). As explained above, the second term in (30) is the one specific to LBD. Thus focus on the second term. When the agents are risk-averse, the marginal utility that type $(\theta_1+\varepsilon)$ obtains from this increase in consumption is equal to

$$v'(c_1(\theta_1+\varepsilon))\delta\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1,y_1(\theta_1)}\left[-I_1^2(\tilde{\theta},y_1(\theta_1))\psi_\theta(y_2(\tilde{\theta}),\tilde{\theta}_2)\right] \tag{31}$$

Next note that incentive compatibility requires that the utility of all types above $\theta_1+\varepsilon$ be increased by the amount in (31). The total cost to the planner (in consumption units) of providing this extra utility to the latter types, accounting for the heterogeneity in the latter types' marginal utility of consumption, and normalized by the weighted disutility cost $\psi_y(y_1(\theta_1),\theta_1)f_1(\theta_1)$ to obtain an absolute measure, is thus equal to

$$\frac{\int_{\theta_1+\varepsilon}^{\bar{\theta}_1}\frac{1}{v'(c_1(s))}\frac{dF_1(s)}{1-F_1(\theta_1+\varepsilon)}}{\psi_y(y_1(\theta_1),\theta_1)}\frac{1-F_1(\theta_1+\varepsilon)}{f_1(\theta_1)}\left[v'(c_1(\theta_1+\varepsilon))\delta\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1,y_1(\theta_1)}\left[-I_1^2(\tilde{\theta},y_1(\theta_1))\psi_\theta(y_2(\tilde{\theta}),\tilde{\theta}_2)\right]\right]$$

Taking the limit as ε goes to zero and rearranging yields the term $RA(\theta_1)\Omega(\theta_1)$ in (29).

Next, consider the term $D(\theta_1)\Omega(\theta_1)$ in (29). The correction term $D(\theta_1)$ controls for the higher Pareto weights the planner assigns to all agents whose initial productivity is above $\underline{\theta}_1$, relative to the benchmark with Rawlsian preferences for redistribution. Other things equal, this term naturally contributes to lower relative 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 $v'(c_1(\theta_1))$ comes with the benefit of relaxing the redistribution constraint (4). The benefit for the planner in revenue terms is equal to[31]

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

Arguments similar to those above for the explanation of the term $RA(\theta_1)\Omega(\theta_1)$ then imply that the reduction in the total cost of providing all types above θ_1 with the extra utility in (31), normalized by $\psi_y(y_1(\theta_1),\theta_1)f_1(\theta_1)$, is equal to the term $D(\theta_1)\Omega(\theta_1)$ in (29).

[31]Observe that $\int_{\theta_1}^{\bar{\theta}_1}q(s)\frac{dF_1(s)}{1-F_1(\theta_1)}$ is the weighted-average benefit of increasing by one util the expected lifetime utility of each 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 $\int_{\underline{\theta}_1}^{\bar{\theta}_1}\frac{1}{v'(c_1(s))}dF_1(s)$, 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 util 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 $\int_{\underline{\theta}_1}^{\bar{\theta}_1}\frac{1}{v'(c_1(s))}dF_1(s)$. The product of the two terms $\int_{\theta_1}^{\bar{\theta}_1}q(s)\frac{dF_1(s)}{1-F_1(\theta_1)}$ and $\int_{\underline{\theta}_1}^{\bar{\theta}_1}\frac{1}{v'(c_1(s))}dF_1(s)$ thus represents the weighted-average benefit, in revenue terms, of increasing by one util the expected lifetime utility of all period-1 types above θ_1.

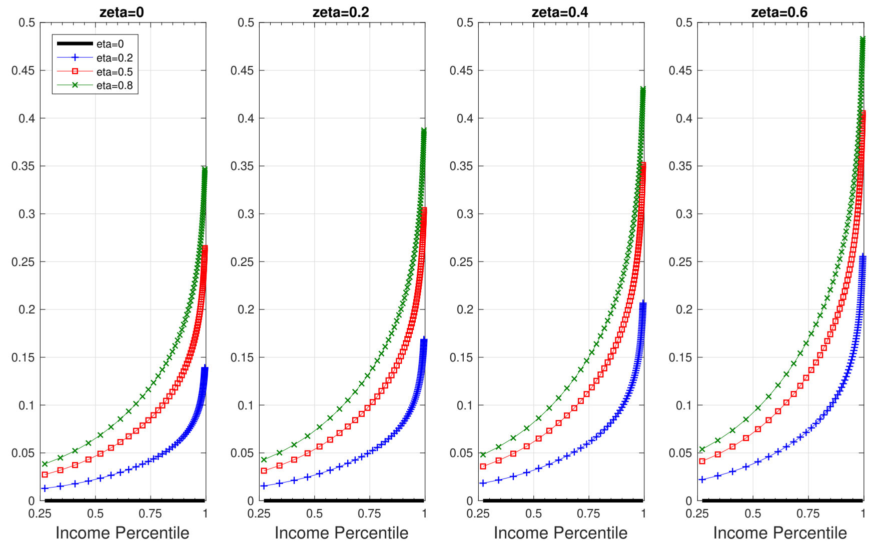

Figure 2: Risk aversion effects on $\hat{W}_1$ in Utilitarian Pareto-lognormal case ($\rho = 1$; $\phi = 2$)

In the online Supplementary Material, we discuss how the correction term $RA - D$ interacts with the uncorrected LBD term Ω in the determination of the period-1 relative 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 reduce the expected 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 relative 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 relative wedges depends critically on the shape of the skills distribution.[32] To illustrate this point, Figures 2 and 3 depict the period-1 relative 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

[32]For a related discussion, see also Golosov et at. (2016).

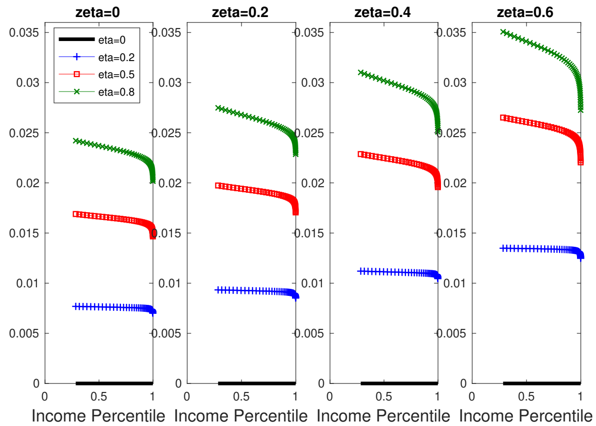

Figure 3: Effects of risk aversion on $\hat{W}_1$ in Utilitarian lognormal case ($\rho = 1$; $\phi = 2$)

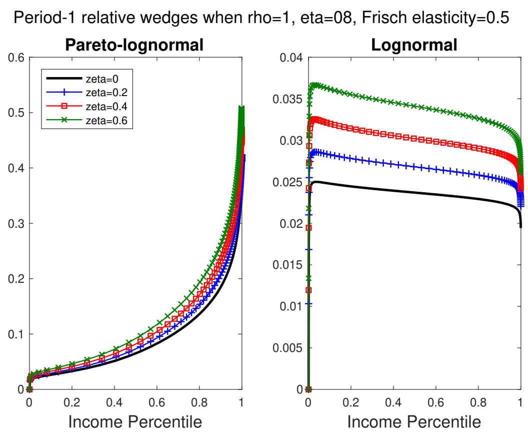

Figure 4: Effects of LBD on $\hat{W}_1$ in Utilitarian Pareto-lognormal and lognormal case ($\eta = .8$; $\phi = 2$)

the case where θ_1 and ε_2 are drawn from the Lognormal distribution in Farhi and Werning (2013).[33] When the agents are risk averse, relative 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 (although only mildly) the progressivity of the period-1 relative wedges, whereas the opposite is true in the Lognormal case, as can be seen from Figure 4 (which depicts, as an example, the value of risk aversion $\eta = 0.5$ at the middle of the range considered above). These observations will be important for the interpretation of the quantitative results in the next section.

[33] Specifically, 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).

5 Quantitative Analysis

We now turn to the quantitative implications of our analysis.[34] 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 relative 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, yet quasi-optimal, tax codes in which taxes are (history-independent). In Subsection 5.4, we study comparative statics of such quasi-optimal tax codes with respect to the intensity of the LBD effects and the exogenous persistence of the workers' productivity. In Subsection 5.5, we illustrate the importance of the stochasticity of the LBD effects for the structure of quasi-optimal tax codes. Finally, in Subsection 5.6, we isolate the role of LBD by conducting a counterfactual analysis where we compare the quasi-optimal tax code in the calibrated economy to its counterpart in an economy that features the same productivity process as in the calibrated economy but where the latter is exogenous.

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 at., 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 output 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 $\hat{\beta}_s / \sum_{r=1}^{20} \hat{\beta}_r$, $s = 1, ..., 20$, declining over time, and with each $\widehat{\beta}_s$ being a positive scalar. This specification is 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 $(\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_{20})$ is proportional to $(1, \beta, ..., \beta^{19})$, where β is the annual discount factor, and the latter is equal to 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.[35]

[34] Details on the computations can be found in the online Supplementary Material.

[35] See the online Supplementary Material for a formal proof of the equivalence between the two economies.

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

To calibrate the model, we first fix a few parameters to the levels typically assumed in the literature (see Table 1).[36] 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 (capturing the initial human capital) and where ε_1 is a random variable described below. Period-2 productivity, instead, is given by

$$\theta_2 = \theta_1^{\rho} \left(\frac{\sum_{s=1}^{20} \beta^{s-1} y_s}{\sum_{s=1}^{20} \beta^{s-1}} \right)^{\zeta} \varepsilon_2, \tag{32}$$

where ε_2 is a random variable described below. The specification in (32) is the same as in (2), except for the fact that y_1 is replaced by

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

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. We truncate the distribution at the 1^{st} percentile and set μ so that the mean of the truncated distribution is equal to one. The parameter σ governs the variance of the distribution for the given tail parameter ξ.[37]

[36] The 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 at. (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.

