Taxing Capital?
The Importance of How Human Capital is Accumulated[*]

Adam Blandin[†]   William B. Peterman[‡]

June 5, 2019

Abstract

This paper shows that in a life-cycle framework the optimal tax on capital crucially depends on how human capital is accumulated. We focus on three cases common to the macroeconomic literature: (i) Learning-*By*-Doing (LBD), (ii) Learning-*Or*-Doing (LOD), and (iii) exogenous accumulation. First, we show analytically in a simple two-period model that endogenizing human capital introduces novel motives for the government to tax capital when it cannot directly condition taxes on age, and moreover that these motives differ depending on whether LBD or LOD is assumed. We then quantify differences in optimal capital taxes using a rich life-cycle framework that features heterogeneity in learning ability and initial human capital. With proportional taxes the optimal capital tax is 16pp higher with LBD than with exogenous accumulation, but 2pp lower with LOD than with exogenous accumulation. We show that heterogeneity in learning ability strengthens the novel channels introduced by human capital, resulting in a larger gap in capital taxes between LBD and LOD relative to a case with homogeneous workers. Finally, we show that although allowing the government to make labor taxes progressive reduces the gap in optimal capital taxes, a sizable gap still persists.

JEL: E24, E62, H21.

Key Words: Optimal Taxation, Capital Taxation, Human Capital.

1 Introduction

A classic result in the optimal taxation literature is that it is not optimal to tax capital in an infinitely-lived agent model (Chamley (1986) and Judd (1985)). In contrast, a host of work has shown that in life cycle models the optimal tax on capital need not be zero because in these models a tax on capital provides a way for the government to indirectly condition taxes on age (see Erosa and Gervais (2002), Garriga (2001), Smyth (2006), Conesa et al. (2009), and Peterman (2013)). Recent work by Karabarbounis (2016) and da Costa and Santos (2018) demonstrates that the government's desire to use this type of age-dependent tax can increase when workers accumulate human capital endogenously, and Peterman (2016a) finds that when age-dependent taxes are disallowed, incorporating one specific form of endogenous human capital accumulation can have a considerable effect on the optimal capital tax.

However, there are a variety of ways that the macroeconomic literature incorporates human capital accumulation. In this paper, we establish both analytically and quantitatively that the optimal capital tax crucially depends not only on whether human capital is accumulated endogenously versus exogenously, but also on the form of endogenous human capital accumulation.

Specifically, we analyze optimal capital taxes across three human capital technologies which are common in the macroeconomic literature: exogenous accumulation, Learning-By-Doing (LBD), and Learning-Or-Doing (LOD).[1] Under exogenous human capital accumulation, workers take their life-cycle human capital profiles as given. Under LBD, an agent acquires human capital endogenously by working, which means they learn and earn labor income simultaneously. Under LOD, which is also referred to as Ben-Porath skill accumulation or on-the-job training, workers accumulate human capital by spending some of their non-leisure time training.[2]

We begin by deriving the intuition for the relationship between endogenous human capital and optimal taxes in a simple overlapping generations model with homogeneous 2-period-lived workers whose utility function is both separable and homothetic with respect

---

[1]Examples of life-cycle analyses that include these three forms of human capital accumulation are Conesa et al. (2009), Conesa and Krueger (2006), Huggett et al. (2011)Hansen and İmrohoroğlu (2009), Imai and Keane (2004), Chang et al. (2002), Jones et al. (1997), Jones and Manuelli (1999), Guvenen et al. (2009), Kuruscu (2006), Kapicka (2006), and Kapicka (2009).

[2]This paper does not evaluate the effect of formal education on optimal tax policy but instead focuses on human capital acquired after an individuals begins working. Although, the quantitative model is calibrated to exclude time spent in school, the mechanisms by which LOD changes the optimal tax policy would be similar for formal education. For a discussion of the effects of formal education on optimal taxation see Jacobs and Bovenberg (2009).

to each consumption and hours worked. The government's goal is to set capital taxes and age-dependent labor taxes to maximize the welfare of the worker. This model is a useful reference point because Garriga (2001) finds that, in this setting, optimal taxes do not depend on age and optimal capital taxes are zero when human capital is exogenous.[3] We show that endogenous human capital creates novel motives for the government to condition labor taxes on age, but that these motivations can differ qualitatively depending on whether we assume LBD or LOD. Further, we show that this can lead to a different optimal tax on capital when age-dependent taxes are disallowed.

Introducing LBD affects optimal taxes because it alters incentives to work over the life-cycle. With LBD work has two benefits: it increases contemporaneous earnings, and increases future human capital. In this 2-period model, because this latter benefit is only relevant when workers are young, LBD causes the worker to supply labor *less elastically when young* and more elastically when old. As a result, labor taxes distort workers' choices less when they are young relative to when they are old, so the government optimally taxes workers more when they are young. A positive tax on capital can mimic such an optimal tax that decreases with age. We refer to this channel as the "elasticity channel".

Under LOD the elasticity channel is also present, but it works in the opposite direction. In this case, training is a substitute for working, as both increase lifetime income, but only young workers train. This causes workers to supply labor *more elastically when young*. Therefore, with LOD the elasticity channel induces the government to impose labor taxes that increase with age. However, under LOD there exists a second force affecting optimal taxes, which we label the "savings channel." In this model workers can save in two ways: by buying assets or by training. When, due to the elasticity channel, labor taxes increase with age, the after-tax return to training is distorted downwards. The savings channel arises because the government wants to equalize the before-tax return to the two methods of saving. Therefore, the savings channel works against the elasticity channel. As a result, whether taxes increase or decrease with age in the LOD model and how the optimal capital tax is affected when age-dependent taxes are disallowed is unclear.

After deriving these analytic results, we quantify the strength of these channels by solving for optimal taxes in rich computational setting where human capital is accumulated

---

[3]A host of work demonstrates a similar set of results in a two generation model with a single cohort. Two important examples are Atkinson and Stiglitz (1976), and Deaton (1979).

either exogenously, with LBD, or with LOD. Agents enter the model at age 20 and work until retirement at age 65, at which point they begin collecting Social Security benefits. Agents die stochastically between age 20 and 100. In all three models, agents are heterogeneous in their initial human capital and learning abilities, which translates to heterogeneity in the initial level and steepness of life-cycle wage profiles. The model parameters governing this heterogeneity are set to match key moments for the US. In particular, we set the joint distribution over learning abilities and initial human capital to replicate the mean and variance of earnings at different ages over the working life. Additionally, the calibrated models closely match the age profile of mean savings in the US economy, even though this is untargeted.

We search for optimal labor and capital taxes within our three calibrated models. When we restrict the government to proportional taxes on labor and capital, we find that relative to the exogenous model, LBD leads to an increase in the optimal capital tax because of the elasticity channel. By contrast, LOD leads to a decrease in the optimal capital tax because the elasticity channel, working in the opposite direction, is much stronger than the savings channel. Overall, there is a 19 pp gap between the optimal capital tax in the LBD and LOD models.

We find that heterogeneity is an important driver of this gap because it interacts with the tax considerations generated by human capital. In particular, heterogeneity dramatically reduces the strength of the savings channel in the LOD model. When workers are homogeneous, the government taxes capital, in part, to induce young workers to save via human capital as opposed to saving via physical capital. With heterogeneous workers, however, the young workers with the greatest potential for saving via human capital (workers with high learning ability) are also the workers with the least desire to save via physical capital (because they have the steepest earnings profiles). As a result, when workers have heterogeneous learning abilities in the LOD model taxing capital has a small overall effect on human capital decisions, which implies a weak savings channel and a lower optimal tax on capital.

Finally, we highlight strong interactions between progressive labor taxes, capital taxes, and endogenous human capital accumulation. Optimal labor taxes are more progressive in LOD than in LBD. One reason is that progressive taxes ease borrowing constraints for low income workers, and there are more borrowing constrained workers with LOD because young

workers must sacrifice labor income to train. Further, in addition to redistributing resources across workers within the same cohort, progressive taxes also redistribute resources *over the life-cycle*, from older to younger workers, since income tends to rise with age (see earlier work by Erosa and Gervais (2002), Garriga (2001), Conesa and Krueger (2006), and Gervais (2012)). That is, a progressive tax mimics a labor tax that is increasing in age. This works against the optimal age-dependence of taxes in the LBD model, but aligns with the optimal age-dependence of taxes in the LOD model, leading to more progressive taxes in the latter.

Importantly, we find that allowing the government to make labor taxes progressive somewhat reduces the gap in optimal capital taxes between LBD and LOD. It does so for two reasons. First, in the LOD model the government would like labor taxes to increase with age, but capital taxes implicitly tax younger workers more. However, progressive taxes mimic an age-dependent tax that is increasing in age, which softens the blow of capital taxes. Second, progressive taxes impose lower average tax rates on low income workers, which require the government to raise additional revenue. The higher degree of progressivity in the LOD model leads to larger increases in marginal tax rates on labor and capital to raise this additional revenue. Combined, these two forces reduce the gap in the optimal tax on capital between LBD and LOD from 19pp to 8pp.

This paper is related to a sizeable literature that argues it can be optimal to tax capital in order to mimic a set of missing tax instruments (see Correia (1996), Armenter and Albanesi (2009), and Jones et al. (1997)). Specifically, we combine two related strands of research within this literature. The first strand quantitatively analyzes optimal capital taxation within structural life-cycle models with exogenous human capital accumulation. For example, Conesa et al. (2009) find that the optimal tax on capital is large, motivated, in part, by the desire to mimic a labor income tax which is decreasing in age. This type of tax is optimal because in their model workers supply labor more elastically as they age.[4] Using a similar model setting, Peterman (2013) points out that another reason Conesa et al. (2009) find a large optimal capital tax is that they do not allow the government to tax accidental bequests at a different rate from ordinary capital income. Moreover, Cespedes and Kuklik (2012) find that when a non-linear mapping between hours and wages is incorporated into a similar model hours become more persistent and the optimal capital tax falls significantly.

---

[4]This is because (i) in their model labor falls with age, and (ii) they adopt a utility function where the Frisch elasticity is decreasing in labor.

We extend these studies by determining how incorporating different forms of endogenous human capital accumulation affect optimal capital taxes in these types of life cycle models.

The second strand of research analyzes the impact of endogenous human capital on optimal taxes but excludes life cycle considerations.[5] For example, Jones et al. (1997) and Judd (1999) characterize optimal taxes in an infinitely lived agent model in which agents are required to use market goods to acquire human capital similar to ordinary capital. They find that if the government can distinguish between pure consumption and human capital investment, then, similar to a model with exogenous human capital accumulation, it is not optimal to distort either human or physical capital accumulation in the long run. Reis (2007) shows in a similar model that if the government cannot distinguish between consumption and human capital investment, then the optimal capital tax is also zero as long as the level of capital in the economy does not influence the relative productivity of human capital. Chen et al. (2010) study a similar setting that also includes labor search, and find that endogenous human capital increases optimal capital tax because the positive tax on capital unravels the labor frictions induced by search.

In addition to ours, three other papers also bridge these two literatures and examine the impact of endogenous human capital on optimal taxes within a life-cycle framework. Crucially, our paper is the first to feature heterogeneity in learning ability across workers, which we show has a first order effect on optimal taxes. Further, each of these existing works focus on different aspects of optimal taxation than the current paper. Karabarbounis (2016) characterizes optimal taxes that can depend on age, assets, and filing status within a LBD model. He finds large welfare gains from allowing taxes to vary with these tags, and, like us, argues that heterogeneity in the elasticity of labor supply across workers is crucial. da Costa and Santos (2018) study optimal age-dependent labor income taxes in the presence of endogenous human capital. Rather than focusing on how the optimal capital tax changes in order to mimic an age-dependent tax with different forms of human capital accumulation, they focus on the case where taxes can be conditioned on age. Instead, we focus on the role that capital taxes can play in mimicking age-dependent taxes. Finally, Peterman (2016a) finds that adding LBD causes the optimal tax on capital to increase when age-dependent

---

[5]Our paper follows a large literature that uses the Ramsey approach to optimal taxation. A parallel literature in dynamic public finance also examines optimal taxation with endogenous human capital. Important examples of this include Golosov et al. (2003), Stantcheva (2015), and Stantcheva (2017).

taxes are not allowed, but does not compare optimal taxes in LBD to optimal taxes in LOD.

2 A Simple Model of Taxation with Human Capital

We present a tractable overlapping generations model in which agents live for 2 periods to demonstrate that the presence of endogenous human capital can induce a welfare-maximizing government to impose age-dependent labor income taxes. Garriga (2001) shows that if the period utility function is additively separable and homothetic in each consumption and labor, then optimal labor income taxes do not depend on age. That paper assumed an exogenous human capital model. Here we show that if, alternatively, human capital is accumulated endogenously, then optimal labor income taxes do depend on age. Further, we show that the age dependence of taxes qualitatively depends on whether human capital is accumulated via a LBD model or via a LOD model. We conclude by demonstrating that taxes on capital income can imperfectly mimic age-dependent labor taxes if the latter are not feasible.

Our simple model abstracts from several important features for analytical tractability, including retirement, population growth, heterogeneity, progressive taxes, mortality risk, and market-clearing factor prices. These features are included in the full computational model.

2.1 Simple Model Environment

Time is discrete, and each period an identical worker is born. Workers live for two periods. There is one consumption good c whose price is normalized to one. At each age $s \in \{1, 2\}$ the worker has human capital h_s. The worker supplies a portion of their labor endowment, l, in a competitive labor market. Workers can save a portion of their age 1 earnings as capital, k_1. The net rate of return on capital is r. Workers cannot borrow.

Human capital at age 1, h_1, is simply an endowment, and is the same for all generations. For simplicity we normalize $h_1 = 1$. In the exogenous human capital model, h_2 is also an endowment. In the LBD version of the model, h_2 depends on period 1 labor, l_1. In the LOD version h_2 depends on period 1 human capital investment, x_1. Investment is distinct from labor because workers are not paid a wage for investment time. In the LBD model we assume that $h_2(l_1)$ is strictly increasing and strictly concave in l_1, with infinite marginal product at $l_1 = 0$. In the LOD model we make the analogous assumptions for $h_2(x_1)$ and x_1.

A government levies taxes in order to finance government consumption stream G. The government imposes linear labor income taxes at each age, τ_1 and τ_2, and linear capital income taxes, τ_k. The government budget constraint and aggregate resource constraint in the economy are, respectively,

$$G = \tau_1 l_{1,t} + \tau_2 h_{2,t-1} l_{2,t-1} + \tau_k r k_{1,t-1} \tag{1}$$

$$c_{1,t} + c_{2,t-1} + k_{1,t} - k_{1,t-1} + G = r k_{1,t-1} + l_{1,t} + h_{2,t-1} l_{2,t-1} \ , \tag{2}$$

where the first subscript denotes a worker's age and the second subscript denotes the period in which the worker was born. We suppress the latter subscript for the remainder of this section since all generations are identical.

