Optimal income taxation with asset accumulation

Abraham, Arpad and Koehne, Sebastian and Pavoni, Nicola

European University Institute, Institute for International Economic Studies (IIES), Stockholm University, Bocconi University, IGIER, IFS, and CEPR

3 May 2012

Optimal Income Taxation with Asset Accumulation

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

May 3, 2012

Abstract

Several frictions restrict the government's ability to tax assets. First of all, it is very costly to monitor trades on international asset markets. Moreover, agents can resort to non-observable 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 observability have important consequences for the taxation of labor income. Using a dynamic moral hazard model of social insurance, we find that optimal labor income taxes typically become less progressive when assets are imperfectly observed. We evaluate the effect quantitatively in a model calibrated to U.S. data.

Keywords: Optimal Income Taxation, Capital Taxation, Asset Accumulation, Progressivity.
JEL: D82, D86, E21, H21.

1 Introduction

The existence of international asset markets implies that taxation authorities do not have perfect (or low cost) 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)

Even when agents choose not to hide their wealth abroad, they have access to number of non-observable storage technologies at home, both in developed and developing countries. For example, agents can

[*]European University Institute, Florence. E-mail address: Arpad.Abraham@EUI.eu
[†]IIES, Stockholm University. E-mail address: sebastian.koehne@iies.su.se
[‡]Bocconi University, IGIER, IFS, and CEPR. E-mail address: nicola.pavoni@unibocconi.it

accumulate cash, gold, or durable goods. These assets typically bring low returns, but may nonetheless impose important restrictions 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 imperfectly observable. We contrast two stylized environments. In the first one, 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. When absolute risk-aversion is convex, we find that optimal consumption in the scenario with hidden assets moves in a less concave (or more convex) way with labor income. In this sense, the optimal allocation becomes less progressive in that scenario. This finding can be easily rephrased in terms of the progressivity of labor income taxes, since our model allows for a straightforward decentralization: optimal allocations can be implemented by letting agents pay nonlinear taxes on labor income and linear taxes on assets (Gottardi and Pavoni 2011).[1] Our results show that marginal labor income taxes should become less progressive when the government's ability to tax/observe asset holdings is imperfect.

We derive our results in a simple dynamic model of social insurance. A continuum of ex-ante identical agents influence their labor incomes by exerting effort. Labor income realizations are not perfectly controllable and effort is private information. This creates a moral hazard problem. The social planner thus faces a trade-off between insuring agents against idiosyncratic income uncertainty on the one hand and the associated disincentive effects on the other hand. In addition, agents have access to a risk-free asset, which gives them limited means for self-insurance. In this model, the planner wants to distort agents' asset decisions, because asset accumulation provides insurance against the labor income shocks and thereby reduces the incentives to exert effort.[2]

Using the first-order approach (Abraham, Koehne and Pavoni 2011), we can switch from the observable asset case to the scenario with hidden asset accumulation by adding the agent's Euler equation as a constraint to the principal's optimization problem. This constraint crucially changes the allocation of consumption across income states. 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. Due to the Euler equation, it becomes important how such changes in utility affect the agent's *marginal* utility. One can show that a marginal increase of utility in a state with consumption c reduces the agent's marginal utility in that state by $-u''(c)/u'(c)$.[3] This relaxes the Euler equation and thereby modifies how the gains of allocating utility vary in the cross-section. Obviously, the Euler equation affects the

[1] In the scenario with hidden assets, the tax rate on assets is zero, of course.

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

[3] To increase $u(c)$ by ε, c has to be increased by $\varepsilon/u'(c)$. Using a first-order approximation, this changes the agent's marginal utility by $u'(c) - u'(c + \varepsilon/u'(c)) \approx -\varepsilon u''(c)/u'(c)$.

costs and benefits of allocating utility also by changing the shadow costs of the remaining constraints of the principal's problem. However, we show that the former effect is key. If absolute risk-aversion is convex, we thus find that optimal consumption becomes a more convex function of labor income when asset accumulation is not observable. Put differently, marginal taxes on labor income become less progressive when asset income cannot be sufficiently taxed.

Intuitively, imperfect observability of assets entails that the planner can rely less on consumption frontloading (Rogerson 1985, Golosov, Kocherlakota and Tsyvinski 2003) to create incentives. As a consequence, the cross-sectional structure of consumption needs to be modified. Notice, however, that there are two distinct ways of creating stronger incentives using the consumption cross-section. One possibility is to increase the rewards for high performance, so that consumption becomes a more convex function of income, loosely speaking. The second option is to impose more severe punishments for low income realizations, which makes consumption more concave instead. We find that the curvature of the agent's coefficient of absolute risk aversion determines which of the two possibilities dominates. Clearly, optimal incentives are shaped by the planner's costs and benefits of allocating utility across income states, and these costs and benefits include a component that relates to marginal changes in the agent's Euler equation. As explained above, this implies that the curvature of absolute risk aversion determines how asset accumulation changes the optimal incentive scheme.

In a quantitative exercise, we estimate some of the key parameters of the model. We use consumption and income data from the PSID (Panel Study of Income Dynamics) as adapted by Blundell, Pistaferri and Preston (2008) and postulate that the data is generated by a tax system in which labor income taxes are set optimally given an asset income tax rate of 40%.[4] Using the implied parameters, we compute the optimal allocation when asset income taxation is unrestricted and compare it to the data. Under unrestricted asset taxation, the progressivity of the optimal allocation increases sizably. The welfare gain of unrestricted asset taxation varies with the coefficient of relative risk aversion and amounts to 1.3% in consumption equivalent terms for our benchmark calibration. The required asset income tax rates are implausibly high, however, being close to one hundred per cent or above for all specifications. This suggests that imperfect asset observability/taxability is the empirically relevant case for the United States.

To the best of our knowledge, this is the first paper that explores optimal income taxation in a framework where assets are imperfectly observable. Recent work on dynamic Mirrleesian economies analyzes optimal income taxes when assets are observable/taxable without frictions; see Golosov, Troshkin, and Tsyvinski (2011), and Farhi and Werning (2011). In those works, the reason for asset

[4] 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).

taxation is very similar to our model and stems from disincentive effects associated with the accumulation of wealth. While the Mirrlees (1971) framework focuses on redistribution in a population with heterogeneous skills that are exogenously distributed, our approach highlights the social insurance (or ex-post redistribution) aspect of income taxation. In spirit, our model is therefore closer to the works by Varian (1980) and Eaton and Rosen (1980). With respect to the nonobservability of assets, our model is related to Golosov and Tsyvinski (2007), who analyze capital subsidies/distortions in a dynamic Mirrleesian economy with private insurance markets and hidden asset trades.

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 time-varying labor supply elasticities and borrowing constraints, they argue that capital income taxes and progressive labor income taxes are two alternative ways of mimicking age-dependent taxation. They then use numerical methods to determine the efficient relation between the two instruments. Interestingly, in the present environment capital taxes play an entirely different role and we obtain very different conclusions. While in Conesa, Kitao and Krueger (2009) capital income taxes and progressive labor income taxes are substitutable instruments, in our model they are complements. Laroque (2010) derives analytically 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 re-distributional target. The role of capital taxes is to facilitate the use of such re-distributional instrument in the presence of informational asymmetries. Hence we obtain a complementarity between capital taxes and labor income tax progressivity.

Finally, our paper is 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), among other things (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. The present paper emphasizes the link between income tax progressivity and the availability of savings technologies.

The paper proceeds as follows: Section 2 describes the setup of the model. Section 3 presents the main result of the paper: hidden asset accumulation makes optimal consumption schemes less progressive. In Section 4, we explore alternative concepts of concavity/progressivity. Section 5 explores the quantitative importance of our results, while Section 6 concludes and considers a couple of extensions to the model.

2 Model

Consider a benevolent social planner (the principal) whose objective is to maximize the welfare of its citizens. The (small open) economy consists of a continuum of ex-ante identical agents who live for two periods, $t = 0, 1$, and can influence their date-1 labor income realizations by exerting effort. The planner designs an allocation to insure them against idiosyncratic risk and provide them appropriate incentives for working hard. 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, 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 v is strictly increasing and convex. The agent's discount factor is denoted by $\beta > 0$.

Technology and endowments The technological process can be seen as the production of human capital through costly effort, where human capital represents any characteristic that determines the agent's productivity. At date $t = 0$, the agent has a *fixed* endowment y_0. At date $t = 1$, the agent has a stochastic income $y \in Y := [\underline{y}, \overline{y}]$. 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 $e_0 \geq 0$. 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 for 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 income realization y. An allocation $(c_0, c(\cdot), e_0)$ is called *feasible* if it satisfies the planner's *budget constraint*

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

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

Second best The agent's savings technology is observable (and contractable) for the planner. Hence, without loss of generality, we can assume that the planner directly controls consumption. A *second best* allocation is an allocation that maximizes ex-ante welfare[5]

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

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

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

2.1 Decentralization and the first-order approach

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. 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 and the cumulative distribution function of income is log-convex in effort.[6]

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.

[5] Although for pure notational simplicity we consider the case with a continuum of output levels, we do not discuss in detail the technicalities related to the existence of a solution and the existence of the multipliers in infinite dimensional spaces. We can provide details; alternatively, the reader can read the model as one with a large but finite number of output levels.

