

Optimal Income Taxation with Asset Accumulation*

Árpád Ábrahám[†], Sebastian Koehne^{‡§}, Nicola Pavoni[¶]

December 5, 2014

Abstract

Several frictions restrict the government's ability to tax assets. First, it is very costly to monitor trades on international asset markets. Second, agents can resort to nonobservable low-return assets such as cash, gold or foreign currencies if taxes on observable assets become too high. This paper shows that limitations in asset taxation have important consequences for the taxation of labor income. Using a dynamic moral hazard model of social insurance, we find that optimal labor income taxes become less progressive when governments face limitations in asset taxation. We evaluate the quantitative effect of imperfect asset taxation for two applications of our model.

Keywords: Optimal Income Taxation, Capital Taxation, Progressivity.

JEL: D82, D86, E21, H21.

*The paper has benefitted from helpful discussions at numerous seminars and conferences. We particularly thank Varadarajan V. Chari, Mikhail Golosov, Chris Sleet, and Oleg Tsyvinski for detailed and insightful comments. Sebastian Koehne gratefully acknowledges financial support from Knut och Alice Wallenbergs Stiftelse (grant no. 2012.0315).

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

[‡]Institute for International Economic Studies (IIES), Stockholm University, and CESifo.

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

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

1 Introduction

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

According to a study by the Tax Justice Network, in 2010 more than \$21 trillion of global private financial wealth was invested in offshore accounts and not reported to the tax authorities. Moreover, even when agents choose not to hide their wealth abroad, they have access to number of nonobservable storage technologies at home. For example, agents can accumulate cash, gold, or durable goods. These assets bring lower returns, but nonetheless impose limits for the collection of taxes on assets that are more easily observed.

Motivated by these considerations, this paper explores optimal tax systems in a framework where assets are 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 in the scenario with hidden assets, optimal consumption 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. Our results show that marginal labor income taxes become less progressive when the government's ability to tax/observe asset holdings is imperfect.

We derive our results in a tractable dynamic model of social insurance. A continuum of ex ante identical agents influence their labor incomes by exerting effort. Labor income is subject to uncertainty and effort is private information. This creates a moral hazard problem. The social planner 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 a means for self-insurance.

In this model, the planner wants to distort agents' asset decisions, because asset accumulation provides insurance against the idiosyncratic income uncertainty and thereby reduces the incentives to exert effort.¹ When fully capable of doing so, the planner uses capital income taxation to deter the agent from accumulating assets and labor income taxation to balance consumption insurance and effort incentives optimally. When capital income taxation is limited, the planner is forced to use labor income taxation also to reduce the agent's incentive to save. Efficiency requires that, for each income state, the costs of increasing the agent's utility by a marginal unit equal the benefits of doing so. A marginal increase of utility in a state with consumption c reduces the agent's marginal return of savings in that state by $Ra(c)$, where $a(c)$ is the absolute risk aversion of the agent at consumption c in that state and R is the asset return.² Hence, relative to the perfect capital taxation case, there is an additional social return to allocating utility to a given state which is proportional to the level of absolute risk aversion of the agent. Therefore, unless absolute risk aversion is constant or linear, limits to capital taxation have direct implications for the curvature of optimal consumption.³ In particular, whenever absolute risk aversion is convex, the planner finds it optimal to generate an additional convexity (or regressivity) of consumption.

The paper also evaluates the quantitative impact of capital taxation on income tax progressivity and on welfare. The identification and estimation of the key parameters of our framework is not straightforward because the technology that determines how effort affects future income is not directly observable. In the paper, we use two different identification strategies (and two distinct applications) to overcome this issue. The first application uses consumption and income data from the PSID (Panel Study of Income Dynamics) as adapted by Blundell, Pistaferri and Preston (2008). The main identification assumption is 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 This strategy has the advantage that all the fundamental parameters can be identified using only one cross-sectional joint distribution of consumption and income. The second application interprets our model as a model of unobservable human capital accumulation. Naturally, in this application, agents make their human capital accumulation decision assuming a particular (progressivity of the) tax system. Given this,

¹See Diamond and Mirrlees (1978), Rogerson (1985), and Golosov, Kocherlakota and Tsyvinski (2003).

²A marginal increase of utility in a state with consumption c reduces the agent's marginal utility in that state by $-u''(c)/u'(c)$. To increase $u(c)$ by ε , indeed, c has to be increased by $\varepsilon/u'(c)$. The first-order effect on the agent's marginal utility is therefore given by $-\varepsilon u''(c)/u'(c)$.

³Of course, the very same argument implies that both the level and slope of the optimal consumption allocation might also be affected by the properties of the absolute risk aversion.

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 exogenous change of the tax system allows us to identify the underlying parameters of the technology governing the effect of human capital accumulation on the (life-time) earnings distribution. We use the Economic Recovery Tax Act of 1981 as such exogenous event and we compare the distributions of life-time earnings of cohorts who presumably made their most important human capital accumulation decisions before and after this tax reform was enacted. We use the NLSY79 (1979 wave of the National Longitudinal Survey of the Youth) data set for these calculations.

After the estimation, in both applications, we compare the optimal allocations of perfect asset taxation (observable assets) and limited asset taxation (hidden assets). Under perfect asset taxation, the progressivity of the optimal allocation increases sizably in both applications. The welfare gain of perfect asset taxation varies with the coefficient of relative risk aversion and amounts to 1.3–1.5% in consumption equivalent terms in the benchmark cases of the two applications. However, the required asset income tax rates are implausibly high, 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.