[37] A Pareto-lognormal distribution G has support $(0, \infty)$, density $g(\epsilon) = \frac{\xi}{\epsilon^{1+\xi}} \exp(\xi\mu + \xi^2 \frac{\sigma^2}{2}) \Phi(\frac{log(\epsilon) - \mu - \xi\sigma^2}{\sigma})$, and cdf $G(\epsilon) = \Phi(\frac{log(\epsilon) - \mu}{\sigma}) - \frac{1}{\epsilon^{\xi}} \exp(\xi\mu + \xi^2 \frac{\sigma^2}{2}) \Phi(\frac{log(\epsilon) - \mu - \xi\sigma^2}{\sigma})$, 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.

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

$$T(y) = y - e^{\alpha_0} y^{1-0.181} \tag{33}$$

of the existing US income tax code from Heathcote et. al. (2017), with the parameter $\alpha_0 = -0.1005$ set so that the total tax revenues are normalized to zero.[38] 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.[39] 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 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' life cycle, using year 21 as the cutoff age.[40] Table 2 reports the calibrated parameters, the target moments, and the absolute percentage deviations of the model-generated moments from the target moments.[41] 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 relative wedges

Given the above calibration, we then characterize the optimal relative wedges for the calibrated economy.

We assume that the planner has Utilitarian preferences for redistribution.[42] We solve the pri-

[38]Golosov et. al. (2016) use a similar estimation, but from the 2014 version of the Heathcote et. al. paper. Namely, they assume that $T(y) = y - e^{\alpha_0} y^{1-0.151}$.

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

[40]Contrary 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.

[41]As 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.

[42]The 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

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

mal to the planner's problem as described in Section 2. That is, we identify the allocations that maximize the sum of the agents' expected lifetime utility subject to the constraint that expected intertempotral 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 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 (13) are satisfied.

We first report that reforming the existing tax code by adopting the optimal one would yield an increase in the agents' utility equal to that generated by a 4.04% equiproportionate increase in consumption at each history, starting from the allocations under the existing US tax code, while keeping output choices constant.[43] This appears to be a significant improvement, although it is lower than the figure that one would obtain in a counterfactual economy that is identical to the calibrated one except for the fact that the productivity process is exogenous as is the case when there are no LBD effects (see the discussion in Section 5.6).

We now turn to the optimal relative wedges in the calibrated economy. Period-1 relative 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 relative wedge, focusing on the top 75% of the distribution. The inverse-U shape of the relative wedges appears to follow from the fact that the calibrated Pareto-lognormal distribution has a Pareto right tail only asymptotically, i.e., for productivity levels exceeding the 99.9^{th} 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, relative wedges have an inverted U-shape (see Golosov et al. (2016) as well as the

(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.

[43]This 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 $\Delta V = 0.057685$. Therefore, when translated in consumption terms, it is as if the reform yields an equiproportionate increase in consumption at every history equal to $(e^{\Delta V})^{\frac{1}{1+\delta}} - 1 = 0.040403$, that is 4.04%.

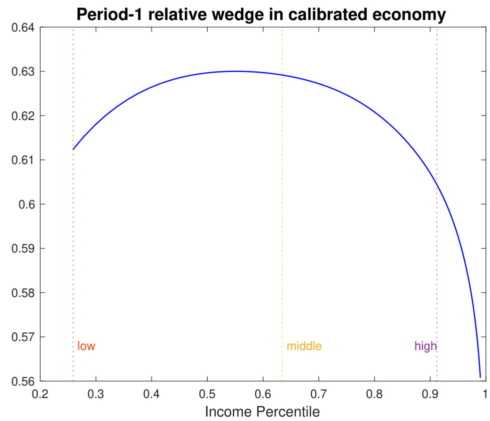

Figure 5: Period-1 optimal relative wedge. Vertical lines indicate period-1 income percentiles for middle earnings (approximately equal to the mean earnings), low earnings (approximately equal to half of the mean earnings) and high earnings (approximately equal to twice the mean earnings).

	Period-1 wedge	Period-2 wedge
Low earnings	0.6124	0.9654
Middle earnings	0.6292	1.0269
High earnings	0.6044	0.9849

Table 3: Optimal relative wedges for selected histories

discussion at the end of the previous section).

The figure also highlights three selected income percentiles corresponding to low, middle, and high earnings. The middle earnings correspond to earnings approximately equal to the period-1 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 relative 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 relative wedges also have an inverse-U shape. However, contrary to their period-1 counterparts, period-2 relative wedges decrease after the 30^{th} percentile. Period-1 relative wedges thus exhibit progressivity over a range of income percentiles for which the period-2 relative wedges are regressive (e.g., between the 30^{th} and 60^{th} percentile). Also note that, at any percentile of the period-2 distribution, period-2 relative wedges are increasing in period-1 incomes, across the selected histories described above.

Next, we discuss the dynamics of the relative wedges. Table 3 reports the relative wedges for three particular histories of productivity shocks. 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 these

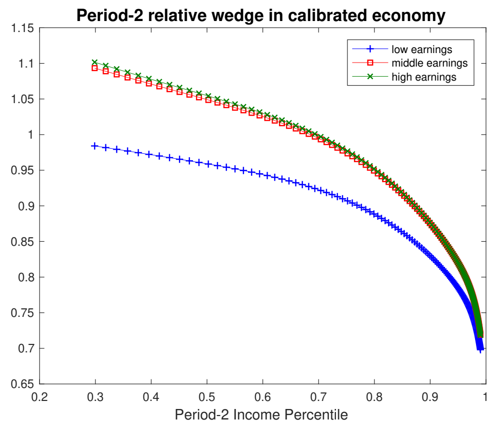

Figure 6: Period-2 relative wedges as function of period-2 income percentiles for low, middle, and high period-1 earnings, where middle earnings correspond to mean period-1 earnings, low earnings correspond to half of the mean of period-1 earnings, and high earnings correspond to twice the mean of period-1 earnings.

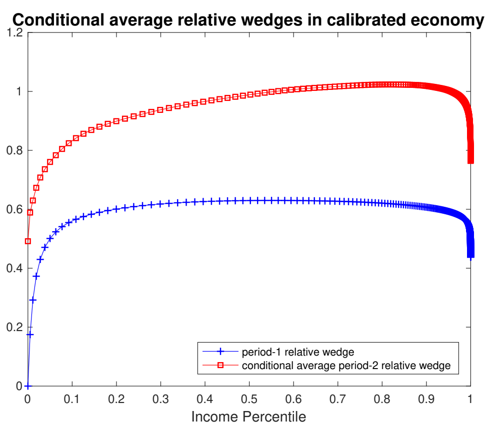

Figure 7: Period-1 and conditional period-2 relative wedges as function of period-1 income percentile

histories, relative wedges increase over the life cycle. While there exist histories for which the period-2 relative wedge is lower than the corresponding period-1 relative wedge, the conditional average period-2 relative wedge is higher than the corresponding period-1 relative wedge at any given period-1 income percentile, as can be seen from Figure 7 (that is, $\mathbb{E}[\hat{W}_2(\theta_1,\tilde{\theta}_2)|\theta_1, y_1(\theta_1)] > \hat{W}_1(\theta_1))$. The average period-1 relative wedge (across all histories) is equal to 0.5984, whereas the average period-2 relative wedge (also across all histories) is equal to 0.9507.

The property that relative wedges tend to increase over the life cycle is consistent with the results in Farhi and Werning (2013) and Golosov et al. (2016). In the calibrated economy, agents are highly risk averse (the coefficient of relative risk aversion is equal to 1) and are exposed to significant risk (the variance of the period-2 productivity shock is equal to 0.41). It is known that, in such circumstances, in the absence of LBD, relative wedges increase over the life cycle. While distorting income downwards in either period helps reducing the workers' information rents, the distortions in period 2 also help reducing the volatility of period-2 earnings and hence of period-2 consumption. As a result, the planner distorts the income choices of the old workers more than those of the young ones (see, for example, Farhi and Werning, 2013, and Golosov et al., 2016).[44] Our results show that, when the intensity of the LBD effects is of the magnitude of the calibrated economy, the same dynamics obtain in the presence of LBD.[45]

Next, consider the progressivity of the relative 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 the calibrated Pareto-tail parameter ξ only asymptotically. As discussed at the end of the previous section, under such a distribution, relative wedges should not be expected to be progressive over the entire range of income percentiles.[46]

5.3 Approximate Optimality of History-invariant Tax Codes

We now turn to the question of whether simple age-dependent by history-invariant tax codes can yield most of the welfare gains from the optimal reform of the current US tax code. Consider the following class of tax codes

$$T_1(y_1) = -B + y_1 - e^{\alpha_{0,1}} y_1^{1-\alpha_1} \qquad \text{and} \qquad T_2(y_2) = y_2 - e^{\alpha_{0,2}} y_2^{1-\alpha_2}.$$

Observe that the tax rates in period t are increasing in α_t. The special case of linear taxes corresponds to $\alpha_1 = \alpha_2 = 0$, in which case the constant tax rates are equal to $1 - e^{\alpha_{0,1}}$ and $1 - e^{\alpha_{0,2}}$ 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 $B = 0$, $\alpha_1 = \alpha_2 = 0.181$ and $\alpha_{0,1} = \alpha_{0,2} = \alpha_0$, with α_0 chosen so that average tax revenues are normalized to zero.

