Optimal income taxation when asset taxation is limited

Árpád Ábrahám, Sebastian Koehne, Nicola Pavoni

PII: S0047-2727(16)00036-0
DOI: doi: 10.1016/j.jpubeco.2016.02.003
Reference: PUBEC 3650

To appear in: *Journal of Public Economics*

Received date: 14 July 2015
Revised date: 23 February 2016
Accepted date: 24 February 2016

Please cite this article as: Ábrahám, Árpád, Koehne, Sebastian, Pavoni, Nicola, Optimal income taxation when asset taxation is limited, *Journal of Public Economics* (2016), doi: 10.1016/j.jpubeco.2016.02.003

Optimal income taxation when asset taxation is limited

Árpád Ábrahám[*], *Sebastian Koehne*[†], *Nicola Pavoni*[‡]

February 23, 2016

Abstract

Several frictions restrict the government's ability to tax assets. First, it is very costly to monitor trades on international asset markets. Second, agents can resort to nonobservable low-return assets such as cash, gold or foreign currencies if taxes on observable assets become too high. This paper shows that limitations in asset taxation have important consequences for the taxation of labor income. We study a simple dynamic moral hazard model of social insurance with observable and nonobservable saving decisions. We find that optimal labor income taxes become *less progressive* when the ability to tax savings is limited.

Keywords: Optimal Income Taxation, Capital Taxation, Progressivity.

JEL: D82, D86, E21, H21.

[*]European University Institute, Florence. E-mail address: Arpad.Abraham@EUI.eu

[†]Corresponding author. Institute for International Economic Studies (IIES), Stockholm University, SE-10691 Stockholm, Sweden, Phone: +46 8 16 35 64, Fax: +46 8 16 14 43, E-mail address: sebastian.koehne@iies.su.se

[‡]Bocconi University, IGIER, IFS, and CEPR. E-mail address: nicola.pavoni@unibocconi.it

1 Introduction

The existence of international asset markets implies that taxation authorities do not have perfect control over agents' wealth and consumption. This creates an important obstacle for tax policy:

> "In a world of high and growing capital mobility there is a limit to the amount of tax that can be levied without inducing investors to hide their wealth in foreign tax havens." (*Mirrlees Review* 2010, p.916)

According to a study by the Tax Justice Network, in 2010 more than $21 trillion of global private financial wealth were invested in offshore accounts and not reported to the tax authorities. These concerns motivated the recent legal steps by the most developed economies to crack down on off-shore bank accounts and the traditional norms of bank secrecy. However, even when agents choose not to hide their wealth abroad, they have access to a number of nonobservable storage technologies at home. For example, agents can accumulate cash, gold, or durable goods. These assets bring lower returns, but nonetheless impose limits for the collection of taxes on assets that are more easily observed.

Motivated by these considerations, this paper explores optimal tax systems in a framework where assets are nonobservable. Our findings show that labor income taxes become less progressive when the government's ability to observe and tax assets is limited. Throughout the paper, we measure the progressivity of labor income taxes based on the concavity of the associated consumption allocation. Tax systems and allocations are closely related in our framework because constrained efficient allocations can be implemented with nonlinear taxes on labor income and linear taxes on assets. Because of this relationship, our measure of progressivity captures the slope of the marginal tax rate on labor income.[1] We obtain our results by contrasting two stylized environments. In the first, consumption and assets are observable (and contractable) for the government. In the second environment, these choices are private information. We compare the constrained efficient allocations of the two scenarios. Assuming that absolute risk aversion is convex, we find that in the scenario with nonobservable assets, optimal consumption moves in a less concave way with labor income. In this sense, the optimal allocation becomes less progressive when assets are nonobservable.

[1] For example, a constant marginal tax rate on labor income results in a linear relationship between income and consumption. If the marginal tax rate increases with labor income, the relationship becomes concave. According to our definition, tax systems with rising marginal tax rates on labor income are, thus, more progressive than systems with a constant marginal tax rate.

We study a tractable dynamic model of social insurance. A continuum of ex ante identical agents influence their labor incomes by exerting effort. Labor income is subject to uncertainty and effort is private information. This creates a moral hazard problem. The social planner faces a trade-off between insuring agents against idiosyncratic income uncertainty and the associated disincentive effects. In addition, agents have access to a risk-free asset, which gives them a means for self-insurance.

In this model, the planner wants to manipulate agents' asset decisions, because asset accumulation provides an insurance against the idiosyncratic income uncertainty and thereby reduces the incentives to exert effort.[2] When fully capable of doing so, the planner uses asset taxation to deter the agent from accumulating assets and labor income taxation to balance consumption insurance and effort incentives optimally. When asset taxation is limited, the planner is forced to use labor income taxation also to reduce the agent's incentive to save. Efficiency requires that, for each income state, the costs of increasing the agent's utility by a marginal unit equal the benefits of doing so. A marginal increase in utility in a state with consumption c reduces the agent's marginal return to savings in that state by $Ra(c)$, where $a(c)$ is the absolute risk aversion of the agent at consumption c in that state and R is the asset return.[3] Hence, under limited asset taxation, there is an additional social return to allocating utility to a given state and this return is proportional to the level of absolute risk aversion of the agent. Therefore, unless absolute risk aversion is constant or linear, limits to asset taxation have direct implications for the curvature of optimal consumption.[4] In particular, whenever absolute risk aversion is convex, the planner finds it optimal to generate an additional convexity (or reduced progressivity) of consumption when asset taxation is limited.

The paper also illustrates the quantitative impact of asset taxation on optimal labor tax progressivity. The calibration of the key parameters of our framework is not straightforward because the technology that determines how effort affects future income is not directly observable. We use consumption and income data from the PSID (Panel Study of Income Dynamics) as adapted by Blundell, Pistaferri and Preston (2008). In the calibration exercise, we assume that the data is generated by a tax system where labor income taxes are set

[2]See Diamond and Mirrlees (1978), Rogerson (1985), and Golosov, Kocherlakota and Tsyvinski (2003).

[3]A marginal increase of utility in a state with consumption c reduces the agent's marginal utility in that state by $-u''(c)/u'(c)$. To increase $u(c)$ by ε, indeed, c has to be increased by $\varepsilon/u'(c)$. The first-order effect on the agent's marginal utility is therefore given by $-\varepsilon u''(c)/u'(c)$.

[4]Naturally, the very same argument implies that both the level and slope of the optimal consumption allocation might also be affected by the properties of absolute risk aversion.

optimally given the existing asset income tax rate.[5] This calibration strategy has the advantage that all fundamental parameters can be identified using only one cross-sectional joint distribution of consumption and income.

Based on the calibrated parameters, we compare the optimal allocations of optimal asset taxation (observable assets) and limited asset taxation (hidden assets). Under optimal asset taxation, the progressivity of the optimal allocation increases sizeably. Our measure of the curvature of consumption increases by 4% for log utility, and between 24% and 49% for relative risk aversion levels of 2 and 3. The tax rates on asset income in the scenario with observable assets are high and exceed 90% for all specifications. Although our numerical simulations are only illustrative, they suggest that a limited capacity to tax assets can have a substantial impact on the optimal progressivity of income taxes.

The paper proceeds as follows. Section 2 surveys the related literature. Section 3 describes the setup of the model. Section 4 presents the theoretical results of the paper. We show that hidden asset accumulation leads to optimal consumption schemes that are less progressive. Section 5 illustrates the quantitative implications of our results. Section 6 discusses model limitations and viable extensions. The appendix collects all proofs that are omitted from the main text and provides further details on the calibration strategy for Section 5.

2 Related literature

To the best of our knowledge, this is the first paper that explores optimal income taxation in a framework where assets are nonobservable. Recent work on dynamic Mirrleesian economies analyzes optimal income taxes when assets are observable/taxable without frictions; see Farhi and Werning (2013) and Golosov, Troshkin and Tsyvinski (2013). While the Mirrlees (1971) framework focuses on redistribution in a population with heterogeneous skills that are exogenously distributed, our approach highlights the social insurance aspect of income taxation. In spirit, our model is therefore closer to the works by Varian (1980) and Eaton and Rosen (1980). Kapicka and Neira (2015) also study optimal taxation in a moral hazard model. Human capital is nonobservable (similar to the effort decision in our framework) but consumption and savings are observable in their model.

With respect to the nonobservability of assets, our model is related to the contribution by Golosov and Tsyvinski (2007) who analyze capital taxation in a dynamic Mirrleesian

[5] We set the asset income tax rate to 40%. This rate is in line with U.S. effective tax rates on capital income calculated by Mendoza, Razin and Tesar (1994) and Domeij and Heathcote (2004).

economy with private insurance markets and hidden asset trades. Our model provides a systematic comparison between second best and third best allocations. In particular, we analyze how the progressivity of optimal allocations depends on the observability of asset holdings.

An entirely different link between labor income and capital income taxation is explored by Conesa, Kitao and Krueger (2009). Using a life-cycle model with age-dependent labor supply elasticities and borrowing constraints, they argue that capital income taxes and progressive labor income taxes are two alternative ways of providing age-dependent insurance against idiosyncratic shocks. They use numerical methods to determine the efficient relation between the two instruments. Interestingly, capital taxes play an entirely different role in the present environment and we obtain very different conclusions. While in the framework of Conesa et al. (2009), capital income taxes and progressive labor income taxes are substitutable instruments, in our model they are complements. Laroque (2011) analytically derives a similar substitutability between labor income and capital income taxes, restricting labor taxation to be nonlinear but homogenous across age groups. In both these cases, the substitutability arises because exogenously restricted labor income taxes are, in general, imperfect instruments to perform redistribution. In our (fully-optimal taxation) environment, labor income taxes can achieve any feasible redistributional target. The role of capital taxes is to facilitate the use of such a redistributional instrument in the presence of informational asymmetries. Hence, we obtain a complementarity between capital taxes and labor income tax progressivity.

Our paper is also related to the literature on optimal tax progressivity in static models. This literature highlights the roles of the skill distribution (Mirrlees, 1971), the welfare criterion (Sadka, 1976) and earnings elasticities (Saez, 2001). For a recent survey on the issue, see Diamond and Saez (2011). However, dynamic considerations and, in particular, asset decisions are absent in those works.

3 Model

Consider a benevolent social planner (the principal) whose objective is to maximize the welfare of its citizens. The economy consists of a continuum of ex ante identical agents who live for two periods, $t = 0, 1$, and can influence their period-1 labor income realizations by exerting effort. The planner designs an allocation to insure them against idiosyncratic risk and provide them with appropriate incentives to exert effort. The planner's budget must be (intertemporally) balanced.

Preferences. The agent derives utility from consumption $c_t \geq \underline{c} \geq -\infty$ and effort $e_t \geq 0$ according to $u(c_t, e_t)$, where u is a concave, twice continuously differentiable function which is strictly increasing and strictly concave in c_t, and strictly decreasing and (weakly) concave in e_t. We assume that consumption and effort are complements: $u''_{ec}(c_t, e_t) \geq 0$. This specification of preferences includes both the additively separable case, $u(c, e) = u(c) - v(e)$, and the case with monetary costs of effort, $u(c - v(e))$, assuming that v is strictly increasing and convex. The agent's discount factor is denoted by $\beta > 0$.

Technology and endowments. The technological process can be interpreted as the production of human capital through costly effort, where human capital represents any characteristic that determines the agent's labor income. At date $t = 0$, the agent has a *fixed* endowment y_0. At date $t = 1$, the agent has a stochastic income $y \in Y := [\underline{y}, \overline{y}]$.[6] The realization of y is publicly observable, while the probability distribution over Y is affected by the agent's unobservable effort level e_0 that is exerted at $t = 0$. The probability density of this distribution is given by the smooth function $f(y, e_0)$. As in most of the the optimal contracting literature, we assume *full support*, that is $f(y, e_0) > 0$ for all $y \in Y$ and all $e_0 \geq 0$. There is no production or any other action at $t \geq 2$. Since utility is strictly decreasing in effort, the agent exerts effort $e_1 = 0$ at date 1. In what follows, we therefore use the notation $u_1(c) := u(c, 0)$ for date-1 utility.

The agent has access to a linear savings technology that allows him to transfer qb_0 units of date-0 consumption into b_0 units of date-1 consumption. The savings technology is observable to the planner.

Allocations. An *allocation* $(\boldsymbol{c}, e_0)$ consists of a consumption scheme $\boldsymbol{c} = (c_0, c(\cdot))$ and a recommended effort level e_0. The consumption scheme has two components: c_0 denotes the agent's consumption in period $t = 0$ and $c(y)$, $y \in Y$, denotes the agent's consumption in period $t = 1$ conditional on the realization y. An allocation $(c_0, c(\cdot), e_0)$ is called *feasible* if it satisfies the planner's *budget constraint*

$$y_0 - c_0 + q \int_{\underline{y}}^{\overline{y}} (y - c(y)) f(y, e_0)\, dy - G \geq 0, \tag{1}$$

[6] Although for pure notational simplicity we consider the case with a continuum of income levels, we do not discuss the technicalities related to the existence of solutions in infinite dimensional spaces. We can provide details on this issue. Alternatively, the reader can interpret the model as a framework with a large but finite number of output levels.

where G denotes government consumption and q is the rate at which the planner and the agent transfer resources over time.