Workers have preferences over consumption c and non-leisure time $l + x$ in each period represented by a $u(c, l + x)$. [6] Workers discount period 2 utility at rate β, and maximize lifetime utility, subject to a lifetime budget constraint and non-borrowing constraints:

$$\max_{c_1,c_2,l_1,l_2,x_1,s_1 \geq 0} \quad u(c_1, l_1 + x_1) + \beta u(c_2, l_2 + x_2) \tag{3}$$

$$s.t. \quad c_1 + k_1 = (1 - \tau_1) l_1 \ ; \tag{4}$$

$$c_2 = (1 - \tau_2) h_2(\cdot) l_2 + (1 + (1 - \tau_k) r) k_1 \ . \tag{5}$$

2.2 Solution to the Worker's Problem

Below we write the worker's first order conditions. We denote the Lagrange multiplier for the age s budget constraint by ρ_s. For brevity we denote $u(c_s, l_s + x_s)$ by u_s.

$$\frac{\partial u_1/\partial l_1}{\partial u_1/\partial c_1} = (1 - \tau_1) + (\rho_2/\rho_1)(1 - \tau_2)(\partial h_2/\partial l_1) l_2 \ ; \tag{6}$$

$$\frac{\partial u_1/\partial x_1}{\partial u_1/\partial c_1} = (\rho_2/\rho_1)(1 - \tau_2)(\partial h_2/\partial x_1) l_2 \ ; \tag{7}$$

$$\frac{\partial u_2/\partial l_2}{\partial u_2/\partial c_2} = (1 - \tau_2) h_2 \ ; \tag{8}$$

$$\frac{\partial u_1/\partial c_1}{\partial u_2/\partial c_2} = (\rho_2/\rho_1)\beta(1 + r(1 - \tau_k)) \ . \tag{9}$$

We write in black the terms and equations that are common to all versions of the

---

[6]Note that workers do not train in age 2 of the LOD model because there is no age 3 to reap the rewards of investing.

model, in blue the terms that only apply to the LBD version, and in red the terms that only apply to the LOD version. Equation (6) describes the ratio of the first order conditions on l_1 and c_1. In the LBD model the additional term in blue appears because labor at age 1 also increases human capital at age 2. Equation (7) describes the ratio of the first order conditions on x_1 and c_1, which pertains only to the LOD model in period 1.

2.3 Optimal Taxes

This section solves for the optimal tax policy (τ_1, τ_2, τ_k). We use the primal approach, solving first for the optimal allocation, then using this allocation to back out optimal taxes. The details of the derivations are in Appendix A. We denote the Lagrange multiplier on the consumption-labor implementability constraint by λ and the multiplier on the additional consumption-investment implementability constraint (in the LOD model) by η.

To derive optimal taxes we assume the following functional form for utility:

$$u(c, l+x) = \frac{c^{1-\sigma_1}}{1-\sigma_1} - \chi \frac{(l+x)^{1+1/\sigma_2}}{1+1/\sigma_2} \; . \tag{10}$$

Garriga (2001) showed that optimal taxes do not depend on age with this utility function and exogenous human capital accumulation. Therefore, assuming this utility function allows us to isolate the impact of endogenous human capital on optimal age-dependent taxes.

Optimal labor taxes (τ_1, τ_2) satisfy the following conditions in each of the three versions:

$$\frac{1-\tau_2^{exog}}{1-\tau_1^{exog}} = \frac{1+\lambda(1+1/\sigma_2)}{1+\lambda(1+1/\sigma_2)} = 1 \tag{11}$$

$$\frac{1-\tau_2^{lbd}}{1-\tau_1^{lbd}} = \frac{\left[1+\lambda(1+1/\sigma_2)\left(1-\frac{l_1(\partial h_2/\partial l_1)}{h_2}\right)\right]\left[1+\frac{l_2 h_2}{1+r(1-\tau_k)}\right]}{1+\lambda(1+1/\sigma_2)+\lambda l_2^{1+1/\sigma_2}l_1^{1-1/\sigma_2}h_2^{-1}\left[\frac{\partial h_2/\partial l_1}{h_2}-\frac{\partial^2 h_2}{\partial l_1^2}\right]} - \frac{l_2(\partial h_2/\partial l_1)}{1+r(1-\tau_k)} > 1 \tag{12}$$

$$\frac{1-\tau_2^{lod}}{1-\tau_1^{lod}} = \frac{1+\lambda\left(1+\frac{l_1}{\sigma_2(l_1+x_1)}\right)+\frac{\eta h_2}{\sigma_2(l_1+x_1)}}{1+\lambda(1+1/\sigma_2)-\eta(\partial h_2/\partial x_1)(1+1/\sigma_2)} \neq 1 \; . \tag{13}$$

Equation (11) shows that in the exogenous model, optimal labor taxes do not depend on age, consistent with Garriga (2001). Taxes do not depend on age because period utility is (i) homothetic in both consumption and labor, and (ii) separable in both consumption and labor. In contrast, Equations (12) and (13) show that in the endogenous human capital models, optimal labor taxes do depend on age, overturning the result in Garriga (2001). We

now explain the intuition for LBD and LOD separately.

2.3.1 Optimal Taxes in the Learning By Doing Version

In the exogenous human capital version, the Frisch elasticity of labor at both ages is simply $\varepsilon_F^{exog} = \sigma_2$. The Frisch elasticity of labor in age 2 of the LBD model is also σ_2, since there are no human capital considerations in age 2. However, in period 1 of the LBD model, the Frisch elasticity is given by

$$\varepsilon_F^{lbd} = \frac{\sigma_2}{1 + \frac{h_2}{h_1(1+r)}\left[(\partial h_2/\partial l_1) - l_1\sigma_2(\partial^2 h_2/\partial l_1^2)\right]} < \sigma_2 \ , \tag{14}$$

where the less-than symbol results from $\partial h_2/\partial l_1 > 1$ and $\partial^2 h_2/\partial l_1^2 < 1$.

   The lower Frisch elasticity in the LBD model is because now l_1 has the additional benefit of increasing h_2. Thus, with LBD labor is supplied less elastically by younger workers relative to older workers. This creates an incentive for the government to tax the labor income of younger workers at a relatively higher rate. We use the term "elasticity channel" to describe the effect on optimal tax policy caused by a change in the Frisch elasticity from including endogenous human capital.

2.3.2 Optimal Taxes in the Learning Or Doing Version

In period 1 of the LOD model, the Frisch elasticity is given by

$$\varepsilon_F^{lod} = \sigma_2\left(\frac{l_1 + x_1}{l_1}\right) > \sigma_2 \tag{15}$$

The age 1 Frisch elasticity in the LOD model differs from the exogenous model because now there are two uses of time: labor and investment. Workers can now defer labor income to age 2 by reducing l_1 and increasing x_1, which increases h_2. This means that age 1 labor is more elastic with respect to τ_1 in the LOD model than age 2 labor, which incentivizes the government to keep τ_1 lower relative to τ_2. Thus, the "elasticity channel" in the LOD model works in the opposite direction as in the LBD model.

   The LOD model also features a second channel that impacts optimal taxes, which

causes the inclusion of two terms

$$\frac{\eta h_2}{\sigma_2(l_1 + x_1)} \, , \tag{16}$$

$$- \eta(\partial h_2/\partial l_1)(1 + 1/\sigma_2) \, , \tag{17}$$

in the numerator and denominator of Equation (13), respectively. The savings channel arises from the intertemporal link created because agents can now save in two ways: first by holding assets, and second by training. Quantitatively, we find that η is positive and these additional terms cause the right hand side of (13) to increase. In this case, the planner is incentivized to keep τ_2 low relative to τ_1, so as to not discourage the worker from "saving by investing" in age 1. Intuitively, the elasticity channels leads the social planner to tax labor income when an agent is old at a relatively higher rate. However, this discourages agents from accumulating human capital and encourages savings with physical capital. The savings channel leads the social planner to offset some of this discouragement of savings via human capital. Therefore, the two channels may have opposing effects on the optimal tax policy, and the overall effect is ambiguous.

2.4 The Connection Between Age-Dependent Taxes and Capital Taxes

Positive capital income taxes can imperfectly mimic age dependent taxes, which is important if the government is not able to condition taxes on age. The connection between optimal age-dependent labor taxes and capital taxes is well-established; see e.g. Erosa and Gervais (2002) and Garriga (2001) for in-depth analyses. Here we briefly provide intuition for how optimal capital taxes can depend on the mode of human capital accumulation (exogenous, LBD, or LOD) when age-dependent labor taxes are not feasible.

To begin, we combine the worker's two budget constraints (4) and (5):

$$c_1 + \left(\frac{1}{1 + (1 - \tau_k)r} \right) c_2 = (1 - \tau_1)l_1 + \left(\frac{1 - \tau_2}{1 + (1 - \tau_k)r} \right) l_2 h_2(\cdot) \, . \tag{18}$$

The above equation reveals how a positive capital tax can mimic labor taxes that decrease with age: both a decrease in τ_2 and an increase in τ_k raise the after-tax price of age 2 labor income without affecting the after-tax price of age 1 labor income. However, this equation also makes clear that the capital tax only mimics age-dependent taxes *imperfectly*: raising

τ_k also raises the price of consumption in age 2 relative to age 1.

The effect is that capital taxes introduce wedges into the workers' optimality conditions in ways that mimic age dependent taxes:[7]

$$\frac{\partial u_1/\partial l_1}{\partial u_2/\partial l_2} = \frac{\beta(1+r(1-\tau_k))}{h_2}\left(\frac{1-\tau_1}{1-\tau_2}\right) + \left(\frac{\partial h_2}{\partial l_1}\right)\left(\frac{l_2}{h_2}\right); \tag{19}$$

$$\frac{\partial u_1/\partial l_1}{\partial u_1/\partial x_1} = \frac{\beta(1+r(1-\tau_k))}{l_2}\left(\frac{1-\tau_1}{1-\tau_2}\right)\left(\frac{1}{\partial h_2/\partial l_1}\right). \tag{20}$$

Equation (19) shows that the government can introduce a wedge into the intertemporal Euler equation for labor by either increasing the ratio τ_1/τ_2 or by increasing τ_k. Similarly, (20) shows that in the LOD model the government can introduce a wedge into the age 1 labor/investment optimality condition by doing the same. Given that we have shown that the optimal age profile of labor taxes can depend qualitatively on the mode of human capital accumulation, these equations show that when age dependent taxes are infeasible, then the optimal tax on capital may also strongly depend on the mode of human capital accumulation.

3   Quantitative Model

We now construct richer, computational versions of the exogenous-human-capital, LBD, and LOD models, which we use to quantify the impact of endogenous human capital for the optimal capital tax. The models in this section include several features that we abstracted from in the previous section, including retirement, mortality risk, heterogeneity in productivity and learning ability, progressive taxes, and are solved for in general equilibrium.

3.1   Demographics and Endowments

Time is discrete: agents enter the model when they start working at the age of 20, and can live to a maximum age of J: thus, the model is populated with J-19 overlapping generations. Conditional on being alive at age j, Ψ_j is the probability of an agent living to age $j+1$. If an agent dies with assets, the assets are confiscated by the government and distributed equally to all the living agents as transfers (TR). All agents retire exogenously at age j_r.

---

[7]Equation (19) is obtained by taking the ratio of (6) and (8), while equation (20) is the ratio of (6) and (7). Again, we write in blue the pieces of the conditions specific to the LBD model, and we write in red the pieces specific to the LOD model.

In each period a cohort of new agents is born. The size of the cohort born in each period grows at rate ν. Thus, the cohort shares, $\{\mu_j\}_{j=20}^J$, are given by,

$$\mu_j = \frac{\Psi_{j-1}}{1+\nu}\mu_{j-1}, \text{ for } j = 20, ..., J, \tag{21}$$

where μ_{20} is normalized such that

$$\sum_{j=20}^J \mu_j = 1. \tag{22}$$

3.2 Preferences

Workers are endowed with one unit of productive time each period. In the exogenous and LBD models, workers divide time between labor, l, and leisure, $1-l$. In the LOD model, workers divide time between labor, l, investment, x, and leisure, $1-l-x$. Workers are only paid a wage for labor time.

Utility over consumption and leisure is time-separable and discounted, conditional on surviving, at rate β,

$$\sum_{j=20}^J \Psi_{j-1}\beta^j u(c_j, l_j + x_j) \ .$$

3.3 Human Capital Accumulation

At birth agents are endowed with initial human capital h_{20} and a learning ability α which determines their skill in accumulating new human capital. Initial endowments are heterogeneous across individuals.

In the exogenous model, an agent's period $j+1$ human capital h_{j+1} is exogenous and pre-determined. In the LBD model, h_{j+1} is a function of learning ability α, as well as existing human capital and the agent's labor choice, $h_{j+1} = H^{lbd}(h_j, l_j; \alpha)$. In the LOD model, h_{j+1} is a function of learning ability α, existing human capital, and the agent's investment choice, $h_{j+1} = H^{lod}(h_j, x_j; \alpha)$.

3.4 Government Policies

The government has two fiscal instruments to finance its unproductive consumption, G.[8] First, the government taxes capital income, $y_k \equiv r(a + TR)$, according to a tax schedule $T^k[y_k]$. r denotes the net return on savings, and a denotes assets held at the start of the period. Second, the government taxes each individual's taxable labor income, $y_l \equiv wh_j l_j (1 - \tau^{ss}/2)$ (where w denotes the wage rate), according to a tax schedule $T^l[y_l]$. Note that in the model, as in current US law, employer contributions to Social Security, $(\tau^{ss}/2)wh_j l_j$, are not subject to income taxes.

We impose three restrictions on the labor and capital income tax policies. First, we assume human capital is unobservable, meaning that the government cannot tax human capital accumulation. Second, we assume the rates cannot be age-dependent. Third, both of the taxes are solely functions of the individual's relevant taxable income in the current period.

In addition to raising resources for consumption in the unproductive sector, the government runs a pay-as-you-go (PAYGO) Social Security system.[9] The government pays a common benefit SS to all retired agents. SS is set such that retired agents receive an exogenously set fraction, b, of the average income of all working individuals. Benefits financed by taxing labor income at a flat rate, τ^{ss}, which is set to balance Social Security system budget each period. The Social Security system is not considered part of the tax policy that the government optimizes.

3.5 The Agent's Problem

At the start of a period agents differ along four dimensions: an age j and a state vector $z = (a; h; \alpha)$, where a denotes quantity of physical assets the agent brings into the period, h denotes human capital, and α denotes learning ability. The distribution over states z at each age j is given by the function $\Lambda_j(z)$. All individuals are born at age 20 with zero assets, so $\Lambda_{20}(z)$ reduces to a bivariate distribution over $(h_{20}; \alpha)$.

Working age agents $(j < j_r)$ split their time between labor l, human capital investment

---

[8]As opposed to assuming G is unproductive, including G such that it enters the agent's utility function in an additively separable manner will result in the same optimal tax policies, as long as G is held fixed.