Related literature. 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 Farhi and Werning (2013) and Golosov, Troshkin, and Tsyvinski (2013). In those works, the reason for asset 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 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 age-dependent labor supply elasticities and borrowing constraints, they argue that capital income taxes and progressive labor income taxes are two alternative ways of providing age-dependent insurance against idiosyncratic shocks. They 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 sub-

stitutable 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 of 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 redistributive target. The role of capital taxes is to facilitate the use of such redistributive 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). 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 explores how the access to saving technologies changes the progressivity of the optimal tax scheme.

The paper proceeds as follows. Section 2 describes the setup of the model. Section 3 presents the main theoretical results of the paper. We show that hidden asset accumulation leads to optimal consumption schemes that are less progressive according to several progressivity concepts. Section 4 explores the quantitative importance of our results. Section 5 provides concluding remarks. The appendix collects all proofs that are omitted from the main text and provides further details on the estimation strategy for Section 4.

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 exerting effort. The planner's budget must be (intertemporally) balanced.

Preferences. The agent derives utility from consumption $c_t \geq \underline{c} \geq -\infty$ and effort $e_t \geq 0$ according to $u(c_t, e_t)$, where u is a concave, twice continuously differentiable function which is strictly increasing and strictly concave in c_t , 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 and, ultimately, labor income. At date $t = 0$, the agent has a *fixed* endowment y_0 . At date $t = 1$, the agent has a stochastic income $y \in Y := [\underline{y}, \bar{y}]$. 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 $y \in Y$, 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 to the planner.

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

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

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

2.1 Observable assets and 'second best' allocations

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

Second best. A *second best* allocation is an allocation that maximizes ex-ante welfare⁵

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

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

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

and the *incentive compatibility constraint* for effort

$$e_0 \in \arg \max_e u(c_0, e) + \beta \int_{\underline{y}}^{\bar{y}} u_1(c(y))f(y, e) dy. \quad (3)$$

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

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

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

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

First, we note that an after-tax asset price \tilde{q} is equivalent to a capital wedge τ^K (tax on the gross return) that is constant across agents, $\tau^K = 1 - q/\tilde{q}$, or a linear capital income tax

⁵Although for pure notational simplicity we consider the case with a continuum of output levels, we do not discuss the technicalities related to the existence of a solution 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.

at rate $\tau^K/(1 - q)$. Notice moreover 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 not surprising, because the timing of tax collection is irrelevant by Ricardian equivalence.

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

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

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

The second first-order condition (6) determines the agent's asset decision exclusively based on consumption levels and the after-tax asset 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.

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. 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.

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

$$\frac{\lambda}{u'_c(c_0, e_0)} = 1 + \mu \frac{u''_{ec}(c_0, e_0)}{u'_c(c_0, e_0)}, \quad (7)$$

$$\frac{\lambda q}{\beta u'_1(c(y))} = 1 + \mu \frac{f_e(y, e_0)}{f(y, e_0)}, \quad y \in [\underline{y}, \bar{y}], \quad (8)$$

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

2.2 Hidden assets and ‘third best’ allocations

Whereas 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 describe 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 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.

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 $1 - \frac{q}{\beta} \left(\frac{c(y)}{c_0} \right)^\sigma$, 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.

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

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

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

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

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

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

$$q^n u'_c(c_0, e_0) - \beta \int_{\underline{y}}^{\bar{y}} u'_1(c(y)) f(y, e_0) dy \geq 0. \quad (11)$$

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

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

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

If agents face this tax system and have access to the nonobservable savings technologies at 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 an interior solution, the first-order conditions of the Lagrangian with respect to consumption are

now:

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

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

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

Proposition 2. *Suppose that the FOA is valid and let $(c_0, c(\cdot), e_0)$ be a third best allocation that is interior. Then there exists a number $\bar{q} > q$ such that equations (12) and (13) characterizing the consumption scheme are satisfied with $\xi > 0$ whenever $q^n < \bar{q}$.*

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

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

3 Theoretical results on progressivity

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

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

We call the allocation *regressive* if $c'(y)$ is increasing in y .

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

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

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

Proposition 3 (Sufficient conditions for progressivity/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.*

Note that in the previous proposition, 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

⁸Other cases where the progressivity/regressivity does not differ between second best and third best are when a has the same shape as $1/u'_1$ (quadratic utility) and when a is linear (and hence increasing).

obtain second best schemes that are *proportional*, since $1/u_1'(c) = c$ is both concave and convex. Interestingly, since absolute risk aversion $a(c) = 1/c$ is convex, third best schemes are *regressive* in this case.⁹

3.1 Rankings of progressivity for linear likelihood ratios

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

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

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

Proof. Given validity of the FOA, by equations (8) and (13) the second and third best con-

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 .

sumption schemes $c^{sb}(y)$ and $c^{tb}(y)$ are characterized as follows:

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

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

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

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

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

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

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 there is a return of 1, since the agent's utility is increased by one unit. Furthermore, increasing the agent's utility also relaxes the incentive constraint for effort, generating a return of $\mu \frac{f_e(y, e_0)}{f(y, e_0)}$.¹¹ Finally, by increasing $u_1(c(y))$ the planner alleviates the savings motive of the

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

¹¹Of 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.

agent. Since the return to one unit of saving in state y is given by $u'_1(c(y))$, the gain of a unit increase in $u_1(c(y))$ is measured by $\zeta a(c(y))$, where ζ is the multiplier of the agent's Euler equation and $a(c) = -u''_1(c)/u'_1(c)$. That is, $a(c)$ is the appropriate measure for the gains of relaxing the Euler equation. In other words, the social gains of deterring the agent from saving in a given state are proportional to the agent's absolute risk aversion in that state.