3.1 Observable assets and second best allocations

As a benchmark case, we assume that the agent's savings technology is observable (and contractable) for the planner. In this case, we can assume without loss of generality that the planner directly controls consumption.

Second best. A *second best* allocation is an allocation that maximizes ex-ante welfare

$$\max_{(\mathbf{c},e_0)} \; u(c_0,e_0)+\beta\int_{\underline{y}}^{\overline{y}} u_1(c(y))f(y,e_0)\,dy$$

subject to $c_0 \geq \underline{c}$, $c(y) \geq \underline{c}$, $e_0 \geq 0$, the planner's budget constraint

$$y_0 - c_0 + q\int_{\underline{y}}^{\overline{y}} (y-c(y))f(y,e_0)\,dy - G \geq 0, \tag{2}$$

and the *incentive compatibility constraint* for effort

$$e_0 \in \arg\max_{e} \; u(c_0,e)+\beta\int_{\underline{y}}^{\overline{y}} u_1(c(y))f(y,e)\,dy. \tag{3}$$

Any second best allocation can be generated as an equilibrium outcome of a competitive environment where agents exert effort and save/borrow subject to appropriate taxes on income and assets. To simplify the analysis, we assume throughout this paper that the first-order approach (FOA) is valid. This enables us to characterize the agent's choice of effort e_0 and assets b_0 based on the associated first-order conditions (in inequality or equality form). When the FOA holds, second best allocations can be decentralized by imposing a *linear* tax on assets, complemented by suitably defined nonlinear labor income taxes and transfers.

Proposition 1 (Decentralization). *Suppose that the FOA is valid and let $(c_0, c(\cdot), e_0)$ be a second best allocation that is interior: $c_0 > \underline{c}$, $c(y) > \underline{c}$, $y \in Y$, $e_0 > 0$. Then, there exists a tax system* *consisting of income transfers* $(\tau_0, \tau(\cdot))$ *and an after-tax asset price* $\tilde{q}\ (> q)$ *such that*

$$\begin{aligned} c_0 &= y_0 + \tau_0, \\ c(y) &= y + \tau(y), \quad y \in Y, \\ (e_0, 0) &\in \arg\max_{(e,b)} \; u(y_0 + \tau_0 - \tilde{q}b, e) + \beta \int_{\underline{y}}^{\overline{y}} u_1(y + \tau(y) + b) f(y, e)\, dy. \end{aligned} \tag{4}$$

That is, there exists a tax system $(\tau_0, \tau(\cdot), \tilde{q})$ *that decentralizes the allocation* $(c_0, c(\cdot), e_0)$*.*

First, we note that an after-tax asset price $\tilde{q}$ is equivalent to a capital wedge τ^K (tax on the gross return $1/q$) that is constant across agents, $\tau^K = 1 - q/\tilde{q}$, or a linear capital income tax at rate $\tau^K/(1-q)$. Moreover, note that we have normalized the asset holdings to $b_0 = 0$ in the above proposition. This is without loss of generality, since there is an indeterminacy between τ_0 and b_0. The planner can generate the same allocation with a system $(\tau_0, \tau(\cdot), \tilde{q})$ and $b_0 = 0$ or with a system $(\tau_0 - \tilde{q}\varepsilon, \tau(\cdot) + \varepsilon, \tilde{q})$ and $b_0 = \varepsilon$ for any value of ε. This indeterminacy is not surprising, because the timing of tax collection is irrelevant by Ricardian equivalence.

Proposition 1 is intuitive and the proof is omitted. It is efficient to tax the savings technology, because savings provide intertemporal insurance when the agent plans to shirk. The reason why a linear tax on assets is sufficient to obtain the second best allocation becomes apparent once we replace the incentive constraint (4) with the associated first-order conditions:

$$u'_e(y_0 + \tau_0, e_0) + \beta \int_{\underline{y}}^{\overline{y}} u_1(y + \tau(y)) f_e(y, e_0)\, dy \;\geq\; 0, \tag{5}$$

$$\tilde{q}u'_c(y_0 + \tau_0, e_0) - \beta \int_{\underline{y}}^{\overline{y}} u'_1(y + \tau(y)) f(y, e_0)\, dy \;\geq\; 0. \tag{6}$$

The second first-order condition (6) determines the agent's asset decision taking optimal effort as given, while (5) is the optimality condition for the agent's effort decision taking optimal assets as given. In this sense, the planner can essentially ignore the problem of *"joint deviations"* when taxing asset trades. That is the essence of the validity of the first-order approach (FOA). It is now clear that by choosing a sufficiently large value for $\tilde{q}$, the planner can circumvent the first-order condition for the asset decision and obtain the second best allocation.

Sufficient conditions for the validity of the FOA in this setup are given in Abraham, Koehne and Pavoni (2011). Specifically, the FOA is valid if the agent has nonincreasing absolute risk aversion (NIARA) and the cumulative distribution function of income is log-convex in effort. As discussed by Abraham et al. (2011), both conditions have broad empirical support. First, virtually all estimations of u reveal NIARA; see Guiso and Paiella (2008) for example. The condition on the distribution function can be interpreted as a restriction on the agent's Frisch elasticity of labor supply. This restriction is satisfied as long as the Frisch elasticity is smaller than unity. In fact, most empirical studies find values for this elasticity between 0 and 0.5; see Domeij and Floden (2006), for instance.

Besides allowing for a very natural decentralization, the FOA also generates a sharp characterization of second best consumption schemes. Assuming that consumption is interior, the first-order conditions of the Lagrangian with respect to consumption are:[7]

$$\frac{\lambda}{u_c'(c_0,e_0)} = 1+\mu\frac{u_{ec}''(c_0,e_0)}{u_c'(c_0,e_0)}, \tag{7}$$

$$\frac{\lambda q}{\beta u_1'(c(y))} = 1+\mu\frac{f_e(y,e_0)}{f(y,e_0)}, \quad y\in[\underline{y},\overline{y}], \tag{8}$$

where λ and μ are the (nonnegative) Lagrange multipliers associated with the budget constraint (2) and the first-order version of the incentive constraint (3), respectively.

3.2 Hidden assets and third best allocations

Savings technologies such as domestic bank accounts, pension funds or houses may be observable at moderate costs, but there are many alternative ways of transferring resources over time that are more difficult to monitor. For instance, agents may open accounts at foreign banks or they may accumulate cash, gold or durable goods. These technologies typically bring low returns (or involve transaction costs of various sorts), but are prohibitively costly to observe for tax authorities. Hence, if the after-tax return of the observable savings technology, $1/\tilde{q}$, becomes too low, agents have a strong incentive to use nonobservable assets to run away from taxation.

Notice that, even though we describe a particular decentralization mechanism in this paper, the above problem is general. Decentralizations with income-dependent asset taxes (Kocherlakota 2005), for instance, make the savings technology unattractive by lowering the after-tax return specifically in low-income states. In this case, the *average* asset tax can be zero. However, agents would still prefer to save on a hidden asset market as long as the

[7] A sufficient condition for interiority is, for example, $u_c'(c,0)=0$ for all $c>\underline{c}$ in combination with the Inada condition $\lim_{c\to\underline{c}} u_c'(c,0)=\infty$.

return on this market is sufficiently close to the observable asset return, because the risk-adjusted expected return of hidden assets dominates the risk-adjusted after-tax return of observable assets.

This motivates the study of optimal allocations and decentralizations when agents have access to a nonobservable savings technology. We assume that the nonobservable technology is linear and transfers $q^n \geq q$ units of date-0 consumption into one unit of date-1 consumption.

Third best. Using the FOA, we define a *third best* allocation as an allocation $(c_0, c(\cdot), e_0)$ that maximizes ex-ante welfare

$$\max_{(\mathbf{c}, e_0)} \; u(c_0, e_0) + \beta \int_{\underline{y}}^{\overline{y}} u_1(c(y)) f(y, e_0)\, dy$$

subject to $c_0 \geq \underline{c}$, $c(y) \geq \underline{c}$, $e_0 \geq 0$, the planner's budget constraint

$$y_0 - c_0 + q \int_{\underline{y}}^{\overline{y}} (y - c(y)) f(y, e_0)\, dy - G \geq 0 \tag{9}$$

and the first-order incentive conditions for effort and nonobservable savings

$$u_e'(c_0, e_0) + \beta \int_{\underline{y}}^{\overline{y}} u_1(c(y)) f_e(y, e_0)\, dy \;\; \geq \;\; 0, \tag{10}$$

$$q^n u_c'(c_0, e_0) - \beta \int_{\underline{y}}^{\overline{y}} u_1'(c(y)) f(y, e_0)\, dy \;\; \geq \;\; 0. \tag{11}$$

Obviously, in our terminology the notion *second best* refers to allocations that are constrained efficient given the nonobservability of effort, while the term *third best* refers to allocations that are constrained efficient given the nonobservability of effort and assets/consumption. Moreover, note that we have written the agent's Euler equation (11) in inequality form. Proposition 2 below shows that this inequality is binding as long as the nonobservable asset is not too expensive as compared to the observable asset.

To decentralize a third best allocation $(c_0, c(\cdot), e_0)$, we define taxes/transfers $(\tau_0, \tau(\cdot))$ on labor income and an after-tax price $\tilde{q}$ of the observable asset as follows:

$$\begin{aligned}\tau_0 &= c_0 - y_0,\\ \tau(y) &= c(y) - y, \quad y \in Y,\\ \tilde{q} &= q^n.\end{aligned}$$

If agents face this tax system and have access to the nonobservable savings technologies at the rate q^n, the resulting allocation will obviously be $(c_0, c(\cdot), e_0)$.

Once more, we can use the FOA to characterize the consumption scheme. Assuming an interior solution, the first-order conditions of the Lagrangian with respect to consumption are now:

$$\frac{\lambda}{u'_c(c_0,e_0)} = 1 + \mu\frac{u''_{ec}(c_0,e_0)}{u'_c(c_0,e_0)} + \xi q^n \frac{u''_{cc}(c_0,e_0)}{u'_c(c_0,e_0)}, \tag{12}$$

$$\frac{\lambda q}{\beta u'_1(c(y))} = 1 + \mu\frac{f_e(y,e_0)}{f(y,e_0)} + \xi a(c(y)), \quad y \in [\underline{y}, \overline{y}], \tag{13}$$

where $a(c) := -u''_1(c)/u'_1(c)$ denotes absolute risk aversion, and λ, μ and ξ are the (nonnegative) Lagrange multipliers associated with the budget constraint (9), the first-order condition for effort (10), and the Euler equation (11), respectively.

Proposition 2. *Suppose that the FOA is valid and let $(c_0, c(\cdot), e_0)$ be a third best allocation that is interior. Then, there exists a number $\bar{q} > q$ such that equations (12) and (13) characterizing the consumption scheme are satisfied with $\xi > 0$ whenever $q^n < \bar{q}$. That is, the Euler equation is binding if the return on the nonobservable asset is not too low.*

We provide the proof of Proposition 2, as well as all other omitted proofs, in Appendix A. Proposition 2 states that if the return on the nonobservable savings technology $1/q^n$ is sufficiently high (although possibly lower than the return on observable savings), the agent's Euler equation will be binding in the planner's problem. To simplify the exposition, we set $q^n := q$ from now on, so that the returns of the nonobservable and observable savings technologies coincide. All our results will be independent of this particular choice of q^n and only rely on the fact the Euler equation is binding for the planner in that case.

Comparing the characterization of third best consumption schemes, (12), (13), to the characterization of second best consumption schemes, (7), (8), we notice that the difference between the two environments is closely related to the effect of the agent's Euler equation (11) and the associated Lagrange multiplier ξ. We discuss the implications of this finding in detail in the next section.

4 Theoretical results on progressivity

We are interested in the shape of second best and third best consumption schemes $c(y)$. As we saw above, this shape is related one-to-one to the curvature of labor income taxes in the associated decentralization.

Definition 1. We say that an allocation $(c_0, c(\cdot), e_0)$ is *progressive* if $c'(y)$ is decreasing in y. We call the allocation *regressive* if $c'(y)$ is increasing in y.

Recall that $\tau(y) = c(y) - y$ denotes the agent's transfer in labor income state y; hence the negative of $\tau(y)$ represents the labor income tax. Definition 1 implies that whenever a consumption scheme is progressive (regressive), we have a tax system with increasing (decreasing) marginal taxes $-\tau'(y)$ on labor income supporting it.

In a progressive system, taxes are increasing more quickly than income. At the same time, for the states when the agent is receiving a transfer, transfers are increasing more slowly than income is decreasing. The opposite happens when we have a regressive scheme. Intuitively, if the scheme is progressive, incentives are provided more by imposing "penalties" for low income realizations, since consumption decreases relatively quickly when income decreases. Regressive schemes, in contrast, put more emphasis on "rewards" for high income levels than "punishments" for low income levels.

We can find sufficient conditions for the progressivity or regressivity of optimal allocations by exploiting the optimality conditions for consumption. The curvature of consumption in the second best allocation depends on the shape of the inverse marginal utility and the likelihood ratio function, as shown by equation (8). The same forces are at work in the third best allocation, as shown by equation (13), but the curvature of absolute risk aversion becomes an additional factor. This allows us to establish the following sufficient conditions.