[9] Peterman (2013) shows that in a similar model, excluding Social Security can generate unrealistic life cycle profiles and can impact optimal tax policy.

x, and leisure $1 - l - x$. Individuals also choose consumption c and assets saved for the next period a'. We state the agent's problem recursively:

$$V_j(z) = \max_{c,a',l,x} \quad u_j(c, \ l + x) + \beta V_{j+1}(z') \tag{23}$$

$$s.t. \quad c + a' = y - T^k[y_k] - T^l[y_l] - T^{ss} \ ; \tag{24}$$

$$y = a(1+r) + whl + TR(1+r) \ ; \tag{25}$$

$$y_k = r(a + TR) \ ; \tag{26}$$

$$y_l = wh_j l_j (1 - \tau^{ss}/2) \ ; \tag{27}$$

$$T^{ss} = \tau^{ss} wh_j l_j \ ; \tag{28}$$

$$h' = H(h, \ \cdot \ ; \alpha) \ ; \tag{29}$$

$$l, x \geq 0; \quad l + x \leq 1 \ ; \tag{30}$$

$$a' \geq 0 \ . \tag{31}$$

Equation (24) is the period budget constraint facing the worker, where y is pre-tax income from each labor and capital, T^k is capital income tax liabilities, T^l is labor income tax liabilities, and T^{SS} is Social Security tax liabilities. Equation (29) governs the evolution of human capital. In the LBD version of the model, (29) is $H^{lbd}(h, l; \alpha)$; in the LOD version of the model, (29) is $H^{lod}(h, x; \alpha)$; in the exogenous model, h' is taken as given.

If $j \geq j_r$ the individual is retired and must set $l = x = 0$. Retirees solve:

$$V_j(z) = \max_{c,a'} \quad u(c, \ 1) + \beta V_{j+1}(z') \tag{32}$$

$$s.t. \quad c + a' = y + SS - T^k[y_k] \ ; \tag{33}$$

$$y = a(1+r) + TR(1+r) \ ; \tag{34}$$

$$y_k = r(a + TR) \ ; \tag{35}$$

$$a' \geq 0 \ . \tag{36}$$

Equation (33) is the period budget constraint facing a retiree, where SS_t is the Social Security retirement benefit.

3.6 Firm

Competitive firms operate a production function, F. The aggregate resource constraint is

$$C + K_{t+1} - (1 - \delta_k)K + G \leq F(K, H), \tag{37}$$

where K, C, and H represent the aggregate capital stock, aggregate consumption, and aggregate labor (measured in efficiency units), respectively. The depreciation rate for physical capital is δ_k.

3.7 Definition of Stationary Competitive Equilibrium

Given a Social Security replacement rate b, government expenditures G, and a sequence of population shares $\{\mu_j\}_{j=20}^J$, a Stationary Competitive Equilibrium consists of a collection of agent choices for each age j and state z, $\{c_j(z), k_j'(z), l_j(z), x_j(z)\}$, distributions of individuals $\Lambda(z) = (\Lambda_j(z))$, a production plan for the firm (H, K), a government labor tax schedule T^l, a government capital tax schedule T^k, a Social Security tax rate τ^{ss}, a period utility function u, Social Security benefits SS, prices (w, r), and transfers TR such that:

1. Given prices, policies, transfers, and benefits, the agent choices solve (23)-(35).
2. Prices r and w satisfy
$$r = F_1(K, H) , \tag{38}$$
$$w = F_2(K, H) . \tag{39}$$

3. Social Security policies satisfy
$$SS = b \frac{wH}{\sum_{j=20}^{j_r-1} \mu_j} \tag{40}$$

   and
$$\tau_{ss} = \frac{SS}{w \sum_{j=20}^{j_r-1} \mu_j}. \tag{41}$$

4. Transfers are given by
$$TR = \sum_{j=20}^J \mu_j(1 - \Psi_j)a_{j+1}. \tag{42}$$

5. Government balances its budget
$$G = \left(\sum_{j=20}^J \mu_j \int T^k[r(a + TR)]\Lambda_j(dz) \right) + \left(\sum_{j=20}^{j_r-1} \mu_j \int T^l[whl_j(z)(1 - \tau^{ss}/2)]\Lambda_j(dz) \right). \tag{43}$$

15

6. Capital, labor, and output markets clear

$$H = \sum_{j=20}^{J} \mu_j \int hl_j(z)\Lambda_j(dz) \,, \tag{44}$$

$$K = \sum_{j=20}^{J} \mu_j \int a\Lambda_j(dz) \,, \tag{45}$$

$$C = \sum_{j=20}^{J} \mu_j \int c_j(z)\Lambda_j(dz) \,, \tag{46}$$

$$G + C + K(1+\nu) = F(K,H) + K(1-\delta^k) \tag{47}$$

4   Parameter Values in the Baseline Economies

This section sets parameter values for the three baseline economies: the exogenous human capital model, the LBD model, and the LOD model. Parameter values and targets are summarized in Tables 1 and 2.

4.1   Demographics

Agents enter the model at age 20 and exogenously retire at age 65. We set conditional survival probabilities using estimates by Bell and Miller (2002). Individuals who survive until age 100 die for certain the next period. We assume an annual population growth rate of 1.1%.

4.2   Preferences

Agents have time-separable preferences over consumption and non-leisure time:

$$u(c, l+x) = \frac{c^{1-\sigma_1}}{1-\sigma_1} - \chi \frac{(l+x)^{1+\frac{1}{\sigma_2}}}{1+\frac{1}{\sigma_2}} \,. \tag{48}$$

For the exogenous human capital model this utility function implies that the Frisch labor supply elasticity is a constant. This allows us to isolate the effects of each of the channels on the optimal tax policy. Following Conesa et al. (2009) and da Costa and Santos (2018), we exogenously set $\sigma_1 = 2$, which generates a standard degree of risk aversion. [10]

---

[10] We follow Peterman (2013) by choosing a utility function which is additively separable in consumption and leisure. This eliminates any feedback between labor choices and the elasticity of labor except that which operates though the human capital

Table 1: Parameters Set Independently

	Parameter	Value	Target
Demographics	Retire Age: j_r	45	By Assumption
	Max Age: J	80	By Assumption
	Surv. Prob: Ψ_j	Bell and Miller (2002)	Data
	Pop. Growth: n	0.011	Data
Preferences	Risk aversion: σ_1	2	Conesa and Krueger (2006)
	Frisch Elasticity: σ_2	0.5	Frisch$=\frac{1}{2}$
The Firm	α	0.360	Data
	δ	0.083	$\frac{I}{Y} = 25.5\%$
	A	1	Normalization
Government Policies	λ_1	0.036	Guner et al. (2014)
	b	0.5	Conesa and Krueger (2006)

Note: This table displays values for parameters set independently for the baseline model economies. Since agents enter the model at age 20, retirement j_r corresponds to age 65 and the max age J corresponds to age 100. See text for further details.

We set σ_2, β, and χ to match moments generated endogenously by the model. These values are determined jointly along with other parameters, but we discuss the moments that the parameters are closely linked to. Past micro-econometric studies (such as Altonji (1986), MaCurdy (1981), and Domeij and Floden (2006)) estimate a Frisch elasticity between 0 and 0.5. However, more recent research has shown that these estimates may be downward biased.[11] Therefore, we set σ_2 to generate a Frisch elasticity at the upper bound, 0.5. Following Conesa and Krueger (2006), we set β to produce a capital-to-output ratio of 2.7. We set χ so that in the baseline model agents spend on average one third of their time on non-leisure activities.

4.3 Firm

We assume that the production function $F(K, H)$ is Cobb-Douglas with capital share θ set to .36. The depreciation rate is set to .083, which produces an investment to output ratio of .255.

---

channels discussed in Section 2. We note that if we were to incorporate TFP growth, the model would need to be modified to be made consistent with balanced growth; see Peterman and Sager (2018) for a discussion of how this can be done.

[11]Reasons for this downward bias include : utilizing weak instruments; not accounting for borrowing constraints; disregarding the life cycle effect of endogenous-age specific human capital; omitting correlated variables such as wage uncertainty; ignoring secondary earners; and not accounting for labor market frictions. See Imai and Keane (2004), Domeij and Floden (2006), Pistaferri (2003), Chetty (2009), Peterman (2016b), Wallenius (2011), and Contreras and Sinclair (2008).

Table 2: Parameters Set Jointly

	Parameter	Exog.	LBD	LOD	Target
Preferences	Conditional Discount: β	0.995	0.991	0.997	$K/Y = 2.7$
	Disutility of Labor: χ	41.8	73.6	72.4	$\overline{h_j + n_j} = \frac{1}{3}$
Human Capital	Mean Initial HC: μ_h	1	1	1	Normalization
	SD Initial HC: σ_h	0.569	0.532	0.555	Earnings profiles*
	Mean Ability: μ_α	-0.890	-0.807	-1.052	Earnings profiles*
	SD Ability: σ_α	0.162	0.174	0.338	Earnings profiles*
	Correlation: $\rho_{h,\alpha}$	0.990	0.969	0.990	Earnings profiles*
	Depreciation Rate: δ_h	$-$	0.096	0.073	Earnings profiles*
	Curvature on h: ϕ_1	$-$	0.580	0.280	Earnings profiles*
	Curvature on x or l: ϕ_2	$-$	0.843	0.300	Earnings profiles*
Government Policies	λ_0	0.819	0.819	0.820	Balance Budget

Note: This table displays values for parameters set jointly for the baseline model economies. * Earnings profiles include the following moments: the life-cycle peak of mean earnings, mean earnings at age 30, mean earnings at age 64, and the variance of earnings at ages 20, 30, and 64. Note that the exogenous model does not contain the human capital parameters δ_h, ϕ_1, or ϕ_2 because we take human capital profiles directly from the LBD model.

4.4 Government Policies and Tax Functions

For calibration we use an off-the-shelf tax function estimated by Guner et al. (2014):

$$T(y, \bar{y}; \lambda_0, \lambda_1) = y \left(1 - \lambda_0(y/\bar{y})^{-\lambda_1}\right) , \qquad (49)$$

where y represents the sum of labor and capital income and $\bar{y}$ is average income. We impose their estimates for λ_1, which mainly determines the progressivity of income taxes. We endogenously set λ_0 to balance the government's budget each period.[12]

Following Conesa and Krueger (2006), we set government consumption G to equal 17 percent of output in the baseline economy. When searching for the optimal tax policy we hold G constant.

The government also runs an old-age Social Security system. Social Security benefits are set so that the replacement rate b is 50 percent.[13] The payroll tax τ_{ss} balances the Social Security budget each period.

---

[12] The authors do not fit separate tax functions for labor and capital income. Accordingly, we use a uniform tax system on the sum of both sources of income when calibrating the model, but allow for separate tax policies on the two sources of income when solving for the optimal policy. Note also that with this specification a larger value for λ_0 implies a lower tax rate.

[13] The replacement rate matches the rate in Conesa and Krueger (2006).

4.5 Human Capital Accumulation

In the LBD and LOD models, human capital is accumulated according to

$$H^{lod}(h, x; \alpha) = (1 - \delta_h)h + \alpha h^{\phi_1} x^{\phi_2} \tag{50}$$

$$H^{lbd}(h, l; \alpha) = (1 - \delta_h)h + \alpha h^{\phi_1} l^{\phi_2} . \tag{51}$$

δ_h is the depreciation rate of existing human capital. α is the individual's learning ability. ϕ_1 determines the curvature of new human capital with respect to existing human capital. ϕ_2 determines the curvature of new human capital with respect to investment x or labor l in the LOD and LBD model, respectively. Note that the only difference in functional forms between H^{lod} and H^{lbd} is the relevant time input.

We impose that the joint distribution over initial human capital and learning ability at age 20 $\Lambda_{20}(h, \alpha)$ is log-Normal, i.e. $\log(h_{20}, a) \sim N(\mu_h, \mu_a, \sigma_h, \sigma_a, \rho_{h,a})$ where (μ_h, μ_a) are means for $\log(h_{20}, a)$; (σ_h, σ_a) are standard deviations for $\log(h_{20}, a)$; and $\rho_{h,a}$ is the correlation between $\log(h_{20}, a)$. We discretize the endowment space for (h_1, a) into a 50-by-7 grid. We then use the Tauchen (1986) procedure to assign population weights to each gridpoint given endowment parameters.
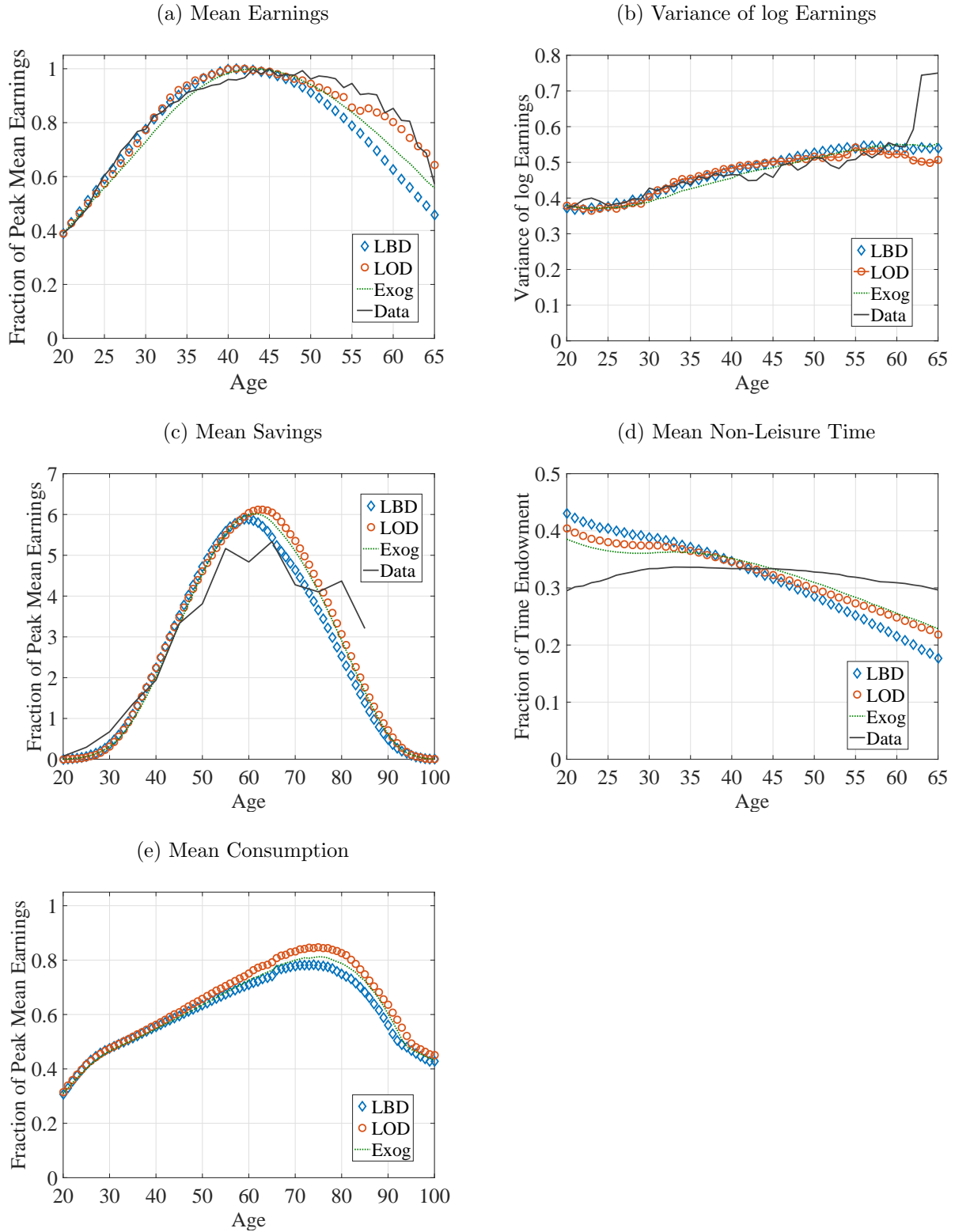
In both LBD and LOD we jointly target δ_h, ϕ_1, ϕ_2, as well as $(\mu_h, \mu_\alpha, \sigma_h, \sigma_\alpha, \rho_{h,\alpha})$, to replicate a number of moments from the life-cycle profile for mean earnings and the life-cycle profile for the variance of log earnings. There is a fairly tight link between human capital parameters and life-cycle earnings moments. All else equal, increasing the mean, μ_h, and variance, σ_h, of initial human capital increases the mean and variance of earnings *early* in the working life, respectively. Increasing the mean, μ_α, and variance, σ_α, of learning ability increases the mean and variance of earnings *late* in the working life, respectively. A correlation $\rho_{h,\alpha}$ close to one produces a positively sloped earnings variance profile early in the life-cycle, whereas a low or negative correlation produces a negatively-sloped earnings variance profile early on. The reason is that if the correlation between initial human capital and learning ability is low, then there are many workers with low initial human capital and high ability whose earnings eventually catch up to the earnings of workers with high initial human capital and low ability. For a more thorough discussion of these points see Blandin (2018).

