

Optimal Income Taxation with Asset Accumulation

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

May 16, 2011

Abstract

Several frictions restrict the government's ability to tax capital income. First of all, it is very costly to monitor trades on international asset markets. Moreover, agents can resort to non-taxable low-return assets such as cash, gold or foreign currencies if taxes on capital income become too high. This paper shows that limitations to capital taxation 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 capital taxation is restricted. We evaluate the effect quantitatively in a model calibrated to U.S. data and find evidence that restrictions to capital income taxation appear to be binding in the United States.

Keywords: Optimal Income Taxation, Capital Taxation, Asset Accumulation, Progressivity.

JEL: D82, D86, E21, H21.

1 Introduction

The progressivity of the income tax code is a central issue in the economic literature and the public debate. While we observe progressive tax systems in virtually all developed countries nowadays, theoretical insights on the optimal degree of progressivity remain limited. Previous research has shown that the skill distribution (Mirrlees 1971), the welfare criterion (Sadka 1976), and earnings elasticities (Saez 2001) play important roles in this context.

The existing analyses of optimal income tax progressivity have largely focused on models where labor is the only source of income. In the present paper, we argue that the optimal shape of labor income taxes cannot be determined in isolation from the tax code on capital income. Specifically,

^{*}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

we show that when the government's ability to tax capital income is limited, then optimal labor income taxes become less progressive.

Limitations for capital income taxation arise whenever the government does not have perfect (or low cost) control over agents' wealth and consumption. This situation seems very relevant for developed economies with access to international asset markets:

“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)

A similar problem applies to transitional and developing countries, where foreign currency, gold, or other non-observable assets are often used for self-insurance. Motivated by these considerations, we contrast two stylized environments in this paper. In the first one, consumption and asset decisions are perfectly 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 asset decisions 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 capital income (compare Gottardi and Pavoni 2010).¹ The decreased concavity of optimal consumption with respect to labor income then translates into marginal tax rates on labor income that increase less quickly.

We derive our results in a two-period 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, which 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 bond, which gives them a limited means for self-insurance. In this model, the planner wants to distort agents' asset decisions, because the bond provides insurance against the realization of labor income and thereby reduces the incentives to exert effort.²

Using the first-order approach (Abraham, Koehne and Pavoni 2010), we can switch from the observable asset case to the scenario with hidden asset accumulation by adding a Euler equation for

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

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

the agent to the principal's optimization problem. This constraint crucially changes the allocation of consumption across income states. Efficiency requires that for each 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 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 capital income cannot be sufficiently taxed.

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 a capital income tax rate of 40%.⁴ Using the implied parameters, we compute the optimal allocation when capital income taxation is unrestricted and compare it to the data. Under unrestricted capital taxation, the progressivity of the optimal allocation increases sizeably. Moreover, welfare increases by between 0.001% and 13.23% in consumption equivalent terms, depending on the coefficient of relative risk aversion. The required capital income tax rates are implausibly high, however, exceeding one hundred per cent for all specifications. This indicates that limitations to capital income taxation may be binding in the United States.

To the best of our knowledge, this is the first paper that examines how limits to capital taxation affect the optimal labor income tax code. Recent work on dynamic Mirrleesian economies studies optimal income taxation when capital taxation is unrestricted; see Golosov, Troshkin, and Tsyvinski (2009), and Farhi and Werning (2010). In that literature, the reason for capital taxation is 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, 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).

³To 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 $u'(c) - u'(c + \varepsilon/u'(c)) \approx -\varepsilon u''(c)/u'(c)$.

⁴This 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).

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, 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 we obtain a very different conclusion. While Conesa, Kitao and Krueger (2009) argue that capital income taxes and progressive labor income taxes are essentially substitutes, 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 exogenous taxes are in general imperfect instruments to perform redistribution. Since wealth is typically positively correlated to skill, it is optimal to compensate the imperfect distribution coming from labor income taxation with capital income taxation.

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.

2 Model

Consider a benevolent social planner (the principal) whose objective is to maximize the welfare of the 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 offers a tax/transfer system 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 $\beta > 0$.

Production and endowments At date $t = 0$, the agent has a *fixed* endowment y_0 . At date $t = 1$, the agent has a stochastic income $y \in Y := [\underline{y}, \bar{y}]$. 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, e_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.

Allocations An *allocation* (\mathbf{c}, e_0) consists of a consumption scheme $\mathbf{c} = (c_0, c(\cdot))$ 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*

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

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

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

$$\max_{(c, e_0)} u(c_0, e_0) + \beta \int_{\underline{y}}^{\bar{y}} u_1(c(y)) f(y, e_0) dy \quad (2)$$

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

$$y_0 - c_0 + q \int_{\underline{y}}^{\bar{y}} (y - c(y)) f(y, e_0) dy - G \geq 0 \quad (3)$$

and the *incentive compatibility constraint* for effort

$$e_0 \in \arg \max_e u(c_0, e) + \beta \int_{\underline{y}}^{\bar{y}} u_1(c(y)) f(y, e) dy. \quad (4)$$

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 trade bonds facing an appropriate tax system. 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 bond trades 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 (2010). 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.⁵

We interpret q as the price the planner faces when buying or (short-)selling risk-free bonds on a competitive market. When the FOA holds, second best allocations can be decentralized by letting agents trade in the same bond market subject to a *linear* tax, given suitably defined nonlinear labor income taxes; compare Gottardi and Pavoni (2010).

Proposition 1 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 bond price $\tilde{q} (> q)$ such that

$$\begin{aligned} c_0 &= y_0 + \tau_0, \\ c(y) &= y + \tau(y), \quad y \in Y, \\ (e_0, 0) &\in \arg \max_{(e,b)} u(y_0 + \tau_0 - \tilde{q}b, e_0) + \beta \int_{\underline{y}}^{\bar{y}} u_1(y + \tau(y) + b) f(y, e_0) dy. \end{aligned} \quad (5)$$

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

Proof. See Gottardi and Pavoni (2010). **Q.E.D.**

The above result is intuitive. It is efficient to tax the bond, because the bond provides intertemporal insurance when the agent plans to shirk. The reason why a linear tax on bond trades is sufficient to obtain the second best becomes apparent once we replace the incentive constraint (5) by the associated first-order conditions

$$u'_e(y_0 + \tau_0, e_0) + \beta \int_{\underline{y}}^{\bar{y}} u_1(y + \tau(y)) f_e(y, e_0) dy \geq 0, \quad (6)$$

$$\tilde{q} u'_c(y_0 + \tau_0, e_0) - \beta \int_{\underline{y}}^{\bar{y}} u'_1(y + \tau(y)) f(y, e_0) dy \geq 0. \quad (7)$$

The second condition determines the agent's asset decision based exclusively 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.

⁵As argued by Abraham, Koehne, and Pavoni (2010), both conditions have quite a broad empirical support. First, virtually all estimations for 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.

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 the Ricardian equivalence result.

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:⁶

$$\frac{\lambda}{u'_c(c_0, e_0)} = 1 + \mu \frac{u''_{ec}(c_0, e_0)}{u'_c(c_0, e_0)}, \quad (8)$$

$$\frac{\lambda q}{\beta u'_1(c(y))} = 1 + \mu \frac{f_e(y, e_0)}{f(y, e_0)}, \quad y \in [\underline{y}, \bar{y}], \quad (9)$$

where $a(c) := -u''_1(c)/u'_1(c)$ denotes absolute risk aversion, while λ and μ are the (nonnegative) Lagrange multipliers associated with the budget constraint (3) and the first-order version of the incentive constraint (4), respectively.