Proposition 3 (Sufficient conditions for progressivity). *Suppose that the FOA is justified and that second best allocations and third best allocations are interior.*

(i) *If the likelihood ratio function* $l(y,e) := \frac{f_e(y,e)}{f(y,e)}$ *is concave in* y *and* $\frac{1}{u_1'(c)}$ *is convex in* c*, second best allocations are progressive. If, in addition, absolute risk aversion* $a(c)$ *is decreasing and concave, third best allocations are progressive as well.*

(ii) *On the other hand, if* $l(y, e)$ *is convex in* y *and* $\frac{1}{u_1'(c)}$ *is concave in* c*, second best allocations are regressive. If, in addition, absolute risk aversion* $a(c)$ *is decreasing and convex, third best allocations are regressive as well.*

Note that in the previous proposition, consumption will be increasing in income if the likelihood ratio function $l(y, e)$ is increasing in y. Proposition 3 implies that CARA utilities with concave likelihood ratios lead to progressive schemes, both in the second best and the third best.[8] In the second best, progressive schemes are also induced by concave likelihood ratios and CRRA utilities with $\sigma \geq 1$, since $\frac{1}{u_1'(c)} = c^{\sigma}$ is convex in this case. For logarithmic utility with linear likelihood ratios, we obtain second best schemes that are *proportional*, since $1/u_1'(c) = c$ is both concave and convex. Interestingly, since absolute risk aversion $a(c) = 1/c$ is convex, third best schemes are *regressive* in this case.[9]

4.1 Rankings of progressivity for linear likelihood ratios

Proposition 3 above studied the curvature of consumption in an absolute sense. However, we are particularly interested in *relative* statements that compare the shape of consumption between second best and third best allocations. The current and the next section will provide such comparisons. We will find a general pattern for all utility functions with convex absolute risk aversion: when assets are observable (second best), the allocation has a more concave relationship between labor income and consumption. In other words, observability of assets calls for more progressivity in the labor income tax system.

In order to formalize this insight, we note that consumption patterns in moral hazard models are generally obtained as functions of the likelihood ratio $l(\cdot, e)$, see e.g. Holmstrom (1979). The most common way of measuring concavity/progressivity, however, is to study how consumption changes as a function of *income*. If likelihood ratios are linear in income, then the curvature of consumption as a function of the likelihood ratio (the natural outcome of a moral hazard model) is identical to the curvature of consumption as a function of income (the typical way of measuring progressivity in the applied literature). In other cases, the curvatures are related monotonically, but they are not exactly identical. Linear likelihood ratios are thus a natural starting point for studying progressivity in moral hazard models.

[8] Other cases where the progressivity/regressivity does not differ between second best and third best emerge when a has the same shape as $1/u_1'$ (quadratic utility) and when a is linear (and hence increasing).

[9] More precisely, consumption is characterized by $\frac{\lambda q}{\beta} c(y) - \xi \frac{1}{c(y)} = 1 + \mu\, l(y, e)$ in this case. Since the left-hand side is concave in c and the right-hand side is linear in y, the consumption scheme $c(y)$ must be convex in y.

Proposition 4 (Ranking of progressivity). *Suppose that the FOA is justified and that second best and third best allocations are interior. Suppose that u_1 has convex absolute risk aversion and that the likelihood ratio $l(y,e)$ is increasing and linear in y. Under these conditions, if the third best allocation is progressive, the second best allocation is progressive as well (but not vice versa). On the other hand, if the second best allocation is regressive, the third best allocation is regressive as well (but not vice versa).*

Proof. Given the validity of the FOA, by equations (8) and (13), the second and third best consumption schemes $c^{sb}(y)$ and $c^{tb}(y)$ are characterized as follows:

$$g^{sb}\left(c^{sb}(y)\right) = 1 + \mu^{sb}\, l\left(y, e_0^{sb}\right), \text{ where } g^{sb}(c) := \frac{\lambda^{sb} q}{\beta u_1'(c)}, \tag{14}$$

$$g^{tb}\left(c^{tb}(y)\right) = 1 + \mu^{tb} l\left(y, e_0^{tb}\right), \text{ where } g^{tb}(c) := \frac{\lambda^{tb} q}{\beta u_1'(c)} - \xi^{tb} a(c), \text{ with } \xi^{tb} > 0. \tag{15}$$

Since $l\left(y, e_0^{tb}\right)$ is linear in y by assumption, concavity of c^{tb} is equivalent to convexity of g^{tb}. Moreover, since $a(c)$ is convex in c by assumption, convexity of g^{tb} implies convexity of $g^{sb} = \left(g^{tb} + \xi^{tb} a\right) \lambda^{sb} / \lambda^{tb}$ (but not vice versa). Finally, notice that convexity of g^{sb} is equivalent to concavity of c^{sb}, since $l\left(y, e_0^{sb}\right)$ is linear in y. This establishes the first part of the proposition. The second part can be seen analogously. **Q.E.D.**

Many well-known probability distributions generate linear likelihood ratio functions as assumed in Proposition 4. One example is the exponential distribution with mean $\varphi(e)$ or, more generally, the Gamma distribution with mean $\varphi(e)$ for any shape parameter $k > 0$ and any increasing function φ. Another example is the normal distribution with mean $\varphi(e)$ and fixed variance (truncated to the compact interval $[\underline{y}, \overline{y}]$).[10] Moreover, we note that the linear likelihood property is unrelated to the validity of the first-order approach, since the latter imposes conditions on the curvature of φ or, equivalently, on the convexity of $u(c,e)$ as a function of effort e.

In order to obtain a clearer intuition of Proposition 4, we further examine the planner's first-order condition (13), namely

$$\frac{\lambda q}{\beta u_1'(c(y))} = 1 + \mu \frac{f_e(y, e_0)}{f(y, e_0)} + \xi a(c(y)).$$

This expression equates the discounted present value (normalized by $f(y, e_0)$) of the costs

[10] An example for discrete output spaces is the Poisson distribution with mean e.

and benefits of increasing the agent's utility by one unit in state y. The increase in utility costs the planner $\frac{q}{\beta u_1'(c(y))}$ units in consumption terms. Multiplied by the shadow price of resources λ, we obtain the left-hand side of the above expression. In terms of benefits, first there is a return of 1, since the agent's utility is increased by one unit. Furthermore, increasing the agent's utility also relaxes the incentive constraint for effort, generating a return of $\mu\frac{f_e(y,e_0)}{f(y,e_0)}$.[11] Finally, by increasing $u_1(c(y))$, the planner alleviates the agent's savings motive. Since the return to one unit of saving in state y is given by $u_1'(c(y))$, the gain of a unit increase in $u_1(c(y))$ is measured by $\xi a(c(y))$, where ξ is the multiplier of the agent's Euler equation and $a(c) = -u_1''(c)/u_1'(c)$. That is, $a(c)$ is the appropriate measure for the gains of relaxing the Euler equation. In other words, the social gains of deterring the agent from saving in a given state are proportional to the agent's absolute risk aversion in that state.

The novel term $\xi a(\cdot)$ in the planner's first-order condition captures the impact of nonobservable savings. To gain some intuition, suppose that we hold all other parameters and the multipliers λ and μ as fixed. Then, the impact of absolute risk aversion $a(\cdot)$ on the progressivity of optimal consumption is immediately visible in the planner's first-order condition. For CARA utility, or generally whenever absolute risk aversion is linear, the relative reduction of the agent's marginal utility per unit of utility, measured by $a(c) = -u''(c)/u'(c)$, changes linearly with consumption. For CARA utility, hidden saving therefore affects the level and slope, but not the curvature of consumption. For the widespread case of convex absolute risk aversion, however, the first-order condition suggests that the convexity of $a(\cdot)$ raises the convexity of optimal consumption. This intuition is confirmed by our formal proof that accounts for the endogeneity of the Lagrange multipliers. For convex absolute risk aversion, it cannot happen that the third best allocation is progressive while the second best allocation is not (Proposition 4). This provides a clear sense in which second best allocations are *more progressive* than third best allocations.

Another common approach to compare the progressivity/concavity of functions is to explore concave transformations. Recall that a function f_1 is a *concave transformation* of a function f_2 if there is an increasing and concave function v such that $f_1 = v \circ f_2$. For the case of logarithmic utility, we are able to rank the progressivity of the second and third best allocation in the sense of concave transformations.[12]

[11] If the increase in consumption is done in a state with a negative likelihood ratio, this represents a cost since the incentive constraint is, in fact, tightened.

[12] For NIARA utilities, we can more generally show that second best consumption is a *quasi-concave* transformation of third best consumption. Yet, since consumption is typically monotonic for both cases (Abraham et al., 2011), such a result does not generate a meaningful ranking.

Proposition 5 (Logarithmic utility). *Suppose (as in Proposition 4) that the FOA is justified, second best and third best allocations are interior, and the likelihood ratio $l(y,e)$ is increasing and linear in y. In addition, suppose that u_1 is logarithmic. Then, second best consumption is a concave transformation of third best consumption.*

Proof. For logarithmic utility, we have $u_1'(c) = a(c) = 1/c$. By equations (14) and (15), we can link second best and third best consumption as follows:

$$\frac{\lambda^{sb} q}{\beta} c^{sb}(y) - \mu^{sb}\, l\left(y, e_0^{sb}\right) + \mu^{tb} l\left(y, e_0^{tb}\right) = \frac{\lambda^{tb} q}{\beta} c^{tb}(y) - \frac{\xi^{tb}}{c^{tb}(y)}. \tag{16}$$

Since $l\left(y, e_0\right)$ is linear in y by assumption, equation (14) shows that $c^{sb}(y)$ is linear in y. Thus, all expressions on the left-hand side of (16) are linear in y and hence, linear in $c^{sb}(y)$. Since the right-hand side of (16) is concave in $c^{tb}(y)$, the result immediately follows. **Q.E.D.**

4.2 Rankings of progressivity for nonlinear likelihood ratios

For nonlinear likelihood ratios, two separate channels determine the contrast between consumption allocations in the second best (observable assets) scenario and the third best (hidden assets) scenario. First, as pointed out in the analysis of the planner's first-order conditions for consumption, the efficient way of relaxing the agent's Euler equation generates state-dependent returns in the third best that are proportional to the coefficient of absolute risk aversion. If absolute risk aversion is nonlinear, this has a direct influence on optimal progressivity. Second, the implemented effort level may change from the second best to the third best, which means that the role of income as an effort signal can differ between the two scenarios. This can indirectly affect the progressivity. The remainder of this section will mainly focus on the first channel. That is, in the spirit of Grossman and Hart (1983), we will analyze how the *implementation* of a given effort level e_0 depends on the economic environment. All propositions that follow describe how the curvature of the efficient allocation of consumption changes in the presence of hidden savings for any given effort level that the planner aims to implement in the two scenarios. Towards the end of the section, we discuss how our results hold when we also take into account changes in the implemented effort levels.

We denote the consumption allocation that optimally implements a given effort $e_0 > 0$ by $(\tilde{c}_0^{sb}, \tilde{c}^{sb}(\cdot))$ for the scenario with observable saving and by $(\tilde{c}_0^{tb}, \tilde{c}^{tb}(\cdot))$ for the case of hidden saving. For nonlinear likelihood ratios, we can rank the progressivity of such allocations in a way that is very similar to Proposition 4. As usual, we maintain the assumption that the FOA is justified and that second best and third best consumption levels are interior.

Proposition 6. *Consider the problem of implementing a given effort level e_0 under observable assets and under hidden assets. Suppose that u_1 has convex absolute risk aversion. If the optimal implementation under hidden assets, $\tilde{c}^{tb}$, is a concave transformation of the likelihood ratio function $l(\cdot, e_0)$, then the optimal implementation under observable assets, $\tilde{c}^{sb}$, is a concave transformation of $l(\cdot, e_0)$. On the other hand, if $\tilde{c}^{sb}$ is a convex transformation of $l(\cdot, e_0)$, then $\tilde{c}^{tb}$ is a convex transformation of $l(\cdot, e_0)$.*

The previous result generates a sense in which the consumption scheme implementing e_0 in the case of observable assets is *more progressive* than the scheme in the case of hidden assets. This result is analogous to Proposition 4 for the case of nonlinear likelihood ratios. We can also derive an analogue to Proposition 5. To this end, let us consider the class of HARA (or linear risk tolerance) utility functions, namely

$$u_1(c) = \rho\left(\eta + \frac{c}{\gamma}\right)^{1-\gamma} \quad \text{with } \rho\frac{1-\gamma}{\gamma} > 0 \text{ and } \eta + \frac{c}{\gamma} > 0.$$

For this class, absolute risk aversion is a convex function given by $a(c) = \left(\eta + \frac{c}{\gamma}\right)^{-1}$. Special cases of the HARA class are CRRA functions, CARA functions, and quadratic utility.