To construct the empirical life-cycle profiles for earnings and hours used in calibration, we estimate age effects from a sample of workers in the 1980-2013 Panel Study of Income Dynamics (PSID). We restrict our sample to heads of households aged 20-65 who report annual work hours exceeding 500 hours and annual labor earnings exceeding $5,000 in 2009 dollars.[14] We provide further estimation and calibration details in Appendix B. We plot the resulting life-cycle profiles in Figure 1.

In the exogenous human capital model we assume that workers face the same individual human capital profiles generated by the baseline LBD human capital model, but that these human capital profiles are exogenous and fixed (i.e., labor has no impact on human capital). When we search for optimal taxes in Section 5, the human capital profiles in the exogenous model do not change across policy regimes. Badel et al. (2018) use a similar procedure to isolate the role of human capital.

---

[14]The sources of labor income are "wages and salary", "commission", "overtime", "professional practice income", "tips", "additional income", and "all other labor income".

Figure 1: Life-cycle Outcomes: Models and Data

(a) Mean Earnings

(b) Variance of log Earnings

(c) Mean Savings

(d) Mean Non-Leisure Time

(e) Mean Consumption

The figure displays profiles for several variables over the working life. Mean earnings, savings, and consumption are expressed as a fraction of the life-cycle maximum for mean earnings. The variances of log earnings are expressed in log points. Mean labor is expressed as a fraction of an annual time endowment, which is set to 5,840.

21

4.6 Properties of the Baseline Model

Figure 1 plots life cycle profiles from each US data and each baseline model economy. Model parameters are calibrated to match life-cycle profiles for the mean and variance of earnings in the US, and the models do so successfully. One notable exception is that the models do not produce the sharp uptick in the variance of log earnings after age 62 in the data, which is primarily driven by a growing number of workers "gradually retiring" and is outside the scope of this analysis.[15]

While the models are calibrated to match the overall mean of non-leisure time in the data, we do not target the age profile of non-leisure time. The models produce a too-large decline in mean non-leisure time over the life-cycle. This occurs because the return to accumulating human capital steadily declines as workers approach retirement, which induces workers to work or invest less. One way of reducing this downward slope would be to introduce a penalty to part time work, as in Aaronson and French (2004).

We calibrate the models to match the aggregate capital to output ratio, but do not target the life-cycle profile of capital (savings). The fact that the age profiles of mean savings in the models are close to the age profiles in the data is an important validation of the model.

5 Quantitative Analysis

5.1 The Optimal Taxation Problem

The government's goal is to maximize a Social Welfare function which we take as the expected utility of an agent prior to entering the model at age 20:

$$\int \left(u\left(c_{20}(z), l_{20}(z) + x_{20}(z)\right) + \sum_{s=21}^{J-j_r-1} \left[\beta^s \prod_{q=21}^{s} (\Psi_q) u\left(c_s(z), l_s(z) + x_s(z)\right) \right] \right) \Lambda_{20}(dz) . \quad (52)$$

The government chooses the steady state tax policy to maximize the welfare of an agent who will enter the economy in its new steady state. That is, the government does not take into account the welfare of agents alive during the transition to the new steady state. There are several reasons why the welfare change of an agent alive during the transition will differ from the welfare change of an agent who enters the economy in the new steady state.

---

[15]See Peterman (2012) for a thorough discussion of the effect of endogenous retirement on optimal taxes.

Transitional agents are less able to respond to the new tax policy relative to agents born after the reform because they have already accumulated stocks of physical and human capital in response to the baseline tax policies and factor prices. Additionally, because aggregate stocks of physical and human capital only gradually converge to the new steady state level, agents alive during the transition will experience a different sequence of interest rates and wage rates than agents who enter during the new steady state. Finally, we note that if one were to explicitly incorporate the transitional welfare implications of a new policy, one would also want to consider the optimal use of government debt in order to implement a gradual transition in tax rates.[16]

We use the Social Welfare function to conduct a series of optimal taxation exercises designed to quantify the various forces at work. In Sections 5.2 and 5.3 we search for the optimal proportional tax rates on labor τ_l and capital τ_k. In Section 5.4 we allow the government to impose labor taxes according to the functional form in (49), which nests a proportional tax but also allows for progressive taxation of labor income.[17]

Collectively our exercises deliver two main findings. First, the optimal tax on capital is in every case substantially higher in the LBD model relative to the LOD model. Second, within-cohort heterogeneity has important effects on optimal taxes, and in particular it has important effects on the gap in the optimal capital tax between LBD and LOD.

5.2 Proportional Labor Taxes

We begin by examining optimal tax policies when the government is restricted to using a flat tax on capital and labor, which allows us to isolate the effect of human capital accumulation on the optimal capital tax. Later we allow the government to use progressive labor taxes in order to see how the optimal capital tax and human capital accumulation interact with progressive taxation.

The top panel of Table 3 outlines the optimal taxes when the government is restricted

---

[16]Related to these points, see Röhrs and Winter (2017) and Fehr and Kindermann (2015).

[17]All optimal tax exercises presented here assume that the wage rate and interest rate adjust to clear the markets for labor and capital. However, we have also run these optimal tax exercises under the alternative assumption of fixed factor prices. In these exercises we find a similar gap in optimal capital taxes between LBD and LOD; however, in each case the level of the optimal capital tax is lower. Some intuition for this is that with fixed factor prices a 1pp increase in the capital tax rate lowers the equilibrium after tax return on savings by exactly 1pp. In contrast, with adjusting prices the equilibrium after tax return tends to fall by less than 1pp since the capital to labor ratio tends to fall, which raises the pre-tax return to savings. As a result, in equilibrium, savings tend to respond less to a given increase in the capital tax rate in the closed economy case, which encourages the government to tax capital at relatively higher rates.

Table 3: Optimal Tax Policies

Worker Heterogeneity	Tax Scheme	Tax Parameter	Exog.	LOD	LBD
Heterogeneous	Proportional Labor Taxes	τ_k	0.117	0.094	0.280
		τ_l	0.202	0.207	0.168
Homogeneous	Proportional Labor Taxes	τ_k	0.149	0.186	0.242
		τ_l	0.196	0.188	0.175
Heterogeneous	Progressive Labor Taxes	τ_k	0.216	0.282	0.363
		λ_0^l	0.842	0.854	0.868
		λ_1^l	0.460	0.408	0.348

The table displays optimal taxes in the exogenous, LOD, and LBD models. All exercises search for the optimal proportional capital tax along with an optimal labor tax. In the cases where labor taxes are required to be proportional, we refer to the optimal average/marginal labor tax as τ_l. In the case where labor taxes can be progressive according to the functional form in (49), we refer to the two optimal labor tax parameters as λ_0^l and λ_1^l.
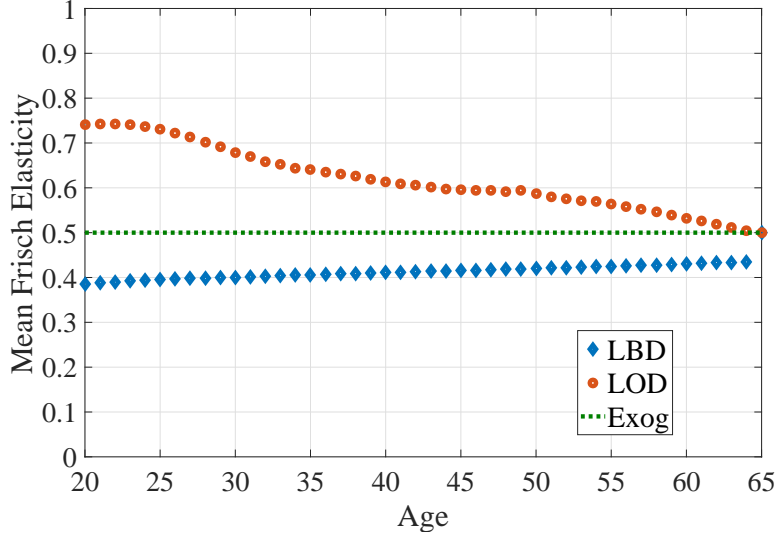
to using proportional taxes. We find an optimal tax on capital of 11.7% in the exogenous model, 28.0% in the LBD model, and 9.4% in the LOD model. Thus, there is a considerable variation in the optimal capital tax depending on how human capital is accumulated.

Relative to the exogenous model, we find that the optimal tax on capital is significantly higher in the LBD model. As described in Section 2, a reason for the larger tax on capital in the LBD model is the "elasticity channel". In particular, in the exogenous model, workers of all ages have a Frisch elasticity equal to σ_2; see Figure 2.[18] In the LBD model there is an additional return to working (learning) which lowers the Frisch elasticity for workers. However, this additional return declines to zero as workers approach retirement, which means workers in the LBD model have a Frisch elasticity that is increasing in age. The optimal capital tax is higher in the LBD model in order to implicitly tax workers at a higher rate when they are younger and have lower Frisch elasticities.

Relative to the exogenous model, the optimal capital tax in the LOD model is slightly lower. In the LOD model, both the elasticity and savings channels affect the optimal capital tax. As we show in Section 2, the elasticity channel lowers the optimal capital tax in the LOD model, the opposite direction as in the LBD model, because in the LOD model younger workers have a higher Frisch elasticity than older workers: Figure 2 shows that the LOD Frisch elasticity decreases from above 0.7 at age 20 to 0.5 at age 64. Conversely, the savings

---

[18]The Frisch profiles are determined under each model's respective optimal tax policies.

Figure 2: Frisch Elasticity Over the Life-Cycle

The figure displays profiles for the mean Frisch labor elasticity over the working life. See Appendix A for derivations of these elasticities for each model.

channel increases the optimal capital tax because a higher capital tax will lead agents to lower their saving-via-assets and increase their saving-via-training. Overall, this elasticity channel dominates and the optimal capital tax declines relative to the exogenous optimal.

5.3 The Effect of Heterogeneity

In order to isolate the effect of heterogeneity, we solve for the optimal tax policy in a model that excludes within cohort heterogeneity in skills and learning ability. See Appendix B for details of how we adjust our calibration procedure for the homogeneous case.

The middle panel of Table 3 details the optimal proportional taxes in this simplified setting. Although there is still a large gap between the optimal capital taxes in the LBD and LOD model, this gap decreases from 19pp to 8pp when within cohort heterogeneity is removed. Moreover, we find that in the LOD model without heterogeneity the savings channel dominates the elasticity and the optimal tax on capital is now larger than in the homogeneous exogenous model.

The intuition behind the change in the tax gap is that eliminating worker heterogeneity dramatically reduces the strength of the savings channel in the LOD model. In the homogeneous model the government taxed capital, in part, to induce young workers to save via

Table 4: Percent Change In Variables Under Optimal Taxes, Relative To Baseline Economy

	Exogenous		LBD		LOD	
	Prop.	(Prog.)	Prop.	(Prog.)	Prop.	(Prog.)
Real variables						
Consumption	+1.8%	(−16.4%)	+0.7%	(−19.7%)	+1.9%	(−16.5%)
Output	+2.4%	(−13.1%)	+0.1%	(−17.1%)	+2.6%	(−13.4%)
Capital	+5.9%	(−14.5%)	−1.5%	(−23.6%)	+6.3%	(−15.6%)
Effective Labor	+0.8%	(−12.3%)	+0.9%	(−13.6%)	+0.7%	(−12.2%)
Labor	+1.2%	(−16.1%)	+1.8%	(−8.2%)	+1.1%	(−12.8%)
Training	NA	(NA)	NA	(NA)	+0.4%	(−14.0%)
Prices						
Wages	+1.7%	(+0.8%)	−0.8%	(−4.0%)	+1.8%	(−1.3%)
Interest Rate	−11.0%	(−5.7%)	+5.5%	(+29.1%)	−11.9%	(+9.0%)

This table documents the percent change in economic aggregates and factor prices under optimal taxes relative to the calibrated baseline steady state. For each model, the left column corresponds to optimal taxes when labor taxes are required to be proportional; the right column in parentheses corresponds to optimal taxes when labor taxes are allowed to be progressive. Aggregate human capital is the quantity of efficiency units supplied in the labor market. In the exogenous and LBD models training is not present.

learning as opposed to saving via capital accumulation. With heterogeneous workers, however, the young workers with the greatest potential for saving via learning (workers with high learning ability) are also the workers with the least desire to save via capital accumulation (because they have the steepest earnings profiles). Therefore, taxing capital has a smaller effect on workers' human capital accumulation decisions in the heterogeneous worker model. Overall, the stronger savings channel in the homogeneous model leads the optimal tax on capital to increase relative to the case with heterogeneous workers.

5.4 Progressive Labor Taxes

Our final exercise finds optimal taxes when the government is allowed to make labor taxes progressive. Allowing for progressive taxation means that the government can use an alternative tool to redistribute beside just the flat tax on capital or labor.

The bottom section of Table 3 displays the optimal tax policies with progressive labor taxes. Regardless of the human capital technology, the government uses at least some amount of progressivity in order to redistribute resources away from high income individuals and

towards low income individuals. To pay for this redistribution, the government raises the marginal tax rates on both labor income and on capital income.

Allowing for progressive labor taxes reduces the gap in optimal capital taxes between LBD and LOD somewhat. However, capital taxes remain 8.1pp (29%) higher in the LBD model than the LOD model.[19] The gap in optimal capital taxes between LBD and LOD decreases with progressive labor taxes because the labor tax provides more redistribution with LOD than the LBD model (Figure 3 shows that marginal labor tax rates are more negative for workers earning less than roughly 15% of mean income). This additional redistribution in the optimal LOD policy requires additional revenue, which is partly financed by higher capital taxes.