[6] As discussed by Abraham, Koehne, and Pavoni (2011), both conditions have quite broad empirical support. First, virtually all estimations of u reveal NIARA; see Guiso and Paiella (2008) for example. The condition on the distribution function essentially restricts 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.

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*

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

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

The above result is intuitive and the proof is omitted (compare Gottardi and Pavoni (2011)). 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 becomes apparent once we replace the incentive constraint (4) by the associated first-order conditions

$$
\begin{aligned}
u'_e(y_0 + \tau_0, e_0) + \beta \int_{\underline{y}}^{\overline{y}} u_1(y + \tau(y)) f_e(y, e_0)\, dy &\geq 0, \qquad (5) \\
\tilde{q} u'_c(y_0 + \tau_0, e_0) - \beta \int_{\underline{y}}^{\overline{y}} u'_1(y + \tau(y)) f(y, e_0)\, dy &\geq 0. \qquad (6)
\end{aligned}
$$

The second condition (6) determines the agent's asset decision exclusively based on consumption levels and the price $\tilde{q}$. This means that the planner can essentially ignore the problem of joint deviations when taxing asset trades. It is now clear that by choosing a sufficiently large value for $\tilde{q}$, the planner can in fact ignore this last constraint and obtain the second best allocation.

Notice that we have normalized 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 of course not surprising, because the timing of tax collection is irrelevant by Ricardian equivalence.

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]

$$
\begin{aligned}
\frac{\lambda}{u'_c(c_0, e_0)} &= 1 + \mu \frac{u''_{ec}(c_0, e_0)}{u'_c(c_0, e_0)}, \qquad (7) \\
\frac{\lambda q}{\beta u'_1(c(y))} &= 1 + \mu \frac{f_e(y, e_0)}{f(y, e_0)}, \quad y \in [\underline{y}, \overline{y}], \qquad (8)
\end{aligned}
$$

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

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.

Finally, we note that a tax on the asset price q is equivalent to a tax rate t on the rate of return (constant across agents) given by: $\left(1+\left(\frac{1}{q}-1\right)(1-t)\right)^{-1}=\tilde{q}$.

2.2 Hidden assets and third best allocations

While savings technologies such as domestic bank accounts, pension funds, or houses may be observable at moderate costs, 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 focus on a particular decentralization mechanism in this paper, the above problem is general. Decentralizations that allow asset taxes to depend on the agent's period-1 income realization (Kocherlakota 2005), for instance, can generate zero asset taxes *on average*, but generally require high tax rates for a sizable part of the population.[8]

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

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

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

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

[8] For example, assuming additively separable preferences and CRRA consumption utility, the tax rate on asset holdings in such a decentralization would be $1-\frac{q}{\beta}\left(\frac{c(y)}{c_0}\right)^{\sigma}$, where σ is the coefficient of relative risk aversion. For incentive reasons, $c(y)$ tends to be significantly below c_0 for a range of income levels y, which results in tax rates on assets close to 1 at those income levels. In other words, almost their entire wealth (not just asset income) would be taxed away for those agents.

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

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

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

Obviously, in our terminology the notion 'second best' refers to constrained efficient allocations subject to nonobservability of effort, while the term 'third best' refers to constrained efficient allocations subject to nonobservability of effort and assets/consumption

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 rate q^n, the resulting allocation will obviously be $(c_0, c(\cdot), e_0)$.

Again we can use the FOA to characterize the consumption scheme. Assuming interiority, the first-order conditions of the Lagrangian with respect to consumption are now:

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

$$\frac{\lambda q}{\beta u_1'(c(y))} \quad = \quad 1 + \mu \frac{f_e(y, e_0)}{f(y, e_0)} + \xi a(c(y)), \quad y \in [\underline{y}, \overline{y}], \tag{13}$$

where 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 $q^n < \bar{q}$.*

Proof. Fix q^n. From the Kuhn-Tucker theorem we have $\xi \geq 0$. If $\xi > 0$, we are done. If $\xi = 0$, then 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, since the agent cannot be fully insured when effort is interior. Hence - since from the previous condition we have $\lambda > 0$ - we conclude

$$\beta \int_{\underline{y}}^{\overline{y}} u_1'(c(y)) f(y,e_0)\,dy > q u_c'(c_0,e_0). \tag{14}$$

Clearly, exactly the same allocation delivering condition (14) 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.**

Proposition 2 states that if the return on the nonobservable savings technology $\frac{1}{q^n}$ 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 the particular choice of q^n and rely only on the fact the Euler equation is binding for the planner in that case.[9]

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.

3 Absolute progressivity and linear likelihoods

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 decentralizations.

[9] The quantitative analysis in Section 5 suggests that a binding Euler equation is indeed the empirically relevant case.

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 and labor income 'wedge', hence $-\tau(y)$ 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 faster than income does. At the same time, for the states when the agent is receiving a transfer, transfers are increasing slower 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 'large penalties' for low income realizations, since consumption decreases relatively quickly when income decreases. Regressive schemes, by contrast, put more emphasis on rewards for high income levels than punishments for low income levels.

The next proposition provides sufficient conditions for progressivity and regressivity of efficient allocations.

Proposition 3 (Sufficient conditions for progressivity/regressivity) *Assume that the FOA is justified and that second best 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*, then second best allocations are progressive. If, in addition, absolute risk aversion* $a(c)$ *is decreasing and concave, then 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*, then second best allocations are regressive. If, in addition, absolute risk aversion* $a(c)$ *is decreasing and convex, then third best allocations are regressive as well.*

Proof. 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 concavity of u, $\frac{1}{u_1'(\cdot)}$ is always 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:

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

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 $c(\cdot)$ is concave since it is an increasing and concave transformation of a concave function. **Q.E.D.**

Note that in the previous proposition, since the function g is increasing, consumption is increasing as long as 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.[10] 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 $\frac{1}{u_1'(c)} = c$ is both concave and convex. Interestingly, third best schemes are *regressive* in this case (since absolute risk aversion $a(c) = \frac{1}{c}$ is convex).[11]

This particular finding sheds light on a more general pattern under 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. The next result formalizes this insight.

Proposition 4 (Concavity) *Assume that the FOA is justified and let $(c_0, c(\cdot), e_0)$ be an interior, monotonic second best allocation and $(\hat{c}_0, \hat{c}(\cdot), e_0)$ be an interior, monotonic third best allocation, both implementing effort level e_0. Suppose that u_1 has convex absolute risk aversion and that the likelihood ratio $l(y, e_0)$ is linear in y. Under these conditions, if $\hat{c}$ is progressive, then c is as well.*

Proof. Given validity of the FOA, by equations (8) and (13) the consumption schemes $c(y)$ and $\hat{c}(y)$ are characterized as follows:

$$g_{\lambda}(c(y)) = 1 + \mu\, l(y, e_0), \text{ where } g_{\lambda}(c) := \frac{\lambda q}{\beta u_1'(c)}, \tag{15}$$

$$\hat{g}_{\hat{\lambda},\hat{\xi}}(\hat{c}(y)) = 1 + \hat{\mu}\, l(y, e_0), \text{ where } \hat{g}_{\hat{\lambda},\hat{\xi}}(c) := \frac{\hat{\lambda} q}{\beta u_1'(c)} - \hat{\xi} a(c), \text{ with } \hat{\xi} > 0. \tag{16}$$

Since $l(y,e)$ is linear in y by assumption, concavity of $\hat{c}$ is equivalent to convexity of $\hat{g}_{\hat{\lambda},\hat{\xi}}$. Moreover, since $a(c)$ is convex in c by assumption, convexity of $\hat{g}_{\hat{\lambda},\hat{\xi}}$ implies convexity of $g_{\lambda} = \frac{\lambda}{\hat{\lambda}}\left(\hat{g}_{\hat{\lambda},\hat{\xi}} + \hat{\xi} a\right)$. Finally, notice that convexity of g_{λ} is equivalent to concavity of c, since $l(y,e)$ is linear in y. **Q.E.D.**

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

[11] 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.

In order to obtain a clearer intuition of this result, 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 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 of all, since the agent's utility is increased by one unit, there is a return of 1. Furthermore, increasing the agent's utility also relaxes the incentive constraint for effort, generating a return of $\mu \frac{f_e(y,e_0)}{f(y,e_0)}$.[12] Finally, by increasing $u_1(c(y))$ the planner alleviates the saving motive of the agent. This gain, measured by $\xi a(c(y))$, depends crucially on the multiplier ξ of the agent's Euler equation. When assets can be fully taxed (second best), we have $\xi = 0$ and this gain vanishes. By lowering the net return of the asset, the planner is able to circumvent the first-order incentive constraint for assets. However, when asset taxation is ruled out (third best), this constraint is binding and we have $\xi > 0$. Under convex absolute risk aversion, the term $\xi a(c(y))$ is convex. This implies that, ceteris paribus, the benefits of increasing the agent's utility change in a more convex way with labor income. As a consequence, in the third best the agent's utility must also change in a more convex way with labor income, hence consumption becomes more convex in y in this case.