To derive the optimal tax code within this class, we solve for the values of $B, \alpha_{0,1}, \alpha_{0,2}, \alpha_1$ and α_2 that maximize the workers' expected lifetime utility, subject to the constraint that expected intertemporal tax revenues be at least zero (that is, the same level as 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 tax code is given by $B = 0.2603$, $\alpha_{0,1} = -0.4769$, $\alpha_{0,2} = -0.6231$, $\alpha_1 = 0.0055$,

[44] It is also well known that, when the workers are risk neutral and the impulse responses of the workers' productivity in later periods to their initial productivity decline over time (as is the case in most models and also in our calibrated economy), period-1 distortions are more effective at containing the workers' rents than period-2 distortions. As a result, under risk neutrality, distortions decrease over the life cycle. These dynamics are reversed when the level of the agents' risk aversion grows large – see also Garrett and Pavan (2015).

[45] See also Subsections 5.5 and 5.6 for the role of the stochasticity and of the endogeneity of the workers' productivity for the dynamics of optimal taxes.

[46] Note that the shape of the relative wedges for high earnings percentiles in the last two figures is similar to the one in Figure 5 in Golosov et al. (2016).

and $\alpha_2 = -0.0186$. The code is mildly progressive for young workers and mildly regressive for old ones. Interestingly, it yields almost all of the welfare gains of reforming the existing tax code by adopting the fully-optimal code (the one implementing the second-best allocations associated with the fully-optimal relative 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 4.04% at all histories, starting from the allocations under the current US tax code, adopting the quasi-optimal tax code yields an increase in expected lifetime utility equal to a 3.89% equiproportionate increase in consumption. The loss from the quasi-optimal tax code is thus only 0.15% in consumption terms. The income and consumption allocations under the quasi-optimal code are also close to the ones under the fully-optimal code. Namely, the absolute percentage deviation in consumption across histories has mean around 5% and standard deviation around 3%. The absolute percentage deviation in income across histories has a mean around 2.5% and standard deviation around 1.2%. Therefore, the quasi-optimal tax code is approximately optimal.

Importantly, virtually all of the welfare gains from adopting the quasi-optimal tax code can also be generated by adopting a code where taxes are linear. Adopting the optimal linear tax code yields an increase in expected lifetime utility equal to a 3.88% equiproportionate increase in consumption at all histories, starting from the allocations under the current US tax code, which is only 0.01% less than the increase under the quasi-optimal tax code. The tax rates in the optimal linear tax code are equal to 38% and 46% for young and old workers, respectively.[47]

In Figure 8, we plot together the 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). Compared to the current US tax code, the quasi-optimal code features higher tax rates for young workers, whereas, for old workers, features higher tax rates at low income percentiles, but lower tax rates at high income percentiles. The quasi-optimal code is less progressive than the current US one for young workers and is regressive, instead of progressive, for old workers. Finally, under the quasi-optimal tax code, the differential between the average tax rate for old and for young workers is smaller than under the current US tax code. Also note that tax rates in the quasi-optimal tax code are very close to those in the linear optimal tax code.

Next, suppose the planner is constrained to use taxes that are not just history-invariant but also age-independent (that is, $\alpha_{0,1} = \alpha_{0,2}$ and $\alpha_1 = \alpha_2$). The welfare gains, in consumption terms, from replacing the current US tax code with the optimal age-independent one are equal to a 3.80% equiproportionate increase in consumption at all histories, starting from the allocations under the current US tax code, which is only 0.08% less than under the quasi-optimal code.[48] If, in addition, the planner was constrained to use linear taxes, the optimal tax rate would be 41.25%. The welfare gains in consumption terms from replacing the current US tax code with the optimal age-independent

[47] That 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).

[48] The optimal age-independent tax code is given by $B = 0.2624$, $\alpha_0 = -0.5312$ and $\alpha = 0.0022$. Such code is close to an average of the two schedules of the quasi-optimal tax code.

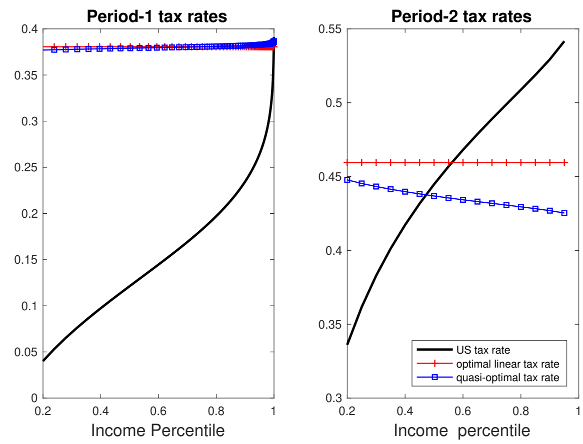

Figure 8: Tax rates as functions of income percentiles

linear one are virtually the same as under the optimal age-independent code.

The lesson we draw from these results is that most of the welfare gains from reforming the current US tax code by adopting the fully-optimal one can also be generated with simple history-invariant tax codes.

We conclude this subsection by relating the period-1 tax rate in the quasi-optimal tax code to the period-1 wedge that such code induces, using the formula in Proposition 2. Because the allocations under the quasi-optimal tax code are virtually the same as under the fully-optimal tax code, the wedges under the quasi-optimal tax code are virtually the same as under the fully-optimal allocations. Also recall that, under the quasi-optimal tax code, the period-2 taxes are invariant in the period-1 incomes. The formula in Proposition 2 then implies that any discrepancy between the period-1 wedge and the period-1 tax rate comes entirely from the effect of the period-1 income on the expected period-2 tax bill, due to the change in the distribution of the period-2 productivity triggered by LBD. Fig 9 depicts the various terms of the formula in Proposition 2. As the figure illustrates, the reason why the period-1 wedge is regressive while the period-1 tax rate is progressive is that, at higher period-1 income percentiles, the effect of LBD on the expected future tax bill is smaller than at lower percentiles, a property that appears to be fairly robust, as discussed in the next subsection.

5.4 Comparative Statics

We now turn to the comparative statics of the quasi-optimal tax code with respect to the intensity of the LBD effects and the workers' skill persistence. In all the cases considered below, we continue to find that the welfare losses from using the quasi-optimal tax code instead of the fully-optimal one are very small (precisely, of the same order as the corresponding losses in the calibrated economy reported above). Furthermore, the allocations under the quasi-optimal tax codes are close to the

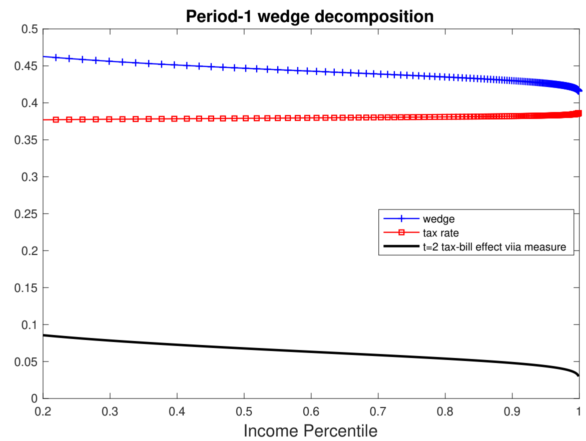

Figure 9: Period-1 tax rates, wedges, and $t = 2$ tax bill effects

second-best ones.[49] We thus expect the comparative statics results below to capture also the response of the fully-optimal tax code to the corresponding changes in the parameters of interest.

5.4.1 Intensity of LBD effects

Figure 9 illustrates the impact of stronger LBD effects on 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 rents 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 (64) in the proof of Proposition 4 in the Appendix). Period-1 consumption drops as the result of the increase in period-1 taxes. Consumption 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 those agents at such percentiles, 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 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

[49]In all cases considered below, adopting the fully-optimal tax code instead of the quasi-optimal one yields welfare gains equivalent to an equiproportionate increase in consumption starting from the allocations in the economy under the current US tax code of at most 0.2%. Moreover, the absolute percentage deviation in consumption across histories has a mean of at most 6% and a standard deviation of at most 3.5%, while the absolute percentage deviation in income across histories has a mean of at most 3.5% and a standard deviation of at most 2%.

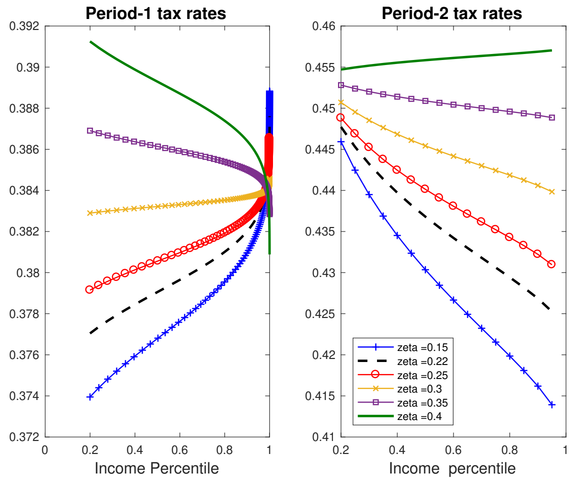

Figure 10: Effects of LBD intensity on quasi-optimal tax code

planner then increases the period-2 marginal tax rates at the top of the period-2 income distribution, because these rates are the most relevant ones for the top period-1 types due to skill persistence. An implication of the above effects is that stronger LBD effects contribute to a lower progressivity of taxes in period 1, but a higher progressivity in period 2.

We also note that, for the levels of LBD intensity considered in Figure 9, reforming the current US tax code by adopting the fully-optimal one generates welfare gains of the same order as in the calibrated economy.[50] 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 9, most of the welfare gains from reforming the current US tax code can also be generated through simple age-dependent linear taxes.[51]

We conclude this part by using the decomposition in Proposition 2 to relate the period-1 tax rates under the quasi-optimal tax code, $\hat{\tau}_1(\theta_1) = \tau_1(y_1(\theta_1))$, to the period-1 wedges, $W_1(\theta_1)$, and the change in the expected period-2 tax bill brought about the change in the period-2 productivity distribution, $\delta \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[\hat{\mathcal{T}}_2(\theta) \right]$. Recall that, under the quasi-optimal tax code, period-2 taxes are invariant to period-1 income, implying that the period-2 tax-bill effect in Proposition 2 originating in the history dependence of the period-2 tax code, $\delta \mathbb{E}^{\lambda[\chi]|\theta_1} \left[\frac{\partial \mathcal{T}_2(y_1(\theta_1), y_2(\theta))}{\partial y_1} \frac{v'(c_2(\theta))}{v'(c_1(\theta_1))} \right]$, is zero.