Lemma 1. *Given a strictly increasing, differentiable function $u_1 : [\underline{c}, \infty) \to R$, consider the two functions defined as follows:*

$$\begin{aligned} g_{\lambda,\mu}(c) &:= \frac{\lambda q}{\mu\beta u_1'(c)} - \frac{1}{\mu}, \\ g_{\hat{\lambda},\hat{\mu},\hat{\xi}}(c) &:= \frac{\hat{\lambda} q}{\hat{\mu}\beta u_1'(c)} - \frac{1}{\hat{\mu}} - \frac{\hat{\xi}}{\hat{\mu}}a(c). \end{aligned}$$

If u_1 belongs to the HARA class with $\gamma \geq -1$, then $g_{\hat{\lambda},\hat{\mu},\hat{\xi}}$ is a concave transformation of $g_{\lambda,\mu}$ for all $\hat{\lambda}, \hat{\xi} \geq 0$, $\lambda, \mu, \hat{\mu} > 0$.

The restriction of $\gamma \geq -1$ in the above result is innocuous to most applications, because it allows for all HARA functions with nonincreasing absolute risk aversion ($\gamma \geq 0$) as well as quadratic utility ($\gamma = -1$), for instance. Lemma 1 enables us to rank the progressivity of consumption in the sense of concave transformations. By the first-order conditions of the implementation problem, the consumption allocations that optimally implement a given effort are characterized as follows:

$$g_{\tilde{\lambda}^{sb},\tilde{\mu}^{sb}}\left(\tilde{c}^{sb}(y)\right) \;=\; l\left(y,e_0\right), \tag{17}$$

$$g_{\tilde{\lambda}^{tb},\tilde{\mu}^{tb},\tilde{\xi}^{tb}}\left(\tilde{c}^{tb}(y)\right) \;=\; l\left(y,e_0\right), \tag{18}$$

where $\tilde{\lambda}^{sb}$ and $\tilde{\mu}^{sb}$ represent the Lagrange multipliers (for the budget constraint and the first-order condition for effort) in the implementation problem under observable assets, whereas $\tilde{\lambda}^{sb}$, $\tilde{\mu}^{sb}$ and $\tilde{\xi}^{tb}$ are the Lagrange multipliers (for the budget constraint, the first-order condition for effort, the Euler equation) in the implementation problem under hidden assets. Because of the link between the optimal consumption schemes in equations (17) and (18), Lemma 1 has the following consequence.

Proposition 7. *Consider (as in Proposition 6) the problem of implementing a given effort level e_0 under observable assets and under hidden assets. Suppose that u_1 belongs to the HARA class with $\gamma \geq -1$. Then, there exists a monotonic function g such that $g \circ \tilde{c}^{sb}$ is a concave transformation of $g \circ \tilde{c}^{tb}$. In particular, if u_1 is logarithmic, $\tilde{c}^{sb}$ is a concave transformation of $\tilde{c}^{tb}$.*

Proof. Let $g(\cdot) := g_{\tilde{\lambda}^{sb},\tilde{\mu}^{sb}}(\cdot)$ and note that g is an increasing function. By Lemma 1 and equations (17) and (18), there exists a concave function h such that $\tilde{c}^{sb}$ and $\tilde{c}^{tb}$ are related as follows:

$$g(\tilde{c}^{sb}(y)) = h \circ g(\tilde{c}^{tb}(y)).$$

For logarithmic utility, g is an affine function, which implies that the composition $g^{-1} \circ h \circ g$ is concave whenever h is concave. Hence, for logarithmic utility, $\tilde{c}^{sb} = g^{-1} \circ h \circ g \circ \tilde{c}^{tb}$ is a concave transformation of $\tilde{c}^{tb}$. **Q.E.D.**

Proposition 7 shows that for HARA utilities, $\tilde{c}^{sb}$ is a concave transformation of $\tilde{c}^{tb}$ (after a change of variables). In this sense, optimal consumption is *more progressive* in the case of observable savings than in the case of hidden savings for any given effort level that the planner aims to implement. Proposition 7 generalizes Proposition 5 to the class of non-logarithmic HARA utilities and nonlinear likelihood ratios.[13]

All our results for nonlinear likelihood ratios generalize when we take into account changes in the implemented effort levels provided that $l\left(l^{-1}\left(y,e_0^{tb}\right),e_0^{sb}\right)$ is concave. The

[13]The same generalization of Proposition 5 exists for HARA utilities and linear likelihood ratios (allowing for changes in effort).

last condition is satisfied if the likelihood ratio in the third best is a convex transformation of the likelihood ratio in the second best. In fact, a weaker condition is sufficient. As shown by the line of proof of Proposition 7, it is sufficient that $l\left(\cdot, e_0^{sb}\right) \circ l^{-1}\left(\cdot, e_0^{tb}\right) \circ h$ is concave, where h is a strictly concave function. This condition is satisfied whenever $l\left(y, e_0^{tb}\right)$ is "not too concave" relative to $l\left(y, e_0^{sb}\right)$. How much the curvature of the likelihood ratio differs between the two scenarios is impossible to predict without detailed knowledge of the density function $f(y, e)$. We illustrate the role of the density function in our quantitative exploration. We find that the likelihood ratio induced by the effort for hidden assets is more convex (less concave) than that for observable assets. Therefore, our theoretical insights on nonlinear likelihood ratios are, in fact, further strengthened through the variation of effort between the two allocations.

5 Quantitative exploration

In this section, we parametrize our model to illustrate the quantitative effects of limited asset taxation. The quantitative exploration serves multiple purposes. First, we complement our theoretical results. For example, recall that the theoretical results on nonlinear likelihood ratios compare two allocations that implement the same effort level. In this section, we allow effort to change between the two scenarios. The second target of this exercise is to quantitatively explore how a limited possibility of taxing assets affects optimal allocations and, consequently, optimal labor income taxes.

Third, we discuss and implement a calibration strategy to recover the fundamental parameters of our model. In dynamic private information models, the standard strategy is to use cross-sectional and longitudinal income data to recover the underlying shock process (see, for example, Farhi and Werning (2013) and Golosov et al. (2013)) assuming that agents face a stylized form of the existing tax and transfer system. Given that the income process is partially endogenous in our environment, this approach would not fully identify the deep parameters. In particular, it would not provide sufficient information about the effect of effort on the distribution of income.

For the calibration, we assume that the joint distribution of consumption and income is generated by a constrained efficient allocation. The advantage of this approach is that a single cross section of consumption and income (or consumption and income growth) suffices to identify all fundamental parameters. Gayle and Miller (2009) use a similar identification strategy to estimate dynamic moral hazard models of executive compensation. Following this strategy, all key parameters are naturally identified based on the optimality conditions of the model.

Consider the following interpretation of our model: agents face income shocks, they exert unobservable work effort and they can use a saving technology with a gross return given by $1/\tilde{q}$, where $\tilde{q}$ is the *after-tax* asset price. In order to estimate the model parameters, we use consumption and income data and postulate that the data is generated by a specification of the model where capital income is taxed at an exogenous rate of 40%. Equivalently, the after-tax asset price is given by $\tilde{q} = \frac{q}{0.6+0.4q}$.[14] Note that the capital income tax of 40% is in line with U.S. effective tax rates on capital income as calculated by Mendoza et al. (1994) and Domeij and Heathcote (2004). We estimate the key parameters of the model by matching joint moments of consumption and income in an appropriately cleaned cross-sectional data. Then, we use the estimated parameters and solve the (counterfactual) model with optimal capital taxes, assuming a full capacity to observe and tax capital. The final outcome is a comparison of the optimal labor income taxes between the two scenarios, with a special attention to the change in progressivity.

Data. We use PSID (Panel Study of Income Dynamics) data for 1992 as adapted by Blundell et al. (2008). This data source contains consumption data and income data at the household level. The consumption data is imputed using food consumption (measured at the PSID) and household characteristics using the CEX (Survey of Consumption Expenditure) as a basis for the imputation procedure. Household data is useful for two reasons: (i) Consumption can be credibly measured at the household level only; (ii) taxation is mostly determined at the family level (which is typically equivalent to the household level) in the United States. We use total consumption expenditure as the measure of consumption.[15]

In our model, we have ex ante identical individuals who face the same (partially endogenous) process of income shocks. However, in the data, income is influenced by observable factors such as age, education and race. We want to control for these characteristics in order to make income shocks comparable across individuals. For this purpose, we postulate the following process for income: $y^i = \phi(X^i)\eta^i$, where y^i is household i's income, X^i are observable household characteristics (a constant, age, education and race of the household head), and η^i is our measure of the cleaned income shock. In order to isolate η^i, we regress $\log(y^i)$ on X^i. The predicted residual $\hat{\eta}^i$ of this regression is our estimate of the income shock.

[14] In line with our terminology from Section 3.2, the constrained efficient allocation with an after-tax asset price of $\tilde{q}$ is defined as the third best allocation under the assumption that agents have access to a nonobservable asset at the price $q^n = \tilde{q}$.

[15] We performed a sensitivity analysis based on nondurable consumption data. The results were qualitatively the same but the quantitative effects of optimal capital taxes were somewhat less pronounced.

The next objective is to find the consumption function. To be able to relate it to the income measure η^i, we postulate that the consumption function is also multiplicatively separable: $c^i = g^0(Z^i)g^1\left(\phi(X^i)\right) c\left(\eta^i\right)$, where Z^i are household characteristics that affect consumption, but (by assumption) do not affect income, such as the number of kids and beginning of period household assets. Our target is to identify $c\left(\eta\right)$, the pure response of consumption to the income shock. To isolate this effect, we first run a separate regression of $\log(c^i)$ on X^i and Z^i. The predicted residual of this regression is $\hat{\varepsilon}^i$. Then, we use a flexible functional form to obtain $c\left(\cdot\right)$. In particular, we estimate the following regression: $\log(\hat{\varepsilon}^i) = \sum_{j=0}^4 \gamma_j \left(\log(\hat{\eta}^i)\right)^j$. Hence, in the notation of our model, the estimate of the consumption function is given by $\hat{c}\left(y\right) = \exp\left(\sum_{j=0}^4 \hat{\gamma}_j \left(\log(y)\right)^j\right)$.

Model specification. For the quantitative exploration of our model, we move to a formulation with discrete income levels. We assume that we have N levels of second-period income, denoted by y_s, $s = 1, \ldots, N$, with $y_s > y_{s-1}$. This implies that the density function of income, $f(y,e)$, is replaced by probability weights $p_s(e)$, with $\sum_{s=1}^N p_s(e) = 1$ for all e. For the estimation of the parameters, we impose further structure. We assume

$$p_s(e) = \exp(-e)\pi_s^l + (1 - \exp(-e))\pi_s^h,$$

where π^h and π^l are probability distributions on the set $\{y_1, \ldots, y_N\}$. In addition to tractability, this formulation has the advantage that it satisfies the requirements for the applicability of the first-order approach.[16]

In order to account for heterogeneity in the data, we allow for heterogeneity in the initial endowments, specify a unit root process for income shocks, and choose preferences to be homothetic. In particular, we assume:

$$u(c,e) = \frac{c^{1-\gamma}\left(1-e\right)^{\gamma-\sigma}}{1-\gamma} \text{ with } \sigma > \gamma \geq 1 \text{ and } 1 > e > 0.$$

Here, γ measures the coefficient of risk aversion and the period utility is given by $\frac{c^{1-\gamma}}{1-\gamma}$ after the initial period.[17] The homothetic specification is useful for our empirical strategy for two reasons. First, as demonstrated by Proposition 8 in Appendix B, we are entitled to use the income and consumption residuals $\hat{\varepsilon}^i$ and $\hat{\eta}^i$ computed above as inputs for our calibration

[16]Note that we do not need to impose the monotone likelihood ratio (MLR) property, because the proof of the validity of the first-order approach only requires consumption to be monotone (see Abraham et al. (2011) for details).

[17]When $\gamma = 1$, we assume that preferences take a logarithmic form.

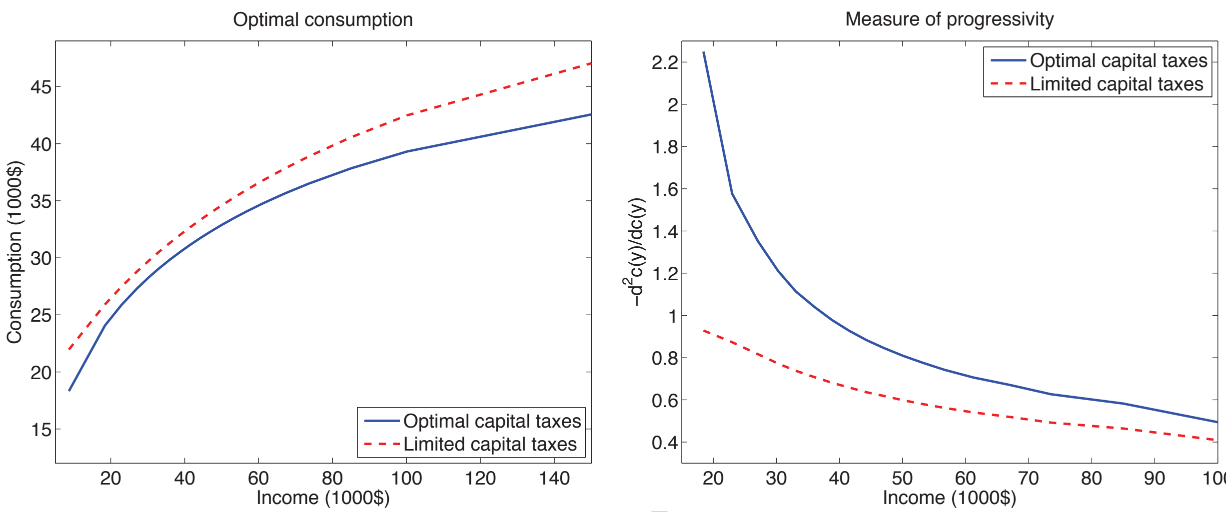
procedure. The other key advantage of the homothetic model is that we can estimate the probability distribution and all other parameters assuming that effort does not change across agents. Hence, the first-order conditions and expectations are evaluated at the same level of effort e_0^*.