A closely related intuition for equation (13) can be obtained by rewriting it as follows:

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

On the right-hand side, we have the (rescaled) likelihood ratio. As in the static moral hazard problem, this function governs the allocation of utility across income states y. The only change compared to the static problem is the term $\xi a(c(y))$ on the left-hand side. This term stems from the agent's Euler equation and modifies the planner's costs of allocating utility across states. In the static model, allocating utility only generates a direct resource cost to the planner. This cost, captured by the discounted inverse marginal utility, is also present here. In addition, allocating utility to state y affects the intertemporal structure of the consumption scheme, which creates an additional cost due to the agent's Euler equation.

[12] Of course, 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.

4 General results on progressivity

Since at least Holmstrom (1979), it is well known that consumption patterns under moral hazard are crucially influenced by the shape of the likelihood ratio function $l(\cdot, e)$. Stated in more negative terms, one can always find functions $l(\cdot, e)$ so that the shape of consumption is almost arbitrary. To make the impact of asset observability on the shape of optimal consumption easier to observe, we have therefore normalized the curvature of the likelihood ratio by assuming linearity in Proposition 4.

In this section, we study how the observability of assets changes the curvature of the consumption scheme for arbitrary likelihood ratio functions. As usual, we assume that the FOA is justified and that $(c_0, c(\cdot), e_0)$ and $(\hat{c}_0, \hat{c}(\cdot), e_0)$ are interior, monotonic second best and third best allocations, respectively, implementing the same effort level e_0.

Probably the most well known ranking in terms of concavity in economics is that dictated by concave transformations (e.g., Gollier 2001).

Definition 2 We say that f_1 is a concave (convex) transformation of f_2 if there is an increasing and concave (convex) function v such that $f_1 = v \circ f_2$.

Proposition 5 *Assume that u_1 has convex absolute risk aversion. Then, if $\hat{c}$ is a concave transformation of l, then c is a concave transformation of l. Conversely, if c is a convex transformation of l, then $\hat{c}$ has the same property.*

Proof. Recall that we have

$$g_\lambda\left(c(y)\right) = 1 + \mu\, l\left(y, e_0\right), \tag{17}$$

$$\hat{g}_{\hat{\lambda},\hat{\xi}}\left(\hat{c}(y)\right) = 1 + \hat{\mu}\, l\left(y, e_0\right), \tag{18}$$

where the functions g_λ and $\hat{g}_{\hat{\lambda},\hat{\xi}}$ are defined as in (15) and (16), respectively. First, suppose that $\hat{c}$ is a concave transformation of l. Since the right-hand side of (18) is a positive affine transformation of l, this implies that $\hat{g}_{\hat{\lambda},\hat{\xi}}$ is convex. Now, notice that convexity of $\hat{g}_{\hat{\lambda},\hat{\xi}}$ implies that $g_\lambda\left(c\right) = \frac{\lambda}{\hat{\lambda}}\left(\hat{g}_{\hat{\lambda},\hat{\xi}}(c) + \hat{\xi} a(c)\right)$ is convex as well (since $a(c)$ is convex by assumption). Hence, using (17), we see that c is a concave transformation of l.

Conversely, suppose that c is a convex transformation of l. Using (17), we see that g_λ is then concave. Convexity of $a(c)$ implies that $\hat{g}_{\hat{\lambda},\hat{\xi}}$ is then also concave, which shows that $\hat{c}$ is a convex transformation of l. **Q.E.D.**

The previous result clearly generates a sense in which c is 'more progressive' than $\hat{c}$. Note that this finding generalizes Proposition 4 to arbitrary shapes of the likelihood ratio function l. As a drawback, we can rank the curvature of c and $\hat{c}$ only when, for example, $\hat{c}$ is a concave transformation of l. We will now reduce the set of possible utility functions to facilitate such comparisons.

Let us consider the class of HARA (or linear risk tolerance) utility functions, namely

$$u_1(c) = \rho \left(\eta + \frac{c}{\gamma} \right)^{1-\gamma}$$

$$\text{with } \rho \frac{1-\gamma}{\gamma} > 0, \text{ and } \eta + \frac{c}{\gamma} > 0.$$

For this class, we have $a(c) = \left(\eta + \frac{c}{\gamma} \right)^{-1}$. Hence, absolute risk aversion is convex. Special cases of the HARA class are CRRA, CARA, and quadratic utility (e.g., see Gollier 2001).

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

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

Then, if u_1 *belongs to the HARA class with* $\gamma \geq -1$, *then* $\hat{g}_{\hat{\lambda},\hat{\xi}}$ *is a concave transformation of* g_λ *for all* $\hat{\lambda}, \hat{\xi} \geq 0$, $\lambda > 0$.

Proof. If u belongs to the HARA class, we obtain

$$\hat{g}_{\hat{\lambda},\hat{\xi}}(c) = \frac{\hat{\lambda}}{\lambda} g_\lambda(c) - \hat{\xi} a(c) = \frac{\hat{\lambda}}{\lambda} g_\lambda(c) - \hat{\xi} \lambda^{\frac{1}{\gamma}} \kappa \left(g_\lambda(c) \right)^{-\frac{1}{\gamma}}, \text{ with } \kappa = \left[\frac{\gamma q}{\beta \rho (1-\gamma)} \right]^{\frac{1}{\gamma}} > 0.$$

In other words, we have

$$\hat{g}_{\hat{\lambda},\hat{\xi}}(c) = h\left(g_\lambda(c) \right), \text{ where } h(g) = \frac{\hat{\lambda}}{\lambda} g - \hat{\xi} \lambda^{\frac{1}{\gamma}} \kappa \, g^{-\frac{1}{\gamma}}.$$

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

The restriction $\gamma \geq -1$ in the above result is innocuous to most applications and it allows for all HARA functions with nonincreasing absolute risk aversion as well as quadratic utility, for instance.

Recall that second best and third best consumption schemes are characterized as follows:

$$\begin{aligned} g_\lambda(c(y)) &= 1 + \mu \, l(y, e_0), \\ \hat{g}_{\hat{\lambda},\hat{\xi}}(\hat{c}(y)) &= 1 + \hat{\mu} \, l(y, e_0). \end{aligned}$$

For logarithmic utility, g_λ is linear. Lemma 1 therefore has the following consequence.

Corollary *Suppose u_1 is logarithmic. Then c is a concave transformation of $\hat{c}$.*

Proof. By Lemma 1, there exists a concave function $\tilde{h}$ such that c and $\hat{c}$ are related as follows:

$$c(y) = \tilde{g}^{-1} \circ \tilde{h} \circ \tilde{g}\left(\hat{c}(y)\right),$$

where $\tilde{g}(c) = \frac{1}{\mu}\left(\frac{\lambda q}{u'(c)} - 1\right)$ is increasing. For logarithmic utility, $\tilde{g}$ is an affine function, which implies that the composition $\tilde{g}^{-1} \circ \tilde{h} \circ \tilde{g}$ is concave whenever $\tilde{h}$ is concave. **Q.E.D.**

To state the consequences of Lemma 1 for general HARA functions, we introduce the concept of G-convexity (e.g., see Avriel et al., 1988), which is widely used in optimization. A function f is G-convex if once we transform f with G we get a convex function. More formally:

Definition 3 Let f be a function and G an increasing function mapping from the image of f to the real numbers. The function f is called *G-convex (G-concave)* if $G \circ f$ is a convex (concave) function.

This concept generalizes the standard notion of convexity. It is easy to see that a function f is convex if and only if it is G-convex for any increasing affine function G. Moreover, it can be shown that if G is concave and f is G-convex then f must be convex, but the converse is false.[13]

Lemma 2 *Assume u_1 belongs to the HARA class with $\gamma \geq -1$. Then c is g_λ-convex (g_λ-concave) if and only if $\hat{c}$ is $\hat{g}_{\hat{\lambda},\hat{\xi}}$-convex ($\hat{g}_{\hat{\lambda},\hat{\xi}}$-concave).*[14]

Proof. Recall that consumption is determined as follows:

$$\begin{aligned} g_\lambda\left(c(y)\right) &= 1 + \mu\, l\left(y, e_0\right), \\ \hat{g}_{\hat{\lambda},\hat{\xi}}\left(\hat{c}(y)\right) &= 1 + \hat{\mu}\, l\left(y, e_0\right). \end{aligned}$$

As a consequence, we can relate the two consumption functions as follows:

$$\frac{1}{\mu}\Big(g_\lambda\left(c(y)\right) - 1\Big) = \frac{1}{\hat{\mu}}\Big(\hat{g}_{\hat{\lambda},\hat{\xi}}\left(\hat{c}(y)\right) - 1\Big). \tag{19}$$

Now the result follows from the simple fact that convexity/concavity is preserved under positive affine transformations. **Q.E.D.**

[13] For example, suppose $f(x) = x^2$ and $G(\cdot) = \log(\cdot)$, then $G(f(x)) = 2\log(x)$, which is obviously not convex.

[14] In fact, this statement is not only true for concavity and convexity, but more generally for any property defined with respect to the transformations g_λ and $\hat{g}_{\hat{\lambda},\hat{\xi}}$.