Figure 11 (top row) illustrates how the period-1 wedges and the period-1 tax rates vary with the

[50]Namely, these gains are equivalent to those brought by an equiproportionate 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 $\zeta = 0.15$, $\zeta = 0.22$, $\zeta = 0.25$, $\zeta = 0.3$, $\zeta = 0.35$, and $\zeta = 0.4$.

[51]Precisely, 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 $\zeta = 0.15$, $\zeta = 0.22$, $\zeta = 0.25$, $\zeta = 0.3$, $\zeta = 0.35$, 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.

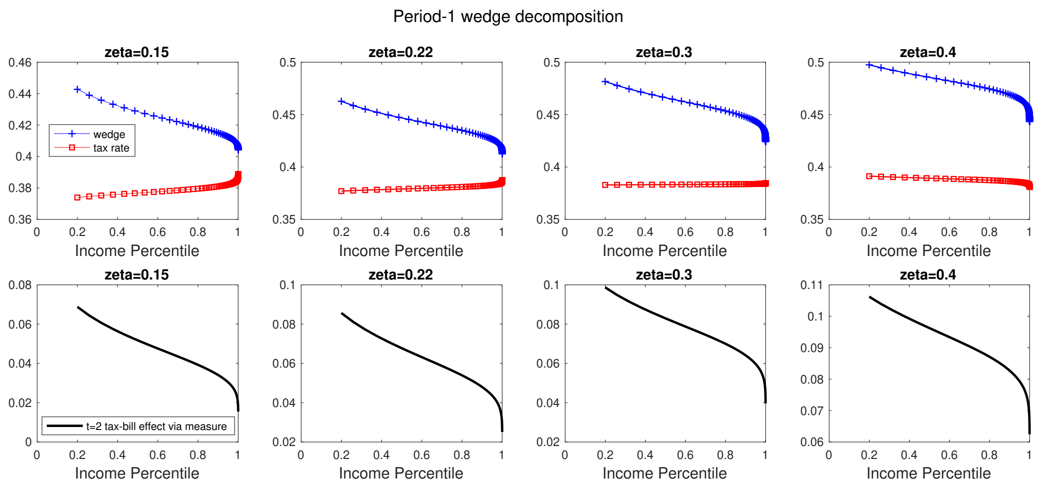

Figure 11: Period-1 tax rates, wedges, and $t = 2$ tax bill effects for different LBD intensities

income percentiles, for different levels of the LBD intensity. As the figure shows, while the period-1 wedges are regressive, the period-1 tax tares are progressive for low levels of LBD but regressive for high levels of LBD, consistently with what reported in Figure 9. The bottom row depicts the period-2 tax bill effect, which coincides with the vertical distance between the period-1 wedge and the period-1 tax rate in the top row.

Period-2 tax bill effects are positive at all percentiles and for all LBD intensities, and are larger the stronger the LBD effects. That the period-2 tax bill effects are positive reflects the fact that period-2 tax rates are everywhere positive, as reported in Figure 9, which, in turn, implies that the total period-2 tax bill is increasing in period-2 earnings. Higher period-1 incomes, by shifting the distribution of period-2 earnings towards higher levels (via the effect on the period-2 productivity distribution) thus come with an increase in expected period-2 taxes. Furthermore, because period-2 tax rates are increasing in the intensity of the LBD effects for all period-2 income percentiles, stronger LBD effects imply a larger period-2 tax bill effect. Other things equal, this channel thus contributes to lower period-1 tax rates (recall that period-1 tax rates are equal to the difference between period-1 wedges and the period-2 tax bill effect). That period-1 tax rates increase with the intensity of the LBD effects, as documented in Figure 9, is thus not a priori obvious, given that both the period-1 wedges and the period-2 tax bill effects increase with the intensity of the LBD effects.

Importantly, that period-1 tax rates become less progressive as the intensity of the LBD effects increases, as documented in Figure 11, originates in the regressivity of the period-2 tax bill effects. This, in turn, originates in the concavity of the LBD effects in period-1 income (equivalently, in the diminishing marginal effects of LBD on period-2 productivity), which is due to the fact that $\zeta < 1$

in all the cases considered in Figure 11. To see this, let

$$Y(\theta_1, y_1, \theta_2) = \left.\frac{\partial z_2(\theta_1, y_1, \varepsilon_2)}{\partial y_1}\right|_{\varepsilon_2 = \frac{\theta_2}{\theta_1^\rho y_1^\zeta}} = \frac{\zeta\theta_2}{y_1}$$

denote the impulse response of $\theta_2 = z_2(\theta_1, y_1, \varepsilon_2)$ to y_1, i.e., the marginal variation of θ_2 induced by a marginal variation in y_1, holding fixed the shock ε_2 that, together with θ_1 and y_1, leads to θ_2. While under the specification in (2), the impact of LBD on period-2 productivity is stronger for more productive period-1 types than for less productive ones,[52] the impulse response of θ_2 to y_1 is decreasing in y_1, due to the diminishing returns to LBD. Next observe that the period-2 tax bill effect is equal to[53]

$$\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)}\left[\hat{\mathcal{T}}_2(\tilde{\theta})\right] = \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)}\left[\frac{\partial\hat{\mathcal{T}}_2(\theta_1,\tilde{\theta}_2)}{\partial\theta_2}\frac{\zeta\tilde{\theta}_2}{y_1(\theta_1)}\right] \tag{34}$$

Because more productive period-1 types generate higher period-1 incomes than less productive ones, the period-2 tax bill effect can be decreasing in θ_1 even if $\mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)}\left[\frac{\partial\hat{\mathcal{T}}_2(\theta_1,\tilde{\theta}_2)}{\partial\theta_2}\tilde{\theta}_2\right]$ is increasing in θ_1. Figure 11 indicates that this is indeed the case for the parameters under consideration. Under the specification of our calibrated economy, the regressivity of the period-2 tax bill effect is sufficiently strong to counter the regressivity of the period-1 wedges and make the period-1 tax rates progressive for low levels of the intensity of the LBD effects, but not for high ones.

5.4.2 Skill Persistence

Next, consider variations in skill persistence. Figure 12 depicts the period-1 and period-2 tax rates in the quasi-optimal tax code, for different values of the skill-persistence parameter ρ. As the figure shows, an increase in skill persistence leads to higher marginal tax rates in both periods, except for young workers at high percentiles. The increase in period-2 tax rates for old workers at low percentiles 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, which, in turn, implies higher rents. Because period-2 handicaps are increasing in period-2 incomes, to contain the increase in the agents' rents, the planner optimally increases the period-2 tax rates so as to reduce the period-2 incomes. This effect is not specific to economies with LBD; it is a common feature of dynamic economies with persistent productivity. The increase in period-1 tax rates for young workers at low percentiles, instead, follows from two mechanisms that operate in the same direction. First, 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

[52]By this, we mean the following. Holding the shock ε_2 constant, the marginal effect of an increase in y_1 on θ_2 is increasing in θ_1.

[53]Note that in (34) we use the fact that, given any Lipschitz continuous function $J(\theta_2)$, and any kernel $F_2(\theta_2|\theta_1, y_1)$, $\frac{\partial}{\partial y_1}\mathbb{E}^{\lambda[\chi]|\theta_1, y_1}\left[J(\tilde{\theta}_2)\right] = \mathbb{E}^{\lambda[\chi]|\theta_1, y_1}\left[Y(\theta_1, y_1, \tilde{\theta}_2)\frac{\partial J(\tilde{\theta}_2)}{\partial\theta_2}\right]$.

discussed above for the comparative statics with respect to the intensity of the LBD effects). Second, an increase in period-2 handicaps increases the benefit of distorting downwards the period-1 incomes so as to shift the distribution of period-2 productivity towards levels commanding lower rents (this is the novel mechanism due to LBD discussed throughout the entire paper).

The above effects are at work at all percentiles. However, the impact of these effects on tax rates at high percentiles is small under the distribution of the productivity shocks in the calibrated economy. The reason is that, under this distribution, the relative value of distorting the labor supply (in either period) of the most productive 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. Therefore, 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.

To understand the effects of higher skill persistence on tax rates at high percentiles, observe that, under the specification behind Figure 12, for high productivity levels in period 1, the LBD effect is higher the higher the skill persistence is. Thus, for young workers at high period-1 income percentiles, the efficiency cost of making them less productive in period 2 by reducing their period-1 labor supply is very high when the persistence effects are strong (as captured by high values of ρ). When persistence effects grow large, the planner then optimally reduces the tax rates for the most productive period-1 types to make such types highly productive also in period 2. This effect, which is specific to economies with LBD, explains why (a) period-1 tax rates can go down at high percentiles as skill persistence increases, and (b) why the regressivity of the period-1 taxes increases as persistence increases. Making young workers at high percentiles work hard 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 (due to skill persistence). This increase in period-2 tax rates is small because, as mentioned above, the relative value of distorting the labor supply of the most productive agents is small under the calibrated distribution of the productivity shocks. However, when combined with the other effects discussed in the previous paragraph, this channel leads to an increase in period-2 tax rates at high percentiles that is slightly more pronounced than at lower percentiles. As a result, the regressivity of the period-2 taxes is slightly lower when persistence is higher.