Calibration. As a first step, we fix some parameters. First, we set $q = .96$ to match a yearly real interest rate of 4%, which is the historical average return on real assets in the United States. We explore a range of coefficients of relative risk aversion for consumption, with $\gamma = 3$ being our baseline case. This value is in line with estimation results based both on survey data (e.g., Barsky, Juster, Kimball and Shapiro, 1997) and actual choices (e.g., Paravisini, Rappoport and Ravina, 2015). We also consider lower values of risk aversion (e.g., Chetty, 2006).[18] For the income process, we set $N = 20$ and choose the medians of the 20 percentile groups of cleaned income for the income levels $\eta_1, ..., \eta_{20}$. The homothetic model allows us to normalize income in the initial period.

Given the fixed parameters, we determine the preference parameters (β, σ) and the probability weights $\{\pi_s^h, \pi_s^l\}_{s=1}^N$ that determine the likelihood ratios. We estimate these parameters using a method of moments estimator to match specific empirical moments for consumption and income in the data. The optimality conditions of the model give us a sufficient number of restrictions to estimate all parameters. In particular, we use the planner's optimality conditions for first and second period consumption, the planner's optimality condition for effort, and the agent's optimality conditions for effort and assets. Finally, we obtain the parameter G for government consumption as a residual of the estimation procedure implied by the government's budget constraint. Further details on the estimation strategy are provided in Appendix B.1.

Simulation results. We use the preset and estimated parameters of the above model (exogenous capital taxes) to determine the allocation for the counterfactual scenario with optimal capital taxes—assuming a full observability of capital. Figure 1a displays second-period consumption for this scenario together with the second-period consumption function of the benchmark. It is obvious from the picture that the level of second-period consumption is

[18] There is little consensus on what is the most plausible value for the coefficient of relative risk aversion. Barsky et al. (1997) use survey responses to hypothetical situations of participants in the Health and Retirement Study and estimate a coefficient between 3 and 4 on average. Paravisini et al. (2015) use actual investment data (in the 'Lending Club' platform in the U.S.) and find an average value slightly below 3. Chetty (2006) explores evidence on labor supply responses and argues for a coefficient between 1 and 2. Cohen and Einav (2007) use choice data on insurance contracts and estimate levels of (absolute) risk aversion that are higher than those obtained by Chetty.



(a) Consumption (b) Income tax progressivity

Figure 1: Optimal consumption with optimal and limited capital taxation. Notes: Figure 1a displays constrained efficient allocations in the calibrated models with limited capital taxes (capital income tax rate of 40 percent) and optimal capital taxes (assuming a full observability of capital). Figure 1b shows the associated curvature of consumption, measured as the absolute value of $c''(y)/c'(y)$. In both figures, the solid line represents the model with optimal capital taxes and the dashed line represents the model with limited capital taxes.

higher in the case with limited/exogenous capital taxes (capital income tax rate of 40%). This is not surprising because optimal capital taxes in general imply frontloaded consumption (Rogerson 1985, Golosov et al. 2003). Note that optimal capital taxes are associated with an inverse Euler equation, whereas the scenario with limited capital taxes is characterized by a standard Euler equation. By inspecting the planner's optimality condition for consumption (13), we note that the Euler equation generates an additional positive return on the right-hand side ($\xi a(\cdot) > 0$). Intuitively, holding the other multipliers and parameters fixed, this term suggests that the marginal consumption utility in the second period is lower, and consumption is thus higher, in the case with limited capital taxes.

We also observe that, since consumption is concave for the two scenarios, optimal labor income taxes are progressive for both allocations. To compare the progressivity of the two allocations, we use the concavity measure $-c''(y)/c'(y)$ to quantify the progressivity of consumption.[19] If this progressivity measure of an allocation is uniformly higher than that of a second allocation, the first allocation is a concave transformation of the second (assuming

[19] In addition to the obvious analogy to absolute risk aversion, the advantage compared with the concavity measure $c''(y)$ is that it makes functions with different slopes $c'(y)$ more comparable.

Risk aversion	1	2	3	4
Average measure of progressivity ($-c''(y)/c'(y)$)				
Optimal capital taxes (endog. weights)	0.670	0.801	0.962	1.098
Optimal capital taxes (exog. weights)	0.671	0.805	0.977	1.135
Limited capital taxes	0.644	0.644	0.644	0.644
Welfare gain from optimal capital taxes (% of consumption)				
	0.061	0.294	1.303	3.351
Optimal capital wedge (%)				
$\tau^K = 1 - q/\tilde{q}$	3.77	20.17	39.65	55.02

Table 1: Quantitative measures of progressivity, welfare gains and capital wedges. Notes: The table summarizes some key statistics of constrained efficient allocations in the calibrated models with limited capital taxes (capital income tax rate of 40 percent) and with optimal capital taxes (assuming a full observability of capital). For the model with optimal capital taxes, we calculate the average measure of progressivity using the associated income distribution (endogenous weights) and using the income distribution from the model with limited capital taxes (exogenous weights). Welfare gains from optimal capital taxes are expressed as consumption equivalent variation in percent.

that both allocations are monotonically increasing). In Figure 1b, we have plotted this measure of progressivity for the optimal consumption scheme when capital taxes are limited and when they are optimal. The pattern is clear: the model with optimal capital taxes results in a uniformly more concave (progressive) consumption scheme as compared to the case when capital taxes are limited. The differences are particularly large for lower levels of income.

In Table 1, we have quantified the graphical observations from Figures 1a and 1b and have checked the robustness to alternative levels of risk aversion. The figures are drawn using a (baseline) coefficient of relative risk aversion equal to 3. Our results are qualitatively the same for all risk aversion levels, but there are significant quantitative differences. In particular, the effect of capital taxation on progressivity is increasing in the level of risk aversion.[20] For logarithmic utility, our measure of progressivity increases by only 4% when we switch from the allocation with limited capital taxes to that with optimal capital taxes. For coefficients of relative risk aversion of 2 and 3, the progressivity of income taxes increases by 24 and 49 percent, respectively. Note that the change in measured progressivity comes from two sources. First, as shown by Figure 1a, the concavity of the optimal consumption function $c(y)$ changes. Second, the distribution of income changes, as effort is different under optimal capital taxes as compared to the benchmark case. For this reason, we calculate the measure of progressivity both with and without this second effect (endoge-

[20]See Appendix B.2 for an extended discussion of the role of relative risk aversion in our model.

nous versus exogenous weights). Comparing the third and fourth rows of Table 1, we notice that the changing effort mitigates the increase in progressivity in a non-negligible way only for higher risk aversion levels. This also implies that effort is indeed higher when optimal capital taxes are levied. In turn, higher effort implies a higher weight on high income realizations where the progressivity differences are lower (see Figure 1b). In any case, this second indirect effect through effort is small and, hence, the difference in the progressivity measure is still increasing in risk aversion.

We obtain a similar message if we consider the welfare gains from optimal capital taxation in consumption equivalent terms (presented in row 7 of Table 1). The gains are small for the logarithmic case, sizeable for the intermediate cases of relative risk aversion of 2 and 3, and considerable for high values of risk aversion. We have also displayed the implied capital wedges, calculated as $\tau^K = 1 - q/\tilde{q}$, where $\tilde{q}$ is the after-tax asset price in the optimal capital tax scenario. Notice that τ^K is indeed the tax rate on the gross return, not on capital income. The 40 percent tax on capital income in the benchmark model is equivalent to a capital wedge of 1.6 percent.[21] It turns out that the capital wedges in the scenario with observable assets are much higher than this number for all risk aversion levels. Even for logarithmic utility, the capital wedges imply a tax rate on capital income of around 90 percent.[22]

Finally, we examine the role of endogenous effort for the change in progressivity. Figure 2 plots the likelihood ratio function implied by the estimated parameters. We note that the likelihood ratio function becomes more concave for higher effort levels. Moreover, effort in the second best allocation (0.48) is higher than in the case with limited capital taxes (0.33). Hence, the change of the likelihood ratio contributes to the lower degree of progressivity in the third best. Note that this effect goes in the same direction as our insights on the convex cross-sectional returns of relaxing the Euler equation. Therefore, the change in effort between the second best and third best reinforces our theoretical results concerning the progressivity of the consumption allocations.

[21] Recall that the capital wedge τ^K is equivalent to a tax rate on capital income given by $t = \tau^K/(1-q)$.

[22] Admittedly, our model is very stylized. Golosov et al. (2013), however, study a dynamic Mirrlees model with logarithmic utility and a full observability of assets. The capital wedges (and the associated capital income taxes) are similar to those we find for the logarithmic case. Farhi and Werning (2013) study a similar Mirrlees model with logarithmic utility and obtain tax rates on capital income that are smaller than ours. In their calibration, the private information friction seems to be less severe.

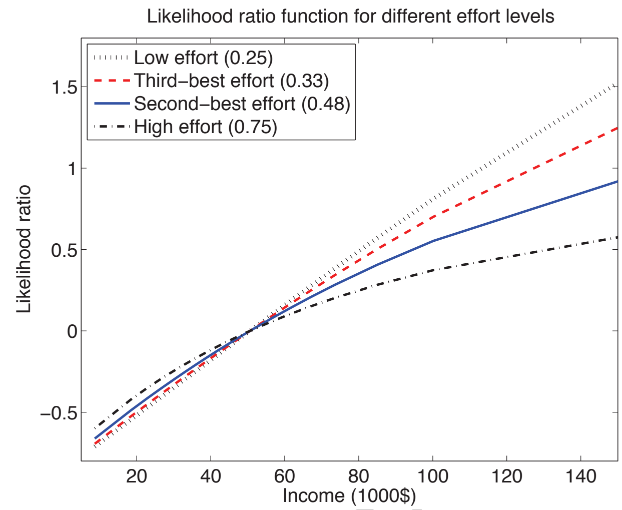


Figure 2: Likelihood ratio function in the calibrated model. Notes: The likelihood ratio function is evaluated at four different effort levels: a low effort benchmark (dotted line), effort in the constrained efficient allocation with limited capital taxes (dashed line), effort in the constrained efficient allocation with optimal capital taxes (solid line), and a high effort benchmark (dash-dotted line).

6 Discussion of model limitations and extensions

This paper analyzed how limitations to asset taxation change the optimal tax code on labor income. Assuming preferences with convex absolute risk aversion, we found that optimal consumption moves in a more convex way with labor income when asset accumulation cannot be perfectly controlled by the planner. In terms of our decentralization, this implies that taxes on labor income become less progressive when limitations to asset taxation are binding. We complemented our theoretical results with a quantitative illustration based on individual level U.S. data on consumption and income. The results suggest that the effect of imperfect asset taxation on the curvature of the optimal income tax can be sizable, and is very sensitive to the coefficient of relative risk aversion.

In this section, we discuss some of the key assumptions of the analysis and present some possible directions for future research.

General equilibrium. First, we comment on the partial equilibrium nature of our model. In the scenario with optimal capital taxes, the planner modifies the intertemporal consumption profile by increasing consumption in the initial period and reducing it in the second period (relative to the third best allocation). A change of this type is less beneficial in a general equilibrium framework because consumption frontloading becomes more costly when the equilibrium interest rate rises. In order to approximate potential (general equilibrium) limitations in the choice of the intertemporal consumption profile, in Appendix B.3 we consider a version of our model where the planner is constrained by a period-by-period budget constraint. In that framework, when moving from the third best to the second best allocation, the price q adjusts downwards (the interest rate increases) so that borrowing becomes unattractive for the planner. Hence, any intertemporal modification of consumption between the two allocations is prevented by construction. The resulting consumption function in period 1 is graphically represented in Figure 3a in Appendix B.3. The optimal degree of progressivity falls slightly (from 0.962 to 0.874) as compared to the case with a constant q, but remains significantly larger than in the third best allocation (see Figure 3b in Appendix B.3).