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

2.2 Frictions and third best allocations

Several frictions make the implementation of a bond price \tilde{q} too far from q difficult in practice. First of all, trades are extremely costly to monitor in an international asset market. Moreover, agents often have access to alternative ways of transferring resources over time, such as holding cash, foreign currencies, gold, or durable goods, for example. If the after-tax bond return $1/\tilde{q}$ becomes too low, agents will use those alternative storage technologies 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.⁷

This motivates the study of optimal allocations and decentralizations when the planner cannot tax capital income (here: the bond price) at the desired level. To simplify the exposition, we

⁶A sufficient condition for interiority is, for example, $u'_e(c, 0) = 0$ for all $c > \underline{c}$ in combination with the Inada condition $\lim_{c \rightarrow \underline{c}} u'_c(c, 0) = \infty$.

⁷For example, assuming additively separable preferences and CRRA consumption utility, the tax rate on asset holdings in such a decentralization would be $1 - \frac{q}{\beta} \left(\frac{c(y)}{c_0}\right)^\sigma$. 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.

consider the polar case where the bond price cannot be taxed at all, so that agents have access to the same technology for intertemporal transfers as the planner. Our results extend easily to the case where taxation of the bond price is possible, but restricted by an upper bound below the level of \tilde{q} required for Proposition 1.

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

$$\max_{(c, e_0)} u(c_0, e_0) + \beta \int_{\underline{y}}^{\bar{y}} u_1(c(y)) f(y, e_0) dy \quad (10)$$

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

$$y_0 - c_0 + q \int_{\underline{y}}^{\bar{y}} (y - c(y)) f(y, e_0) dy - G \geq 0 \quad (11)$$

and the first-order incentive conditions for effort and bond trades

$$u'_e(c_0, e_0) + \beta \int_{\underline{y}}^{\bar{y}} u_1(c(y)) f_e(y, e_0) dy \geq 0, \quad (12)$$

$$qu'_c(c_0, e_0) - \beta \int_{\underline{y}}^{\bar{y}} u'_1(c(y)) f(y, e_0) dy \geq 0. \quad (13)$$

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

The decentralization of a third best allocation is straightforward. The planner simply sets up a tax/transfer system $(\tau_0, \tau(\cdot))$ for labor income defined as follows:

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

If agents face this tax system and have access to the bond market at price q , 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:

$$\frac{\lambda}{u'_c(c_0, e_0)} = 1 + \mu \frac{u''_{ec}(c_0, e_0)}{u'_c(c_0, e_0)} + \xi \tilde{q} \frac{u''_{cc}(c_0, e_0)}{u'_c(c_0, e_0)}, \quad (14)$$

$$\frac{\lambda q}{\beta u'_1(c(y))} = 1 + \mu \frac{f_e(y, e_0)}{f(y, e_0)} + \xi a(c(y)), \quad y \in [\underline{y}, \bar{y}], \quad (15)$$

where λ , μ and ξ are the (nonnegative) Lagrange multipliers associated with the budget constraint (11), the first-order condition for effort (12), and the Euler equation (13), 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 equations (14) and (15) characterizing the consumption scheme are satisfied with $\xi > 0$.

Proof. 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

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

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

$$\int_{\underline{y}}^{\bar{y}} \frac{\lambda q}{\beta u'_1(c(y))} f(y, e_0) dy = 1.$$

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

$$\frac{\lambda}{u'_c(c_0, e_0)} \geq 1 = \int_{\underline{y}}^{\bar{y}} \frac{\lambda q}{\beta u'_1(c(y))} f(y, e_0) dy \geq \frac{\lambda q}{\beta \int_{\underline{y}}^{\bar{y}} u'_1(c(y)) f(y, e_0) dy},$$

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, we conclude

$$\lambda \beta \int_{\underline{y}}^{\bar{y}} u'_1(c(y)) f(y, e_0) dy > \lambda q u'_c(c_0, e_0),$$

which is incompatible with the agent's Euler equation (13). This shows that ξ cannot be zero.

Q.E.D.

Comparing the characterization of third best consumption schemes, (14), (15), to the characterization of second best consumption schemes, (8), (9), we notice that the difference between the two environments is closely related to the effect of the agent's Euler equation (13) 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 $c(y)$. Clearly, this shape is closely related to the curvature of labor income taxes in the associated decentralizations.

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, 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 , $\frac{1}{u'_1(c)}$ is always increasing. Therefore, if $\frac{1}{u'_1(c)}$ 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 (9) (or equation (15), 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.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 $\frac{1}{u_1'(c)} = c$ 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).⁹

This particular finding sheds light on a more general pattern under convex absolute risk aversion: when capital income can be fully taxed (second best), the allocation has a ‘more concave’ relationship between labor income and consumption. In other words, capital income taxation 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 (9) and (15) the consumption schemes $c(y)$ and $\hat{c}(y)$ are characterized as follows:

$$g_\lambda(c(y)) = 1 + \mu l(y, e_0), \text{ where } g_\lambda(c) := \frac{\lambda q}{\beta u_1'(c)}, \quad (16)$$

$$\hat{g}_{\hat{\lambda}, \hat{\xi}}(\hat{c}(y)) = 1 + \hat{\mu} l(y, e_0), \text{ where } \hat{g}_{\hat{\lambda}, \hat{\xi}}(c) := \frac{\hat{\lambda} q}{\beta u_1'(c)} - \hat{\xi} a(c), \text{ with } \hat{\xi} > 0. \quad (17)$$

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.**

In order to obtain a clearer intuition of this result, we further examine the planner’s first-order condition (15), namely

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

⁸Other 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).

⁹More 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 .

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 $\mu \frac{f_e(y, e_0)}{f(y, e_0)}$.¹⁰ 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 capital income can be fully taxed (second best), we have $\xi = 0$ and this gain vanishes. Intuitively, by lowering the net return of the bond, the planner is able to circumvent the first-order incentive constraint for assets. However, when capital income 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 (15) 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 over 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.

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.

¹⁰Of 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.

To make the impact of capital income taxation 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 capital income taxation 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

$$g_\lambda(c(y)) = 1 + \mu l(y, e_0), \quad (18)$$

$$\hat{g}_{\hat{\lambda}, \hat{\xi}}(\hat{c}(y)) = 1 + \hat{\mu} l(y, e_0), \quad (19)$$

where the functions g_λ and $\hat{g}_{\hat{\lambda}, \hat{\xi}}$ are defined as in (16) and (17), respectively. First, suppose that \hat{c} is a concave transformation of l . Since the right-hand side of (19) 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 $g_\lambda(c) = \frac{\lambda}{\hat{\lambda}} \left(\hat{g}_{\hat{\lambda}, \hat{\xi}}(c) + \hat{\xi} a(c) \right)$ is convex as well (since $a(c)$ is convex by assumption). Hence, using (18), we see that c is a concave transformation of l .