There are two reasons why it is desirable to have more progressive labor taxes in the LOD model relative to the LBD model. First, there are more borrowing constrained young workers in the LOD model relative to the LBD model, despite the fact that the two models produce similar earnings distributions (see Figure 4). This is because training is a substitute for working in the LOD model, which induces many young workers to sacrifice earnings early on in order to raise their future earning potential. Because progressive taxes transfer resources from high earners to low earners, and borrowing constrained workers tend to be low earners, progressive taxes work to ease borrowing constraints. Since there are more borrowing constrained workers in the LOD model than the LBD model, this force is stronger in the former.

Second, in addition to redistributing resources across heterogeneous individuals within the same cohort, progressive taxes also redistribute resources *over the life-cycle* (as emphasized in Gervais (2012)). That is, progressive taxes can mimic an age-dependent tax that is increasing in age: high earners, who tend to be older, face higher taxes than low earners, who tend to be younger. In Section 2 we showed that optimal taxes in the LOD model increase with age, which a progressive tax mimics. In contrast, optimal taxes in the LBD model decrease with age, which is mimicked by a regressive, not a progressive, tax. Thus, the progressive tax in the LOD model raises welfare in two ways: introducing an increasing tax over the life cycle and also redistributing from high income to low income agents. In

---

[19]Similarly, Peterman (2016a) finds that the optimal tax on capital increases in a model with LBD, compared to exogenous human capital accumulation, with heterogeneity. However, that model introduces heterogeneity through idiosyncratic shocks to labor productivity, as opposed to learning ability.

Table 5: Welfare Gains Of Moving To Optimal Taxes

	WELFARE GAINS RELATIVE TO BASELINE TAXES		
Welfare Measure	Exog model	LOD model	LBD model
Ex Ante Welfare (Δ_{CEV})	+8.3%	+7.8%	+4.4%
Level Effect (Δ_{level})	−7.9%	−7.3%	−8.7%
Age Effect (Δ_{age})	+1.1%	+1.3%	+0.8%
Distribution Effect (Δ_{distr})	+15.1%	+13.8%	+12.3%

The table displays the welfare effects of moving from the baseline tax system to the optimal (progressive) tax system in each model. The left column specifies the welfare measure being displayed.

contrast, although the progressive tax raises welfare by redistributing in the LBD model, it lowers welfare by introducing an increasing tax over the lifetime (as opposed to the decreasing tax that is optimal due to the elasticity channel). Overall, these effects result in a more progressive optimal tax in the LOD model relative to the LBD model.

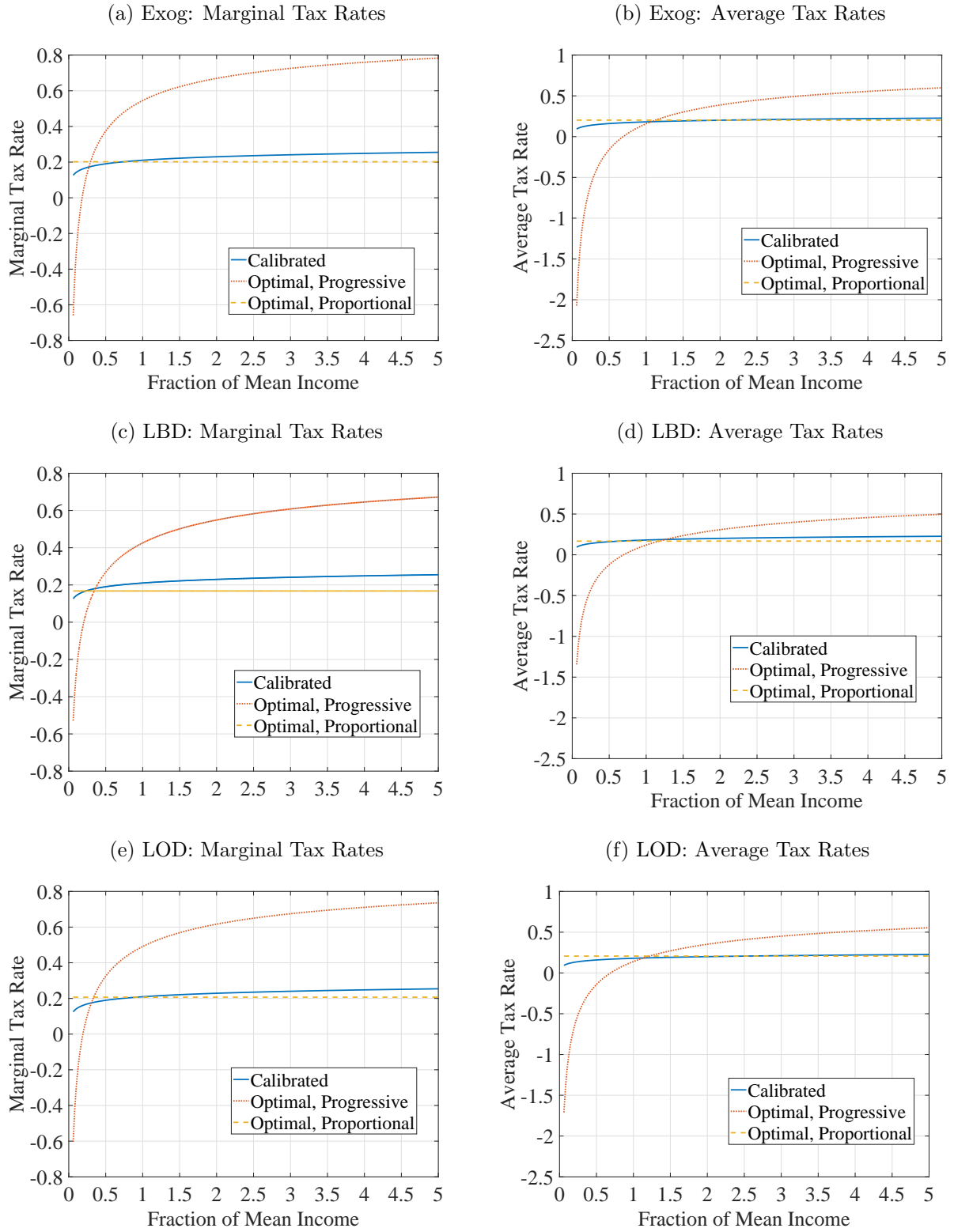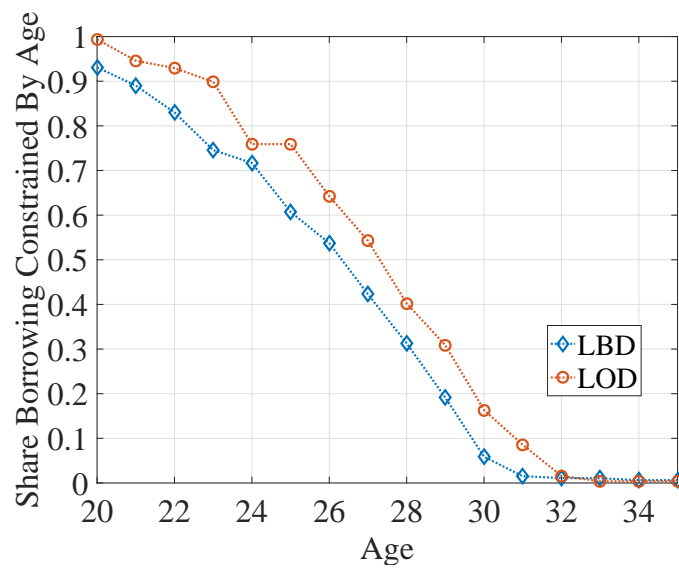Figure 3: Calibrated and Optimal Tax Functions for Labor Income

Figure 4: Borrowing Constraints Over the Life-Cycle

Note: The figure displays the share of workers who are borrowing constrained by age over the early working life. These profiles are obtained from the calibrated LBD and LOD models.

5.5 Impact of Optimal Taxes on the Aggregate Economy

Table 4 documents the steady state impact of optimal taxes, relative to baseline taxes, on economic aggregates and prices (and Figures C1-C3 document the corresponding change in life-cycle profiles in each model). For each model, the left column displays the results for the optimal tax policy with proportional labor taxes, and the right column in parentheses displays the results for the unrestricted (i.e. potentially progressive) labor taxes. Moving to optimal taxes with proportional labor taxes leads to an increase in effective labor (i.e. "efficiency units"), output, and consumption. A key reason for this is that optimal marginal labor tax rates fall for most workers (see Figure 3, panels (a), (c), and (e)). Implementing optimal taxes also increases aggregate capital for the exogenous and LOD models, which have relatively low optimal capital tax rates of 11.7% and 9.4%, respectively. In contrast, aggregate capital falls modestly for the LBD model when optimal capital taxes are implemented, which is unsurprising given its relatively high optimal capital tax of 28%.

Implementing the optimal progressive tax policy leads to large reductions in consumption, capital, labor, and output across all three models. Figure 3 demonstrates that, relative the baseline policy, the optimal policy increases the marginal tax on labor over most of the income distribution. The higher marginal labor tax leads agents to supply less labor, particularly those with high earnings potential. Progressive taxes also induce lower accumulation of human capital (see Guvenen et al. (2013)). With less effective labor supplied, the marginal product of capital falls, which lowers the interest rate and induces agents to save less. All told, reductions in capital and labor lead to less output and consumption.

In the exogenous human capital model, the fall in effective labor is relatively larger than the decrease in capital, because the increase in the marginal labor tax rates is even larger for most of the income distribution. The relatively larger fall in labor leads to an increase in the marginal product of labor (w) and a fall in the marginal product of capital (r). In contrast, in the LOD and LBD models, the fall in effective labor is relatively smaller, and thus there is a decrease in the marginal product of labor and increase in the marginal product of capital.

Table 6: Welfare Cost of Imposing the Wrong Optimal Taxes

| | WELFARE LOSSES RELATIVE TO OPTIMAL TAXES | |
| | LOD model | LBD model |
Welfare Measure	(with LBD taxes)	(with LOD taxes)
Share of Population Experiencing Welfare Loss	0.51	0.74
Ex Ante Welfare (Δ_{CEV})	-0.42%	-0.40%
Level Effect (Δ_{level})	$+2.50\%$	-2.01%
Age Effect (Δ_{age})	$+0.13\%$	-0.07%
Distribution Effect (Δ_{distr})	-2.99%	$+1.65\%$

The table displays the welfare effects of imposing the optimal tax policy solved for with the wrong form of human capital accumulation. The left column specifies the welfare measure being displayed. The middle column displays the welfare costs in the LOD model associated with moving from the optimal LOD policy to the optimal LBD policy. The right column displays the welfare costs in the LBD model associated with moving from the optimal LBD policy to the optimal LOD policy. The share of the population experiencing a welfare loss corresponds to the share of newborns whose lifetime utility in the steady state with the misspecified policy is below their lifetime utility in the steady state with the correct optimal policy.

5.6 Impact of Optimal Taxes on Welfare

For all three models, steady state welfare under the optimal tax regime is substantially higher than in the baseline economy. Across steady states, ex-ante welfare increases by 9.4% in the exogenous model, 8.8% in the LOD model, and 5.4% in the LBD model (measured in consumption equivalent variation).

Following the approach in Peterman and Sager (2018), Table 3 breaks out the ex-ante welfare loss into three components: a level effect, an age effect, and a distribution effect. The level effect captures the impact on welfare arising from changes in the average of consumption and non-leisure time across two economies. The age effect captures the welfare impact from changes in the average age profile of consumption and non-leisure time, holding average consumption and leisure constant. The distribution effect captures the remaining welfare change, primarily changes from the distribution over consumption and non-leisure time within a given cohort.

Qualitatively, each of these effects moves welfare in the same direction for all three human capital models. The level effect is large and negative, consistent with the large decline in aggregate consumption documented in Table 4. The distribution effect is large and positive, which reflects the utilitarian gains from redistributing resources from high income to low income individuals. Finally, the positive age effect reflects the fact that the

government can set the level of the capital tax and the progressivity of the labor income tax to mimic age-dependent taxes, which induces a flatter age-consumption profile.

In Table D1 we also show that the differences in optimal taxes between the LBD and LOD models lead to noticeable welfare consequences. Mistakenly imposing optimal taxes from one model on the other leads to ex-ante welfare losses just under half a percent in CEV terms, and reduces welfare for between 51% and 74% of newborns.

In the case of the LOD model, moving from the optimal LOD taxes to the optimal LBD taxes leads to a less progressive labor tax, a lower marginal tax on labor and a higher tax on capital. The less progressive tax leads to a more unequal distribution of welfare across agents, which results in a negative distribution effect on welfare. This negative effect is partially offset by two opposing effects. First, the higher tax on capital generates a flatter consumption and leisure profile over the lifetime, which increases welfare via the age effect. Second, because labor taxes are both lower and less progressive, workers both work and train more, which leads to a higher level of effective labor units in the economy and a higher level of output. The higher level of output leads to an increase in welfare via the level effect.

In the case of the LBD model, moving from the optimal LBD taxes to the optimal LOD taxes leads to a mirror effect: more progressive labor taxes with a higher marginal rate, and a lower tax on capital. The age effect is negative because the lower tax on capital increases the after-tax return to savings, leading to a steeper consumption profile. In the same direction, the higher marginal tax rate on labor and more progressive tax leads to a lower quantity of effective labor units and lower aggregate output, which generates a negative level effect. However, these negative effects are offset somewhat because the more progressive tax leads to less inequality.

6 Conclusion

This paper analyzes the impact of endogenous human capital on optimal capital taxes. We compare optimal taxes across three cases common to the macroeconomic literature: Learning-By-Doing, Learning-Or-Doing, and exogenous accumulation. We show analytically in a two-period model, in which agents transition from a young to old worker, that endogenous human capital creates novel motives for the government to condition labor taxes

on age, but that these motivations fundamentally depend on whether LBD or LOD is assumed. Intuitively, in a LBD model workers supply labor less elastically when young because labor also increases future human capital which is only beneficial to young agents; this leads the government to tax labor *more* when workers are young relative to when they are old. Alternatively, in a LOD model workers supply labor more elastically when young because labor is a substitute for training when agents are young; this leads the government to tax labor *less* when workers are young relative to when they are old. Because a tax on capital can mimic taxes that decrease with age, these results imply that the optimal capital tax will differ between LBD and LOD when the government is disallowed from using age-dependent taxes.

We then quantify differences in optimal capital taxes using a rich life-cycle framework that features heterogeneity in learning ability and initial human capital. With proportional taxes the optimal capital tax is 16pp higher with LBD than with exogenous accumulation, but 2pp lower with LOD than with exogenous accumulation. We show that heterogeneity in learning ability interacts with the human capital motives identified in the simple model and results in a larger gap in capital taxes between LBD and LOD relative to a case with homogeneous workers. Finally, we show that allowing the government to make labor taxes progressive reduces the gap in optimal capital taxes between LBD and LOD because the progressive tax mimics an income tax that increases with age, which is optimal in the LOD model but not the LBD model.