Proposition 6 *Assume u_1 belongs to the HARA class with $\gamma \geq -1$. If $\hat{c}$ is g_λ-concave then c is g_λ-concave. Conversely, if c is g_λ-convex then $\hat{c}$ is g_λ-convex.*

Proof. Let $\hat{c}$ be g_λ-concave. By Lemma 1, we have $\hat{g}_{\hat{\lambda},\hat{\xi}} = h \circ g_\lambda$ for some increasing and concave function h. Hence, when $\hat{c}$ is g_λ-concave, then $\hat{c}$ must also be $\hat{g}_{\hat{\lambda},\hat{\xi}}$-concave. Now Lemma 2 implies that c is g_λ-concave.

To verify the second statement, let c be g_λ-convex. From Lemma 2, we see that $\hat{c}$ is $\hat{g}_{\hat{\lambda},\hat{\xi}}$-convex, i.e., $\hat{g}_{\hat{\lambda},\hat{\xi}} \circ \hat{c}$ is convex. By Lemma 1, we have $\hat{g}_{\hat{\lambda},\hat{\xi}} = h \circ g_\lambda$ for some increasing and concave function h. Since the inverse of h must be convex, we conclude that $g_\lambda \circ \hat{c} = h^{-1} \circ \hat{g}_{\hat{\lambda},\hat{\xi}} \circ \hat{c}$ is convex. **Q.E.D.**

Proposition 6 shows that whenever $\hat{c}$ satisfies the g_λ-concavity property, then c satisfies this property. In this sense, we note again that c is 'more progressive' than $\hat{c}$.

5 Quantitative analysis

This quantitative exercise serves two purposes. First, we extend our theoretical results. For example, recall that the theoretical results compare two allocations that implement the same effort level. In a calibrated/estimated framework we show that the key result of complementarity between capital taxation and labor income tax progressivity extends to the case where effort is allowed to change between the two scenarios.

The second target of this exercise is to evaluate *quantitatively* how the limited possibility of observing/taxing capital affects optimal labor income taxes. In order to do this, 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 distorted asset price is given by $\tilde{q} = \frac{q}{0.6+0.4q}$. Note that the capital income tax of 40% is in line with U.S. effective tax rates on capital income as calculated by Mendoza, Razin and Tesar (1994) and Domeij and Heathcote (2004). We estimate some of 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 (and postulated) parameters and also solve the model with optimal capital taxes, assuming perfect observability/taxability of capital. The final outcome is a comparison of the optimal labor income taxes between the two scenarios.

5.1 Data

We use PSID (Panel Study of Income Dynamics) data for 1992 as adapted by Blundell, Pistaferri and Preston (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 will use two measures of consumption: non-durable consumption expenditure and total consumption expenditure, the latter being our benchmark case.

In our model, we have ex-ante identical individuals who face the same (partially endogenous) process of income shocks. In the data, however, income is influenced by observable factors such as age, education and race. We want to control for these characteristics to make income shocks comparable across individuals. To do this, 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 residual of this equation $\hat{\eta}^i$ is our estimate of the income shock.

The next objective is to find the consumption function. To be able to relate it to the cleaned income measure η^i, we postulate that the consumption function is multiplicatively separable as well:

$$c^i = g^0(Z^i)g^1\left(\phi(X^i)\right)c\left(\eta^i\right)$$

where Z^i are household characteristics that affect consumption, but (by assumption) do not affect income, such as number of kids and beginning of period household assets. Our target is to identify $c(\eta)$, the pure response of consumption to the income shock. To isolate this effect, we first run separate regression of $\log(c^i)$ on X^i and Z^i. The residual of this equation is $\hat{\varepsilon}^i$. We then use a flexible functional form to obtain $c(\cdot)$. 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 our model's notation, the estimate of the consumption function is given by

$$\hat{c}(y) = \exp\left(\sum_{j=0}^{4} \hat{\gamma}_j \left(\log(y)\right)^j\right).$$

Figure 1 displays the estimated consumption function for both of our measures of consumption. Note that our estimate based on total consumption expenditure displays both significantly more dispersion and a higher overall level.

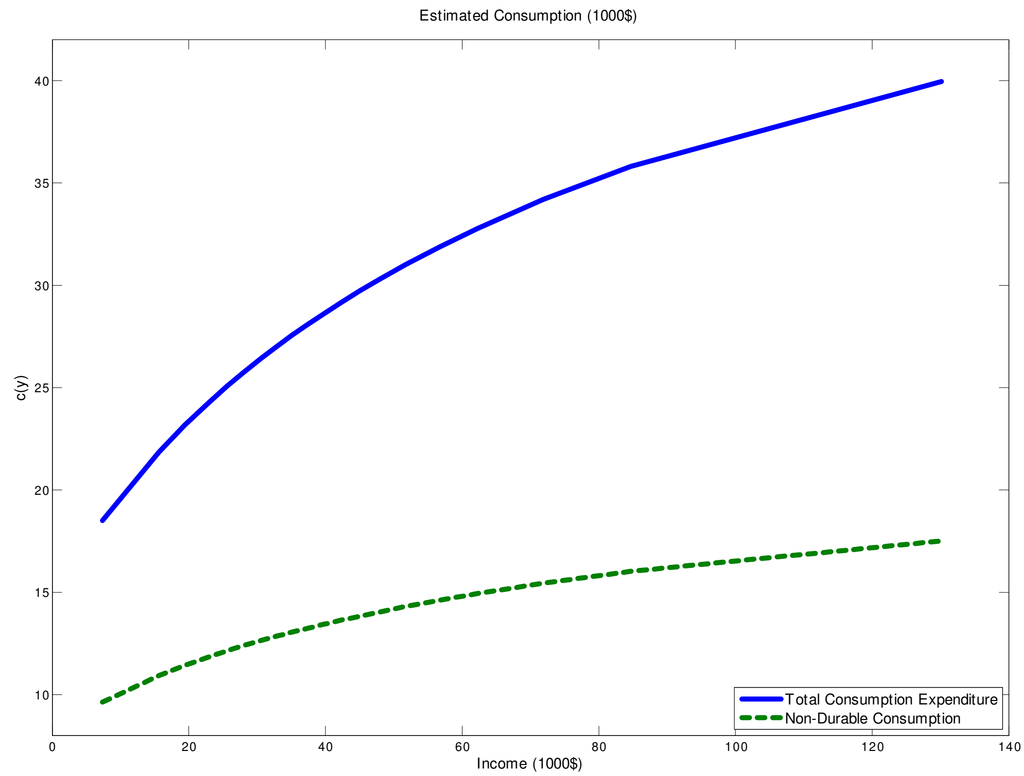

Figure 1: Estimated Consumption Functions

5.2 The empirical specification of the model

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

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

where π^h and π^l are probability distributions on the set $\{y_1, \ldots, y_N\}$ and ρ is a positive scalar. In addition to tractability, this formulation has the advantage that it satisfies the requirements for the applicability of first-order approach given by Abraham, Koehne and Pavoni (2011).[15]

[15] Note that we do not need to impose the *stochastic dominance condition* - which, in our environment, is virtually equivalent to monotone likelihood ratios (MLR) - as in the proof of the validity of the first order approach we only need monotone consumption (see Abraham, Koehne and Pavoni (2011) for details). And as Figure 1 shows this is delivered to us from the data. Note that MLR is a sufficient but not necessary condition for monotone consumption. Nevertheless, as

In order to account for (multiplicative) 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:

$$u(c,e) = \frac{\left[(c)^{\alpha}(v(T-e))^{1-\alpha}\right]^{1-\sigma}}{\alpha(1-\sigma)},$$

where v is a concave function, $\alpha \in (0,1)$ and $\sigma > 0$.[16]

Proposition 7 *Consider the following family of homothetic models with heterogeneous agents:*

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

s.t.

$$\begin{aligned}
\sum_i \left(y_0^i - c_0^i\right) + q \sum_i \sum_s p_s\left(e_0^i\right)\left[y_s^i - c_s^i\right] &\geq G; \\
-\frac{1-\alpha}{\alpha}\frac{v'\left(T-e_0^i\right)}{v\left(T-e_0^i\right)}\left[(c_0^i)^{\alpha}\left(v\left(T-e_0^i\right)\right)^{1-\alpha}\right]^{1-\sigma} &= \beta \sum_s p_s'\left(e_0^i\right)\frac{\left[(c_s^i)^{\alpha}(v(T))^{1-\alpha}\right]^{1-\sigma}}{\alpha(1-\sigma)}; \\
\tilde{q}\frac{\left[(c_0^i)^{\alpha}\left(v\left(T-e_0^i\right)\right)^{1-\alpha}\right]^{1-\sigma}}{c_0^i} &= \beta \sum_s p_s\left(e_0^i\right)\frac{\left[(c_s^i)^{\alpha}(v(T))^{1-\alpha}\right]^{1-\sigma}}{c_s^i};
\end{aligned}$$

with $\beta \in (0,1)$, and $\tilde{q}, q > 0$. Moreover, assume 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 $\gamma > 0$, let the Pareto weights $(\psi^i)_i$ be such that the solution to the above problem delivers period zero consumption $c_0^{*i} = \gamma y_0^i$ for all i. Then there exists $t^* \in R$ 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

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

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

$$\max_{\varepsilon_s, e_0} \frac{\left[(v(T-e_0))^{1-\alpha}\right]^{1-\sigma}}{\alpha(1-\sigma)} + \beta \sum_s p_s(e_0) \frac{\left[(\varepsilon_s)^{\alpha}(v(T))^{1-\alpha}\right]^{1-\sigma}}{\alpha(1-\sigma)};$$

expected, our estimated likelihood ratios will exhibit MLR, that is the estimated probability distributions satisfy: π_s^h/π_s^l increasing in s.