Conversely, suppose that c is a convex transformation of l . Using (18), 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 obviously 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, 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 Given a strictly increasing, differentiable function $u_1 : [\underline{c}, \infty) \rightarrow \mathbb{R}$, consider the two functions defined as follows:

$$g_\lambda(c) := \frac{\lambda q}{\beta u_1'(c)},$$

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

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

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

In other words, we have

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

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 and 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:

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

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

For logarithmic utility, g_λ is linear. The lemma therefore has the following consequence.

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

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

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

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 the above lemma 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.¹¹

Proposition 6 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).¹²

Proof. Recall that consumption is determined as follows:

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

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

$$\frac{1}{\mu} (g_\lambda(c(y)) - 1) = \frac{1}{\hat{\mu}} (\hat{g}_{\hat{\lambda}, \hat{\xi}}(\hat{c}(y)) - 1). \quad (20)$$

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

Corollary 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.

¹¹For example, suppose $f(x) = x^2$ and $G(\cdot) = \log(\cdot)$, then $G(f(x)) = 2 \log(x)$, which is obviously not convex.

¹²In 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 $\hat{g}_{\hat{\lambda}, \hat{\xi}}$.

Proof. Let \hat{c} be g_λ -concave. By the Lemma, 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 Proposition 6 implies that c is g_λ -concave

To verify the second statement, let c be g_λ -convex. From Proposition 6, 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 the Lemma, 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.

The corollary 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 of capital taxation and 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 the effect of capital taxation on labor income taxes *quantitatively*. In order to do this, we use consumption and income data and postulate that the data is generated by the model where capital income is taxed at a rate of 40%. Equivalently, the distorted bond 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. 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 1991 and 1992 as adapted by Blundell, Pistaferri and Preston (2008). This data file 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 factors such as age, education and innate ability. We want to control for these characteristics to obtain a clean measure of income. To do this, we postulate the following process for income:

$$y_t^i = \phi_t(X_t^i)\theta^i\nu_t^i,$$

where y_t^i is household i 's income at time t , X_t^i are observable household characteristics (a constant, age, education and race of the household head), θ^i are unobservable fixed effects at the household level (like average ability of the couple) and ν_t^i is a shock to the income of the household.¹³ Furthermore, we assume that this shock evolves according to a geometric random walk $\nu_t^i = \nu_{t-1}^i\eta_t^i$ (e.g., see Blundell et al. (2008)). The term η_t^i is our measure of the cleaned income shock.

In order to isolate η_t^i , we first run separate regressions of $\log(y_t^i)$ on X_t^i for $t = 1991$ and $t = 1992$. If we call the residual of this equation u_t^i , then this is our estimate of $\log(\theta^i\nu_t^i)$. Now, to estimate the shock η_t^i , we use the equation $\nu_t^i = \nu_{t-1}^i\eta_t^i$ and set

$$\hat{\eta}_t^i = \exp(u_t^i - u_{t-1}^i).$$

The next objective is to find the consumption function. To be able to relate it to the cleaned income measure η_t^i , we postulate that the consumption function is multiplicatively separable as well. Let $h_t^i := (\eta_0^i, \eta_1^i, \dots, \eta_t^i)$ be the history of individual income innovations up to period t :

$$\begin{aligned} c_t^i &= g_t^0(Z_t^i)g_t^1(\phi(X_t^i))g^2(\theta^i)c_t(h_t^i) \\ &= g_t^0(Z_t^i)g_t^1(\phi(X_t^i))g^2(\theta^i)c_{t-1}(h_{t-1}^i)\varepsilon(\eta_t^i) \end{aligned}$$

where Z_t^i are household characteristics that affect consumption, but (by assumption) do not affect income, such as number of kids and household assets.¹⁴ Our target is to identify $\varepsilon(\eta_t^i)$, the pure response of consumption to the income shock.

To isolate this effect, we first run separate regression of $\log(c_t^i)$ on X_t^i and Z_t^i for $t = 1991$ and $t = 1992$. The residual of this equation is called ω_t^i and estimates $\log(g^2(\theta^i)c_t(h_t^i))$. Then, using the assumption $c_t(h_t^i) = c_{t-1}(h_{t-1}^i)\varepsilon(\eta_t^i)$, we obtain our estimate of consumption in the data as:

$$\hat{\varepsilon}_t^i = \exp(\omega_t^i - \omega_{t-1}^i).$$

¹³We allow for time dependence of ϕ to capture possible aggregate shocks.

¹⁴Again, we allow for time dependence of g^0 and g^1 to capture possible aggregate shocks.

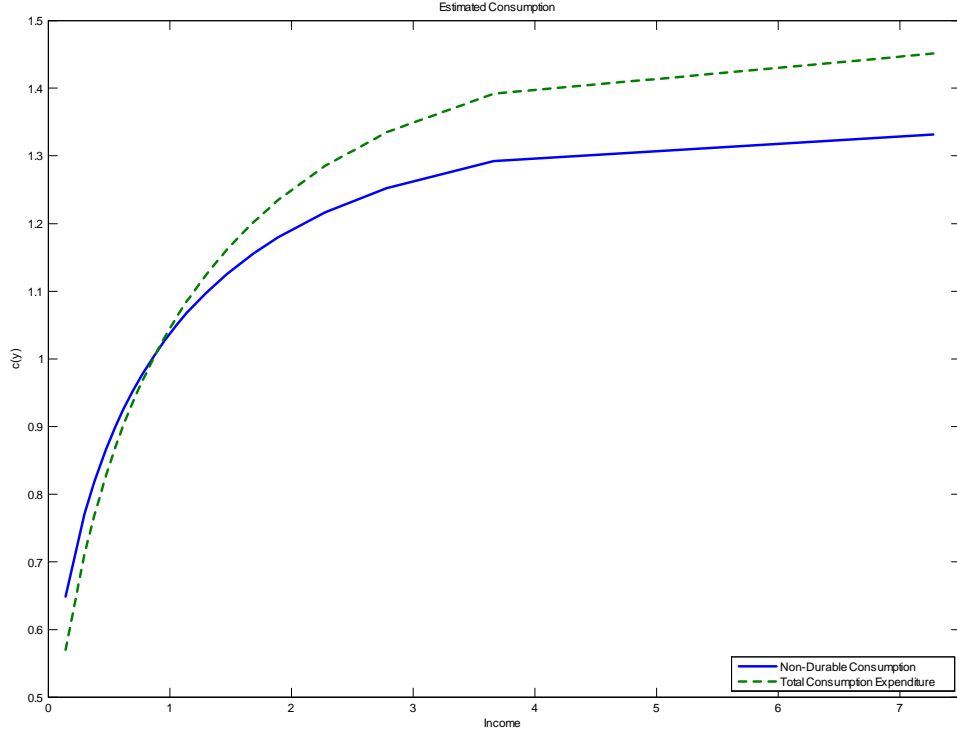


Figure 1: Estimated Consumption Functions

In the final stage, we use a flexible functional form to obtain $\varepsilon(\cdot)$. In particular, we estimate the following regression:

$$\log(\hat{\varepsilon}_t^i) = \sum_{j=0}^4 \gamma_j (\log(\hat{\eta}_t^i))^j.$$

Hence, in our model's notation, the estimate of the consumption function is given by

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

Figure 1 displays the estimated consumption function for both of our measures of consumption. Note that our estimate based on total consumption expenditure displays more dispersion and less progressivity.