Both our partial and general equilibrium frameworks assume that the intertemporal allocation of resources by the planner and by the agents takes place through financial assets and not through investment in productive capital. Now, we discuss that, qualitatively, this is without loss of generality. In order to see this, recall that the population has a measure of one and hours are supplied inelastically. Let $N_0 = y_0$ and $N_1(e) = \int_{\underline{y}}^{\overline{y}} y f(y,e) dy$ be the total supply of labor in the two periods (in efficiency units). Then, total production in period t equals $F(K_t, N_t)$, where K_0 is given. Resource feasibility at $t = 0$ implies $K_1 = (1-\delta)K_0 + F(K_0, N_0) - c_0 - G_0$, where G_0 represents government consumption in the initial period. In this general framework, labor income in period-1 state y equals $\tilde{y}(y, w_1) = w_1 y$, where $w_1 = F_N(K_1, N_1)$. Recall that our implementation exercise defines taxes as $\tau(\tilde{y}(y)) = \tilde{y}(y) - \tilde{c}(\tilde{y}(y)) = w_1 y - \tilde{c}(w_1 y)$. Hence, it is immediate that the linearity of $\tilde{y}$ in y implies that the curvature properties of $c(y)$ carry over to the newly defined consumption function $\tilde{c}(\tilde{y})$. Therefore, we expect to retain all our qualitative results regarding the progressivity of allocations for this extension as well.[23]

Finally, note that the motivation for asset taxation is the same independently of whether we consider partial or general equilibrium or whether we consider financial assets or productive capital. To see this, note that savings taxes in our model have a Pigouvian nature and are set in order to internalize a negative 'informational externality' generated by capital accumulation by the private sector. In other words, without the tax on savings, agents

[23]Whenever the production function is strictly concave with respect to capital, the allocation in the general production economy will be an intermediate outcome between the case with fixed q considered in the main text (which corresponds to a linear production function $F(K, N) = K/q + N$) and the case with flexible q but no intertemporal reallocation described in Appendix B.3. Similar to the latter case, the intertemporal price adjusts in this case. However, similar to the baseline case, an intertemporal reallocation of resources is possible through investing in productive capital, mitigating the change in the interest rate as compared to the case described in Appendix B.3.

would save too much and the society would accumulate too much capital. As discussed, for example, by Golosov et al. (2003), this reasoning applies irrespective of whether savings take the form of productive capital or a simple storage technology. In particular, the spirit of capital taxation in our model is very different from classical Ramsey taxation. Taxes are not motivated in order to finance government spending—they actually improve efficiency.

Standard intensive margin. The model we presented here is one of action moral hazard, similar to that of Varian (1980) and Eaton and Rosen (1980). This framework has the important advantage of tractability. Although a common interpretation of this model is that of insurance, we believe that it conveys a number of general principles for optimal taxation that also apply to models of redistribution under asymmetric information on productivity. In our model, the periodic income y is the result of previously supplied effort and is subject to some uncertainty. Natural interpretations of the outcome y include the result of job search activities and the monetary consequences of a promotion or a demotion, i.e., of a better or worse match (within the same firm or into a new firm). For self-employed individuals, y can be seen as the earnings from entrepreneurial activity. It would not be difficult to include a standard intensive margin of labor supply in our model at $t = 1$. Suppose, for simplicity, that the utility function takes an additively separable form $u_1(c) - v(n)$, where n represents hours of work. If we now interpret y as productivity, total income becomes $I = yn$. Clearly, our analysis would not change at all if both y and I (or n) were observable. The case where the government can only observe I is that of Mirrlees (1971). In the latter case, the intensive-margin incentive constraints would take the familiar form:

$$\frac{dc(y)}{dy}u_1'(c(y)) = \frac{v'(n(y))}{y}\frac{dI(y)}{dy}.$$

The analysis of the intensive margin is standard. If we assume no bunching, the validity of the FOA for effort, and use the envelope theorem, we obtain the formula for third best allocations as:

$$\frac{q\lambda}{\beta u'(c(y))} = 1 + \mu l(y;e) + \xi a(c(y)) - \frac{d\phi(y)/dy}{\beta f(y;e)},$$

where the multiplier $\phi(y)$ associated with the intensive-margin incentive constraint is related to the Spence-Mirrlees condition and the labor supply distortion, and satisfies $\phi\left(\underline{y}\right) = \phi(\overline{y}) = 0$. The comparison between restricted and unrestricted asset taxation once more amounts to considering the cases with $\xi > 0$ and $\xi = 0$, respectively. Although the forces at play are the same as above, analytic results with an intensive margin (and private information on y) are complicated by the fact that the schedule of multipliers ϕ changes when allocations are compared.

Longer time horizons. In general, it seems difficult to conjecture to what extent our results will generalize to multi-period settings. For example, the cost of funds λ and the shadow cost of incentives μ and savings ξ are complicated objects that depend on many details of the environment. To a large extent, these prices are the key determinants of the curvature of consumption and progressivity. Hence, extending our model to multiple periods will be an important task for future research. However, there are two challenges with this endeavor. First, we already observed the technical complications associated with extending the first-order approach to more general environments (Abraham et al., 2011). Second, while a numerical verification of the first-order approach is possible in multi-period settings, the computation of such models is very difficult in practice because the domain of the problem becomes notoriously ill-behaved when nonobservable assets are present. For this reason, existing numerical approaches relied on binary outcomes (Abraham and Pavoni, 2008), which precludes a discussion of progressivity.

A simple way of studying long-run dynamics is to embed our model in an overlapping generations framework with dynastic preferences through "warm glow" motives for bequests. Assume that preferences in the last period are $u_1\left(c^{\theta}w^{1-\theta}\right)$, with $\theta \in (0,1)$. Here, c is consumption as above, while w represents wealth transfers to future generations (bequests). Given the net income $y+\tau(y)$ in the last period, the agent solves:[24]

$$\begin{aligned} &\max_{w,c\geq 0}\ u_1\left(c^{\theta}w^{1-\theta}\right) \\ s.t.\qquad & c+w = y+\tau(y). \end{aligned}$$

The chosen functional form implies that expenditures on c and w will be fixed proportions of the disposable income: $\hat{c}(y) = \theta(y+\tau(y))$ and $\hat{w}(y) = (1-\theta)(y+\tau(y))$. Hence, this model with bequests is equivalent to our original model with a utility function $\tilde{u}(c) = u_1(\alpha c)$, where u_1 is our original utility function and $\alpha = \theta^{\theta}(1-\theta)^{1-\theta}$ is a constant. Clearly, none of our theoretical results will change, since the convexity of absolute risk aversion is invariant to this modification. It is relatively straightforward to embed such model into a fully dynastic framework. When y is observed, w is easily computable as a (deterministic) function of $y+\tau$, since the warm glow mechanics do not leave space for strategic considerations in the inter-generational transfer of wealth. Then, w would play the role of the initial

[24]Note that there are no reasons to impose capital taxes at $t=1$ in order to alleviate incentives.

endowment y_0 for the next generation. Naturally, this framework generates heterogeneity in the initial endowments. However, the link between initial consumption c_0 and the endowment y_0 would be dictated by distributional motives alone (i.e., no incentive constraint for effort would play any role here), similar to the role of heterogeneity in our quantitative exploration (see Proposition 8 in Appendix B).

Acknowledgements

Earlier drafts of this paper were circulated under the title "Optimal Income Taxation with Asset Accumulation." The paper has benefitted from helpful discussions at numerous seminars and conferences. We particularly thank Claus Thustrup Kreiner, two anonymous referees, Varadarajan V. Chari, Mikhail Golosov, Chris Sleet, and Oleg Tsyvinski for insightful comments. Sebastian Koehne gratefully acknowledges financial support from the Knut och Alice Wallenbergs Stiftelse (grant 2012.0315). Nicola Pavoni thanks the MIUR (Prin prot. 2010NE9L9Z-004) for financial support.

References

[1] ABRAHAM, A., S. KOEHNE, AND N. PAVONI (2011): "On the First-Order Approach in Principal-Agent Models with Hidden Borrowing and Lending," *Journal of Economic Theory*, 146, 1331–1361.

[2] ABRAHAM, A., AND N. PAVONI (2008): "Efficient Allocations with Moral Hazard and Hidden Borrowing and Lending: A Recursive Formulation," *Review of Economic Dynamics*, 11, 781–803.

[3] BARSKY, R. B., JUSTER, F. T., KIMBALL, M. S. AND SHAPIRO, M. D. (1997), "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study," *Quarterly Journal of Economics,* 112(2), 537-579.

[4] BLUNDELL R, PISTAFERRI L., PRESTON I. (2008): "Consumption inequality and partial insurance." *American Economic Review*, 98: 1887–921.

[5] CHETTY, R. (2006): "A New Method of Estimating Risk Aversion," *American Economic Review,* 96(5): 1821-1834.

[6] COHEN, A. AND EINAV, L. (2007), "Estimating Risk Preferences from Deductible Choice?," *American Economic Review,* 97(3), 745-788.

[7] CONESA, J. C., S. KITAO, AND D. KRUEGER (2009): "Taxing Capital? Not a Bad Idea After All!," *American Economic Review*, 99 (1), 25-48.

[8] DIAMOND, P. A., AND J. A. MIRRLEES (1978): "A model of social insurance with variable retirement," *Journal of Public Economics*, 10(3), 295–336.

[9] DIAMOND, P. A., AND E. SAEZ (2011): "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25(4), 165-190.

[10] DOMEIJ, D., AND M. FLODEN (2006): "The Labor-Supply Elasticity and Borrowing Constraints: Why Estimates are Biased," *Review of Economic Dynamics*, 9(2), 242–262.

[11] DOMEIJ, D., AND J. HEATHCOTE (2004): "On The Distributional Effects Of Reducing Capital Taxes," *International Economic Review*, 45(2), 523–554.

[12] EATON, J., AND H. S. ROSEN (1980): "Optimal Redistributive Taxation and Uncertainty," *Quarterly Journal of Economics*, 95(2), 357–64.

[13] FARHI, E., AND I. WERNING (2013): "Insurance and Taxation over the Life Cycle," *Review of Economic Studies*, 80(2), 596-635.

[14] GAYLE G.-L., AND R. MILLER (2009): "Has Moral Hazard Become a More Important Factor in Managerial Compensation?" *American Economic Review* , 99(5), 1740–1769.

[15] GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKI (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70(3), 569–587.

[16] GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (2013): "Redistribution and Social Insurance," *mimeo.*

[17] GOLOSOV, M., AND A. TSYVINSKI (2007): "Optimal Taxation with Endogenous Insurance Markets," *Quarterly Journal of Economics*, 122(2), 487–534.

[18] GROSSMAN, S. J., AND O. D. HART (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51(1), 7–45.

[19] GUISO, L., AND M. PAIELLA (2008): "Risk Aversion, Wealth, and Background Risk," *Journal of the European Economic Association*, 6(6), 1109–1150.

[20] HOLMSTROM, B. (1979): "Moral hazard and observability," *Bell Journal of Economics*, 10(1), 74–91.

[21] KAPICKA, M., AND J. NEIRA (2015): "Optimal Taxation with Risky Human Capital," *mimeo.*

[22] KOCHERLAKOTA, N. R. (2005): "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica*, 73(5), 1587–1621.

[23] LAROQUE, G. (2011): "On Income and Wealth Taxation in A Life-Cycle Model with Extensive Labour Supply," *Economic Journal*, 121(554), F144-F161.

[24] MENDOZA, E., A. RAZIN, AND L. TESAR (1994): "Effective tax rates in macroeconomics: Cross-country estimates of tax rates on factor incomes and consumption," *Journal of Monetary Economics*, 34(3), 297–323.

[25] MIRRLEES, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(114), 175–208.

[26] PARAVISINI, D., V. RAPPOPORT AND E. RAVINA (2015): "Risk Aversion and Wealth: Evidence from Person-to-Person Lending Portfolios," *Management Science*, forthcoming.

[27] ROGERSON, W. (1985), "Repeated Moral Hazard," *Econometrica*, 53: 69-76.

[28] SADKA, E. (1976): "On Progressive Income Taxation," *American Economic Review*, 66(5), 931–35.

[29] SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–29.

[30] VARIAN, H. R. (1980): "Redistributive taxation as social insurance," *Journal of Public Economics*, 14(1), 49–68.