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

Another common approach to compare the progressivity/concavity of functions is to explore concave transformations. Recall that a function f_1 is a *concave transformation* of a function f_2 if there is an increasing and concave function v such that $f_1 = v \circ f_2$.

For the case of logarithmic utility, we are able to rank the progressivity of the second and third best allocation in the sense of concave transformations.¹²

Proposition 5 (Logarithmic utility). *In addition to the assumptions from Proposition 4, suppose that u_1 is logarithmic. Then second best consumption is a concave transformation of third best consumption.*

Proof. For logarithmic utility, we have $u'_1(c) = a(c) = 1/c$. By equations (14) and (15), we can link second best and third best consumption as follows:

$$\frac{\lambda^{sb} q}{\beta} c^{sb}(y) - \mu^{sb} l(y, e_0^{sb}) + \mu^{tb} l(y, e_0^{tb}) = \frac{\lambda^{tb} q}{\beta} c^{tb}(y) - \frac{\zeta^{tb}}{c^{tb}(y)} \quad (16)$$

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

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

3.2 Rankings of progressivity for nonlinear likelihood ratios

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

With slight abuse of notation, we denote the consumption allocation that optimally implements a given effort $e_0 > 0$ by $(c_0^{sb}, c^{sb}(\cdot))$ for the scenario with observable saving and by $(c_0^{tb}, c^{tb}(\cdot))$ for the case of hidden saving. As usual, we assume that the FOA is justified and that second best and third best consumption levels are interior.

For nonlinear likelihood ratios, we can rank the progressivity of allocations in a way that is very similar to Proposition 4.

Proposition 6. *Assume that u_1 has convex absolute risk aversion. Then, if c^{tb} is a concave transformation of $l(\cdot, e_0)$, then c^{sb} is a concave transformation of $l(\cdot, e_0)$. On the other hand, if c^{sb} is a convex transformation of $l(\cdot, e_0)$, then c^{tb} is a convex transformation of $l(\cdot, e_0)$.*

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

HARA (or linear risk tolerance) utility functions, namely

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

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

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

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

$$g_{\lambda, \mu}(c) := \frac{\lambda q}{\mu \beta u_1'(c)} - \frac{1}{\mu'}$$

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

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

The restriction of $\gamma \geq -1$ in the above result is innocuous to most applications, because it allows for all HARA functions with nonincreasing absolute risk aversion ($\gamma \geq 0$) as well as quadratic utility ($\gamma = -1$), for instance.

Lemma 1 enables us to rank the progressivity of consumption in the sense of concave transformations. Recall that the consumption allocations that optimally implement a given effort are characterized as follows:

$$g_{\lambda^{sb}, \mu^{sb}}(c^{sb}(y)) = l(y, e_0), \quad (17)$$

$$g_{\lambda^{tb}, \mu^{tb}, \xi^{tb}}(c^{tb}(y)) = l(y, e_0). \quad (18)$$

Due to the link between second best and third best consumption schemes in equations (17) and (18), Lemma 1 has the following consequence.

Proposition 7. *Suppose that u_1 belongs to the HARA class with $\gamma \geq -1$. Then there exists a monotonic function g such that $g \circ c^{sb}$ is a concave transformation of $g \circ c^{tb}$. In particular, if u_1 is logarithmic, c^{sb} is a concave transformation of c^{tb} .*

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

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

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

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

All our results for nonlinear likelihood ratios generalize when we take into account changes in the implemented effort levels provided that $l(l^{-1}(y, e_0^{tb}), e_0^{sb})$ is concave. The last condition is satisfied if the likelihood ratio in the third best is a convex transformation of the likelihood ratio in the second best. In fact, a weaker condition is sufficient. As the line of proof of Proposition 7 shows, it is sufficient that $l(\cdot, e_0^{sb}) \circ l^{-1}(\cdot, e_0^{tb}) \circ h$ is concave, where h is a strictly concave function. This condition is satisfied whenever $l(y, e_0^{tb})$ is “not too concave” relative to $l(y, e_0^{sb})$. How much the curvature of the likelihood ratio differs between the two scenarios is impossible to predict without detailed knowledge of density function $f(y, e)$. We will try to make progress on this issue in our quantitative analysis. In both of our quantitative applications, the likelihood ratio induced by the effort for hidden assets will be more convex (less concave) than the one for observable assets. Therefore, our theoretical insights on nonlinear likelihood ratios are, in fact, further strengthened through the variation of effort between the two allocations.

4 Applications

In this section, we apply our model to study two macroeconomic problems from a quantitative perspective. The quantitative analysis serves multiple purposes. First, we complement our theoretical results. For example, recall that the theoretical results on nonlinear likelihood ratios compare two allocations that implement the same effort level. In this section,

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

we allow effort to change between the two scenarios.

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

Below, we present two applications showing two very different ways to overcome this problem. The first approach (see Section 4.1) assumes that the joint distribution of consumption and income is generated by the third-best allocation. The advantage of this approach is that a single cross section of consumption and income (or consumption and income growth) suffices to identify all the fundamental parameters. Gayle and Miller (2009) use a similar identification strategy to estimate dynamic moral hazard models of executive compensation. The strong assumption behind this approach is that the data is generated by a constrained efficient allocation. The key advantage is that by using the necessary optimality conditions all the key parameters of the model are naturally identified.