For any level of persistence considered in Figure 12, reforming the current US tax code by adopting the fully-optimal one generates welfare gains of the same order as in the calibrated economy.[54] Interestingly, the higher the degree of skill persistence, the larger these welfare gains are. Moreover, as in the calibrated economy, most of these welfare gains can also be generated through simple age-dependent linear taxes.[55]

[54]Namely, these gains are equivalent to those brought about by an equiproportionate 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 $\rho = 0.3$, $\rho = 0.35$, $\rho = 0.4$, $\rho = 0.45$, $\rho = 0.5$, $\rho = 0.55$, and $\rho = 0.6$.

[55]To be precise, adopting the optimal linear tax code yields welfare gains equivalent to an equiproportionate increase

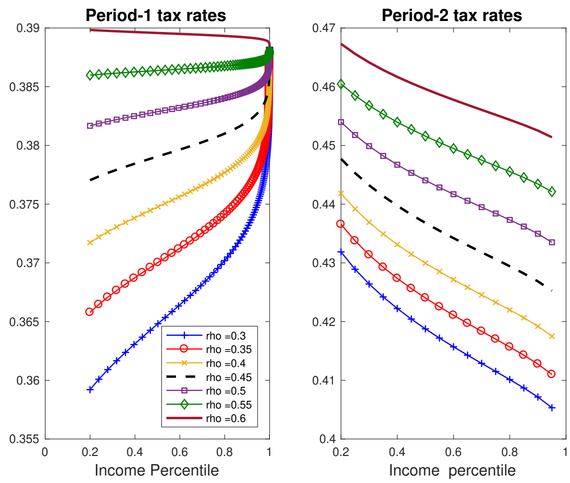

Figure 12: Effects of skill persistence on quasi-optimal taxes

5.5 Importance of the stochasticity of the LBD effects

Next, consider the role played by the stochasticity of the LBD effects. Figure 13 shows how the optimal period-1 tax rates vary when one varies the variance of the period-2 shocks, ε_2, holding the Pareto tail parameter ξ constant.

As the figure shows, more stochasticity comes with (i) a reduction in period-1 tax rates, with the exception of very high percentiles, (ii) an increase in period-2 tax rates, (iii) an increase in the progressivity of period-1 taxes, and (iv) a reduction in the progressivity of period-2 taxes. The reason for these comparative statics is the one anticipated in the Introduction. More stochasticity implies more volatility in the income and consumption of old workers. To contain such volatility, the planner increases the period-2 tax rates. The increase helps reducing period-2 output (for all levels of the period-2 productivity) which in turn dampens the volatility of period-2 earnings. This mechanism is the same as in Farhi and Werning (2013) and Golosov et al (2016).

In an economy with LBD, however, the planner has also another instrument to dampen the volatility of period-2 earnings. By inducing the workers to cut on their period-1 labor supply, the planner reduces the volatility of the workers' period-2 productivity, which in turn helps reducing the volatility of the workers' period-2 earnings. This novel effect is stronger for more productive period-1 types than for less productive ones. At high percentiles, where the volatility of the period-2 earnings is the highest, the planner then optimally responds to an increase in stochasticity by

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 $\rho = 0.3$, $\rho = 0.35$, $\rho = 0.4$, $\rho = 0.45$, $\rho = 0.5$, $\rho = 0.55$, 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 equiproportionate 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.

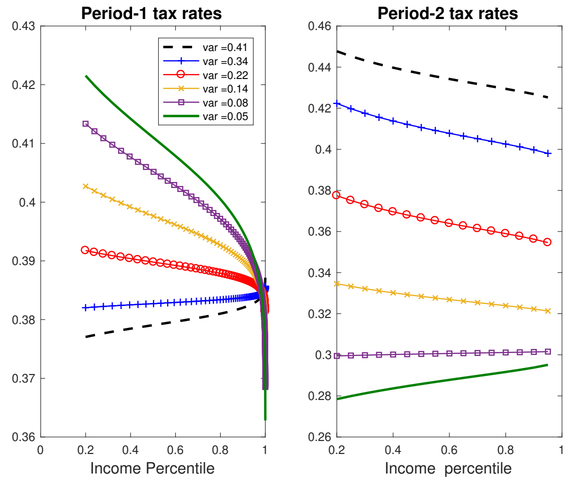

Figure 13: Effects of reductions in stochasticity of LBD on quasi-optimal taxes

raising tax rates in both periods. At lower percentiles, instead, the increase in the period-2 tax rates suffices to dampen the volatility of the period-2 earnings. That for lower percentiles the increase in the period-2 tax rates is more pronounced than for higher percentiles then reflects the fact that, at higher percentiles, volatility is already dampened through the increase in the period-1 tax rates. Furthermore, because the increase in period-2 tax rates comes with a reduction in period-2 handicaps/rents, it also reduces the benefit of distorting downwards the period-1 incomes to economize on expected rents (this the mechanism discussed at length in the paper). At low percentiles, this effect then leads to a reduction in period-1 tax rates. Together, the above effects explain properties (i)-(iv) described above. Interestingly, as anticipated in the Introduction, both the progressivity of taxes within each period and the dynamics of taxes over the life cycle are reversed when one compares the quasi-optimal tax code in the calibrated economy with the one in an economy with vanishing stochasticity. That tax rates are higher for young workers than for old ones when stochasticity vanishes is consistent with the findings in Best and Kleven (2013) and Kapicka (2015) who document a declining tax rates over the life cycle.

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.[56]

These findings underscore the importance for tax design of accounting for the stochastic effects of LBD found in the data, and hence the focus of this paper.

[56]In 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 equiproportionate increase in consumption starting from the allocations under the US tax code of at most 2%.

Symbol	Value	Target Moment	Data	Absolute Percentage Deviation
$\widehat{\rho}$	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

5.6 Isolating the Role of LBD: Counterfactual Analysis

We conclude this section by discussing the role that LBD plays for relative 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 $\theta_2 = h_2\theta_1^{\widehat{\rho}}\varepsilon_2$, where h_2 and $\hat{\rho}$ are positive scalars. All the parameters of the model are the same as in the previous subsections, except for ζ and ρ which are replaced by $\zeta = 0$ 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 $\left(\frac{\theta_1^{\rho}y_1(\theta_1)^{\zeta}-h_2\theta_1^{\widehat{\rho}}}{\theta_1^{\rho}y_1(\theta_1)^{\zeta}}\right)^2$ 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 $\widehat{\rho} = 0.6722$ 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 (under the current US tax code) 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 without LBD.[57]

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 relative wedges, (b) the value of reforming the US 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

[57]In the economy without LBD effects, the parameter configuration in Table 4 need not minimize the sum of the squared percentage deviations of the targeted moments. However, the deviations under such parametrization are small and hence the calibration appears satisfactory. The selected parametrization has the advantage that the values of the parameters h_1, σ, and ξ are the same as in the calibrated economy and the productivity distributions are also very similar between the two economies. This permits us to isolate the implications of the endogeneity of the distribution from all the rest.

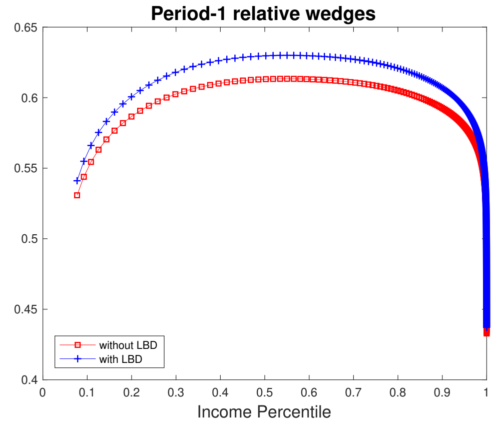

Figure 14: Period-1 relative wedges with and without LBD

current US tax code by adopting the fully-optimal one yields welfare gains that are equivalent to those brought about by a 4.62% equiproportionate increase in consumption at all histories, starting from the allocations under the existing US tax code.[58] This figure is 14.35% larger than the corresponding one in the calibrated economy with LBD. Ignoring LBD, thus leads to a sizable overestimation of the benefits of reforming the current tax code.

Next, consider the relative wedges. In this alternative economy, the period-1 relative wedges are distinctively lower than in the corresponding economy with LBD, for all income percentiles.[59] This result, which is illustrated in Figure 14, is consistent with the discussion in the previous subsections.

Importantly, that wedges are higher with LBD does not imply that taxes are also higher, as one can see from Figure 15. 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. As the figure shows, period-1 tax rates are lower in the calibrated economy with LBD than in the counterfactual one, whereas, for old workers, they are lower at low income percentiles but higher at high income percentiles. Furthermore, the period-1 quasi-optimal tax code is progressive with LBD, whereas is regressive without it. In period 2, 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. Finally, the differential between the average tax rate for old and for young workers is larger in the economy with LBD than in the counterfactual one without it.

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 progres-

[58]As in the calibrated economy with LBD, the quasi-optimal tax code in the counterfactual economy 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 0.19% equiproportionate increase in consumption at all histories, starting from the allocations under the current US tax code. The allocations under the quasi-optimal code are also close to the second-best ones. Namely, the absolute percentage deviation in consumption across histories has a mean around 5% and a standard deviation around 2%, while the corresponding moments for income are 2% and 1.5%.

[59]The same is true for the wedges.

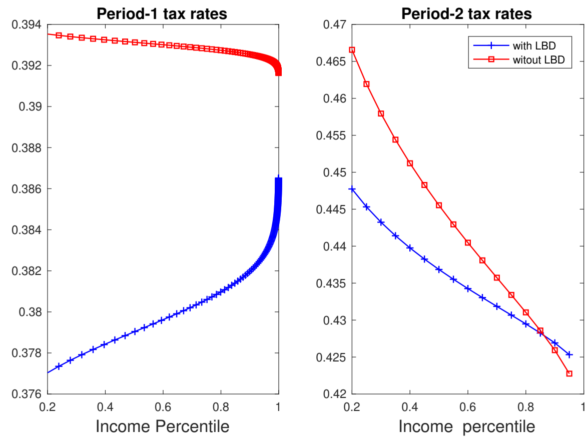

Figure 15: Quasi-optimal income tax rates with and without LBD

sive marginal tax rates, whereas higher persistence commands higher and more regressive marginal tax rates. As Figure 15 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 differences indicate that reforming the current US tax code while accounting for LBD calls for very different proposals than the ones suggested while ignoring LBD.