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 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 $\{y_1, \dots, y_N\}$. 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 (2010).¹⁵

In order to account for (multiplicative) heterogeneity in the data, we allow for heterogeneity in the initial endowments, and choose preferences to be homothetic. In particular, we assume:

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

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:

$$\begin{aligned} & \max_{c_0^i, c_1^i, e_0^i} \sum_i \pi^i \left\{ \frac{[(c_0^i)^\alpha (v(T - e_0^i))^{1-\alpha}]^{1-\sigma}}{1 - \sigma} + \beta \sum_s p_s(e_0^i) \frac{[(c_s^i)^\alpha (v(T))^{1-\alpha}]^{1-\sigma}}{1 - \sigma} \right\} \\ & s.t. \sum_i (y_0^i - c_0^i) + q \sum_i \sum_s p_s(e_0^i) [y_s^i - c_s^i] \geq G^*; \\ & - (1 - \alpha)(1 - \sigma) \frac{v'(T - e_0^i)}{v(T - e_0^i)} \frac{[(c_0^i)^\alpha (v(T - e_0^i))^{1-\alpha}]^{1-\sigma}}{1 - \sigma} = \beta \sum_s p'_s(e_0^i) \frac{[(c_s^i)^\alpha (v(T))^{1-\alpha}]^{1-\sigma}}{(1 - \sigma)}; \\ & \tilde{q} \frac{\alpha}{c_0^i} [(c_0^i)^\alpha (v(T - e_0^i))^{1-\alpha}]^{1-\sigma} = \beta \sum_s p_s(e_0^i) \frac{\alpha}{c_s^i} [(c_s^i)^\alpha (v(T))^{1-\alpha}]^{1-\sigma}; \end{aligned}$$

with $\beta \in (0, 1)$, and $\tilde{q}, q > 0$. Moreover, assume income follows: $y_s^i = y_0^i \eta_s$.¹⁷ Then, for each given vector of income levels in period zero $\{y_0^i\}_i > 0$ and any scalar $\gamma > 0$, there exists a

¹⁵Note 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 et al. (2010) for details). And as Figure 1 shows this is delivered to us from the data. Note that MLR is only a sufficient condition for monotone consumption. Nevertheless, as expected, our estimated likelihood ratios will exhibit MLR, that is the estimated probability distributions satisfy: $\frac{\pi_s^h}{\pi_s^l}$ increasing in s .

¹⁶Where, obviously, when $\sigma = 1$ we assume preferences take a log-log form.

¹⁷We could assume $y_s^i = \Gamma (y_0^i)^\rho \eta_s$, and then we would get different Pareto weights generating $c_0^i = \hat{\gamma} (y_0^i)^\rho$. For example, the model with full insurance in period zero and *i.i.d.* shocks is doable as well, but this is trivially true even without homotheticity.

vector of Pareto weights $\{\pi^{*i}\}_i > 0$ and a number $\tau^* \in \mathbb{R}$ such that a solution to the above optimal allocation problem is such that

$$\begin{aligned} c_0^{*i} &= \gamma y_0^i \text{ for all } i \\ c_0^{*i} &= c_0^* \text{ for all } i \\ c_s^{*i} &= c_0^{*i} \varepsilon_s^* \text{ for all } i \text{ and a random variable } \varepsilon^*. \end{aligned}$$

The Pareto weights are related to each other as follows:

$$\pi^{*i} = (\gamma y_0^i)^{1-(1-\sigma)\alpha} \text{ for all } i > 0,$$

and e_0^* and ε_s^* are a solution to the following ‘normalized’ problem

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

s.t

$$\begin{aligned} \frac{1}{\gamma} - 1 + q \sum_s p_s(e_0) \left[\frac{\eta_s}{\gamma} - \varepsilon_s \right] - \tau^* &\geq 0 \\ -(1-\alpha)(1-\sigma) \frac{v'(T - e_0)}{v(T - e_0)} \frac{\left[(v(T - e_0))^{1-\alpha} \right]^{1-\sigma}}{1-\sigma} &= \beta \sum_s p'_s(e_0) \frac{\left[(\varepsilon_s)^\alpha (v(T))^{1-\alpha} \right]^{1-\sigma}}{(1-\sigma)} \\ \tilde{q} \left[(v(T - e_0))^{1-\alpha} \right]^{1-\sigma} &= \beta \sum_s p_s(e_0) \frac{1}{\varepsilon_s} \left[(\varepsilon_s)^\alpha (v(T))^{1-\alpha} \right]^{1-\sigma}, \end{aligned}$$

while τ^* is given by $G^* = \tau^* \sum_i \gamma y_0^i$.

Proof. See Appendix. **Q.E.D.**

A few remarks are now in order.

Note first that the Pareto weights π^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. The proposition states that they can be estimated from income and consumption data, up to a constant π^* . This is however not within the set of targets of this paper.¹⁸

The previous result is important for our empirical strategy for at least two main reasons. First, the proposition suggests that we can use the consumption innovations as computed in the previous

¹⁸The relationship between period zero consumption c_0^i and labor income y_0^i follows a ‘fixed-effect hypothesis’. For example, assume the following simple behavioral assumption

$$c_0^i = \delta_1 b_{-1}^i + \delta_2 y_0^i.$$

By ignoring the initial level of assets we might neglect something potentially important. Here we postulate a rela-

section as inputs in our estimation/calibration exercise. More precisely, the proposition suggests that we can use the values $\hat{c}_s^* = \frac{\hat{c}_s^i}{\hat{c}_0^i}$ as consumption inputs regardless of the actual value for \hat{c}_0^i , where $\log(\hat{c}_0^i)$ and $\log(\hat{c}_s^i)$ are in fact residuals from the regressions we described in the previous section, that is, their empirical counterparts are ω_{1991}^i and ω_{1992}^i , respectively. Note that in the data we have $\mathbf{E} \left[\log \left(\frac{\hat{c}_s^i}{\hat{c}_0^i} \right) \right] = 0$, where the expectation is taken with respect to shocks and individuals. Clearly, this is by construction, because with our pooled data we cannot distinguish common individual trends in consumption from aggregate shocks. The constants in the regressions will capture both. In any case, we find that the estimated ratio of the intercepts in these regressions is approximately one.

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^* .

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 of relative risk-aversion for consumption to 3, that is $(1 - \sigma)\alpha = 3$, in line with recent estimation results by Paravisini, Rappoport, and Ravina (2010).¹⁹ We normalize $T = 1$ and $\gamma = 1$, 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 η_1, \dots, η_{20} . To be consistent with this choice and with Proposition 7 we set $y_0 = 1$.

Given this choice of parameters, the rest of the 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.²⁰

tionship between financial wealth and ability as follows

$$b_{-1}^i = \phi y_0^i$$

hence, in the data we should find

$$c_0^i = (\delta_1 \phi + \delta_2) y_0^i.$$

Hence, this model would generate the same data as postulated in the proposition, with $\gamma = \delta_1 \phi + \delta_2$. To be consistent with our empirical strategy recovering income and consumption innovations, we assume $\gamma = 1$.

¹⁹We 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 five but, as we will see, the differences between the two scenarios are more pronounced if risk aversion is larger.

²⁰This choice turned out to be irrelevant, because we obtained a practically perfect fit.