Appendices

A Proofs omitted from the main text

Proof of Proposition 2. Fix q^n. From the Kuhn-Tucker theorem, we have $\xi \geq 0$. If $\xi > 0$, we are done. If $\xi = 0$, the first-order conditions of the Lagrangian read

$$\begin{aligned} \frac{\lambda}{u_c'(c_0, e_0)} &= 1 + \mu \frac{u_{ec}''(c_0, e_0)}{u_c'(c_0, e_0)}, \\ \frac{\lambda q}{\beta u_1'(c(y))} &= 1 + \mu \frac{f_e(y, e_0)}{f(y, e_0)}, \quad y \in [\underline{y}, \overline{y}]. \end{aligned}$$

Since $f(y, e)$ is a density, integration of the last line yields

$$\int_{\underline{y}}^{\overline{y}} \frac{\lambda q}{\beta u_1'(c(y))} f(y, e_0)\, dy = 1.$$

Using $\mu \geq 0$ and the assumption $u_{ec}'' \geq 0$, we obtain

$$\frac{\lambda}{u_c'(c_0, e_0)} \geq 1 = \int_{\underline{y}}^{\overline{y}} \frac{\lambda q}{\beta u_1'(c(y))} f(y, e_0)\, dy \geq \frac{\lambda q}{\beta \int_{\underline{y}}^{\overline{y}} u_1'(c(y)) f(y, e_0)\, dy},$$

where the last inequality follows from Jensen's inequality. This inequality is, in fact, strict, because the agent cannot be fully insured when effort is interior. Since we have $\lambda > 0$ from the previous condition, we conclude

$$\beta \int_{\underline{y}}^{\overline{y}} u_1'(c(y)) f(y, e_0)\, dy > q u_c'(c_0, e_0). \tag{19}$$

Clearly, exactly the same allocation delivering condition (19) is obtainable for all q^n by ignoring the agent's Euler equation. If we now define $\bar{q} > q$ such that

$$\beta \int_{\underline{y}}^{\overline{y}} u_1'(c(y)) f(y, e_0)\, dy = \bar{q} u_c'(c_0, e_0),$$

it is immediate to see that whenever $q^n < \bar{q}$, the allocation we obtained above ignoring the agent's Euler equation is, in fact, incompatible with (11); hence we must have $\xi > 0$. **Q.E.D.**

Proof of Proposition 3. We only show (i), since statement (ii) can be seen analogously. Define

$$g(c) := \frac{\lambda q}{\beta u_1'(c)} - \xi a(c).$$

By the concavity of u, the function $\frac{1}{u_1'(\cdot)}$ is increasing. Therefore, if $\frac{1}{u_1'(\cdot)}$ is convex and $\xi = 0$ (or $\xi > 0$ and $a(\cdot)$ decreasing and concave), then $g(\cdot)$ is increasing and convex. Given the validity of the FOA, equation (8) (or equation (13), respectively) shows that second best (third best) consumption schemes are characterized as follows:

$$g\left(c(y)\right) = 1 + \mu\, l\left(y, e_0\right),$$

where, by assumption, the right-hand side is a positive affine transformation of a concave function. By applying the inverse function of $g(\cdot)$ to both sides, we see that $c\left(\cdot\right)$ is concave because it is an increasing and concave transformation of a concave function. **Q.E.D.**

Proof of Proposition 6. Using the Lagrange multipliers of the implementation problem, we define the functions

$$\tilde{g}^{sb}\left(c\right) := \frac{\tilde{\lambda}^{sb} q}{\beta u_1'(c)} \quad , \quad \tilde{g}^{tb}\left(c\right) := \frac{\tilde{\lambda}^{tb} q}{\beta u_1'(c)} - \tilde{\xi}^{tb} a(c),$$

where $\tilde{\xi}^{tb} > 0$ represents the Lagrange multiplier associated with the Euler equation in the problem with hidden assets. The first-order conditions for consumption are analogous to (14) and (15) and imply

$$\tilde{g}^{sb}\left(\tilde{c}^{sb}(y)\right) = 1 + \tilde{\mu}^{sb}\, l\left(y, e_0\right), \tag{20}$$
$$\tilde{g}^{tb}\left(\tilde{c}^{tb}(y)\right) = 1 + \tilde{\mu}^{tb} l\left(y, e_0\right). \tag{21}$$

First, suppose that $\tilde{c}^{tb}$ is a concave transformation of $l\left(y, e_0\right)$. Since the right-hand side of (21) is a positive affine transformation of $l\left(y, e_0\right)$, this is equivalent to the condition that $\tilde{g}^{tb}$ is convex. Now, since $a(c)$ is convex by assumption, convexity of $\tilde{g}^{tb}$ is sufficient (but not necessary) for $\tilde{g}^{sb}\left(c\right) = \left(\tilde{g}^{tb}(c) + \tilde{\xi}^{tb} a(c)\right) \tilde{\lambda}^{sb} / \tilde{\lambda}^{tb}$ being convex as well. Finally, using (20) we note that $\tilde{g}^{sb}$ is convex if and only if $\tilde{c}^{sb}$ is a concave transformation of $l\left(y, e_0\right)$.

The second part of the proposition follows from similar arguments by exploiting the fact that concavity of $\tilde{g}^{sb}$ is sufficient (but not necessary) for concavity of $\tilde{g}^{tb}$ if absolute risk aversion is convex. **Q.E.D.**

Proof of Lemma 1. After some simple algebra, we obtain

$$g_{\hat{\lambda},\hat{\mu},\hat{\xi}}(c) = \frac{\mu\hat{\lambda}}{\lambda\hat{\mu}} g_{\lambda,\mu}(c) + \frac{1}{\mu} - \frac{1}{\hat{\mu}} - \frac{\hat{\xi}}{\hat{\mu}} a(c).$$

If u belongs to the HARA class, we have

$$a(c) = \left(\frac{q\gamma\lambda}{\beta(1-\gamma)\rho(1+\mu g_{\lambda,\mu}(c))} \right)^{1/\gamma}.$$

Defining $\kappa := (q\gamma)^{1/\gamma}(\beta(1-\gamma)\rho)^{-1/\gamma} > 0$, this implies

$$g_{\hat{\lambda},\hat{\mu},\hat{\xi}}(c) = \frac{\mu\hat{\lambda}}{\lambda\hat{\mu}} g_{\lambda,\mu}(c) + \frac{1}{\mu} - \frac{1}{\hat{\mu}} - \frac{\hat{\xi}}{\hat{\mu}} \lambda^{1/\gamma}\kappa(1+\mu g_{\lambda,\mu}(c))^{-1/\gamma}.$$

Equivalently, we have $g_{\hat{\lambda},\hat{\mu},\hat{\xi}}(c) = h\left(g_{\lambda,\mu}(c)\right)$, where the function h is defined as

$$h(g) = \frac{\mu\hat{\lambda}}{\lambda\hat{\mu}} g + \frac{1}{\mu} - \frac{1}{\hat{\mu}} - \frac{\hat{\xi}}{\hat{\mu}} \lambda^{1/\gamma}\kappa(1+\mu g)^{-1/\gamma}.$$

The second derivative of h with respect to g equals $-\hat{\xi}(1+\gamma)\kappa\lambda^{1/\gamma}\mu^2\hat{\mu}^{-1}\gamma^{-2}(1+\mu g)^{2-1/\gamma}$, which is negative whenever $\gamma \geq -1$. **Q.E.D.**

B Appendix to the quantitative exploration

First, we derive the implications of the homothetic utility function and the unit root process for income.

Proposition 8. *Consider the following family of homothetic models with heterogeneous agents:*

$$\max_{c_0^i, c_s^i, e_0^i} \sum_i \psi^i \left\{ \frac{\left(c_0^i\right)^{1-\gamma}\left(1-e_0^i\right)^{\gamma-\sigma}}{1-\gamma} + \beta \sum_s p_s \left(e_0^i\right) \frac{\left(c_s^i\right)^{1-\gamma}}{1-\gamma} \right\}$$

$$\text{s.t.} \quad \sum_i \left(y_0^i - c_0^i\right) + q \sum_i \sum_s p_s \left(e_0^i\right) \left[y_s^i - c_s^i\right] \geq G;$$

$$(\sigma - \gamma) \frac{\left(c_0^i\right)^{1-\gamma} \left(1 - e_0^i\right)^{\gamma-\sigma-1}}{1-\gamma} = \beta \sum_s p_s' \left(e_0^i\right) \frac{\left(c_s^i\right)^{1-\gamma}}{1-\gamma};$$

$$\tilde{q} \left(c_0^i\right)^{-\gamma} \left(1 - e_0^i\right)^{\gamma-\sigma} = \beta \sum_s p_s \left(e_0^i\right) \left(c_s^i\right)^{-\gamma};$$

with $\beta \in (0,1)$, and $\tilde{q}, q > 0$. Moreover, assume that income follows: $y_s^i = y_0^i \eta_s$. For each given vector of income levels in period zero $\left(y_0^i\right)_i > 0$ and any scalar $\kappa > 0$, let the Pareto weights $(\psi^i)_i$ be such that the solution to the above problem delivers period zero consumption $c_0^{*i} = \kappa y_0^i$ for all i. Then, there exists $t^* \in \mathbb{R}$ and individual specific transfers $t^i = t^* y_0^i$ such that $G = \sum_i t^i$ and the solution to the above problem is

$$\begin{aligned} c_0^{*i} &= \kappa y_0^i \text{ for all } i; \\ e_0^{*i} &= e_0^* \text{ for all } i; \\ c_s^{*i} &= c_0^{*i} \varepsilon_s^* \text{ for all } i; \end{aligned}$$

where e_0^* and ε_s^* are a solution to the following normalized problem:

$$\max_{\varepsilon_s, e_0} \frac{\left(1 - e_0\right)^{\gamma-\sigma}}{1-\gamma} + \beta \sum_s p_s \left(e_0\right) \frac{\left(\varepsilon_s\right)^{1-\gamma}}{1-\gamma};$$

$$\text{s.t.} \quad \frac{1}{\kappa} - 1 + q \sum_s p_s \left(e_0\right) \left[\frac{\eta_s}{\kappa} - \varepsilon_s\right] \geq t^*;$$

$$(\sigma - \gamma) \frac{\left(1 - e_0\right)^{\gamma-\sigma-1}}{1-\gamma} = \beta \sum_s p_s' \left(e_0\right) \frac{\left(\varepsilon_s\right)^{1-\gamma}}{1-\gamma};$$

$$\tilde{q} \left(1 - e_0\right)^{\gamma-\sigma} = \beta \sum_s p_s \left(e_0\right) \left(\varepsilon_s\right)^{-\gamma}.$$

Proof. The linear separability of the planner's problem implies that, given individual transfers t^i, the optimal allocation must solve the following individual contracting problem:

$$V^i = \max_{c_0^i, c_s^i, e_0^i} \psi^i \left\{ \frac{\left(c_0^i\right)^{1-\gamma} \left(1 - e_0^i\right)^{\gamma-\sigma}}{1-\gamma} + \beta \sum_s p_s \left(e_0^i\right) \frac{\left(c_s^i\right)^{1-\gamma}}{1-\gamma} \right\}$$

$$\text{s.t.} \qquad y_0^i - c_0^i + q \sum_s p_s\left(e_0^i\right)\left[y_0^i \eta_s - c_s^i\right] \geq t^i;$$

$$(\sigma - \gamma)\frac{\left(c_0^i\right)^{1-\gamma}\left(1 - e_0^i\right)^{\gamma-\sigma-1}}{1-\gamma} = \beta \sum_s p_s'\left(e_0^i\right)\frac{\left(c_s^i\right)^{1-\gamma}}{1-\gamma};$$

$$\tilde{q}\left(c_0^i\right)^{-\gamma}\left(1 - e_0^i\right)^{\gamma-\sigma} = \beta \sum_s p_s\left(e_0^i\right)\left(c_s^i\right)^{-\gamma};$$

with $\psi^i > 0$. Because preferences are homothetic, the incentive constraints only depend on $\varepsilon_s^i = c_s^i / c_0^i$ and e_0^i. Hence, we can change the choice variables and rewrite the individual contracting problem as

$$V^i = \max_{c_0^i, \varepsilon_s^i, e_0^i} \psi^i\left(c_0^i\right)^{1-\gamma}\left\{\frac{\left(1 - e_0^i\right)^{\gamma-\sigma}}{1-\gamma} + \beta \sum_s p_s\left(e_0^i\right)\frac{\left(\varepsilon_s^i\right)^{1-\gamma}}{1-\gamma}\right\}$$

$$\text{s.t.} \qquad y_0^i - c_0^i + q \sum_s p_s\left(e_0^i\right)\left[y_0^i \eta_s - c_0^i \varepsilon_s^i\right] \geq t^i;$$

$$(\sigma - \gamma)\frac{\left(1 - e_0^i\right)^{\gamma-\sigma-1}}{1-\gamma} = \beta \sum_s p_s'\left(e_0^i\right)\frac{\left(\varepsilon_s^i\right)^{1-\gamma}}{1-\gamma};$$

$$\tilde{q}\left(1 - e_0^i\right)^{\gamma-\sigma} = \beta \sum_s p_s\left(e_0^i\right)\left(\varepsilon_s^i\right)^{-\gamma}.$$

Now fix some individual j. By continuity we can find a transfer t^j such that the solution $\left(c_0^{j*}, e_0^{j*}, \varepsilon_s^{j*}\right)$ of the associated individual problem satisfies $c_0^{j*} = \kappa y_0^j$. By the non-satiation of preferences, t^j is given by

$$t^j = y_0^j - \kappa y_0^j + q y_0^j \sum_s p_s\left(e_0^{j*}\right)\left[\eta_s - \varepsilon_s^{j*}\kappa\right] =: y_0^j t^*.$$

We claim that transfers defined as $t^i := y_0^i t^*$ imply that for all i the contract

$$c_0^{i*} = \kappa y_0^i, \qquad e_0^{i*} = e_0^{j*}, \qquad \varepsilon_s^{i*} = \varepsilon_s^{j*},$$

solves the individual contracting problem. Suppose that the claim is false for some i. By the construction of transfers, the contract $\left(\kappa y_0^i, e_0^{j*}, \varepsilon_s^{j*}\right)$ is incentive-feasible. Hence, if the claim is false, the value V^i must be strictly higher than the one generated by $\left(\kappa y_0^i, e_0^{j*}, \varepsilon_s^{j*}\right)$.