[16] Where, obviously, when $\sigma = 1$ we assume preferences take a logarithmic form.

$s.t$

$$\begin{aligned}
\frac{1}{\gamma}-1+q\sum_{s}p_{s}\left(e_{0}\right)\left[\frac{\eta_{s}}{\gamma}-\varepsilon_{s}\right] &\geq t^{*}; \\
-\frac{(1-\alpha)}{\alpha}\frac{v'\left(T-e_{0}\right)}{v\left(T-e_{0}\right)}\left[\left(v\left(T-e_{0}\right)\right)^{1-\alpha}\right]^{1-\sigma} &= \beta\sum_{s}p'_{s}\left(e_{0}\right)\frac{\left[\left(\varepsilon_{s}\right)^{\alpha}\left(v\left(T\right)\right)^{1-\alpha}\right]^{1-\sigma}}{\alpha\left(1-\sigma\right)}; \\
\tilde{q}\left[\left(v\left(T-e_{0}\right)\right)^{1-\alpha}\right]^{1-\sigma} &= \beta\sum_{s}p_{s}\left(e_{0}\right)\frac{\left[\left(\varepsilon_{s}\right)^{\alpha}\left(v\left(T\right)\right)^{1-\alpha}\right]^{1-\sigma}}{\varepsilon_{s}}.
\end{aligned}$$

Proof. See Appendix. **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 establish such a result formally. 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 due to type-dependent participation constraints in period zero.

Proposition 7 is useful for our empirical strategy for at least two main reasons. First, the proposition suggests that within our empirical model, we are entitled to use the income and consumption residuals as computed in the previous section as inputs in our estimation/calibration exercise. More precisely, the proposition suggests that we can use the values $\hat{\varepsilon}^i$ and $\hat{\eta}^i$ as consumption inputs regardless of the actual value of c^i and y^i. In principle, according this proposition, 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 on the income process. Second, and more importantly, measurement error is known to be large for both income and consumption. This would be largely exacerbated by taking growth rates.

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^*.[17]

5.3 Estimation of model parameters

As a first step, we fix some parameters. First of all, we set $q = .96$ to match a yearly real interest rate of 4%, which is the historical average of return on real assets in the USA. We then set the coefficient

[17] Of course, this also implies that we will partially rely on functional forms for identification.

of relative risk-aversion for consumption to 3, that is $1-(1-\sigma)\,\alpha=3$, in line with recent estimation results by Paravisini, Rappoport, and Ravina (2010).[18] We normalize total time endowment to one ($T=1$) and choose v to be the identity function. For the income process, we set $N=20$ and choose the medians of the 20 percentile groups of cleaned income for the income levels $\eta_1,...,\eta_{20}$. To be consistent with this choice and with Proposition 7 we set $y_0=1$. For expositional simplicity we will assume $\gamma=1$ and hence $c_0^*=y_0$. Note that Proposition 7 implies that for any level of γ we can obtain the optimal consumption allocation by simply rescaling the consumption allocation of this benchmark. The only parameter we need to adjust is t^* or equivalently government consumption G^*.

Given this choice of parameters, the remaining parameters are chosen to match specific empirical moments coming from the data. We use the optimality conditions to design a method of moments estimator for these parameters. We use the identity matrix as a weighting matrix in the estimation.[19]

The first group of remaining parameters of the model are the effort technology parameter ρ and the probability weights $\left\{\pi_s^h,\pi_s^l\right\}_{s=1}^N$ that determine the likelihood ratios. Our target moments for these parameters are $p_s(e_0^*)=1/20$ for all s, where e_0^* is the optimal effort, and $\varepsilon_s^*=\hat{c}\left(\eta_s\right)$, where ε_s^* is the optimal consumption innovation in the model with an exogenous capital income tax rate of 40 per cent, i.e., with $\tilde{q}=\frac{q}{0.6+0.4q}$.

Since the probabilities π_s^l and π_s^h each sum up to one, we have $N-1$ parameters each. Moreover, we have to estimate the parameter ρ. To summarize, we have to estimate $2N-1$ parameters and use the following $2N-1$ model restrictions for these parameters:

$$\begin{aligned} p_s(e_0^*) &= \exp(-\rho e_0^*)\pi_s^l+(1-\exp(-\rho e_0^*))\pi_s^h \text{ for } s=1,...,N-1, && (20)\\ \frac{q}{\beta}\lambda^*(\varepsilon_s^*)^{1-(1-\sigma)\alpha} &= 1+\mu^*\rho\frac{\exp(-\rho e_0^*)\left(\pi_s^h-\pi_s^l\right)}{p_s(e_0^*)}+\xi^*\frac{1-(1-\sigma)\,\alpha}{\varepsilon_s^*} \text{ for } s=1,...,N, && (21)\end{aligned}$$

where (21) is the necessary first-order condition for the optimality of second period consumption. Notice that these equations also include e_0^*, λ^*, μ^* and ξ^*, moreover we have not yet set parameters α and β either. The parameter α is chosen such that the equilibrium level of effort e_0^* equals 1/3, which is roughly the average fraction of working time over total disposable time in the United States. Also notice that, given $p_s(e_0^*)=1/20$ and $\varepsilon_s^*=\hat{c}\left(\eta_s\right)$ for all s, if we sum equation (21) across income levels using weights as $p_s(e_0^*)=1/20$ we obtain

$$\frac{q}{\beta}\lambda^*\frac{1}{20}\sum_{s=1}^{20}\hat{c}\left(\eta_s\right)^{1-(1-\sigma)\alpha}=1+\frac{\xi^*\left(1+\alpha\sigma-\alpha\right)}{20}\sum_{s=1}^{20}\frac{1}{\hat{c}\left(\eta_s\right)}. \tag{22}$$

[18] We have made some sensitivity analysis with respect to the risk aversion parameter. Our results are qualitatively the same for the range of risk aversions between one and four, but the differences between the two scenarios are more pronounced if risk aversion is larger.

[19] This choice turned out to be irrelevant, because we obtained a practically perfect fit for all cases we have considered.

Consequently, the data implies a further restriction between the parameters and endogenous variables $(\beta, \alpha, \lambda^*, \xi^*)$, which we impose directly.

For the remaining variables/parameters, we use the following four optimality conditions, which we require to be satisfied exactly. First, we have the normalized Euler equation ($c_0^* = 1$ is substituted in all subsequent equations):

$$\tilde{q}\left[(1-e_0^*)^{1-\alpha}\right]^{1-\sigma} = \beta\sum_{s=1}^{N} p_s\left(e_0^*\right)\frac{\left[(\varepsilon_s^*)^{\alpha}\right]^{1-\sigma}}{\varepsilon_s^*}. \tag{23}$$

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

$$-\left(1-\alpha\right)\frac{\left[(1-e_0^*)^{1-\alpha}\right]^{1-\sigma}}{1-e_0^*} = \beta\rho\exp(-\rho e_0^*)\sum_{s=1}^{N}\left(\pi_s^h - \pi_s^l\right)\frac{\left[(\varepsilon_s^*)^{\alpha}\right]^{1-\sigma}}{(1-\sigma)}, \tag{24}$$

and the normalized first-order conditions for c_0^*,

$$\frac{\lambda^*}{(1-e_0^*)^{(1-\alpha)(1-\sigma)}} = 1-\xi^*\tilde{q}\left(1+\alpha\sigma-\alpha\right)-\mu^*\frac{(1-\alpha)(1-\sigma)}{(1-e_0^*)}, \tag{25}$$

together with the planner's first-order optimality condition for effort

$$\begin{aligned} q\lambda^*\sum_s p_s'\left(e_0^*\right)\left(\eta_s-\varepsilon_s^*\right)+\mu^*\left(\beta\sum_s p_s''(e_0^*)\frac{(\varepsilon_s^*)^{(1-\sigma)\alpha}}{\alpha(1-\sigma)}-\frac{(1-\alpha)(\alpha+(1-\alpha)\sigma)}{\alpha}\left(1-e_0^*\right)^{-\alpha-(1-\alpha)\sigma-1}\right)+ \\ +\xi^*\left(-\beta\sum_i p_s'(e_0^*)\varepsilon_s^{(1-\sigma)\alpha-1}-\tilde{q}(1-\alpha)(1-\sigma)\left(\varepsilon_s^*\right)^{(1-\sigma)\alpha-1}\left(1-e_0^*\right)^{(1-\alpha)(1-\sigma)-1}\right) = 0. \end{aligned} \tag{26}$$

Finally we obtain from the government's budget constraint the implied government consumption as a function of aggregate income as

$$G^* = q\left(\sum_i \gamma y_0^i\right)\sum_{s=1}^{N} p_s(e_0^*)(\eta_s-\varepsilon_s^*). \tag{27}$$

Here we have used $y_0 - c_0^* = 0$, the unit root process of income and Proposition 7.

We plot the estimated likelihood ratio on Figure 2. As expected (because of the same properties of the estimated consumption function) the likelihood ratio is monotone and concave.