The first group of remaining parameters of the model are ρ and the probability weights $\{\pi_s^h, \pi_s^l\}_{s=1}^N$ 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 $\varepsilon_s^* = \hat{c}(\eta_s)$, where ε_s^* is the optimal consumption innovation in the model with an exogenous capital income tax rate of 0.4.

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:

$$p_s(e_0^*) = \exp(-\rho e_0^*)\pi_s^l + (1 - \exp(-\rho e_0^*))\pi_s^h \text{ for } s = 1, \dots, N - 1, \quad (21)$$

$$\frac{q}{\beta}\lambda^*(\varepsilon_s^*)^{1-(1-\sigma)\alpha} = 1 + \mu^*\rho \frac{\exp(-\rho e_0^*)(\pi_s^h - \pi_s^l)}{p_s(e_0^*)} + \xi^* \frac{1 + \alpha\sigma - \alpha}{\varepsilon_s^*} \text{ for } s = 1, \dots, N. \quad (22)$$

Notice that these equations also include e_0^* , λ^* , μ^* and ξ^* , moreover we have not yet set parameters α and β either. For β , we use the fact that $\gamma = 1$ implies that $c_0^* = 1$. The parameter α is chosen such that the equilibrium level of effort e_0^* equal to $1/3$, which is roughly the average fraction of working time over total disposable time in the United States. For the remaining variables, 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):

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

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

$$-(1 - \alpha) \frac{\left[(1 - e_0^*)^{1-\alpha} \right]^{1-\sigma}}{1 - e_0^*} = \beta \rho \exp(-\rho e_0^*) \sum_{s=1}^N \left(\pi_s^h - \pi_s^l \right) \frac{[(\varepsilon_s^*)^\alpha]^{1-\sigma}}{(1 - \sigma)}, \quad (23)$$

and the normalized first-order conditions for e_0^*

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

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

$$q\lambda^* \sum_s p'_s(e_0^*) (\eta_s - \varepsilon_s^*) + \mu^* \left(\beta \sum_s p''_s(e_0^*) \frac{(\varepsilon_s^*)^{(1-\sigma)\alpha}}{(1 - \sigma)\alpha} - \frac{(1 - \alpha)(\alpha + (1 - \alpha)\sigma)}{\alpha} (1 - e_0^*)^{-\alpha - (1 - \alpha)\sigma - 1} \right) + \quad (24)$$

$$+ \xi^* \left(-\beta \sum_i p'_s(e_0^*) \varepsilon_s^{(1-\sigma)\alpha - 1} - \tilde{q}(1 - \alpha)(1 - \sigma) (\varepsilon_s^*)^{(1-\sigma)\alpha - 1} (1 - e_0^*)^{(1 - \alpha)(1 - \sigma) - 1} \right) = 0.$$

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

$$G^* = q \left(\sum_i \gamma y_0^i \right) \sum_{s=1}^N p_s(e_0^*) (\eta_s - \varepsilon_s^*). \quad (25)$$

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

We plot the 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.

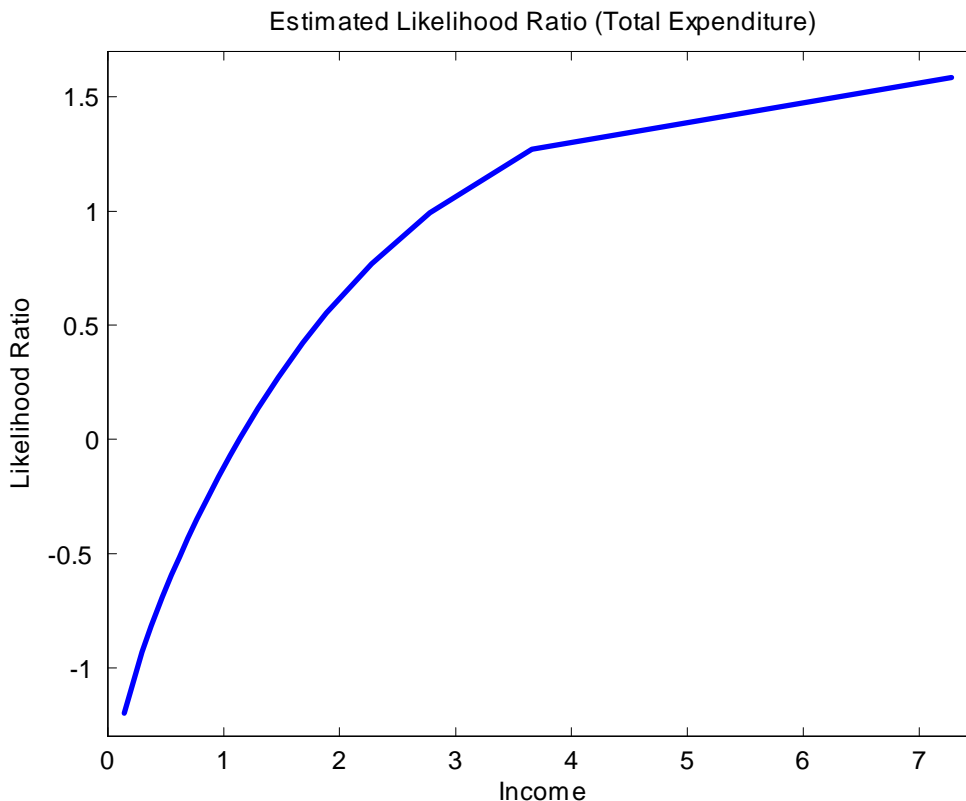


Figure 2: Estimated Likelihood Ratio

5.4 Results

We use the preset and estimated/calibrated parameters of the model above to determine the optimal allocation for the scenario where capital taxes are chosen optimally. 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 restricted capital taxation (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 the Corollary to Proposition 6 which states that if \hat{c} is g_λ -concave then c is g_λ -concave. Moreover, for relative

risk aversion of 3, the function $g_\lambda(c) = \lambda qc^3/\beta$ 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 restricted capital taxation 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.