The second application (see Section 4.2) is based upon the agents' optimal choices *given the existing system of taxes*. Here the difficulty is that identification requires an exogenous change of the tax system, and data on the income distribution both before and after the tax reform. For this exercise, we apply our model to a human capital accumulation problem. Effort is interpreted as the human capital investment decision of young individuals. We take the Economic Recovery Tax Act of President Reagan in 1981 as an exogenous event that has changed the incentives for human capital investment. Then, by comparing the ex post life-time income distributions of cohorts making the investment before and after the tax reforms, we are able to identify the key parameters of the model.

The third target of this exercise is to evaluate *quantitatively* how the limited possibility of taxing capital affects the optimal allocation, and consequently optimal labor income taxes. We will see that, in both applications, the effect of imperfect asset taxation is considerable.

4.1 Application I: Optimal labor income taxation

For this application, we use the following interpretation of our model: agents face income shocks, they exert unobservable work effort and they can use a saving technology with a

gross return given by $1/\tilde{q}$, where \tilde{q} is the *after-tax* asset price. In order to estimate the key parameters of the model, we use consumption and income data and postulate that the data is generated by a specification of the model where capital income is taxed at an exogenous rate of 40%. Equivalently, the after-tax asset price is given by $\tilde{q} = \frac{q}{0.6+0.4q}$.¹⁴ Note that the capital income tax of 40% is in line with U.S. effective tax rates on capital income as calculated by Mendoza, Razin and Tesar (1994) and Domeij and Heathcote (2004). We estimate the key parameters of the model by matching joint moments of consumption and income in an appropriately cleaned cross-sectional data. We then use the estimated parameters and solve the (counterfactual) model with perfect capital taxes, assuming full observability/taxability of capital. The final outcome is a comparison of the optimal labor income taxes between the two scenarios, with a special attention to the change in progressivity.

Data. We use PSID (Panel Study of Income Dynamics) data for 1992 as adapted by Blundell, 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: nondurable 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 in order 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 predicted residual $\hat{\eta}^i$ of this regression is our estimate of the income shock.

The next objective is to find the consumption function. To be able to relate it to the

¹⁴Under the assumption that agents have access to a nonobservable asset with price $q^n = \tilde{q}$, the allocation with a distorted asset price of \tilde{q} coincides with the third best allocation defined in Section 2.2.

cleaned income measure η^i , we postulate that the consumption function is multiplicatively separable as well:

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

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 a separate regression of $\log(c^i)$ on X^i and Z^i . The predicted residual of this regression is $\hat{\varepsilon}^i$. We then use a flexible functional form to obtain $c(\cdot)$. In particular, we estimate the following regression:

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

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

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

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.

Empirical specification. For the quantitative exploration of our model, we move to a formulation with discrete income levels. We assume that we have N levels of second-period income, denoted by y_s , $s = 1, \dots, N$, 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 $\{y_1, \dots, y_N\}$ 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.¹⁵

¹⁵Note that we do not need to impose the monotone likelihood ratio (MLR) property, because 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 given to us by the data. Note that MLR is a sufficient

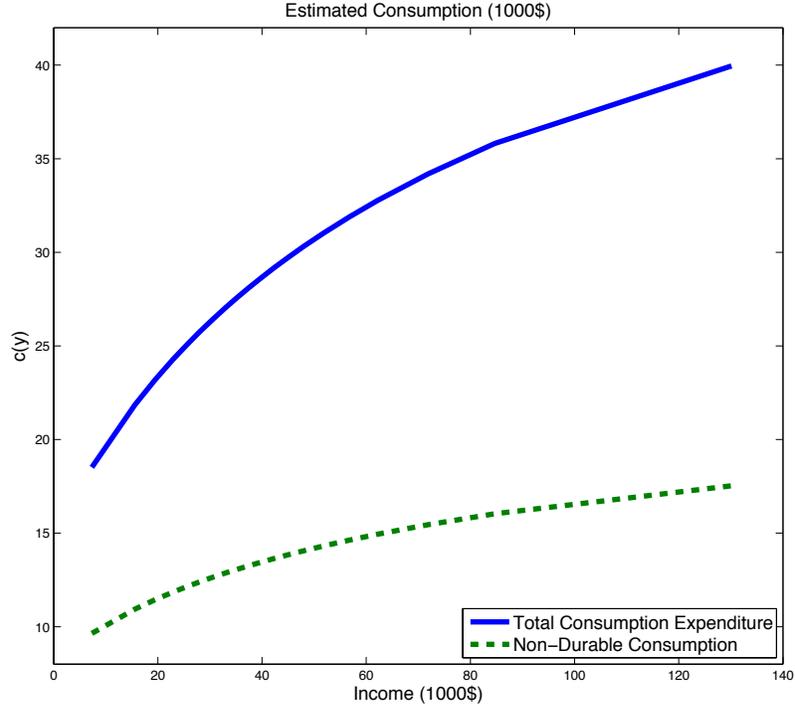


Figure 1: Estimated consumption functions

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

$$u(c, e) = \frac{\left[c^\alpha (v(T - e))^{1-\alpha} \right]^{1-\sigma}}{\alpha(1-\sigma)} \text{ with } 1 > \alpha > 0 \text{ and } \sigma \geq 1,$$

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

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

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

but not necessary condition for monotone consumption. Nevertheless, as expected, our estimated likelihood ratios will exhibit MLR, that is the estimated probability distributions satisfy: π_s^h / π_s^l increasing in s .