6 Conclusions

This paper studies optimal taxation in an economy in which the workers' productivity is stochastic and evolves endogenously over the life cycle as the result of (on-the-job) learning-by-doing. We show that learning-by-doing contributes to higher wedges and alters the relationship between wedges and tax rates under optimal codes. Next, we show that taxes that are invariant to past incomes but age-dependent are approximately optimal.

To isolate the role of learning-by-doing, we compare the quasi-optimal tax code in an economy calibrated to US earnings data to its counterpart in a counterfactual economy without learning-by-doing. We find that the predictions for the level, the progressivity, and the dynamics of taxes are fundamentally different across the two economies. 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.

We believe these insights can guide the debate on 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.[60]

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,” J*ournal 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.

[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] Bergemann, D. and J. Välimäki, (2019): “Dynamic Mechanism Design: An Introduction,” *Journal of Economic Literature*, 57(2), 235-74.

[12] Best, M.C. and H.J. Kleven, (2013): “Optimal Income Taxation with Career Effects of Work Effort,” mimeo, LSE.

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

[13] Blundell, R. and T. MaCurdy, (1999): "Labor Supply: A Review of Alter- native Approaches," Chap. 27, pp. 1559–1695 in Ashenfelter, Orley, and Card, David (eds), *Handbook of Labor Economics*, Vol. 3. North Holland, Elsevier.

[14] Chang, Y., J. Gomes, and F. Schorfheide (2002): "Learning-by-Doing as a Propagation Mechanism," *American Economic Review*, 92(5), 1498-1520.

[15] Conesa, J.C. and D. Krueger, (2006): "On the Optimal Progressivity of the Income Tax Code," *Journal of Monetary Economics*, 53(7), 1425-1450.

[16] 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.

[17] D'Alessandro A., G. Fella and L. Melosi (2018): "Fiscal Stimulus with Learning-By-Doing," mimeo Chicago Federal Reserve.

[18] Dustmann, C. and K. Meghir, (2005): "Wages, Experience, and Seniority," *Review of Economic Studies*, 72, 77-108.

[19] 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.

[20] Farhi, E. and I. Werning, (2013): "Insurance and Taxation over the Life Cycle," *Review of Economic Studies*, 80, 596-635.

[21] Fershtman, D. and A. Pavan, (2017): "Matching Auctions," mimeo, Northwestern University.

[22] Garrett, D. and A. Pavan, (2015): "Dynamic Managerial Compensation: a Variational Approach," *Journal of Economic Theory,* 159, 775-818.

[23] Garrett, D., A. Pavan, and J. Toikka, (2018): "Robust Predictions of Dynamic Screening," mimeo, MIT, TSE, Northwestern.

[24] Golosov, M., A. Tsyvinski, and N. Werquin, (2014): "A Variational Approach to the Analysis of Tax Systems," NBER Working Paper No. 20780.

[25] Golosov, M., M. Troshkin, and A. Tsyvinski, (2016): "Redistribution and Social Insurance," *American Economic Review*, 106(2), 359-86.

[26] Golosov, M., A. Tsyvinski, and I. Werning, (2006): "New Dynamic Public Finance: A User's Guide", NBER Macroeconomic Annual 2006, MIT Press.

[27] Gorry, A. and E. Oberfield, (2012): "Optimal Taxation Over the Life Cycle," *Review of Economic Dynamics*, 15(4), 551–572.

[28] Heathcote, J., K. Storesletten, and G. Violante, (2017): "Optimal Tax Progressivity: An Analytical Framework," *Quarterly Journal of Economics*, 132(4), 1693-1754

[29] Huggett, M., G. Ventura, and A. Yaron, (2011): "Sources of Lifetime Inequality," *American Economic Review*, 101(7), 2923-2954.

[30] Jacobson L., R. LaLonde, and D. Sullivan, (1993): "Earnings Losses of Displaced Workers," *American Economic Review*, 83(4), 685-709.

[31] Kapicka, M., (2006): "Optimal Income Taxation with Human Capital Accumulation and Limited Record Keeping", *Review of Economic Dynamics*, 9, 612–639.

[32] 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.

[33] Kapicka, M., (2015a): "Optimal Mirrleesian Taxation in a Ben-Porath Economy," *American Economic Journal: Macroeconomics*, 7(2), 219-48.

[34] Kapicka, M., (2015b): "Pareto Efficient Income Taxation with Learning by Doing," mimeo, CERGE-EI.

[35] Kapicka, M. and J. Neira, (2016): "Optimal Taxation with Risky Human Capital," mimeo, CERGE-EI.

[36] 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.

[37] Kocherlakota, N. R., (2010): The New Dynamic Public Finance. Princeton University Press.

[38] Krause, A., (2009): "Optimal Nonlinear Income Taxation with Learning-by-doing," *Journal of Public Economics*, 93(9-10), 1098-1110.

[39] Kremer, M., (2001): "Should Taxes be Independent of Age?" mimeo, Harvard University.

[40] 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.

[41] 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.

[42] Lucas, R.E., (1988): "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22(1), 33-42.

[43] Makris, M. and A. Pavan, (2018): "Wedge Dynamics with Endogenous Private Information: A General Recursive Characterization," mimeo, Northwestern University.

[44] Mirrlees, J., (1971): "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies*, 38, 175-208.

[45] Ndiaye, A., (2018): "Flexible Retirement and Optimal Taxation," mimeo, Northwestern University.

[46] Pavan, A., S. Segal, and J. Toikka, (2014): "Dynamic Mechanism Design: A Myersonian Approach," *Econometrica*, 82(2), 601-653.

[47] 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.

[48] Perrault, L., (2018): "Optimal Taxation with Work Experience as a Risky Investment," mimeo, Georgia State University.

[49] Ramsey, F.P., (1927): "A Contribution to the Theory of Taxation," *Economic Journal*, 37(145), 47-61.

[50] Rogerson, W.P., (1985): "Repeated Moral Hazard," *Econometrica*, 53(1), 69-76.

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

[52] Stantcheva, S., (2015): "Learning and (or) Doing: Human Capital Investments and Optimal Taxation," NBER Working Paper 21381.

[53] Stantcheva, S., (2017): "Optimal Taxation and Human Capital Policies over the Life Cycle," *Journal of Political Economy*, 125(6), 1931–1990.

[54] 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.

[55] 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.

[56] 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.

[57] Topel, R., (1991): "Specific Capital, Mobility and Wages: Wages Rise with Seniority," *Journal of Political Economy*, 99(1), 145-176.

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

[59] 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.

[60] 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 $\chi^* = (y^*(\cdot), c^*(\cdot))$ maximizes expected tax revenue $R = \mathbb{E}^{\lambda[\chi]}\left[\sum_t \delta^{t-1}\left(y_t(\tilde{\theta}^t) - c_t(\tilde{\theta}^t)\right)\right]$ subject to the redistribution constraint $\int V_1(\theta_1)q(\theta_1)dF_1(\theta_1) \geq \kappa$.

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

$$\max_{(y_t(\cdot))_{t=1,2}, c_2(\cdot), V_1(\cdot)} \int \{y_1(\theta_1) - C\left(V_1(\theta_1) + \psi(y_1(\theta_1), \theta_1) - \delta \int [v(c_2(\theta)) - \psi(y_2(\theta), \theta_2)]dF_2(\theta_2 \mid \theta_1, y_1(\theta_1))\right)$$
$$+\delta \int [y_2(\theta) - c_2(\theta)]dF_2(\theta_2 \mid \theta_1, y_1(\theta_1))\}dF_1(\theta_1)$$

subject to

$$\int V_1(\theta_1)q(\theta_1)dF_1(\theta_1) - \kappa = 0. \tag{35}$$

Let π be the multiplier of the redistribution constraint (35), which is an integral constraint. At the optimum, the following necessary conditions must hold with $\lambda[\chi^*]$-probability one:

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

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

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

$$-\frac{1}{v'(c_1^*(\theta_1))} + \pi q(\theta_1) = 0, \tag{36}$$

$$\int V_1(\theta_1)q(\theta_1)dF_1(\theta_1) - \kappa = 0 \tag{37}$$

and $c_1(\theta_1) = C\left(V_1(\theta_1) + \psi(y_1(\theta_1), \theta_1) - \delta \int [v(c_2(\theta)) - \psi(y_2(\theta), \theta_2)]dF_2(\theta_2 \mid \theta_1, y_1(\theta_1))\right)$. Rearranging, and using the definition of $LD_1^{\chi}(\theta_1)$, we obtain the conditions in the proposition. Q.E.D.