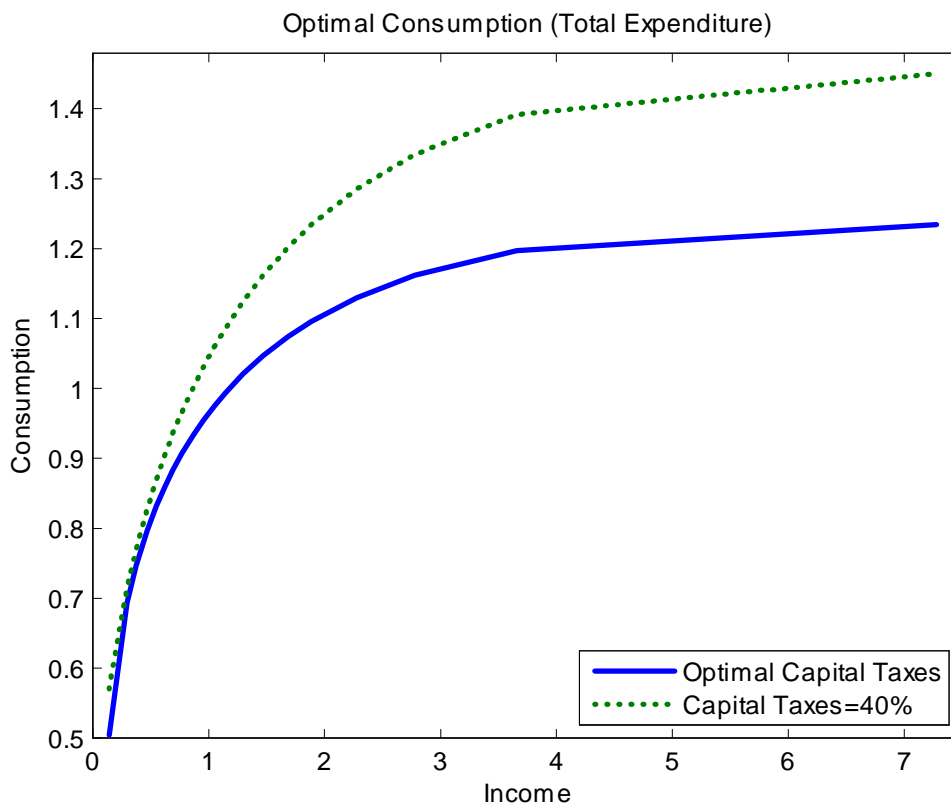


Figure 3: Optimal Consumption with Optimal and Restricted Capital Taxation

To be able to compare progressivity across the two scenarios quantitatively, we have used $-c''(y)/c'(y)$ as a measure of progressivity. This measure is used frequently by applied mathematicians to measure concavity. The obvious 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 level of income (and consumption).

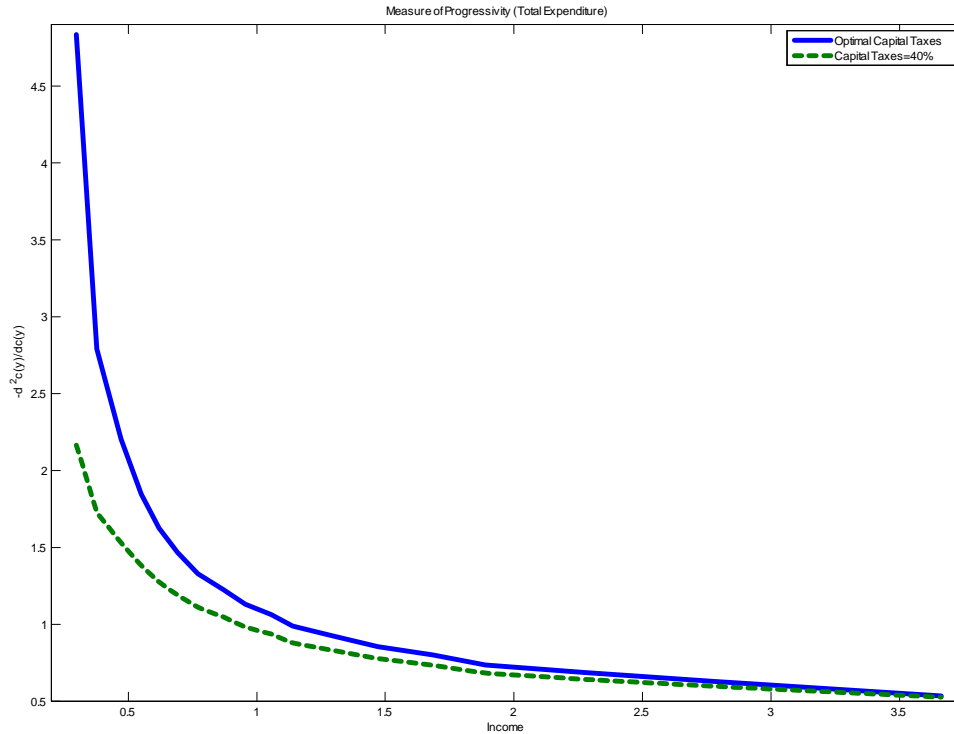


Figure 4: Income Tax Progressivity with Optimal and Restricted Capital Taxation

We have quantified these graphical observations and have checked robustness to 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 is negligible for log utility, but quite large for the other four cases. Also notice, that the average measure of progressivity is ‘flattening out’ as risk version increases. This is partially due to the fact that effort is higher with optimal capital income taxes and the difference between the effort levels is increasing in risk aversion. Higher effort implies a higher weight on high income realizations where the progressivity differences are lower (see Figure 4). This is the underlying reason behind the fact that average progressivity remains relatively constant for high risk aversion levels. To see this more clearly, we also calculated the average progressivity measure using the (equal) weights from the case with restricted capital taxes. As the second row of Table 1 shows, progressivity still increases sizably for high levels of risk aversion in this case.

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 in excess of 100 percent. It is difficult to imagine how such 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 social planner.

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

<i>Risk aversion</i>	1	2	3	4	5
Average measure of progressivity ($-c''(y)/c'(y)$)					
Optimal K taxes (endog. weights)	1.079	1.220	1.390	1.502	1.538
Optimal K taxes (equal weights)	1.079	1.228	1.427	1.591	1.686
K taxes=1.56 (40% on K income)	1.058	1.058	1.058	1.058	1.058
Welfare losses from not taxing capital optimally (%)					
	0.001	0.799	3.130	7.366	13.23
Optimal tax rate on capital (%)					
$\tau^k = \tilde{q}/q - 1$	5.76	35.21	87.56	151.2	208.0

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 (22) for our specification:

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

The direct effect of restricted capital taxation is driven by $\hat{\xi}a(\hat{c})$. 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 $\hat{\sigma}/\varepsilon$, which is also increasing in $\hat{\sigma}$. This higher 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 losses of restricted capital taxation are increasing in risk aversion.

As another robustness check, we also 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 is that non-durable consumption is less dispersed and more concave than total consumption expenditure. 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, it is indeed shown that non-durable consumption is more progressive than total consumption expenditure (recall that the model with restricted capital income 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.

Table 2: Different Consumption Measures

<i>Risk aversion = 3</i>		
	non-durable	total expenditure
Average measure of progressivity ($-c''(y)/c'(y)$)		
Optimal K taxes (endog. weights)	1.286	1.390
Optimal K taxes (equal weights)	1.297	1.427
K taxes=1.56 (40% on K income)	1.150	1.058
Welfare losses from not taxing capital optimally (%)		
	1.157	3.130
Optimal tax rate on capital (%)		
τ^k	49.65	87.56

Hence, we can conclude that the following three main points of this analysis seems to robust to different levels of risk aversion (as far as the coefficient of relative risk aversion is not too low) and to different measures of consumption. (i) Restricted (as opposed to optimal) capital taxation leads to less progressive optimal income taxation. (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 without capital taxation given the effort level from the restricted capital tax case. Intuitively, we disregard the planner's optimality condition regarding effort. Figure 5 displays the results of these calculations

as a function of the likelihood ratio, which is (by construction) the optimal likelihood ratio under restricted capital taxes.