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

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

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 $(y_0^i)_i > 0$ and any scalar $\kappa > 0$, let the Pareto weights $(\psi^i)_i$ be such that the solution to the above problem delivers period zero consumption $c_0^{*i} = \kappa y_0^i$ for all i . Then there exists $t^* \in \mathbb{R}$ and individual specific transfers $t^i = t^* y_0^i$ such that $G = \sum_i t^i$ and the solution to the above problem is

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

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

$$\begin{aligned}
\max_{\varepsilon_s, e_0} & \frac{\left[(v(T-e_0))^{1-\alpha} \right]^{1-\sigma}}{\alpha(1-\sigma)} + \beta \sum_s p_s (e_0) \frac{\left[(\varepsilon_s)^\alpha (v(T))^{1-\alpha} \right]^{1-\sigma}}{\alpha(1-\sigma)}; \\
\text{s.t.} \quad & \frac{1}{\kappa} - 1 + q \sum_s p_s (e_0) \left[\frac{\eta_s}{\kappa} - \varepsilon_s \right] \geq t^*; \\
-\frac{(1-\alpha)}{\alpha} \frac{v'(T-e_0)}{v(T-e_0)} \left[(v(T-e_0))^{1-\alpha} \right]^{1-\sigma} &= \beta \sum_s p'_s (e_0) \frac{\left[(\varepsilon_s)^\alpha (v(T))^{1-\alpha} \right]^{1-\sigma}}{\alpha(1-\sigma)}; \\
\tilde{q} \left[(v(T-e_0))^{1-\alpha} \right]^{1-\sigma} &= \beta \sum_s p_s (e_0) \frac{\left[(\varepsilon_s)^\alpha (v(T))^{1-\alpha} \right]^{1-\sigma}}{\varepsilon_s}.
\end{aligned}$$

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 $t^i = t^* y_0^i$ are indeed optimal. However, because of potential non-concavities 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 8 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 computed above as inputs in our estimation procedure. More precisely, the proposition suggests that we can use the values \hat{e}^i and \hat{y}^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

Estimation. It will be convenient to reparametrize and simplify our utility function as

$$u(c, e) = \frac{c^{1-\gamma} (1-e)^{\gamma-\sigma}}{1-\gamma} \text{ with } \gamma < \sigma \text{ and } \gamma \geq 1 \text{ and } 1-\gamma = \alpha(1-\sigma). \quad (19)$$

This formulation has the advantage that γ directly measures the coefficient of risk aversion and the period utility is given by $\frac{c^{1-\gamma}}{1-\gamma}$ after the first period.

As a first step, we fix some parameters. First, we set $q = .96$ to match a yearly real interest rate of 4%, which is the historical average of return on real assets in the USA. We then set the coefficient of relative risk aversion for consumption to 3, that is $\gamma = 3$, in line with recent estimation results by Paravisini, Rappoport, and Ravina (2010).¹⁸ We normalize the 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 η_1, \dots, η_{20} . To be consistent with this choice and

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

¹⁸We provide a 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.

with Proposition 8 we set $y_0 = 1$. For expositional simplicity, we assume $\kappa = 1$ and hence $c_0^* = y_0$. Note that Proposition 8 implies that for any level of κ 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 the fixed parameters, we determine the preference parameters (β, σ) , the effort technology parameter ρ , and the probability weights $\{\pi_s^h, \pi_s^l\}_{s=1}^N$ that determine the likelihood ratios. We estimate these parameters using a method of moments estimator to match specific empirical moments for consumption and income in the data. The optimality conditions of the model give us a sufficient number of restrictions to estimate all the parameters. In particular, we use the planner's optimality conditions for first and second period consumption, the planner's optimality condition for effort, and the agent's optimality conditions for effort and assets. Finally, we obtain the parameter G^* for government consumption as a residual of the estimation procedure implied by the government's budget constraint. Further details on the estimation strategy are provided in Appendix B.

Constrained efficient allocations. We use the preset and estimated parameters of the above model (exogenous capital taxes) to determine the optimal allocation for the counterfactual scenario with perfect capital taxes—assuming full observability/taxability of capital. Figure 2 displays second-period consumption for this scenario together with the consumption function of the benchmark. It is obvious from the picture that the level of second-period consumption is higher in the case with limited/exogenous capital taxes (tax rate on capital income of 40%). This is not surprising, given that perfect capital taxes in general imply front-loaded consumption (Rogerson 1985, Golosov et al. 2003). Note that perfect capital taxes are associated with an Inverse Euler equation, whereas the scenario with limited capital taxes is characterized by a standard Euler equation. By inspecting the planner's optimality condition for consumption (13) we note that the Euler equation generates an additional positive return on the right-hand side ($\xi a(\cdot) > 0$). Intuitively, holding the other multipliers and parameters fixed, this term suggests that the marginal consumption utility in the second period is lower, and consumption thus higher, in the case with limited capital taxes.