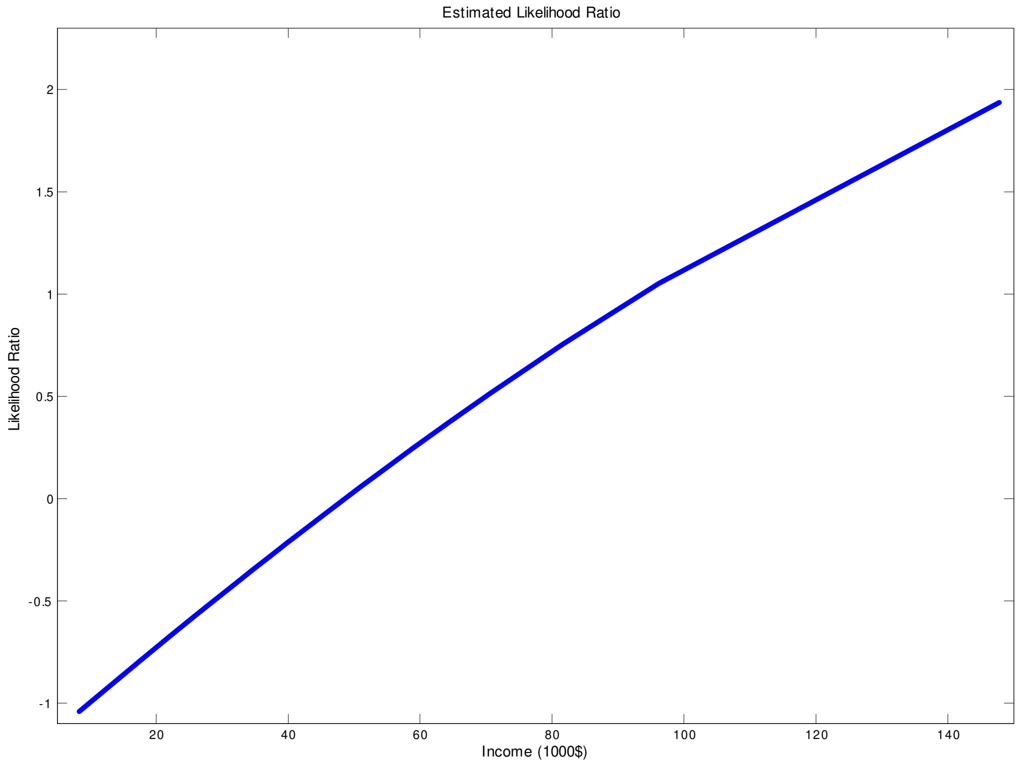

Figure 2: Estimated Likelihood Ratio

5.4 Results

We use the preset and estimated/calibrated parameters of the above model (exogenous capital taxes) to determine the optimal allocation for the scenario where capital taxes are chosen optimally—assuming perfect observability/taxability of capital. Figure 3 displays second-period consumption for this scenario together with the consumption function of the benchmark.

It is obvious from the picture that the average level of second-period consumption is higher in the case with exogenous capital taxes (tax rate on capital income of 40%). This is of course not surprising, given that optimal capital taxes in general imply frontloaded consumption (Rogerson 1985, Golosov et al. 2003).

We also observe that, since consumption is concave for the two cases, optimal labor income taxes are progressive in both scenarios. First note that we can invoke the first part of Proposition 6 stating that if third best consumption $\hat{c}$ is g_λ-concave then second best consumption c is g_λ-concave, too. Moreover, for relative risk aversion of 3, the function $g_\lambda(c) = \lambda q c^3/\beta$ is convex, hence g_λ-concavity implies concavity. However, recall that for the current computations we did not fix effort to be the same across the two allocations, which was a requirement for Proposition 6. On the one hand, this result shows that the endogenous response of effort to imperfect capital taxes does not affect the qualitative results (at least for this set of parameters). On the other hand, we will also show below that the changes in effort (and consequently the likelihood ratio) have a non-negligible quantitative effect.

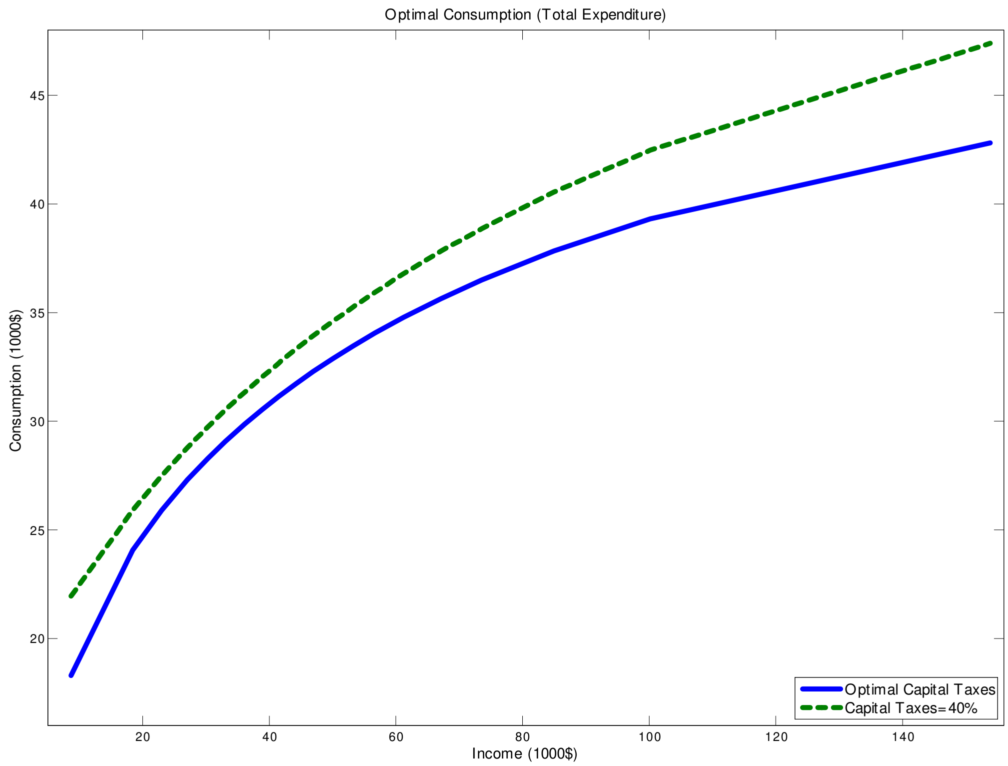

Figure 3: Optimal Consumption with Optimal and Restricted Capital Taxation

To compare progressivity across the two scenarios quantitatively, we use $-c''(y)/c'(y)$ as a measure of progressivity. In addition to the obvious analogy to 'absolute risk aversion', the advantage compared to $c''(y)$ is that it makes functions with different slopes $c'(y)$ more comparable. A higher value of this measure obviously indicates a higher degree of progressivity. On Figure 4, we have plotted this measure of progressivity for the optimal consumption plan for the case when capital taxes are restricted and for the case when they are optimal. The pattern is clear. The model with optimal capital taxes results in a uniformly more concave (progressive) consumption function compared to the case when capital taxes are restricted. The differences are particularly large for lower levels of income (and consumption).