Our findings send two strong messages to researchers working to characterize optimal taxes within life-cycle settings. First, given that *both* of the endogenous human capital models common to the macroeconomic literature generate first order effects on optimal capital taxes, it is important for optimal tax analyses to include endogenous human capital. Second, because LBD and LOD each generate quantitatively (and in some case qualitatively) *different* effects on optimal taxes, it is essential to check robustness across both models.

Relatedly, the large differences in optimal taxes between different endogenous human capital models imply that it would be valuable to have a better understanding of how human capital is accumulated, and which model is more relevant in various settings. That is, we view our results, in part, as a call to action for further investigation into the processes of human capital accumulation themselves. Preliminary steps in this direction have been taken

by Heckman et al. (2002) and Blandin (2018); however, many open questions remain. For example, we suspect that different models of human capital accumulation may be more relevant for some occupations and industries than others. In addition, there may be more general heterogeneity across individuals in how human capital is accumulated which would affect both the efficiency and equity of the tax policy. We are not aware of any existing investigations along these lines.

A   Analytical Derivations

A.1   Primal Approach

In this section we describe the general set up for the primal approach which we use to determine the optimal tax policy.[20] We use a social welfare function that maximizes the expected utility of a newborn and discounts future generations with social discount factor θ,

$$[U(c_{2,0}, l_{2,0})/\theta] + \sum_{t=0}^{\infty} \theta^t [U(c_{1,t}, l_{1,t}) + \beta U(c_{2,t+1}, l_{2,t+1})]. \tag{53}$$

The government maximizes this objective function with respect to two constraints: the implementability constraint and the resource constraint.[21] The implementability constraint is the agent's intertemporal budget constraint, with prices and taxes replaced by his first order conditions (equations (6), (8), and (9))

$$c_{1,t} U_{c_{1,t}} + \beta c_{2,t+1} U_{c_{2,t+1}} + l_{1,t} U_{l_{1,t}} + \beta l_{2,t+1} U_{l_{2,t+1}} = 0. \tag{54}$$

Including this constraint ensures that any allocation the government chooses can be supported by a competitive equilibrium. The resource constraint is

$$c_{1,t} + c_{2,t} + k_{t+1} - k_t + G = r k_t + w(l_{1,t} + l_{2,t} h_2). \tag{55}$$

A.2   Exogenous

The Lagrangian for this specification is

$$\begin{aligned}
\mathcal{L} = & \frac{c_{1,t}^{1-\sigma_1}}{1-\sigma_1} - \chi \frac{l_{1,t}^{1+\frac{1}{\sigma_2}}}{1+\frac{1}{\sigma_2}} + \beta \frac{c_{2,t+1}^{1-\sigma_1}}{1-\sigma_1} - \chi \frac{l_{2,t+1}^{1+\frac{1}{\sigma_2}}}{1+\frac{1}{\sigma_2}} \\
& - \rho_t(c_{1,t} + c_{2,t} + k_{t+1} - k_t + G - r k_t - w(l_{1,t} + l_{2,t} h_2)) \\
& - \rho_{t+1}\theta(c_{1,t+1} + c_{2,t+1} + k_{t+2} - k_{t+1} + G - r k_{t+1} - w(l_{1,t+1} + l_{2,t+1} h_2)) \\
& + \lambda(c_{1,t}^{1-\sigma_1} + \beta c_{2,t+1}^{1-\sigma_1} - \chi l_{1,t}^{1+\frac{1}{\sigma_2}} - \beta \chi l_{2,t+1}^{1+\frac{1}{\sigma_2}})
\end{aligned} \tag{56}$$

where ρ_t is the Lagrange multiplier on the resource constraint in period t and λ is the Lagrange multiplier on the implementability constraint. The first order conditions with respect to labor, capital and consumption are

$$w\rho_t = \chi l_{1,t}^{\frac{1}{\sigma_2}}(1 + \lambda(1 + \frac{1}{\sigma_2})) \tag{57}$$

$$w\rho_{t+1}\theta \, h_2 = \beta \chi l_{2,t+1}^{\frac{1}{\sigma_2}}(1 + \lambda(1 + \frac{1}{\sigma_2})) \tag{58}$$

$$\rho_t = \theta(1+r)\rho_{t+1} \tag{59}$$

---

[20]See Lucas and Stokey (1983) or Erosa and Gervais (2002) for a full description of the primal approach.

[21]The government budget constraint is a third constraint. Due to Walras' Law, we only need to include two of three constraints in the Lagrangian and leave out the government budget constraint.

$$\rho_t = c_{1,t}^{-\sigma_1} + \lambda(1-\sigma_1)c_{1,t}^{-\sigma_1} \tag{60}$$

and

$$\theta\rho_{t+1} = \beta c_{2,t+1}^{-\sigma_1} + \beta\lambda(1-\sigma_1)c_{2,t+1}^{-\sigma_1}. \tag{61}$$

Combining the first order equations for the governments problem with respect to capital and consumption yields

$$\left(\frac{c_{2,t+1}}{c_{1,t}}\right)^{\sigma_1} = \frac{\beta\rho_t}{\rho_{t+1}\theta}. \tag{62}$$

Taking the ratio of the agent's first order conditions under the benchmark utility specification, (6) and (8), gives

$$\frac{1-\tau_2}{1-\tau_1} = \frac{1}{h_2}\left(\frac{c_{1,t}}{c_{2,t+1}}\right)^{-\sigma_1}\left(\frac{l_{2,t+1}}{l_{1,t}}\right)^{\frac{1}{\sigma_2}}. \tag{63}$$

Combining equation 62 and 63 yields

$$\frac{1-\tau_2}{1-\tau_1} = \frac{1}{h_2}\left(\frac{\beta\rho_t}{\rho_{t+1}\theta}\right)\left(\frac{l_{2,t+1}}{l_{1,t}}\right)^{\frac{1}{\sigma_2}}. \tag{64}$$

The ratio of first order equations for the government with respect to young and old hours is

$$\frac{\rho_t\beta}{h_2\rho_{t+1}\theta}\left(\frac{l_{2,t+1}}{l_{1,t}}\right)^{\frac{1}{\sigma_2}} = \frac{1+\lambda(1+\frac{1}{\sigma_2})}{1+\lambda(1+\frac{1}{\sigma_2})}. \tag{65}$$

Combining equation 65 and 64 generates the following expression for labor taxes

$$\frac{1-\tau_2}{1-\tau_1} = \frac{1+\lambda(1+\frac{1}{\sigma_2})}{1+\lambda(1+\frac{1}{\sigma_2})} = 1. \tag{66}$$

Moreover, using the primal approach, the optimal allocation of consumption is represented by the following expression,

$$\left(\frac{c_{1,t}}{c_{2,t+1}}\right)^{-\sigma_1} = \beta(1+r). \tag{67}$$

Assuming the benchmark utility function, the optimal allocation indicated by the primal approach is,

$$\left(\frac{c_{1,t}}{c_{2,t+1}}\right)^{-\sigma_1} = \beta(1+r(1-\tau_k)). \tag{68}$$

Thus, the optimal capital tax is zero. As Garriga (2001) points out, since there is no desire to condition labor income taxes on age in this exogenous model, the optimal tax on capital is zero regardless of whether the government can condition labor income taxes on age.[22]

---

[22]When the government cannot condition labor income taxes on age then the Lagrangian includes an additional constraint,

$$h_2\frac{U_{l_{1,t}}}{U_{c_{1,t}}} = \frac{U_{l_{2,t+1}}}{U_{c_{2,t+1}}}. \tag{69}$$

However, in the analytically tractable model with exogenous human capital accumulation, this constraint is not binding and thus the Lagrange multiplier on this constraint would be equal to zero.

A.3 LBD

The Lagrangian for this LBD specification is modified from the exogenous model. In particular, human capital benefit alters the implementability constraint. Suppressing the arguments of the skills function, the implementability constraint in the LBD model is

$$c_{1,t}U_{c_{1,t}} + \beta c_{2,t+1}U_{c_{2,t+1}} + l_{1,t}U_{l_{1,t}} - \frac{\beta l_{1,t}U_{l_{2,t+1}}l_{t,2}(\partial h_2/\partial l_{1,t})}{h_2} + \beta l_{2,t+1}U_{l_{2,t+1}} = 0. \tag{70}$$

Thus the Lagrangian for the LBD model is

$$\begin{aligned}
\mathscr{L} =& \frac{c_{1,t}^{1-\sigma_1}}{1-\sigma_1} - \chi\frac{l_{1,t}^{1+\frac{1}{\sigma_2}}}{1+\frac{1}{\sigma_2}} + \beta\frac{c_{2,t+1}^{1-\sigma_1}}{1-\sigma_1} - \chi\frac{l_{2,t+1}^{1+\frac{1}{\sigma_2}}}{1+\frac{1}{\sigma_2}} \\
& - \rho_t(c_{1,t} + c_{2,t} + k_{t+1} - k_t + G - rk_t - w(l_{1,t} + l_{2,t}h_{2,t})) \\
& - \rho_{t+1}\theta(c_{1,t+1} + c_{2,t+1} + k_{t+2} - k_{t+1} + G - rk_{t+1} - w(l_{1,t+1} + l_{2,t+1}h_{2,t+1})) \\
& + \lambda(c_{1,t}^{1-\sigma_1} + \beta c_{2,t+1}^{1-\sigma_1} - \chi l_{1,t}^{1+\frac{1}{\sigma_2}} + \frac{\chi\beta l_{2,t+1}^{1+\frac{1}{\sigma_2}}l_{1,t}(\partial h_2/\partial l_{1,t})}{h_2} - \beta\chi l_{2,t+1}^{1+\frac{1}{\sigma_2}}).
\end{aligned} \tag{71}$$

The first order conditions with respect to labor, capital and consumption are

$$\begin{aligned}
w\rho_t =& \chi l_{1,t}^{\frac{1}{\sigma_2}}(1 + \lambda(1 + \frac{1}{\sigma_2})) - \theta\rho_{t+1}l_{2,t+1}(\partial h_2/\partial l_{1,t}) \\
& + \lambda\chi l_{2,t+1}^{1+\frac{1}{\sigma_2}}\beta l_{1,t}\left[\frac{(\partial h_2/\partial l_{1,t})^2}{h_2^2} - \frac{(\partial^2 h_2/\partial l_{1,t}^2)}{h_2}\right]
\end{aligned} \tag{72}$$

$$w\rho_{t+1}\theta h_2 = \beta\chi l_{2,t+1}^{\frac{1}{\sigma_2}}\left[1 + \lambda(1 + \frac{1}{\sigma_2}) + (1 + \frac{1}{\sigma_2})\frac{l_{1,t}(\partial h_2/\partial l_{1,t})\lambda}{h_2}\right] \tag{73}$$

$$\rho_t = \theta(1+r)\rho_{t+1} \tag{74}$$

$$\rho_t = c_{1,t}^{-\sigma_1} + \lambda(1-\sigma_1)c_{1,t}^{-\sigma_1} \tag{75}$$

and

$$\theta\rho_{t+1} = \beta c_{2,t+1}^{-\sigma_1} + \beta\lambda(1-\sigma_1)c_{2,t+1}^{-\sigma_1}. \tag{76}$$

The first order conditions with respect to capital and consumption are the same in the exogenous (59, 60, and 61) and LBD models (74, 75, and 76). Therefore equation 67 still holds for this model and therefore the optimal tax on capital is still zero when the government can condition labor income taxes on age.

Combining the first order equations for the governments problem with respect to capital and consumption yields

$$\left(\frac{c_{2,t+1}}{c_{1,t}}\right)^{\sigma_1} = \frac{\beta\rho_t}{\rho_{t+1}\theta} \tag{77}$$

Taking the ratio of the agent's first order conditions and combining with equation 77 yields

$$\frac{1-\tau_1}{1-\tau_2} = \left(\frac{l_{1,t}}{l_{2,t+1}}\right)^{\frac{1}{\sigma_2}}\left(\frac{\rho_{t+1}\theta h_2}{\beta\rho_t}\right) - \frac{l_{2,t+1}(\partial h_2/\partial l_{1,t})}{1+r(1-\tau_k)}. \tag{78}$$

Combining equations 78, 72 and 73 the ratio of the optimal taxes on labor is,

$$\frac{1-\tau_1}{1-\tau_2} =$$

$$\frac{\left(1+\lambda(1+\frac{1}{\sigma_2})-\lambda(1+\frac{1}{\sigma_2})\frac{l_{1,t}(\partial h_2/\partial l_{1,t})}{h_2}\right)\left(1+\frac{l_{2,t+1}h_2}{1+r(1-\tau_k)}\right)}{1+\lambda(1+\frac{1}{\sigma_2})+l_{2,t+1}^{1+\frac{1}{\sigma_2}}l_{1,t}^{1+\frac{-1}{\sigma_2}}\frac{\lambda}{h_2}\left(\frac{(\partial h_2/\partial l_{1,t})}{h_2}-(\partial h_2/\partial l_{1,t})\right)} - \frac{l_{2,t+1}(\partial h_2/\partial l_{1,t})}{1+r(1-\tau_k)}. \tag{79}$$

A.4 LOD

Since agents have the additional choice variable x_1 in the LOD model, workers have an additional first order condition with respect to this variable. This new first order condition requires an additional constraint in the government's Lagrange that ensures that the allocation the government chooses properly equates an individual's disutility of training when young and working when old. This constraint simplifies to $U_{n_{1,t}}h_2 = \beta U_{h_{2,t+1}}(\partial h_{2,t+1}/\partial x_{1,t})$. We use η as the Lagrange multiplier on this new constraint.