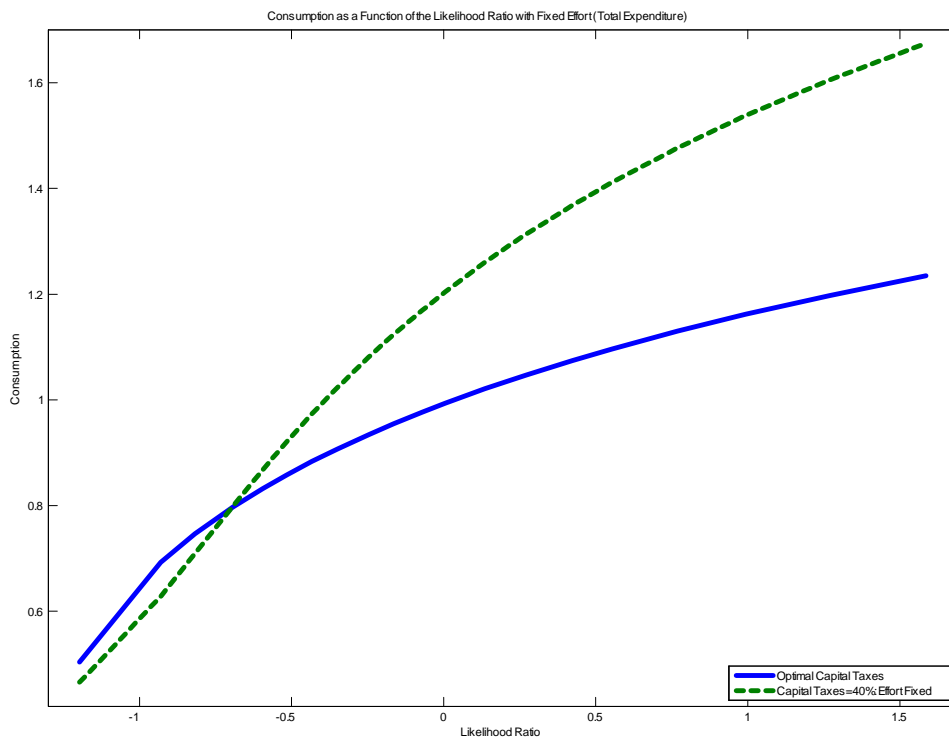


Figure 5: Optimal Consumption as Function of the Likelihood Ratio (Fixed Effort)

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). This is mainly done for 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. Of course, the quantitative analysis might change in this case. A recent example of a quantitative study of optimal income taxation under observable assets (in two periods) is Golosov, Troshkin, and Tsyvinski (2009). In their model, however, capital taxation is unrestricted and thus the government can implement the second best. This creates a possible direction for further research that can be addressed by including ex-ante (unobservable) heterogeneity in our model.²¹

Appendix: Proof of Proposition 7.

The Pareto weights imply transfers across individuals such that we can solve the problem in two stages. Stage one, transfers across individuals. Stage two, separate optimal contracts. Clearly, the transfers must be such that the individual i receives $c_0^{*i} = \gamma y_0^i$. We will chose the Pareto weights according the statement in the proposition, and construct the solution.

Let's start from stage 2. The individual optimal contracting problems is:

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

s.t.

$$y_0^i - c_0^i + q \sum_s p_s(e_0^i) [y_s^i - c_s^i] - \tau^i \geq 0 \quad (\lambda)$$

$$-(1-\alpha) \frac{v'(T - e_0^i)}{v(T - e_0^i)} \left[(c_0^i)^\alpha (v(T - e_0^i))^{1-\alpha} \right]^{1-\sigma} = \beta \sum_s p'_s(e_0^i) \frac{\left[(c_s^i)^\alpha (v(T))^{1-\alpha} \right]^{1-\sigma}}{(1-\sigma)} \quad (\mu^i)$$

$$\tilde{q} \frac{\alpha}{c_0^i} \left[(c_0^i)^\alpha (v(T - e_0^i))^{1-\alpha} \right]^{1-\sigma} = \beta \sum_s p_s(e_0^i) \frac{\alpha}{c_s^i} \left[(c_s^i)^\alpha (v(T))^{1-\alpha} \right]^{1-\sigma} \quad (\xi^i)$$

where τ^i represents the optimal net tranfer to agent i , and $\sum \tau^i = G$. By writing the Lagrangian of the original problem stated in the proposition, and using its additive separable structure, it is easy to show that when τ^i are chosen optimally, λ and for all i , μ^i and ξ^i are exactly as in the original problem in the proposition. Intuitively, the only link across agents is the common budget constraint, which can be accounted for simply by reporting the cost of funds λ .

Now, consider the following 'normalized' problem:

$$\hat{V}^* = \max_{(\varepsilon_s, e_0) \geq 0} \frac{\left[(v(T - e_0))^{1-\alpha} \right]^{1-\sigma}}{1-\sigma} + \beta \sum_s p_s(e_0) \frac{\left[(\varepsilon_s)^\alpha (v(T))^{1-\alpha} \right]^{1-\sigma}}{1-\sigma}$$

²¹The case with observable heterogeneity can be handled quite easily and none of our results would change.

s.t.

$$\begin{aligned} \frac{1}{\gamma} - 1 + q \sum_s p_s(e_0) \left[\frac{\eta_s}{\gamma} - \varepsilon_s \right] - \tau^* &\geq 0; \\ -(1-\alpha)(1-\sigma) \frac{v'(T-e_0)}{v(T-e_0)} \frac{\left[(v(T-e_0))^{1-\alpha} \right]^{1-\sigma}}{1-\sigma} &\leq \beta \sum_s p'_s(e_0) \frac{\left[(\varepsilon_s)^\alpha (v(T))^{1-\alpha} \right]^{1-\sigma}}{(1-\sigma)}; \\ \tilde{q} \left[(v(T-e_0))^{1-\alpha} \right]^{1-\sigma} &= \beta \sum_s p_s(e_0) \frac{1}{\varepsilon_s} \left[(\varepsilon_s)^\alpha (v(T))^{1-\alpha} \right]^{1-\sigma}. \end{aligned}$$

Assume this problem has at least one solution, and let $(\{\hat{\varepsilon}_s^*\}_s, \hat{e}_0^*)$ such a solution at least for the appropriate choice of $\tau^* \in \mathbb{R}$.²² This is our guess for the scale invariant part of the contract.

Since we consider equal treatment across agents with the same income, we replace the index i for y_0 , the only source of heterogeneity.

Now, consider the objective function for the guessed Pareto weights as in the proposition:

$$V^*(y_0) = \max_{c_0, \varepsilon_s, e_0} \frac{\gamma y_0}{(\gamma y_0)^{(1-\sigma)\alpha}} \frac{\left[c_0^\alpha (v(T-e_0))^{1-\alpha} \right]^{1-\sigma}}{1-\sigma} + \frac{\gamma y_0}{(\gamma y_0)^{(1-\sigma)\alpha}} \beta \sum_s p_s(e_0) \frac{\left[(c_0 \varepsilon_s)^\alpha (v(T))^{1-\alpha} \right]^{1-\sigma}}{1-\sigma}$$

s.t.

$$\begin{aligned} y_0 - c_0 + q \sum_s p_s(e_0) [y_0 \eta_s - c_0 \varepsilon_s] - \tau^* \gamma y_0 &\geq 0; \\ -(1-\alpha)(1-\sigma) \frac{v'(T-e_0)}{v(T-e_0)} \frac{\left[(v(T-e_0))^{1-\alpha} \right]^{1-\sigma}}{1-\sigma} &= \beta \sum_s p'_s(e_0) \frac{\left[(\varepsilon_s)^\alpha (v(T))^{1-\alpha} \right]^{1-\sigma}}{(1-\sigma)}; \\ \tilde{q} \left[(v(T-e_0))^{1-\alpha} \right]^{1-\sigma} &= \beta \sum_s p_s(e_0) \frac{1}{\varepsilon_s} \left[(\varepsilon_s)^\alpha (v(T))^{1-\alpha} \right]^{1-\sigma}. \end{aligned}$$