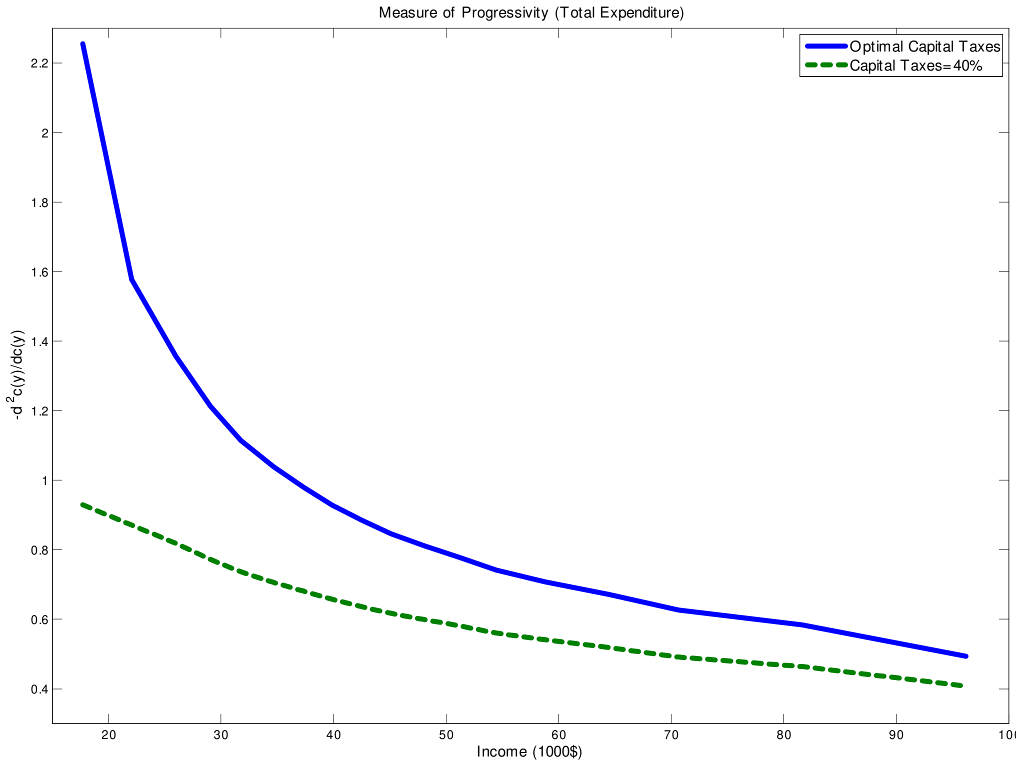

Figure 4: Income Tax Progressivity with Optimal and Restricted Capital Taxation

We have quantified these graphical observations and have checked robustness to alternative levels of risk aversion in Table 1. The results are qualitatively the same for all risk aversion levels, but there are significant quantitative differences. In particular, the difference between the two models is increasing in the level of risk aversion. The difference between the two progressivity measures is negligible for log utility, but quite large for the other three cases (ranging between 20 and 100 percent). Note that the change in measured progressivity is coming from two sources. First, as Figure 3 shows, the concavity of the optimal consumption function ($c(y)$) is changing. Second, the distribution of income changes, as effort is different under optimal capital taxes compared to the benchmark case. For this reason, we calculate the measure of progressivity both with and without this second effect (endogenous vs. exogenous weights). Comparing the first and second 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 4). 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 losses due to restricted capital taxation in consumption equivalent terms (presented in the last row of Table 1). The losses are negligible for the log case, considerable for the intermediate cases, and very large for high values of risk aversion.

We have also displayed the optimal capital taxes, calculated as $\tau^k = \tilde{q}/q - 1$. Notice that τ^k is indeed the tax rate on capital, not on capital income. The 40 percent tax on capital income in the benchmark model is equivalent to a 1.6% tax on capital. It turns out that optimal taxes are much higher than this number for all risk aversion levels, including log utility. The tax rates are actually *implausibly high*. Even in the log case, they imply a tax rate on capital income of around 90 percent. For our benchmark case, the implied tax rate on capital income would be around 1000 percent, or equivalently the after-tax return on savings is -37 percent.[20] It is difficult to imagine how such distortionary taxes can be ever implemented in a world where alternative savings opportunities (potentially with lower return) are available that are not observable and/or not taxable by the government.

Table 1: Quantitative Measures of Progressivity, Welfare Losses and Capital Taxes

Risk aversion	1	2	3	4
Average measure of progressivity $(-c''(y)/c'(y))$				
Optimal K tax (endog. weights)	0.670	0.800	0.963	1.102
Optimal K tax (exog. weights)	0.670	0.804	0.978	1.141
K tax=1.56 (40% on K income)	0.644	0.644	0.644	0.644
Welfare losses from not taxing capital optimally (%)				
	0.035	0.295	1.309	3.372
Optimal capital tax (%)				
$\tau^k = \tilde{q}/q - 1$	3.89	25.15	65.82	123.1

We can get some intuition why the differences are increasing in the risk aversion of the agent $(\hat{\sigma} := 1 - (1 - \sigma)\,\alpha)$ by examining equation (21) for our specification:

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

The direct effect of restricted capital taxation is driven by $\xi^* a(\varepsilon_s^*)$. Note that the higher is $\hat{\sigma}$, the higher is the discrepancy between the Euler equation characterizing the restricted capital taxation case and the inverse Euler characterizing the optimal capital taxation case. This will imply that ξ^* is increasing with $\hat{\sigma}$. Moreover, absolute risk aversion is given by $\hat{\sigma}/\varepsilon_s^*$, which is also increasing in $\hat{\sigma}$. Hence the effect of hidden asset accumulation (or suboptimal capital taxes) is increasing in risk aversion for both

[20] Recall that the after-tax return on capital is given by $1/\tilde{q} - 1$. This is equivalent to a tax rate on capital income defined as $t = 1 - (1/\tilde{q} - 1)/(1/q - 1)$.

of these reasons. The larger discrepancy between the Euler and inverse Euler equations also explains that optimal capital taxes must rise with risk aversion in order to make these two optimality conditions compatible. The same argument also explains why the welfare costs of restricted capital taxation are increasing in risk aversion.

As another robustness check, we examined how the results would change if we use only non-durable consumption as our measure of consumption. As we have seen on Figure 1, the main difference between the two consumption measures is that non-durable consumption is less dispersed (the average slope is significantly lower). Table 2 contains the average measures of progressivity, optimal capital taxes and the welfare losses of restricted capital taxation for the benchmark risk aversion case. First of all, note that our normalized measure of progressivity shows that, although non-durable consumption is flatter, the progressivity is very similar (recall that the model with restricted capital taxation replicates perfectly the consumption allocation for both cases). Second, notice that, with non-durable consumption, we again have a significant increase in progressivity when we impose optimal capital taxes. This once more implies a sizeable welfare gain and a highly implausible tax rate on capital. The only difference is quantitative: all these properties are somewhat less pronounced: for example the increase in progressivity here is 25 percent while it is around 50 percent in the benchmark case. The general message is that whenever the overall level of insurance is higher (consumption responds less to income shocks), imperfect observability/taxability of capital tends to have a smaller effect.

Table 2: Different Consumption Measures

$Risk\ aversion = 3$	non-durable	total expenditure
Average measure of progressivity $(-c''(y)/c'(y))$		
Optimal K tax (endog. weights)	0.849	0.963
Optimal K tax (equal weights)	0.853	0.978
K tax=1.56 (40% on K income)	0.687	0.644
Welfare losses from not taxing capital optimally (%)		
	0.434	1.309
Optimal capital tax (%)		
$\tau^k = \tilde{q}/q - 1$	37.05	65.82

Hence, we can conclude that the following three main points of our analysis are robust to different levels of risk aversion (as far as the coefficient of relative risk aversion is not to low) and to different measures of consumption: (i) Restricted (as opposed to optimal) capital taxation leads to less progressive optimal income taxes. (ii) There are significant welfare losses due to this restriction on capital taxation. (iii) The implied optimal capital taxes are implausibly high.

Finally, we would like to relate the quantitative results to Proposition 5 as well. There we have shown that under convex absolute risk aversion, whenever consumption is concave function of the likelihood ratio in the restricted capital tax case, the same must hold in the model with optimal capital taxes. Recall that this result was obtained assuming constant effort levels across the two scenarios. Therefore we compute the optimal allocation for the scenario with 40 percent capital income taxation given the effort level from the optimal capital tax case. Intuitively, we disregard the planner's optimality condition regarding effort in this case. Figure 5 displays the results of these calculations as a function of the likelihood ratio, which is (by construction) the optimal likelihood ratio under optimal capital taxes.

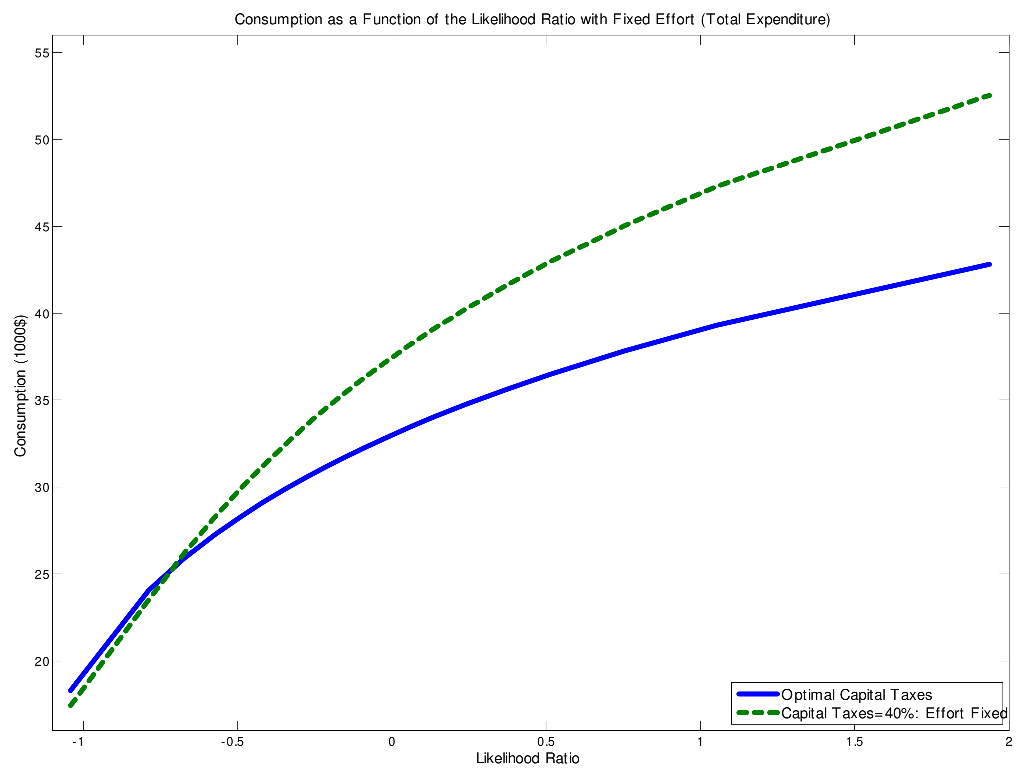

Figure 5: Optimal Consumption as Function of the Likelihood Ratio (Fixed Effort)

This figure is clearly in line with the theoretical results of Proposition 5. First of all, consumption is a concave function of the likelihood ratio in both scenarios. Moreover, consumption under optimal capital taxation is a concave transformation of consumption under restricted capital taxation.