The Lagrangian for the LOD model is

$$\mathcal{L} = \frac{c_{1,t}^{1-\sigma_1}}{1-\sigma_1} - \chi\frac{(l_{1,t}+x_{1,t})^{1+\frac{1}{\sigma_2}}}{1+\frac{1}{\sigma_2}} + \beta\frac{c_{2,t+1}^{1-\sigma_1}}{1-\sigma_1} - \chi\frac{l_{2,t+1}^{1+\frac{1}{\sigma_2}}}{1+\frac{1}{\sigma_2}} \tag{80}$$

$$- \rho_t(c_{1,t}+c_{2,t}+k_{t+1}-k_t+G-rk_t-w(l_{1,t}+l_{2,t}h_{2,t+1}))$$

$$- \rho_{t+1}\theta(c_{1,t+1}+c_{2,t+1}+k_{t+2}-kK_{t+1}+G_{t+1}-rk_{t+1}-w(l_{1,t+1}+l_{2,t+1}h_{2,t+1}))$$

$$+ \lambda(c_{1,t}^{1-\sigma_1}+\beta c_{2,t+1}^{1-\sigma_1}-\chi l_{1,t}^{1+\frac{1}{\sigma_2}}-\beta\chi l_{2,t+1}^{1+\frac{1}{\sigma_2}})$$

$$+ \eta(\chi l_{2,t+1}^{1+\frac{1}{\sigma_2}}(\partial h_{2,t+1}/\partial x_{1,t}) - \chi(l_{1,t}+x_{1,t})^{\frac{1}{\sigma_2}}h_{2,t+1}).$$

The first order conditions with respect to labor, capital, consumption and training are,

$$w\rho_t = \chi(l_{1,t}+x_{1,t})^{\frac{1}{\sigma_2}}\left[1+\lambda\left(1+\frac{l_{1,t}}{\sigma_2(l_{1,t}+x_{1,t})}\right)+\frac{\eta h_{2,t+1}}{\sigma_2(l_{1,t}+x_{1,t})}\right] \tag{81}$$

$$w\rho_{t+1}\theta h_{2,t+1} = \beta\chi l_{2,t+1}^{\frac{1}{\sigma_2}}\left[1+\lambda_2\left(1+\frac{1}{\sigma_2}\right)-\eta\left(1+\frac{1}{\sigma_2}\right)(\partial h_{2,t+1}/\partial x_{1,t})\right] \tag{82}$$

$$\rho_t = \theta(1+r)\rho_{t+1} \tag{83}$$

$$\rho_t = c_{1,t}^{-\sigma_1} + \lambda(1-\sigma_1)c_{1,t}^{-\sigma_1} \tag{84}$$

$$\theta\rho_{t+1} = \beta c_{2,t+1}^{-\sigma_1} + \beta\lambda(1-\sigma_1)c_{2,t+1}^{-\sigma_1} \tag{85}$$

and

$$\theta\rho_{t+1}l_{2,t+1}(\partial h_{2,t+1}/\partial x_{1,t}) = \tag{86}$$

$$\frac{\chi(l_{1,t} + x_{1,t})^{\frac{1}{\sigma_2}}\left(\lambda l_{1,t} + \eta h_{2,t+1} + \sigma_2(l_{1,t} + x_{1,t})(1 + \eta(\partial h_{2,t+1}/\partial x_{1,t}))\right)}{\sigma_2(l_{1,t} + x_{1,t})} \tag{87}$$

$$-\frac{\beta\chi\eta\sigma_2 l_{2,t+1}^{1+\frac{1}{\sigma_2}}(l_{1,t} + x_{1,t})(\partial^2 h_{2,t+1}/\partial x_{1,t}^2)}{\sigma_2(l_{1,t} + x_{1,t})} \tag{88}$$

The first order conditions with respect to capital and consumption are the same in the exogenous (59, 60, and 61) and LOD models (83, 84, and 85). Therefore equation 67 still holds for this model and therefore the optimal tax on capital is still zero when the government can condition labor income taxes on age.

Combining the first order equations for the governments problem with respect to capital and consumption yields

$$\left(\frac{c_{2,t+1}}{c_{1,t}}\right)^{\sigma_1} = \frac{\beta\rho_t}{\rho_{t+1}\theta} \tag{89}$$

Taking the ratio of the agent's first order conditions, (6) and (7), and combining with equation 89 yields

$$\frac{1 - \tau_2}{1 - \tau_1} = \left(\frac{l_{2,t+1}}{l_{1,t} + x_{1,t}}\right)^{\frac{1}{\sigma_2}}\left(\frac{\beta\rho_t}{\rho_{t+1}\theta h_{2,t+1}}\right). \tag{90}$$

Taking the ratio of equations 81 and 82 yields,

$$\left(\frac{l_{2,t+1}}{l_{1,t} + x_{1,t}}\right)^{\frac{1}{\sigma_2}}\left(\frac{\beta\rho_t}{\rho_{t+1}\theta h_{2,t+1}}\right) = \frac{1 + \lambda\left(1 + \frac{l_{1,t}}{\sigma_2(l_{1,t}+x_{1,t})}\right) + \frac{\eta h_{2,t+1}}{\sigma_2(l_{1,t}+x_{1,t})}}{1 + \lambda\left(1 + \frac{1}{\sigma_2}\right) - \eta(\partial h_{2,t+1}/\partial x_{1,t})\left(1 + \frac{1}{\sigma_2}\right)}. \tag{91}$$

Combining equations 90 and 91 generates the following expression for the ratio of the optimal labor taxes,

$$\frac{1 - \tau_2}{1 - \tau_1} = \frac{1 + \lambda\left(1 + \frac{l_{1,t}}{\sigma_2(l_{1,t}+x_{1,t})}\right) + \frac{\eta h_{2,t+1}}{\sigma_2(l_{1,t}+x_{1,t})}}{1 + \lambda\left(1 + \frac{1}{\sigma_2}\right) - \eta(\partial h_{2,t+1}/\partial x_{1,t})\left(1 + \frac{1}{\sigma_2}\right)}. \tag{92}$$

B Calibration Details

B.1 Empirical Life-Cycle Profiles

We follow Huggett et al. (2011): we first group individuals in the PSID into 5-year centered age bins. Then we compute the statistic of interest for each age bin-year combination (j, t), denoted $y_{j,t}$. Finally, we estimate the following fixed effects model for the set of statistics $\{\{y_{j,t}\}_{j=20}^{65}\}_{t=1968}^{2013}$:

$$y_{j,t} = \alpha_j^y + \beta_t^y + \gamma_c^y + \varepsilon_{j,t}^y \tag{B1}$$

where α_j^y, β_t^y, γ_c^y denote age, year, and birth cohort fixed effects for statistic y, respectively, and $\varepsilon_{j,t}^y$ is an error term. To avoid the multicollinearity problem associated with the fact that $c \equiv t - j$, we assume that $\gamma_c = 0 \ \forall c$. The corresponding age effect estimates are plotted in Figure 1.

B.2 Joint calibration procedure

We calibrate 10 parameters jointly: the means of (h_{20}, a), (μ_h, μ_a); the variances of (h_{20}, a), (σ_h, σ_a); the correlation between h_1 and a, $\rho_{h,a}$; the discount rate β, the utility of leisure χ, the depreciation rate of human capital δ_h, and the two elasticity parameters in the human capital production function ϕ_1 and ϕ_2. The procedure is as follows:
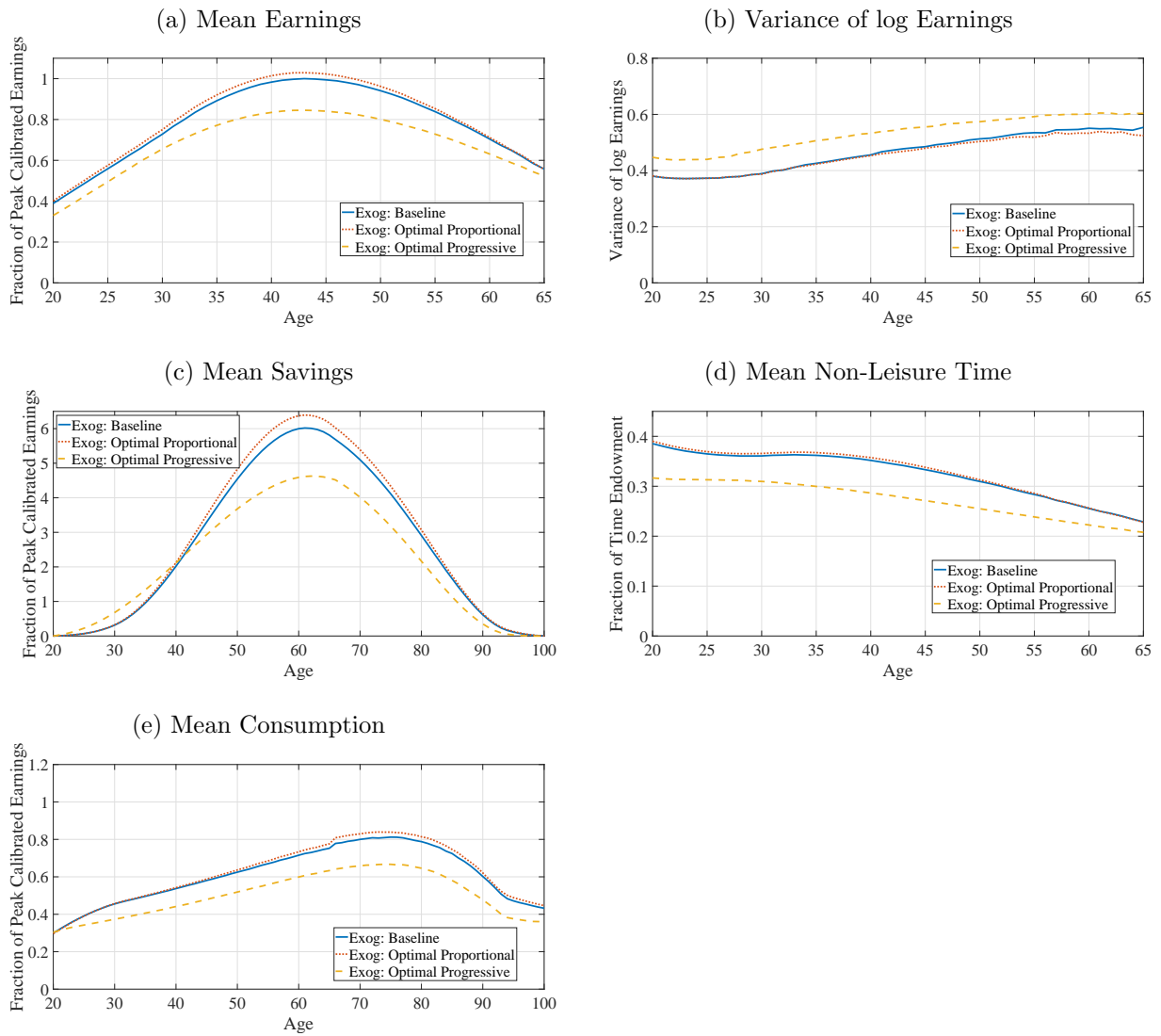
0. Choose initial guesses for the parameters χ, δ_h, ϕ_1, ϕ_2, β, and η_0.

1. Run the model given χ, δ_h, ϕ_1, ϕ_2, β, and η_0. Because equilibrium prices are already determined by the technology parameters, we can solve each individual's problem independently of the initial distribution parameters $(\mu_h, \mu_a, \sigma_h, \sigma_a, \rho_{h,a})$.

2. Given the individual solutions, choose values for $(\mu_h, \mu_a, \sigma_h, \sigma_a, \rho_{h,a})$ to match the following empirical earnings moments: mean earnings at age 20, the peak of mean earnings, the variance of earnings at age 20, the variance of earnings at age 30, and the variance of earnings at age 50.

3. Iterate over values of the human capital elasticities ϕ_1, ϕ_2, and the human capital depreciation rate δ_h, to match mean earnings at age 30 and mean earnings at age 65, repeating Steps 1-2 for each iteration.

4. Match the mean of hours worked in the model and data by iterating over values for the utility of leisure χ, repeating Steps 1-3 for each iteration.

5. Clear the physical capital market by iterating over values of the discount rate β, repeating Steps 1-4 for each iteration.

After the calibration is complete, we re-run the calibration procedure from several different initial guesses to verify uniqueness. The procedure converges to the same final values regardless of the starting values.

Finally, to calibrate the version of our model where workers are homogeneous within a cohort, we simply collapse this procedure by imposing $\sigma_h = \sigma_a = \rho_{h,a} = 0$ and dropping the targeted moments associated with the within-cohort variance of earnings in Step 2.
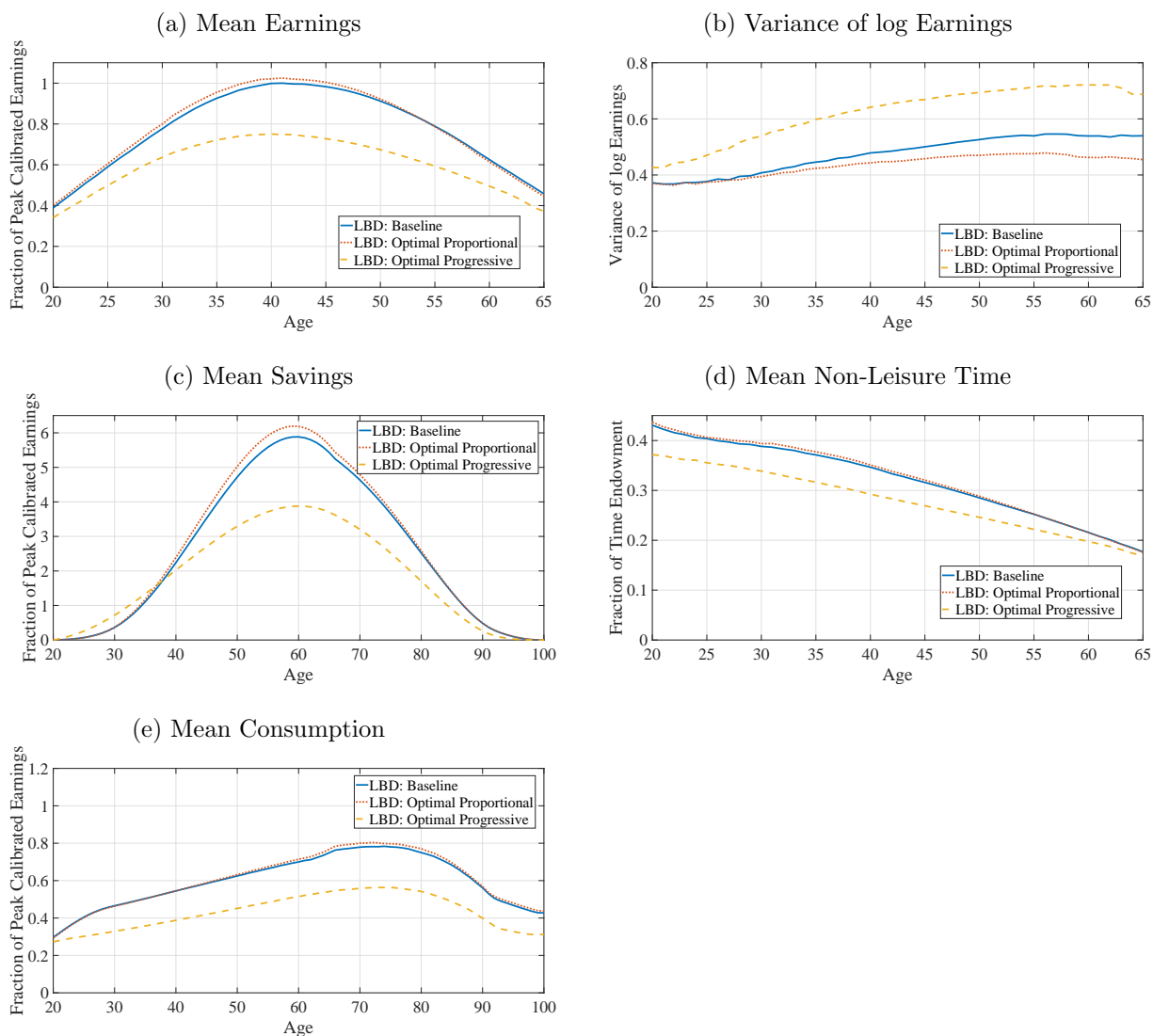
C   Impact of Optimal Taxes on Life-cycle Outcomes

Figure C1: Life-cycle Outcomes and Taxes: Exogenous Model

(a) Mean Earnings

(b) Variance of log Earnings

(c) Mean Savings

(d) Mean Non-Leisure Time

(e) Mean Consumption

The figure displays profiles for several variables over the working life. Mean earnings, savings, and consumption are expressed as a fraction of the life-cycle maximum for mean earnings. The variances of log earnings are expressed in log points. Mean labor is expressed as a fraction of an annual time endowment, which is set to 5,840.