Note that since the IC constraints are homogeneous of degree zero in c_0 , they are indetical to those in the normalized problem. The same argument we made above guarantees that this problem has at least one solution: let $(c_0^*(y_0), e_0^*(y_0), \{\varepsilon_s^*(y_0)\}_s)$ such a solution.²³

²²For example, the set is not empty, as the contract $\varepsilon_s = \kappa \geq 0$ for all s , and - given the parameters, including γ , \tilde{q} , q , and $\{\eta_s\}_s$, if we set

$$\tau^* \leq \frac{1}{\gamma} - 1 + q \sum_s p_s(0) \left[\frac{\eta_s}{\gamma} - \kappa \right]$$

then for an appropriate choice of κ that solves the Euler equation, such contract solves the incentive constraints as well, since

$$-(1-\alpha) \frac{v'(T)}{v(T)} \left[(v(T))^{1-\alpha} \right]^{1-\sigma} \leq \beta \sum_s p'_s(0) \frac{\left[(\kappa)^\alpha (v(T))^{1-\alpha} \right]^{1-\sigma}}{(1-\sigma)} = 0.$$

Existence is hence guaranteed by continuity.

²³A necessary and sufficient condition for the optimality of c_0^* is

$$\frac{\gamma y_0}{(\gamma y_0)^{(1-\sigma)\alpha}} \frac{\alpha(1-\sigma)}{c_0^*} \left\{ \frac{\left[(c_0^*)^\alpha (v(T-e_0^*))^{1-\alpha} \right]^{1-\sigma}}{1-\sigma} + \beta \sum_s p_s(e_0^*) \frac{\left[(c_0^* \varepsilon_s^*)^\alpha (v(T))^{1-\alpha} \right]^{1-\sigma}}{1-\sigma} \right\} = \lambda \left[1 + q \sum_s p_s(e_0^*) \varepsilon_s^* \right]$$

We want to show that using $(\{\hat{\varepsilon}_s^*\}_s, \hat{e}_0^*)$ together with $c_0^*(y_0)$ delivers the optimal value $V^*(y_0)$. Consider the value $c_0^*(y_0) \hat{V}^*$. Clearly, we have $\frac{\gamma y_0}{(\gamma y_0)^{(1-\sigma)\alpha}} (c_0^*(y_0))^{(1-\sigma)\alpha} \hat{V}^* \leq V^*(y_0)$. Now, consider a solution to $V^*(y_0)$. It is easy to see that rescaling such solution with $\frac{\gamma y_0}{(\gamma y_0)^{(1-\sigma)\alpha}} (c_0^*(y_0))^{(1-\sigma)\alpha}$ delivers a guess for \hat{V}^* . In fact, by the definition of \hat{V}^* , we have $\hat{V}^* \geq \frac{V^*(y_0)}{\frac{\gamma y_0}{(\gamma y_0)^{(1-\sigma)\alpha}} (c_0^*(y_0))^{(1-\sigma)\alpha}}$. The two inequalities deliver $V^*(y_0) = \frac{\gamma y_0}{(\gamma y_0)^{(1-\sigma)\alpha}} (c_0^*(y_0))^{(1-\sigma)\alpha} \hat{V}^*$. In particular, $(c_0^*(1))^{(1-\sigma)\alpha} \hat{V}^* = V^*(1)$. It is however easy to see that, since the definition of the problem for \hat{V}^* and $V^*(1)$ are identical, it must be that $c_0^*(1) = 1$.

Given the required allocation, we solve the budget constraint by choosing the value for G as a residual. More precisely, given e_0^* , we have $\tau^* := \frac{1}{\gamma} - 1 + q \sum_s p_s (e_0^*) \left[\frac{\eta_s}{\gamma} - \varepsilon_s \right]$, hence $G^* = \tau^* \sum_i \gamma y_0^i$. **Q.E.D.**

References

- [1] ABRAHAM, A., S. KOEHNE, AND N. PAVONI (2010): “On the First-Order Approach in Principal-Agent Models with Hidden Borrowing and Lending,” *Journal of Economic Theory*, forthcoming.
- [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 Review*, 98: 1887–921.
- [4] CONESA, J. C., S. KITAO, AND D. KRUEGER (2009): “Taxing Capital? Not a Bad Idea After All!,” *American Economic Review*, Vol. 99 (1), 25-48.
- [5] DIAMOND, P. A., AND J. A. MIRRLEES (1978): “A model of social insurance with variable retirement,” *Journal of Public Economics*, 10(3), 295 – 336.
- [6] 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.
- [7] DOMEIJ, D., AND J. HEATHCOTE (2004): “On The Distributional Effects Of Reducing Capital Taxes,” *International Economic Review*, 45(2), 523–554.

Note indeed that if $c_0^* = \gamma y_0$ then the above condition simplifies to

$$\left[(v(T - e_0^*))^{1-\alpha} \right]^{1-\sigma} + \beta \sum_s p_s (e_0^*) [(\varepsilon_s^*)^\alpha (v(T))^{1-\alpha}]^{1-\sigma} = \frac{\lambda}{\alpha} \left[1 + q \sum_s p_s (e_0^*) \varepsilon_s^* \right].$$

- [8] EATON, J., AND H. S. ROSEN (1980): “Optimal Redistributive Taxation and Uncertainty,” *The Quarterly Journal of Economics*, 95(2), 357–64.
- [9] FARHI, E., AND I. WERNING (2010): “Insurance and Taxation over the Life Cycle,” mimeo.
- [10] GOLLIER, C. (2001): *The Economics of Risk and Time*. MIT Press.
- [11] GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKI (2003): “Optimal Indirect and Capital Taxation,” *Review of Economic Studies*, 70(3), 569–587.
- [12] GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (2009), “Optimal Dynamic Taxes,” *mimeo*.
- [13] GOTTARDI, P., AND N. PAVONI (2010): “Ramsey Asset Taxation under Asymmetric Information,” *mimeo*, European University Institute.
- [14] GUISO, L., AND M. PAIELLA (2008): “Risk Aversion, Wealth, and Background Risk,” *Journal of the European Economic Association*, 6(6), 1109–1150.
- [15] HOLMSTROM, B. (1979): “Moral hazard and observability,” *Bell Journal of Economics*, 10(1), 74–91.
- [16] KOCHERLAKOTA, N. R. (2005): “Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation,” *Econometrica*, 73(5), 1587–1621.
- [17] LAROQUE, G. (2010): “On Income and Capital Taxation in a Life Cycle Model with Extensive Labor Supply,” *mimeo*, University College London.
- [18] 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.
- [19] MIRRLEES, J. A. (1971): “An Exploration in the Theory of Optimum Income Taxation,” *Review of Economic Studies*, 38(114), 175–208.
- [20] ROGERSON, W. (1985), “Repeated Moral Hazard,” *Econometrica*, 53: 69-76.
- [21] SADKA, E. (1976): “On Progressive Income Taxation,” *American Economic Review*, 66(5), 931–35.
- [22] SAEZ, E. (2001): “Using Elasticities to Derive Optimal Income Tax Rates,” *Review of Economic Studies*, 68(1), 205–29.

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