6 Concluding remarks

This paper analyzed how restrictions to capital 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 controlled by the planner. In terms of our decentralization, this implies that marginal taxes on labor income become less progressive when restrictions to capital income taxation are binding. We complemented our theoretical results with a quantitative analysis based on individual level U.S. data on consumption and income.

The model we presented here is one of action moral hazard, similar to Varian (1980) and Eaton and Rosen (1980). The framework has the important advantage of tractability. Although a more 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 ex-ante redistribution.

While the standard Mirrlees model focuses on the intensive margin (with notable exceptions, e.g., Chone' and Laroque, 2010), the model we consider here focuses on the extensive margin. The periodic income y is the result of previously supplied effort and is subject to some uncertainty. Natural interpretations for the outcome y include the result of job search activities, 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); or again - for self-employed individuals - y can be seen as earnings from the entrepreneurial activity. It would not be difficult to include an intensive margin into our model in $t = 1$. Suppose, for simplicity, the utility function takes an additive 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 a bit if both y and I were observable, while the case where the government can only observe I is that of Mirrlees (1971).[21]

A Simple Dynastic Model: We conclude by proposing a simple extension of our model that allows for multi-periods with 'dynastic' considerations through 'warm glow' motives for bequests. Assume that in the last period preferences are $u_1(c^{\omega} k^{1-\omega})$, with $\omega \in (0, 1)$. Here, c is consumption as

[21] In this 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, it is not difficult to derive the formula for third-best allocations as:

$$\frac{q\lambda}{\beta u'(c(y))} = 1 + \mu l(y; e) + \xi a(c(y)) - \frac{\frac{d\phi(y)}{dy}}{\beta f(y; e)}, \tag{28}$$

where the multiplier associated to the intensive-margin incentive-constraint $\phi(y)$ is - as usual - related to the Spence-Mirrlees condition and the labor supply distortion,and it satisfies $\phi(\underline{y}) = \phi(\overline{y}) = 0$. The comparison between the case with restricted and unrestricted capital taxation amounts again to considering the cases with $\xi > 0$ and $\xi = 0$ respectively. Although the forces at play are the same as above, an analytic analysis with intensive margin (and private information on y) is complicated by the fact that both λ, μ, and the whole schedule $\phi(\cdot)$ change.

above, while k represents bequest transfers to future generations. Given the net income $y + \tau(y)$ in the last period, the agent solves (note that there are no reasons to impose capital taxes in $t = 1$, at least not in order to alleviate incentives):

$$\begin{aligned} & \max_{k,c \geq 0} \; u_1(c^{\omega} k^{1-\omega}) \\ s.t. \quad & c + k = y + \tau(y) \,. \end{aligned}$$

The chosen functional form implies that expenditures on c and k will be fixed proportions of the disposable income, namely: $c(y) = \omega(y + \tau(y))$, and $k(y) = (1-\omega)(y + \tau(y))$. This model with bequest is hence equivalent to our original model with utility $\tilde{u}(y + \tau) = u_1(A(y + \tau))$ where u_1 is our original utility function and $A = \omega^{\omega}(1-\omega)^{1-\omega}$ is a constant. Clearly, none of our theoretical results changes, since properties such as the convexity of the absolute risk aversion are invariant to this modification. We did not find any sizable quantitative difference either.[22] Since we are interested in the curvature of the consumption function and its changes due to restrictions to capital income taxation, it is intuitive that such extension has little impact on our results.

It is not difficult to see how such model can be embedded into a fully dynastic framework. When y is observed, k is easily computable as a (deterministic) function of $y+\tau$ since the warm glow mechanics does not leave space for strategic considerations in the inter-generational transfer of wealth. Then k would play the role of y_0 for the next generation. Of course, this framework generates heterogeneity in the initial endowments. However, the link between c_0 and y_0 would be dictated by distributional motives alone (i.e., no incentive constraint would play any role here), along the lines of the quantitative section. Details are available upon request.

[22] Details are available upon request. In our robustness computations, ω has been calibrated to match the top bracket. Alternatively, one could set this parameter so that to match the average marginal propensity to consume in the population.

Appendix: Proof of Proposition 7.

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:

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

s.t.

$$\begin{aligned}
y_0^i - c_0^i + q \sum_s p_s\left(e_0^i\right) \left[y_0^i \eta_s - c_s^i\right] &\geq t^i; \\
-\frac{(1-\alpha)}{\alpha} \frac{v'\left(T - e_0^i\right)}{v\left(T - e_0^i\right)} \left[\left(c_0^i\right)^\alpha \left(v\left(T - e_0^i\right)\right)^{1-\alpha} \right]^{1-\sigma} &= \beta \sum_s p_s'\left(e_0^i\right) \frac{\left[\left(c_s^i\right)^\alpha \left(v\left(T\right)\right)^{1-\alpha} \right]^{1-\sigma}}{\alpha(1-\sigma)}; \\
\tilde{q} \frac{\left[\left(c_0^i\right)^\alpha \left(v\left(T - e_0^i\right)\right)^{1-\alpha} \right]^{1-\sigma}}{c_0^i} &= \beta \sum_s p_s\left(e_0^i\right) \frac{\left[\left(c_s^i\right)^\alpha \left(v\left(T\right)\right)^{1-\alpha} \right]^{1-\sigma}}{c_s^i},
\end{aligned}$$

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

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

s.t.

$$\begin{aligned}
y_0^i - c_0^i + q \sum_s p_s\left(e_0^i\right) \left[y_0^i \eta_s - c_0^i \varepsilon_s^i\right] &\geq t^i; \\
-\frac{(1-\alpha)}{\alpha} \frac{v'\left(T - e_0^i\right)}{v\left(T - e_0^i\right)} \left[\left(v\left(T - e_0^i\right)\right)^{1-\alpha} \right]^{1-\sigma} &= \beta \sum_s p_s'\left(e_0^i\right) \frac{\left[\left(\varepsilon_s^i\right)^\alpha \left(v\left(T\right)\right)^{1-\alpha} \right]^{1-\sigma}}{\alpha(1-\sigma)}; \\
\tilde{q} \left[\left(v\left(T - e_0^i\right)\right)^{1-\alpha} \right]^{1-\sigma} &= \beta \sum_s p_s\left(e_0^i\right) \frac{\left[\left(\varepsilon_s^i\right)^\alpha \left(v\left(T\right)\right)^{1-\alpha} \right]^{1-\sigma}}{\varepsilon_s^i}.
\end{aligned}$$

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

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

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

$$\begin{aligned}
c_0^{i*} &= \gamma y_0^i, \\
e_0^{i*} &= e_0^{j*}, \text{ and} \\
\varepsilon_s^{i*} &= \varepsilon_s^{j*},
\end{aligned}$$

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

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

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

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

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

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] Avriel, M., W. E. Diewert, S. Schaible, and I. Zang. (1988): *Generalized Concavity.* Plenum Publishing Corporation, New York.

[3] Blundell R, Pistaferri L., Preston I. (2008): "Consumption inequality and partial insurance." *American Economic Revew*, 98: 1887–921.

[4] Chone P., and G. Laroque, (2010): "Negative Marginal Tax Rates and Heterogeneity," *American Economic Review*, 100(5): 2532-47.

[5] Conesa, J. C., S. Kitao, and D. Krueger (2009): "Taxing Capital? Not a Bad Idea After All!," *American Economic Review*, 99 (1), 25-48.

[6] Diamond, P. A., and J. A. Mirrlees (1978): "A model of social insurance with variable retirement," *Journal of Public Economics*, 10(3), 295–336.

[7] 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.

[8] 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.

[9] Domeij, D., and J. Heathcote (2004): "On The Distributional Effects Of Reducing Capital Taxes," *International Economic Review*, 45(2), 523–554.

[10] Eaton, J., and H. S. Rosen (1980): "Optimal Redistributive Taxation and Uncertainty," *The Quarterly Journal of Economics*, 95(2), 357–64.

[11] Farhi, E., and I. Werning (2011): "Insurance and Taxation over the Life Cycle," mimeo.

[12] Gollier, C. (2001): *The Economics of Risk and Time.* MIT Press.

[13] Golosov, M., N. Kocherlakota, and A. Tsyvinski (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70(3), 569–587.

[14] Golosov, M., M. Troshkin, and A. Tsyvinski (2011), "Optimal Dynamic Taxes," *mimeo.*

[15] Golosov, M., and A. Tsyvinski (2007), "Optimal Taxation with Endogenous Insurance Markets," *The Quarterly Journal of Economics*, 122(2), 487–534.

[16] Gottardi, P., and N. Pavoni (2011): "Ramsey Asset Taxation under Asymmetric Information," *mimeo,* European University Institute.

[17] Guiso, L., and M. Paiella (2008): "Risk Aversion, Wealth, and Background Risk," *Journal of the European Economic Association*, 6(6), 1109–1150.

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

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

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

[21] 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.

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

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

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

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

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