Proof of Proposition 2. Consider the production and saving problem of a worker faced with the labor income tax code $\mathcal{T} = (\mathcal{T}_t(\cdot))$ and rate of return on saving (net of any linear capital tax rate) equal to $r = \frac{1}{\delta} - 1$. For any (θ_1, y_1), let

$$\widehat{V}_1(\theta_1, y_1) = \max_{c_1, \hat{y}_2(\theta_2)} \{v(c_1) - \psi(y_1, \theta_1)$$

$$+\delta \int \left[v\left(\hat{y}_2(\theta_2) - \mathcal{T}_2(y_1, \hat{y}_2(\theta_2)) + (y_1 - T_1(y_1) - c_1)\frac{1}{\delta}\right) - \psi(\hat{y}_2(\theta_2), \theta_2)\right] dF_2(\theta_2|\theta_1, y_1)\}$$

denote the intertemporal utility of a worker of period-1 productivity equal to θ_1 producing income y_1 and then optimally choosing his period-1 consumption c_1 and his period-2 income $\hat{y}_2(\theta_2)$, as a function of his period-2 productivity θ_2. For any (θ_1, y_1), the first-order condition for c_1 is

$$v'(c_1) = \int v'\left(\hat{c}_2(\theta_2; y_1, c_1)\right) dF_2(\theta_2|\theta_1, y_1), \tag{38}$$

whereas, for any $(\theta_1, y_1, \theta_2)$, the first-order condition for $\hat{y}_2(\theta_2)$ is

$$\left[1 - \frac{\partial \mathcal{T}_2(y_1, \hat{y}_2(\theta_2))}{\partial y_2}\right] v'(\hat{c}_2(\theta_2; y_1, c_1)) = \psi_y(\hat{y}_2(\theta_2), \theta_2), \tag{39}$$

where $\hat{c}_2(\theta_2; y_1, c_1) = \hat{y}_2(\theta_2) - \mathcal{T}_2(y_1, \hat{y}_2(\theta_2)) + (y_1 - \mathcal{T}_1(y_1) - c_1)\frac{1}{\delta}$.

Using the envelope theorem, we then have that, for any θ_1, the first-order condition for y_1 is

$$\begin{aligned} -\psi_y(y_1, \theta_1) + \delta \int v'\left(\hat{c}_2(\theta_2; y_1, c_1)\right) \left\{-\tfrac{\partial \mathcal{T}_2(y_1, \hat{y}_2(\theta_2))}{\partial y_1} + \left[1 - \tfrac{\partial \mathcal{T}_1(y_1)}{\partial y_1}\right]\tfrac{1}{\delta}\right\} dF_2(\theta_2|\theta_1, y_1) \\ +\delta \tfrac{\partial}{\partial y_1} \int \hat{V}_2(\theta_2; y_1, c_1) dF_2(\theta_2|\theta_1, y_1) = 0 \end{aligned} \tag{40}$$

where $\hat{V}_2(\theta_2; y_1, c_1) = v(\hat{c}_2(\theta_2; y_1, c_1)) - \psi(\hat{y}_2(\theta_2), \theta_2)$ and where the derivative in the third term in (40) is taken by differentiating the distribution over the period-2 productivity, holding fixed the period-2 continuation utility, $\hat{V}_2(\theta_2; y_1, c_1)$. Using (38), (40) can be rewritten as

$$\begin{aligned} \left[1 - \tfrac{\partial \mathcal{T}_1(y_1)}{\partial y_1}\right] v'(c_1) = \psi_y(y_1, \theta_1) + \delta \int v'\left(\hat{c}_2(\theta_2; y_1, c_1)\right) \left\{\tfrac{\partial \mathcal{T}_2(y_1, \hat{y}_2(\theta_2))}{\partial y_1}\right\} dF_2(\theta_2|\theta_1, y_1) \\ -\delta \tfrac{\partial}{\partial y_1} \int \hat{V}_2(\theta_2; y_1, c_1) dF_2(\theta_2|\theta_1, y_1). \end{aligned} \tag{41}$$

Let $\bar{y}_1(\theta_1)$, $\bar{c}_1(\theta_1)$, and $\bar{y}_2(\theta) = \hat{y}_2(\theta_2; \theta_1, \bar{y}_1(\theta_1))$ denote the solution to the above problem, where $\hat{y}_2(\theta_2; \theta_1, \bar{y}_1(\theta_1))$ is the period-2 income policy that solves the above optimization problem when the period-1 productivity is θ_1 and the period-1 income is $\bar{y}_1(\theta_1)$. For the proposed tax code to implement the allocations under the rule χ, it must be that, for all $\theta = (\theta_1, \theta_2)$, $\bar{y}_1(\theta_1) = y_1(\theta_1)$, $\bar{c}_1(\theta_1) = c_1(\theta_1)$, $\hat{y}_2(\theta_2; \theta_1, \bar{y}_1(\theta_1)) = y_2(\theta)$, and $\hat{c}_2(\theta_2; y_1(\theta_1), c_1(\theta_1)) = c_2(\theta)$, and hence

$$\hat{V}_2(\theta_2; y_1(\theta_1), c_1(\theta_1)) = v(c_2(\theta_2)) - \psi(y_2(\theta), \theta_2) = V_2(\theta_1, \theta_2).$$

It is easy to see that the optimality condition (41) is equivalent to Condition (16) in the main text.

Using (39) and the definition of the period-2 wedges, we then have that, under any tax code implementing the desired allocations, for any $\theta = (\theta_1, \theta_2)$, the condition $\hat{W}_2(\theta) = \hat{\tau}_2(\theta)$ must hold.

Furthermore, using (41) and the definition of the period-1 wedges, we have that

$$W_1(\theta_1) = \frac{\partial \mathcal{T}_1(y_1(\theta_1))}{\partial y_1} + \delta \mathbb{E}^{\lambda[\chi]|\theta_1} \left\{ \frac{v'\big(c_2(\tilde{\theta})\big)}{v'(c_1(\theta_1))} \frac{\partial \mathcal{T}_2(y_1(\theta_1), y_2(\tilde{\theta}))}{\partial y_1} \right\} + \delta \frac{\partial}{\partial y_1} \mathbb{E}^{\lambda[\chi]|\theta_1, y_1(\theta_1)} \left[y_2(\tilde{\theta}) - c_2(\tilde{\theta}) \right]. \tag{42}$$

Next, observe that, for any θ, $y_2(\theta) - c_2(\theta) = \hat{\mathcal{T}}_2(\theta) - \left(y_1(\theta_1) - c_1(\theta_1) - \hat{\mathcal{T}}_1(\theta_1) \right) / \delta$, where recall that $\hat{\mathcal{T}}_1(\theta_1) = \mathcal{T}_1(y_1(\theta_1))$ and $\hat{\mathcal{T}}_2(\theta) = \mathcal{T}_2(y_1(\theta_1), y_2(\theta))$. Because $y_1(\theta_1) - c_1(\theta_1) - \hat{\mathcal{T}}_1(\theta_1)$ is constant in θ_2 and because the derivative in the last term in the right-hand-side of (42) is taken by differentiating the distribution over θ_2 for given period-2 policies, we then have that, under any tax code implementing the desired allocations, for any θ_1, Condition (15) must also hold. Q.E.D.

Proof of Proposition 4. Step 1 characterizes the first-order conditions for the second-best allocations. Step 2 uses the first-order conditions to show that relative 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

$$\frac{\partial V_1(\theta_1)}{\partial \theta_1} = -\psi_\theta(y_1(\theta_1), \theta_1) - \delta \mathbb{E}^{\lambda[\chi]|\theta_1} \left[I_1^2(\tilde{\theta}, y_1(\theta_1)) \psi_\theta(y_2(\tilde{\theta}), \tilde{\theta}_2) \right], \text{ almost all } \theta_1 \in \Theta_1 \tag{43}$$

and

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

The planner's problem can be conveniently reformulated as follows:

$$\max_{y_1(\cdot), V_1(\cdot), \Pi_2(\cdot), Z_2(\cdot)} \int \{ y_1(\theta_1) - C\left(V_1(\theta_1) + \psi(y_1(\theta_1), \theta_1) - \delta \Pi_2(\theta_1)\right) + \delta Q_2(\theta_1, y_1(\theta_1), \Pi_2(\theta_1), Z_2(\theta_1)) \} dF_1(\theta_1)$$

subject to

$$\frac{\partial V_1(\theta_1)}{\partial \theta_1} = -\psi_\theta(y_1(\theta_1), \theta_1) + \delta Z_2(\theta_1), \tag{44}$$

and

$$\int V_1(\theta_1) q(\theta_1) dF_1(\theta_1) - \kappa = 0, \tag{45}$$

where

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

subject to

$$\Pi_2(\theta_1) = \int V_2(\theta) dF_2(\theta_2 \mid \theta_1, y_1(\theta_1)), \tag{46}$$

$$Z_2(\theta_1) = -\int I_1^2(\theta, y_1(\theta_1)) \psi_\theta(y_2(\theta), \theta_2) dF_2(\theta_2 \mid \theta_1, y_1(\theta_1)), \tag{47}$$

and

$$\frac{\partial V_2(\theta)}{\partial \theta_2} = -\psi_\theta(y_2(\theta), \theta_2). \tag{48}$$

This problem consists of two interdependent optimal control problems, one for each period.

We proceed backwards, by solving first the period-2 problem that defines the value function $Q_2(\theta_1, y_1(\theta_1), \Pi_2(\theta_1), Z_2(\theta_1))$. This is an optimal control problem with two integral constraints, (46) and (47). 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 (48).