Figure C2: Life-cycle Outcomes and Taxes: LBD Model

(a) Mean Earnings

(b) Variance of log Earnings

(c) Mean Savings

(d) Mean Non-Leisure Time
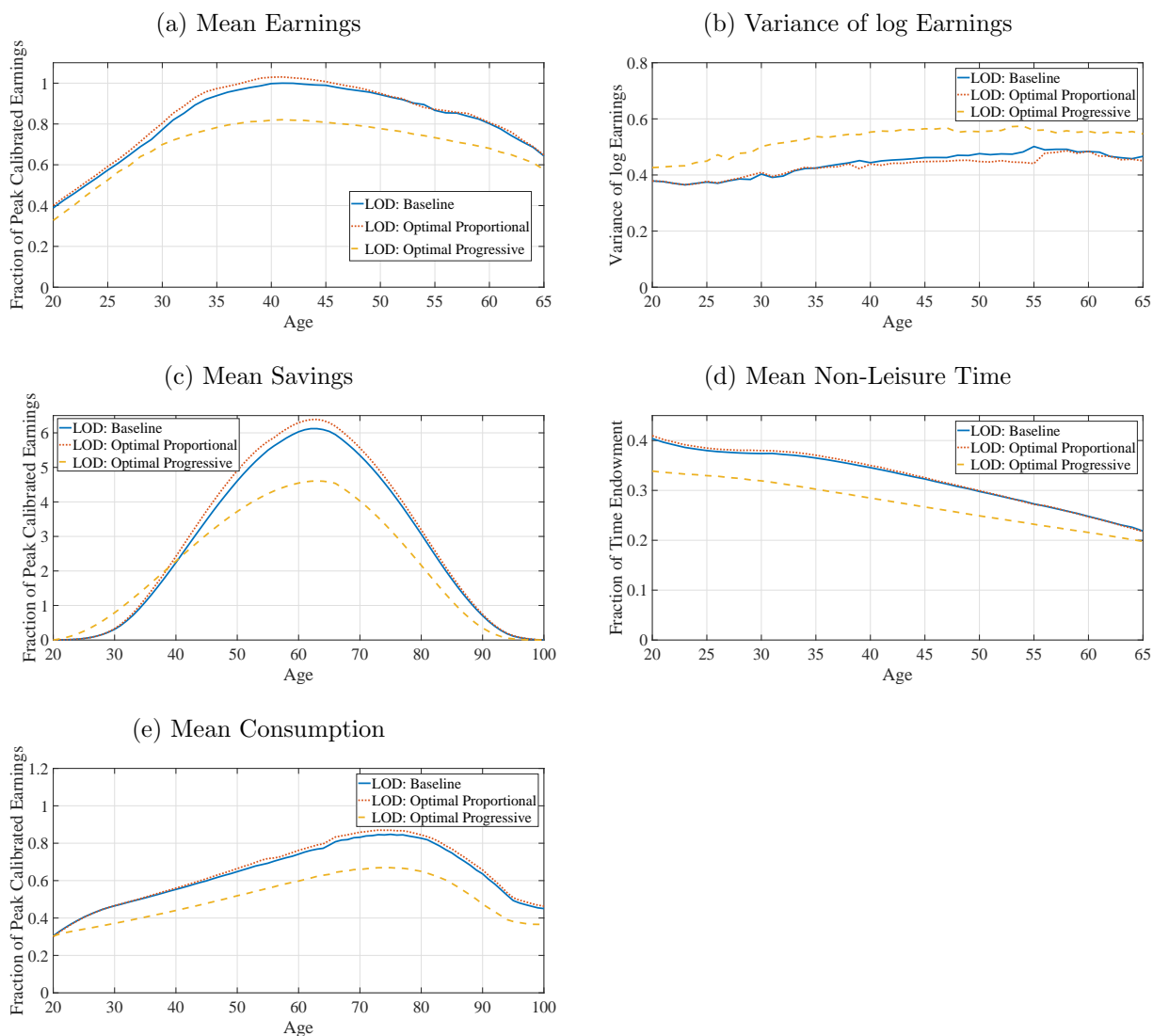
(e) Mean Consumption

The figure displays profiles for several variables over the working life. Mean earnings, savings, and consumption are expressed as a fraction of the life-cycle maximum for mean earnings. The variances of log earnings are expressed in log points. Mean labor is expressed as a fraction of an annual time endowment, which is set to 5,840.

Figure C3: Life-cycle Outcomes and Taxes: LOD Model

(a) Mean Earnings

(b) Variance of log Earnings

(c) Mean Savings

(d) Mean Non-Leisure Time

(e) Mean Consumption

The figure displays profiles for several variables over the working life. Mean earnings, savings, and consumption are expressed as a fraction of the life-cycle maximum for mean earnings. The variances of log earnings are expressed in log points. Mean labor is expressed as a fraction of an annual time endowment, which is set to 5,840.

D   Impact of Optimal Proportional Taxes on Welfare

Table D1: Welfare Gains Of Moving To Optimal Proportional Taxes

Welfare Measure	WELFARE GAINS RELATIVE TO BASELINE TAXES		
	Exog model	LOD model	LBD model
Ex Ante Welfare (Δ_{CEV})	+0.8%	+1.0%	+1.3%
Level Effect (Δ_{level})	+1.4%	+1.1%	1.0%
Age Effect (Δ_{age})	+0.1%	+0.6%	+0.9%
Distribution Effect (Δ_{distr})	−0.7%	−0.8%	−0.6%

The table displays the welfare effects of moving from the baseline tax system to the optimal proportional tax system in each model. The left column specifies the welfare measure being displayed.

References

Aaronson, Daniel and Eric French, "The effect of part-time work on wages: Evidence from the social security rules," *Journal of Labor Economics*, 2004, *22* (2), 329–252.

Altonji, Joseph G., "Intertemporal Substitution in Labor Supply: Evidence from Micro Data," *The Journal of Political Economy*, 1986, *94* (3), S176–S215.

Armenter, Roc and Stefania Albanesi, "Understanding Capital Taxation in Ramsey Models," Working Paper September 2009.

Atkinson, A.B. and J.E. Stiglitz, "The design of tax structure: Direct versus indirect taxation," *Journal of Public Economics*, 1976, *6*, 55–75.

Badel, Alejandro, Mark Huggett, and Wenlan Luo, "Taxing Top Earners: A Human Capital Perspective," *Working Paper*, 2018.

Bell, Felicitie and Michael Miller, "Life Tables for the United States Social Security Area 1900 - 2100," Actuarial Study 120, Office of the Chief Actuary, Social Security Administration 2002.

Blandin, Adam, "Learning by Doing and Ben-Porath: Life-cycle Predictions and Policy Implications," *Journal of Economic Dynamics and Control*, 2018, *90*, 220–235.

Cespedes, Nikita and Michael Kuklik, "Optimal Taxation and Life Cycle Labor Supply Profile," Working Paper December 2012.

Chamley, Christophe, "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica*, May 1986, *54* (3), 607–622.

Chang, Yongsung, Joao Gomes, and Frank Schorfheide, "Learning-By-Doing as a Propagation Mechanism," *American Economic Review*, December 2002, *92* (5), 1498–1520.

Chen, Been-Lon, Hung-Ju Chen, and Ping Wang, "Taxing Capital is Not a Bad Idea Indeed: The Role of Human Capital and Labor-Market Frictions," Working Paper May 2010.

Chetty, Raj, "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," Working Paper 15616, NBER December 2009.

Conesa, Juan Carlos and Dirk Krueger, "On the Optimal Progressivity of the Income Tax Code," *The Journal of Monetary Economics*, October 2006, *53* (7), 1425–1450.

_ , **Sagiri Kitao, and Dirk Krueger**, "Taxing Capital? Not a Bad Idea After All!," *American Economic Review*, 2009, *99* (1), 25–38.

Contreras, Juan and Sven Sinclair, "Labor Supply Response in Macroeconomic Models: Assessing the Empirical Validity of the Intertemporal Labor Supply Response from a Stochastic Overlapping Generations Model with Incomplete Markets," MPRA Paper 10533, University Library of Munich, Germany September 2008.

Correia, Isabel H., "Should Capital Income Be Taxed in the Steady State?," *Journal of Public Economics*, 1996, *60* (1), 147 – 151.

da Costa, Carlos E and Marcelo R Santos, "AGE-DEPENDENT TAXES WITH EN-DOGENOUS HUMAN CAPITAL FORMATION," *International Economic Review*, 2018, *59* (2), 785–823.

Deaton, Angus, "Optimally uniform commodity taxes," *Economics Letters*, 1979, *2* (4), 357 – 361.

Domeij, David and Martin Floden, "The Labor-Supply Elasticity and Borrowing Constraints: Why Estimates are Biased," *Review of Economic Dynamics*, 2006, *9* (2), 242 – 262.

Erosa, Andrés and Martin Gervais, "Optimal Taxation in Life Cycle Economies," *Journal of Economic Theory*, 2002, *105*, 338–369.

Fehr, Hans and Fabian Kindermann, "Taxing capital along the transitionNot a bad idea after all?," *Journal of Economic Dynamics and Control*, 2015, *51*, 64–77.

Garriga, Carlos, "Optimal Fiscal Policy in Overlapping Generations Models," Working Papers in Economics 66, Universitat de Barcelona. Espai de Recerca en Economia January 2001.

Gervais, Martin, "On the optimality of age-dependent taxes and the progressive US tax system," *Journal of Economic Dynamics and Control*, 2012, *36* (4), 682–691.

Golosov, Mikhail, Narayana Kocherlakota, and Aleh Tsyvinski, "Optimal indirect and capital taxation," *The Review of Economic Studies*, 2003, *70* (3), 569–587.

Guner, Nezih, Remzi Kaygusuz, and Gustavo Ventura, "Income taxation of US households: Facts and parametric estimates," *Review of Economic Dynamics*, 2014, *17* (4), 559–581.

Guvenen, Fatih, Burhanettin Kuruscu, and Serdar Ozkan, "Taxation of Human Capital and Wage Inequality: A Cross-Country Analysis," NBER Working Papers 15526, National Bureau of Economic Research, Inc 2009.

_ , _ , **and** _ , "Taxation of Human Capital and Wage Inequality: A Cross-Country Analysis," *Review of Economic Studies*, 2013, *81* (2), 818–850.

Hansen, Gary D. and Selahattin İmrohoroğlu, "Business Cycle Fluctuations and the Life Cycle: How Important Is On-the-Job Skill Accumulation?," *Journal of Economic Theory*, 2009, *144* (6), 2293 – 2309. Dynamic General Equilibrium.

Heckman, James, Lance Lochner, and Ricardo Cossa, "Learning-by-doing vs. on-the-job training: Using variation induced by the EITC to distinguish between models of skill formation," Technical Report, National Bureau of Economic Research 2002.

Huggett, Mark, Gustavo Ventura, and Amir Yaron, "Sources of Lifetime Inequality," *American Economic Review*, 2011, *101* (7), 2923–54.

Imai, Susumu and Michael Keane, "Intertemporal Labor Supply and Human Capital Accumulation," *International Economic Review*, May 2004, *45* (2), 601–641.

Jacobs, Bas and A. Lans Bovenberg, "Human Capital and Optimal Positive Taxation of Capital Income," *International Tax and Public Finance*, 2009, pp. 1–28. 10.1007/s10797-009-9120-5.

Jones, Larry E. and Rodolfo E. Manuelli, "On the Taxation of Human Capital," Unpublished Manuscript, Northwestern University and University of Wisconsin November 1999.

_ , _ , **and Peter E. Rossi**, "On the Optimal Taxation of Capital Income," *Jounal of Economic Theory*, 1997, *73*, 93 – 117.

Judd, Kenneth, "Redistributive Taxation in a Simple Perfect Foresight Model," *Journal of Public Economics*, 1985, *28*, 59–83.

Judd, Kenneth L., "Optimal Taxation and Spending In General Competitive Growth Models," *Journal of Public Economics*, 1999, *71* (1), 1 – 26.

Kapicka, Marek, "Optimal Income Taxation with Human Capital Accumulation and Limited Record Keeping," *Review of Economic Dynamics*, 2006, *9* (4), 612 – 639.

_ , "The Dynamics of Optimal Taxation when Human Capital is Endogenous," Working Paper August 2009.

Karabarbounis, Marios, "A road map for efficiently taxing heterogeneous agents," *American Economic Journal: Macroeconomics*, 2016, *8* (2), 182–214.

Kuruscu, Burhanettin, "Training and Lifetime Income," *The American Economic Review*, 2006, *96* (3), 832–846.

Lucas, Robert E. and Nancy L. Stokey, "Optimal Fiscal and Monetary Policy in an Economy without Capital," *Journal of Monetary Economics*, 1983, *12* (1), 55 – 93.

MaCurdy, Thomas E., "An Empirical Model of Labor Supply in a Life-Cycle Setting," *The Journal of Political Economy*, 1981, *89* (6), 1059–1085.

Peterman, William, "An Extensive Look at Taxes: How does endogenous retirement affect optimal taxation?," 2012.

_ **and Erick Sager**, "Optimal Public Debt with Life Cycle Motives," 2018.

Peterman, William B., "Determining the motives for a positive optimal tax on capital," *Journal of Economic Dynamics and Control*, 2013, *37* (1), 265–295.

_ , "The effect of endogenous human capital accumulation on optimal taxation," *Review of Economic Dynamics*, 2016, *21*, 46 – 71.

_ , "Reconciling Micro and Macro Estimates of the Frisch Labor Supply Elasticity," *Economic Inquiry*, 2016, *54* (1), 100–120.

Pistaferri, Luigi, "Anticipated and Unanticipated Wage Changes, Wage Risk, and Intertemporal Labor Supply," *Journal of Labor Economics*, 2003, *21* (3), 729–754.

Reis, Catarina, "Optimal Taxation with Unobservable Investment in Human Capital," Mimeo, MIT 2007.

Röhrs, Sigrid and Christoph Winter, "Reducing government debt in the presence of inequality," *Journal of Economic Dynamics and Control*, 2017, *82*, 1–20.

Smyth, Seamus, "A Balancing Act: Optimal Nonlinear Taxation in Overlapping Generations Models," Mimeo, Harvard May 2006.

Stantcheva, Stefanie, "Learning and (or) Doing: Human Capital Investments and Optimal Taxation," Working Paper 21381, National Bureau of Economic Research July 2015.

_ , "Optimal Taxation and Human Capital Policies over the Life Cycle," *Journal of Political Economy*, 2017, *125* (6), 1931–1990.

Tauchen, George, "Finite state markov-chain approximations to univariate and vector autoregressions," *Economics letters*, 1986, *20* (2), 177–181.

Wallenius, Johanna, "Human capital accumulation and the intertemporal elasticity of substitution of labor: How large is the bias?," *Review of Economic Dynamics*, 2011, *14* (4), 577–591.