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

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

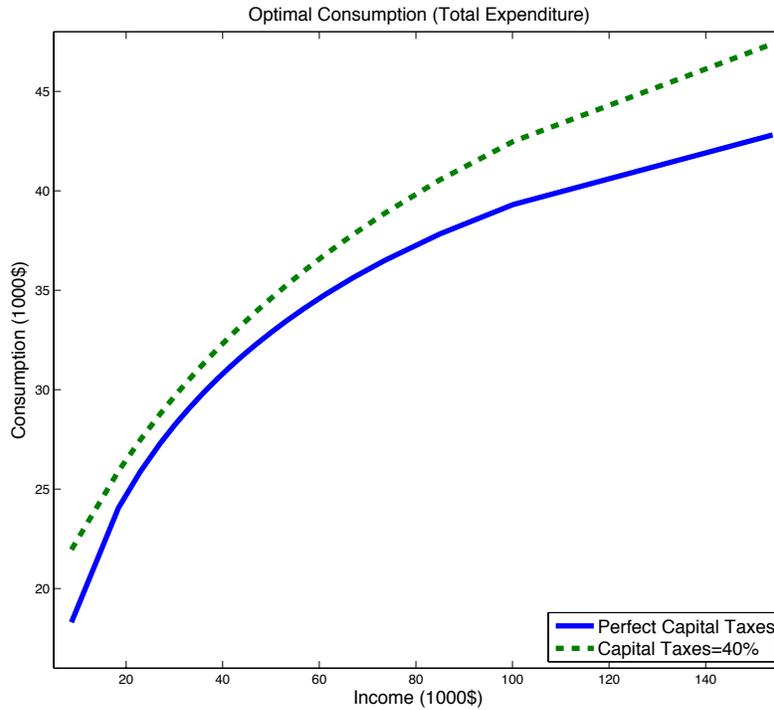


Figure 2: Optimal consumption with perfect and limited capital taxation

second allocation, the first allocation is a concave transformation of the second (assuming that both allocations are monotonically increasing). On Figure 3, we have plotted this measure of progressivity for the optimal consumption scheme when capital taxes are limited and when they are perfect. The pattern is clear: the model with perfect capital taxes results in a uniformly more concave (progressive) consumption scheme compared with the case when capital taxes are limited. The differences are particularly large for lower levels of income.

We have quantified these graphical observations and have checked the 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 logarithmic 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 2 shows, the concavity of the optimal consumption function $c(y)$ changes. Second, the distribution of income changes, as effort is different under perfect capital taxes compared with 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

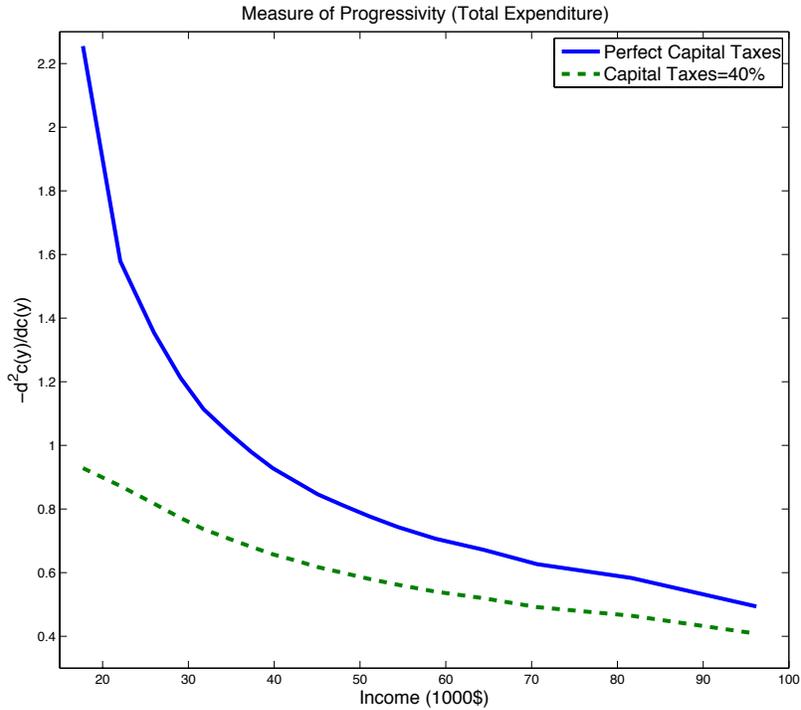


Figure 3: Income tax progressivity with perfect and limited capital taxation

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 perfect capital taxes are levied. In turn, higher effort implies a higher weight on high income realizations where the progressivity differences are lower (see Figure 3). 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 limited capital taxation in consumption equivalent terms (presented in row 7 of Table 1). The losses are negligible for the logarithmic case, considerable for the intermediate cases, and very large for high values of risk aversion.

We have also displayed the implied capital wedges, calculated as $\tau^K = 1 - q/\tilde{q}$, where \tilde{q} is the after-tax asset price in the perfect taxation scenario. Notice that τ^K is indeed the tax rate on the gross return, not on capital income. The 40 percent tax on capital income in the benchmark model is equivalent to a capital wedge of 1.6 percent. It turns out that the capital wedges in the perfect taxation scenario are much higher than this number for all risk aversion levels, including perfect log utility. The wedges are actually *implausibly high*. Even in the

log case, they imply a tax rate on capital income of around 90 percent.²⁰ For our benchmark value of risk aversion, the implied tax rate on capital income is around 1,000 percent.²¹ 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 wedges

Risk aversion	1	2	3	4
Average measure of progressivity ($-c''(y)/c'(y)$)				
Perfect K tax (endog. weights)	0.670	0.800	0.963	1.102
Perfect K tax (exog. weights)	0.670	0.804	0.978	1.141
Limited K tax (40% on K income)	0.644	0.644	0.644	0.644
Welfare loss from limited capital tax (%)				
	0.035	0.295	1.309	3.372
Perfect capital wedge (%)				
$\tau^K = 1 - q/\tilde{q}$	3.74	20.10	39.69	55.18

We can get some intuition why the differences are increasing in the coefficient of relative risk aversion γ by examining the optimality condition for consumption for our specification:

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

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

²⁰Golosov, Troshkin and Tsyvinski (2013) study a dynamic Mirrlees model with logarithmic utility and perfect observability of assets. The capital wedges (and the associated capital income taxes) are similar to the ones we find for the logarithmic case. Farhi and Werning (2013) study a similar Mirrlees model with logarithmic utility and obtain tax rates on capital income that are smaller than ours.