Let $\pi_2(\theta_1)$ and $\xi_2(\theta_1)$ be the multipliers of the two integral constraints (46) and (47) and $\mu_2(\theta)$ the costate variable for the law of motion of $V_2(\theta_1,\theta_2)$. Along with (46), (47), and (48), the following necessary optimality conditions must hold for almost all $\theta_2 \in Supp[F_2(\cdot \mid \theta_1, y_1(\theta_1))]$:

$$1-\frac{\psi_y(y_2(\theta),\theta_2)}{v'(c_2(\theta))}-\mu_2(\theta)\frac{\psi_{\theta y}(y_2(\theta),\theta_2)}{f_2(\theta_2\mid\theta_2,y_1(\theta_1))}+\xi_2(\theta_1)I_1^2(\theta,y_1(\theta_1))\psi_{\theta y}(y_2(\theta),\theta_2)=0, \tag{49}$$

$$\frac{\partial\mu_2(\theta)}{\partial\theta_2}=f_2(\theta_2\mid\theta_1,y_1(\theta_1))\cdot\left\{\frac{1}{v'(c_2(\theta))}+\pi_2(\theta_1)\right\}, \tag{50}$$

along with the boundary conditions

$$\mu_2(\theta_1,\underline{\theta}_2)=0, \tag{51}$$

$$\mu_2(\theta_1,\overline{\theta}_2)=0, \tag{52}$$

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 (44) and π_1 the multiplier associated with the constraint (45). In addition to (44) and (45), the following optimality conditions must hold:

$$\begin{aligned}&1-\tfrac{\psi_y(y_1(\theta_1),\theta_1)}{v'(c_1(\theta_1))}+\delta\tfrac{\partial}{\partial y_1}\textstyle\int\left[y_2(\theta)-c_2(\theta)-\pi_2(\theta_1)V_2(\theta)\right]dF_2(\theta_2\mid\theta_1,y_1(\theta_1))\\&+\delta\xi_2(\theta_1)\tfrac{\partial}{\partial y_1}\textstyle\int I_1^2(\theta,y_1(\theta_1))\psi_\theta(y_2(\theta),\theta_2)dF_2(\theta_2\mid\theta_1,y_1(\theta_1))-\mu_1(\theta_1)\tfrac{\psi_{\theta y}(y_1(\theta_1),\theta_1)}{f_1(\theta_1)}=0,\end{aligned} \tag{53}$$

$$\frac{\partial\mu_1(\theta_1)}{\partial\theta_1}=f_1(\theta_1)\cdot\left\{\frac{1}{v'(c_1(\theta_1))}-\pi_1 q(\theta_1)\right\}, \tag{54}$$

$$\frac{1}{v'(c_1(\theta_1))}+\pi_2(\theta_1)=0, \tag{55}$$

$$\mu_1(\theta_1)+\xi_2(\theta_1)f_1(\theta_1)=0, \tag{56}$$

along with the boundary conditions

$$\mu_1(\underline{\theta}_1)=0, \tag{57}$$

$$\mu_1(\overline{\theta}_1)=0, \tag{58}$$

where $c_1(\theta_1)=C\left(V_1(\theta_1)+\psi(y_1(\theta_1),\theta_1)-\delta\Pi_2(\theta_1)\right)$. 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 properties $\frac{\partial Q_2}{\partial\Pi_2}=\pi_2(\theta_1)$, $\frac{\partial Q_2}{\partial Z_2}=\xi_2(\theta_1)$ and

$$\begin{aligned}\tfrac{\partial Q_2}{\partial y_1}=&\tfrac{\partial}{\partial y_1}\textstyle\int\left[y_2(\theta)-c_2(\theta)\right]dF_2(\theta_2\mid\theta_1,y_1(\theta_1))-\pi_2(\theta_1)\tfrac{\partial}{\partial y_1}\textstyle\int V_2(\theta)dF_2(\theta_2\mid\theta_1,y_1(\theta_1))\\&+\xi_2(\theta_1)\tfrac{\partial}{\partial y_1}\textstyle\int I_1^2(\theta,y_1(\theta_1))\psi_\theta(y_2(\theta),\theta_2)dF_2(\theta_2\mid\theta_1,y_1(\theta_1)).\end{aligned}$$

Now use (54) along with the boundary conditions (57) and (58) to obtain that

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

which implies that

$$\pi_1 = \int_{\underline{\theta}_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(\theta_1))} dF_1(\theta_1), \tag{59}$$

where we used the fact that $\int_{\underline{\theta}_1}^{\bar{\theta}_1} q(\theta_1)dF_1(\theta_1) = 1$. Furthermore, using (54) and (58) 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

$$\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} = \frac{1 - F_1(\theta_1)}{f_1(\theta_1)} \left[\int_{\theta_1}^{\bar{\theta}_1} \frac{1}{v'(c_1(s))} \frac{dF_1(s)}{1 - F_1(\theta_1)} - \pi_1 \int_{\theta_1}^{\bar{\theta}_1} q(s) \frac{dF_1(s)}{1 - F_1(\theta_1)} \right]. \tag{60}$$

Next, use (55) and (56) to rewrite the FOC for $y_1(\theta_1)$ as follows

$$\begin{gathered} 1 - \tfrac{\psi_y(y_1(\theta_1),\theta_1)}{v'(c_1(\theta_1))} + \delta \tfrac{\partial}{\partial y_1} \int \left[y_2(\theta) - c_2(\theta) + \tfrac{V_2(\theta)}{v'(c_1(\theta_1))} \right] dF_2(\theta_2 \mid \theta_1, y_1(\theta_1)) \\ + \left(\tfrac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \delta \tfrac{\partial}{\partial y_1} \int I_1^2(\theta, y_1(\theta_1)) \psi_\theta(y_2(\theta), \theta_2) dF_2(\theta_2 \mid \theta_1, y_1(\theta_1)) + \left(\tfrac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \psi_{\theta y}(y_1(\theta_1), \theta_1) = 0, \end{gathered} \tag{61}$$

with $-\mu_1(\theta_1)/f_1(\theta_1)$ given by (60).

Note that, together, Conditions (50) and (55), imply that

$$\frac{\partial \mu_2(\theta)}{\partial \theta_2} = f_2(\theta_2 \mid \theta_1, y_1(\theta_1)) \cdot \left\{ \frac{1}{v'(c_2(\theta))} - \frac{1}{v'(c_1(\theta_1))} \right\}. \tag{62}$$

Combining (62) with the boundary conditions (51) and (52), we have that

$$0 = \int_{\underline{\theta}_2}^{\bar{\theta}_2} \tfrac{\partial \mu_2(\theta_1,s)}{\partial \theta_2} ds = \int_{\underline{\theta}_2}^{\bar{\theta}_2} \tfrac{1}{v'(c_2(\theta_1,s))} dF_2(s|\theta_1, y_1(\theta_1)) - \tfrac{1}{v'(c_1(\theta_1))}. \tag{63}$$

Next note that (63) yields the familiar Rogerson inverse-Euler condition

$$\frac{1}{v'(c_1(\theta_1))} = \int_{\underline{\theta}_2}^{\bar{\theta}_2} \frac{1}{v'(c_2(\theta_1, s))} dF_2(s|\theta_1, y_1(\theta_1)). \tag{64}$$

Combining (62) and (52) with (64), we have that

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

from which we obtain that

$$\begin{aligned} \tfrac{-\mu_2(\theta_1,\theta_2)}{f_2(\theta_2|\theta_1,y_1(\theta_1))} &= \tfrac{1-F_2(\theta_2|\theta_1,y_1(\theta_1))}{f_2(\theta_2|\theta_1,y_1(\theta_1))} \int_{\theta_2}^{\bar{\theta}_2} \tfrac{1}{v'(c_2(\theta_1,s))} \tfrac{dF_2(s|\theta_1,y_1(\theta_1))}{1-F_2(\theta_2|\theta_1,y_1(\theta_1))} \\ &\quad - \tfrac{1-F_2(\theta_2|\theta_1,y_1(\theta_1))}{f_2(\theta_2|\theta_1,y_1(\theta_1))} \int_{\underline{\theta}_2}^{\bar{\theta}_2} \tfrac{1}{v'(c_2(\theta_1,s))} dF_2(s|\theta_1, y_1(\theta_1)). \end{aligned} \tag{65}$$

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

Consider first the period-2 wedges. Using the FOC for period-2 output (49), we have that the period-2 wedges, under the second-best allocations, are given by

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

Hence, the period-2 relative wedges are given by

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

with $-\mu_2(\theta)/f_2(\theta_2 \mid \theta_2, y_1(\theta_1))$ given by (65) and $-\mu_1(\theta_1)/f_1(\theta_1)$ given by (60). It is easy to see that the formulas for $\hat{W}_2(\theta)$ are the same as in an economy without LBD. Thus, $\hat{W}_2(\theta_1) = \hat{W}_2^{NOLBD}(\theta_1)$.

Next, consider the period-1 wedges. Using the FOC for period-1 output (61) and the definition of the $LD_1^{\chi}(\theta_1)$ function in (5), we have that

$$W_1(\theta_1) \equiv 1 + LD^{\chi}(\theta_1) - \tfrac{\psi_y(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))}$$

$$= -\left(\tfrac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \delta \tfrac{\partial}{\partial y_1} \int I_1^2(\theta, y_1(\theta_1)) \psi_\theta(y_2(\theta), \theta_2) dF_2(\theta_2 \mid \theta_1, y_1(\theta_1)) - \left(\tfrac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \psi_{\theta y}(y_1(\theta_1), \theta_1).$$

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

$$\hat{W}_1(\theta_1) \equiv W_1(\theta_1) / \left(\tfrac{\psi_y(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))} \right)$$

$$= -\frac{\left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \delta \frac{\partial}{\partial y_1} \int I_1^2(\theta, y_1(\theta_1)) \psi_\theta(y_2(\theta), \theta_2) f_2(\theta_2 | \theta_1, y_1(\theta_1)) d\theta_2 + \left(\frac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \psi_{\theta y}(y_1(\theta_1), \theta_1)}{\frac{\psi_y(y_1(\theta_1), \theta_1)}{v'(c_1(\theta_1))}}$$

with $-\mu_1(\theta_1)/f_1(\theta_1)$ given by (60). We thus have that

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

where

$$\hat{W}_1^{NOLBD}(\theta_1) \equiv \left(\frac{\mu_1(\theta_1)}{f_1(\theta_1)} \right) \frac{\psi_{\theta y}(y_t(\theta^t), \theta_t)}{\psi_y(y_t(\theta^t), \theta_t)} v'(c_1(\theta_1))$$

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

$$RA(\theta_1) - D(\theta_1) \equiv \left(\tfrac{-\mu_1(\theta_1)}{f_1(\theta_1)} \right) \gamma_1(\theta_1) v'(c_1(\theta_1)).$$

It is also easy to see that the only term affected by the presence of LBD is $[RA(\theta_1) - D(\theta_1)] \, \Omega(\theta_1)$. Using (59) and (60), 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 equals

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

Q.E.D.