This implies

$$
\begin{aligned}
V^i &> \psi^i \left(\kappa y_0^i\right)^{1-\gamma} \left\{ \frac{\left(1-e_0^{j*}\right)^{\gamma-\sigma}}{1-\gamma} + \beta \sum_s p_s \left(e_0^{j*}\right) \frac{\left(\varepsilon_s^{j*}\right)^{1-\gamma}}{1-\gamma} \right\} \\
&= \frac{\psi^i \left(\kappa y_0^i\right)^{1-\gamma}}{\psi^j \left(\kappa y_0^j\right)^{1-\gamma}} V^j.
\end{aligned}
$$

On the other hand, the contract $\left(c_0^{i*} y_0^j / y_0^i, e_0^{i*}, \varepsilon_s^{i*}\right)$ is incentive-feasible for the individual contracting problem V^j. Hence, we obtain

$$
\begin{aligned}
V^j &\geq \psi^j \left(\frac{c_0^{i*} y_0^j}{y_0^i}\right)^{1-\gamma} \left\{ \frac{\left(1-e_0^{i*}\right)^{\gamma-\sigma}}{1-\gamma} + \beta \sum_s p_s \left(e_0^{i*}\right) \frac{\left(\varepsilon_s^{i*}\right)^{1-\gamma}}{1-\gamma} \right\} \\
&= \frac{\psi^j \left(y_0^j\right)^{1-\gamma}}{\psi^i \left(y_0^i\right)^{1-\gamma}} V^i.
\end{aligned}
$$

Taken together, the two inequalities imply $V^i > V^i$, which is a contradiction. **Q.E.D.**

A few remarks are now in order. It should typically be possible to find a vector of Pareto weights $(\psi^i)_i$ such that the postulated individual-specific transfers $t^i = t^* y_0$ are indeed optimal. However, because of potential non-concavitites in the Pareto frontier, it is difficult to formally establish such a result. We abstract from this subtlety and simply take the existence of such Pareto weights as given for our analysis. Intuitively, the Pareto weights ψ^i are determined by income at time 0. This dependence can be seen as coming from past incentive constraints or from type-dependent participation constraints in period zero.

Proposition 8 demonstrates that we can use the consumption and income residuals $\hat{\varepsilon}^i$ and $\hat{\eta}^i$ as inputs for our calibration procedure. In principle, we could go even further and use residual income and consumption growth in our analysis to identify shocks. We have decided not to follow that approach for two reasons. First, it requires imposing further structure on the consumption functions and the income process. Second, and more importantly, measurement error is known to be large for both income and consumption. This problem would be exacerbated by taking growth rates.

B.1 Estimation

In line with the homothetic specification, we normalize $c_0^* = 1$ and set $y_0 = 1/\kappa$, where κ is the ratio of consumption to income in the data. Given the fixed parameters $(\gamma, q, \tilde{q})$ and income levels $\eta_1, ..., \eta_{20}$, the remaining parameters of the model are the preference parameters (σ, β) and the probability weights $\{\pi_s^h, \pi_s^l\}_{s=1}^N$. Since the probabilities π_s^l and π_s^h each sum up to one, we have $N-1$ parameters each. Hence, in total, there are $2N$ remaining parameters.

Our target moments from the data are $p_s(e_0^*) = 1/20$ for all s, where e_0^* is the optimal effort, and $\varepsilon_s^* = \hat{c}(\eta_s)$, where ε_s^* is the optimal consumption innovation in the model with an exogenous capital income tax rate of 40 percent, i.e., with $\tilde{q} = \frac{q}{0.6+0.4q}$. From the definition of probabilities and the optimality conditions for second-period consumption, we obtain the following $2N-1$ model restrictions:

$$\begin{aligned} p_s(e_0^*) &= \exp(-e_0^*)\pi_s^l + (1-\exp(-e_0^*))\pi_s^h \quad \text{for } s = 1, ..., N-1, & (22) \\ \frac{q}{\beta}\lambda^*(\varepsilon_s^*)^\gamma &= 1 + \mu^* \frac{\exp(-e_0^*)\left(\pi_s^h - \pi_s^l\right)}{p_s(e_0^*)} + \xi^* \frac{\gamma}{\varepsilon_s^*} \quad \text{for } s = 1, ..., N. & (23) \end{aligned}$$

Notice that these equations also include the endogenous variables e_0^*, λ^*, μ^* and ξ^*. We simultaneously solve for these endogenous variables and all unknown parameters by adding the following four model restrictions to the system of equations given by (22) and (23). First, we have the Euler equation:

$$\tilde{q}\,(1-e_0^*)^{\gamma-\sigma} = \beta \sum_{s=1}^{N} p_s\,(e_0^*)\,(\varepsilon_s^*)^{-\gamma}. \tag{24}$$

Then, we can use the first-order incentive compatibility constraint for effort,

$$\frac{\sigma-\gamma}{1-\gamma}\,(1-e_0^*)^{\gamma-\sigma-1} = \beta\exp(-e_0^*)\sum_{s=1}^{N}\left(\pi_s^h - \pi_s^l\right)\frac{(\varepsilon_s^*)^{1-\gamma}}{1-\gamma}, \tag{25}$$

the (normalized) first-order condition for c_0^*,

$$\frac{\lambda^*}{\left(1-e_0^*\right)^{\gamma-\sigma}} = 1 + \mu^* \frac{\sigma-\gamma}{(1-e_0^*)} - \xi^*\tilde{q}\gamma, \tag{26}$$

and the planner's first-order optimality condition for effort,

$$\begin{aligned} 0 &= q\lambda^* \sum_s p_s'\left(e_0^*\right)\left(\frac{\eta_s}{\kappa} - \varepsilon_s^*\right) - \xi^* \left(\tilde{q}(\gamma-\sigma)\left(1-e_0^*\right)^{\gamma-\sigma-1} + \beta \sum_i p_s'(e_0^*)\varepsilon_s^{-\gamma}\right) \\ &\quad + \mu^* \left(\beta \sum_s p_s''(e_0^*)\frac{(\varepsilon_s^*)^{1-\gamma}}{1-\gamma} - \frac{(\gamma-\sigma)(\gamma-\sigma-1)}{1-\gamma}\left(1-e_0^*\right)^{\gamma-\sigma-2}\right). \end{aligned} \tag{27}$$

Moreover, we note that the average fraction of working time over total disposable time in the United States is approximately one-third. Using this observation as a further data moment, we obtain exactly as many moments as the number of unknown variables.

Finally, from the government's budget constraint, we obtain the implied present value of total government consumption as

$$G = \frac{1}{\kappa} - 1 + q\sum_{s=1}^{N} p_s(e_0^*)\left(\frac{\eta_s}{\kappa} - \varepsilon_s^*\right), \tag{28}$$

where we have used the unit root process of income and Proposition 8.

B.2 The role of relative risk aversion

As documented by Table 1 in the main text, the discrepancy between second best and third best allocations depends on the coefficient of relative risk aversion. We can get some intuition for why the differences increase in the coefficient of relative risk aversion γ by examining the optimality condition for third best consumption (Eq. 23) for our specification:

$$\frac{q}{\beta}\lambda^*(\varepsilon_s^*)^{\gamma} - \xi^*\frac{\gamma}{\varepsilon_s^*} = 1 + \mu^*\rho\frac{\exp(-\rho e_0^*)\left(\pi_s^h - \pi_s^l\right)}{p_s(e_0^*)} \qquad \text{for } i = 1, \dots, N.$$

The direct effect of limited capital taxation is driven by $\xi^* a(\varepsilon_s^*)$. Note that the higher is γ, the higher is the discrepancy between the Euler equation characterizing the limited capital taxation case and the inverse Euler characterizing the optimal capital taxation case. This will imply that ξ^* is increasing with γ. Moreover, absolute risk aversion is given by $a(\varepsilon_s^*) = \gamma/\varepsilon_s^*$, which is also increasing in γ. Hence, the effect of hidden asset accumulation (or limited capital taxes) is increasing in risk aversion for both these reasons. The larger discrepancy between the Euler and inverse Euler equations also explains that the capital taxes must rise with risk aversion in order to make these two optimality conditions compatible. The same argument explains why the welfare costs of limited capital taxation are increasing in risk aversion.

B.3 General equilibrium (flexible interest rate)

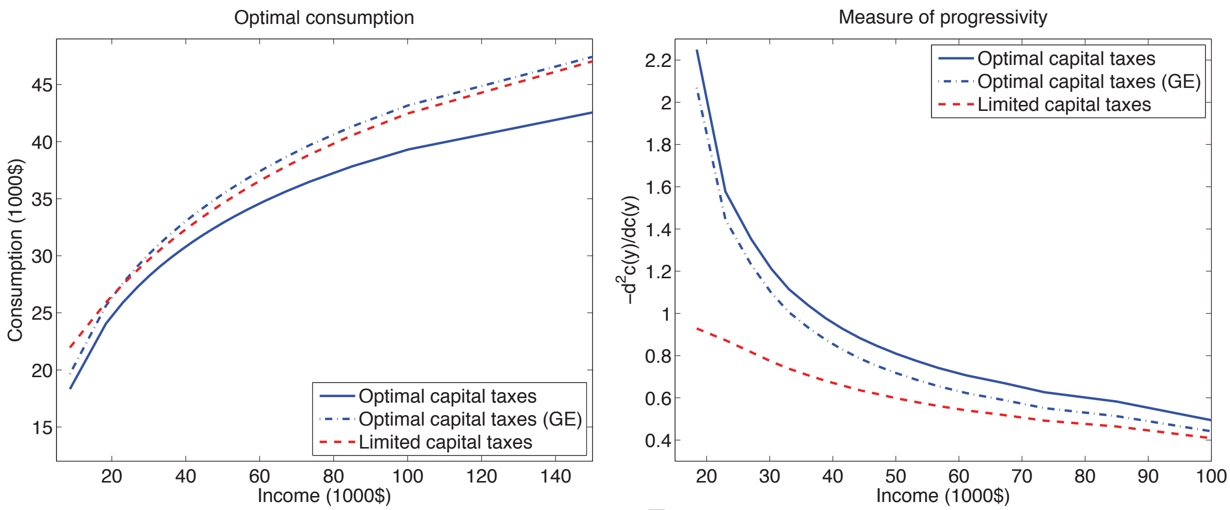
In this subsection, we explore the sensitivity of our results on progressivity to the assumption of a fixed interest rate (partial equilibrium). To relax this assumption, we first decompose the present value of government consumption (28) into separate components for each period:

$$G = G_0 + qG_1 \quad \text{where} \quad G_0 = \frac{1}{\kappa} - 1 \quad \text{and} \quad G_1 = \sum_{s=1}^{N} p_s(e_0^*) \left(\frac{\eta_s}{\kappa} - \varepsilon_s^* \right). \tag{29}$$

This timing of government consumption implies that the scenario with limited capital taxes can be interpreted as a general equilibrium setup, since the aggregate resource constraints are satisfied period by period. Put differently, q is a market clearing price for intertemporal assets such that no intertemporal trade occurs in equilibrium.

In the counterfactual exercise of the main text where we consider optimal capital taxation and observable assets, we impose an *intertemporal* budget constraint and we keep the price q fixed. This implies that the government budget constraints are not satisfied on a period-by-period basis and consumption is frontloaded. In other words, the economy is borrowing in the first period. In this subsection, we abandon the assumption of a fixed price q and we let the intertemporal price adjust such that the resource constraints are satisfied with equality in each period. Given that the planner has an incentive to frontload consumption, it is easy to see that the intertemporal price q needs to decrease (i.e., the interest rate needs to increase) to make borrowing more costly when capital taxation is not limited.

In Figure 3, we plot optimal consumption and the corresponding measure of progressivity for all three specifications (optimal capital taxation with fixed q, flexible q, and limited capital taxation). Given the above arguments, it should not be surprising that the level of consumption has increased as compared to the fixed price case, and is very similar to the consumption level in the scenario with limited capital taxation. However, this change does not affect our results regarding the progressivity of the allocation with optimal capital taxes in a significant way. Although the flexible price case displays somewhat less progressivity than the fixed price case, the allocation with optimal capital taxes and flexible prices is clearly more progressive than the allocation with limited capital taxes for all levels of income. In fact, for the majority of income levels, the progressivity of the allocation with optimal capital taxes is very similar for fixed intertemporal prices and flexible intertemporal prices. The



(a) Consumption

(b) Income tax progressivity

Figure 3: Optimal consumption with optimal and limited capital taxation. Notes: Figure 3a displays constrained efficient allocations in the calibrated models with limited capital taxes (capital income tax rate of 40 percent) and optimal capital taxes (assuming a full observability of capital). Figure 3b shows the associated curvature of consumption, measured as the absolute value of $c''(y)/c'(y)$. In both figures, the solid line displays the model with optimal capital taxes and a fixed intertemporal price, the dash-dotted line represents the model with optimal capital taxes and a flexible intertemporal price, and the dashed line represents the model with limited capital taxes.

average degree of progressivity falls somewhat (from 0.962 to 0.874) when the intertemporal price is flexible, but remains significantly larger than in the allocation with limited capital taxation (0.644).

Research highlights

- We study optimal labor income taxes when capital income taxation is limited
- We use a dynamic moral hazard model with observable and nonobservable assets
- Limited observability of assets reduces the possibility to tax capital
- Optimal labor income taxes become less progressive when tax on capital is limited
- We illustrate the quantitative impact of capital taxation on labor tax progressivity