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

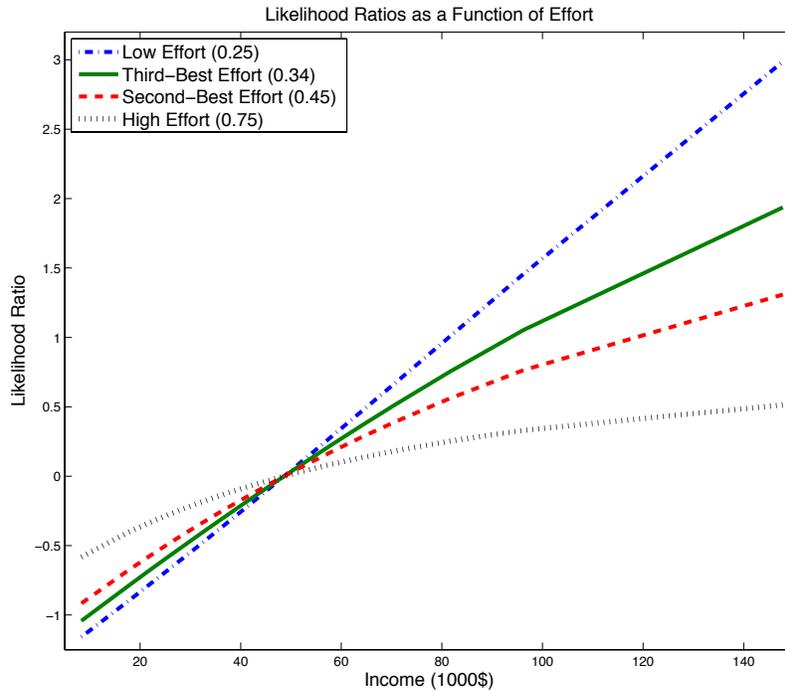


Figure 4: Likelihood ratio function implied by the estimated parameters of the labor income taxation model

argument explains why the welfare costs of limited capital taxation are increasing in risk aversion.

As another robustness check, we examine in Appendix C how the results change if we use nondurable consumption as our measure of consumption. With nondurable consumption, we again have a significant increase in progressivity when we impose perfect 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 is 25 percent, whereas 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.

Finally, we examine the role of endogenous effort for the change in progressivity. Figure 4 plots the likelihood ratio function implied by the estimated parameters. We note that the likelihood ratio function becomes more concave for higher effort levels. Moreover, effort in the second best allocation (perfect capital taxes) is higher than in the third best (limited capital taxes). Hence, the change of the likelihood ratio contributes to the lower degree progressivity in the third best. Note that this effect goes in the same direction as our insights

on the convex cross-sectional returns of relaxing the Euler equation. Therefore, the change in effort between the second- and third-best reinforces our theoretical results concerning the progressivity of the consumption allocations.

4.2 Application II: Human capital accumulation

For this application of the model, we interpret effort e as a human capital investment of the young (between age 16 and 22). We rely on a broad interpretation of investment which includes formal education, on-the-job training and learning-by-doing. In line with our theoretical model, we assume that these early decisions affect the distribution of life-time income agents face later in their life. Naturally, the shape of the tax system—in particular, its progressivity—is going to matter for this decision. We use this observation to identify our parameters.

In 1982 and 1983, the US tax system became significantly less progressive due to President Reagan’s Economic Recovery Tax Act of 1981. This fact is documented in the paper of Gouveia and Strauss (1994), for instance. Gouveia and Strauss approximate the US tax system with a flexible functional form for a number of consecutive years using administrative (tax return) data. Their analysis shows a large drop in progressivity for the years of 1982 and 1983 and then a stabilization of the tax system. The parameter estimates of Gouveia and Strauss have been used by several studies on tax progressivity (e.g., Conesa, Kitao and Krueger, 2009). Guner, Kaygusuz and Ventura (2014) used a similar methodology and obtained that the progressivity of the US tax system roughly remained the same until 2000.

The key assumption for our analysis is that the tax reform was not anticipated. This implies that those agents who had already made their effort (human capital investment) decisions (the ‘old’) assumed that the tax system would remain the same throughout their working life. On the other hand, agents who were still ‘young’ when the reform was introduced, assumed that the new tax system would remain in place for the rest of their working life. Assuming that no other key element of the environment has changed, the differences in (realized cross-sectional) gross life-time income distributions identify the effect of effort on the distribution of life-time income of the individuals.

Below, we first apply our benchmark setup to the human capital investment model. Then, we present the data and the estimation process. Finally, we perform counterfactual exercises to quantitatively evaluate the effect of asset observability.

The human capital model. For this application, we assume agents live for $T + 1$ periods. In period 0, they are young, receive an endowment y_0 , exert human capital effort e_0 , and make a borrowing/saving decision. At the beginning of period 1, they learn the realization of their life-time income and they determine their optimal consumption path for the remaining T periods. This assumption can be interpreted as no uncertainty and no binding constraints after period 1.²² As a consequence, the human capital model is an application of our general setup and periods 1 to T correspond to period 1 of the original setup. In particular, we will denote by y the *gross life-time income* of agents. We assume the same structure and functional form for the probability weights $p_s(\cdot)$ as in the previous subsection for expositional simplicity. We will use the same homothetic preferences as before as well (see equation (19)).

In the model, agents make optimal human capital accumulation decisions taking as given *the existing* capital and labor income taxes. The choice of the optimal path of consumption from period 1 onwards is the solution of the following simple deterministic problem:²³

$$V(\tilde{y}_s^j + b^j) = \max_{\{c_t\}_{t=1}^T} \sum_{t=1}^T \beta^{t-1} \frac{c_t^{1-\gamma}}{1-\gamma} \text{ s.t. } \sum_{t=1}^T \tilde{q}^{t-1} c_t = \tilde{y}_s^j + b^j \text{ for all } 1 \leq s \leq N \text{ and } j = 1, 2.$$

where \tilde{q} is the after-tax price of savings, \tilde{y}_s^j is the net life-time income associated with gross life-time income y_s and tax system j , and b^j is the amount of assets individuals bring into period 1 given tax system (cohort) j , where $j \in \{1, 2\}$.

The problem of the agent *at time 0* is hence given by:²⁴

$$\max_{e,b} \frac{(y_0 - \tilde{q}b)^{1-\gamma} (1-e)^{\gamma-\sigma}}{1-\gamma} + \beta \sum_{s=1}^N p_s(e) V(\tilde{y}_s^j + b).$$

Since agents of different cohorts face different mappings from gross to net life-time incomes, the model induces different effort choices between the two cohorts. This results in different

²²Another interpretation is that they will learn only their expected life-time income in period 1, but there are complete financial markets from period 1 onwards.

²³It is easy to see that

$$V(\tilde{y}_s^j + b^j) = \Gamma \frac{(\tilde{y}_s^j + b^j)^{1-\gamma}}{1-\gamma}, \quad \text{where } \Gamma = \frac{1 - \left(\beta \left(\frac{\beta}{\tilde{q}}\right)^{1/\gamma}\right)^T}{1 - \beta \left(\frac{\beta}{\tilde{q}}\right)^{1/\gamma}} \left(\frac{1 - \tilde{q} \left(\frac{\beta}{\tilde{q}}\right)^{1/\gamma}}{1 - \left(\tilde{q} \left(\frac{\beta}{\tilde{q}}\right)^{1/\gamma}\right)^T} \right)^{1-\gamma}.$$

²⁴Note that we assume that agents have identical initial income y_0 at $t = 0$.

distributions of gross life-time income, which are given by

$$p_s(e^j) = \pi_s^h + \exp(-\rho e^j) (\pi_s^l - \pi_s^h) \quad \text{for all } s. \quad (20)$$

Notice that

$$p_s(e^2) - p_s(e^1) = (\exp(-\rho e^1) - \exp(-\rho e^2)) (\pi_s^h - \pi_s^l). \quad (21)$$

Equation (21) provides the main intuition of our identification: given ρ , for any two effort levels e^1 and e^2 taken by the individuals of different cohorts, the difference between the life-time income distributions identifies the difference between the two base probability distributions π_s^h and π_s^l for all s . Then, from equation (20) we can recover the actual value of these probabilities.

Estimation strategy and data. Our estimation strategy has two main elements. Some parameters we set exogenously. This includes both some normalizations and some parameters where we can use external information. The remaining parameters are directly estimated using the structural model described above.

We set exogenously the following parameters. We consider $N = 20$ levels of life time income. We consider one period as five years and set $T = 8$. We set the intertemporal discount price to $q = (1/1.04)^5$, which corresponds to a pre-tax rate of return of 4 percent annual. The after-tax asset price is $\tilde{q} = 0.885$ and corresponds to a 40 percent tax on interest income.²⁵ The coefficient of relative risk aversion is set at $\gamma = 3$ as before. The remaining parameters are the vectors $\Theta := (\rho, \beta, \sigma, y_0)$ and $\{\pi_s^l, \pi_s^h\}_{s=1}^S$. These parameters are set by matching the life-time income distributions from the NLSY79 data and the (average) level of effort given by the American Time Use Surveys 1985.²⁶ Further details on these moments are provided in Appendix D.

The identification procedure requires data on the dispersion of life-time utilities of two cohorts. For this, we need to find individuals who made their effort (human capital investment decisions) before the tax reform has happened. We will label this cohort as the *old* group. We also need to find a group of individuals who made their key effort decision after the reform was implemented. We need to observe these individuals for a long enough time to be able to compute a relatively reliable measure of life-time income. Luckily, the 1979 wave of the National Longitudinal Survey of the Youth (NLSY79) will satisfy these requirements. The sample contains a large set of individuals who are between 14 and 22 in 1979.

²⁵Formally we set $\tilde{q} = \frac{q}{0.6+0.4q}$. This corresponds to a capital wedge $\tau^K = 1 - q/\tilde{q}$ of 7.1 percent.

²⁶We thank Mark Aguiar for providing us with the data they have compiled for Aguiar and Hurst (2007).

We consider as the *young* group those who were between 14 and 18 in 1979 (17 and 21 in 1982) and we consider the *old* group as the ones who were between 19 and 22 in 1979 (22 and 25 in 1982). This way we maximize the sample size, as we use all cohorts.^{27,28}

After having obtained the annual income data for all individuals for every year (when the agent was alive), we calculate the discounted present value of gross and net life-time income. For discounting, we use the sequence of the nominal interest rates. Note that this also corrects for inflation. For the calculation of net life-time income, we first compute the 'hypothetically expected' net annual income by applying the tax functions estimated by Gouveia and Strauss (1994). In particular, for the old cohort, we use the 1981 tax parameters for all years; for the young cohort, we use the 1983 parameters. This choice is in line with our key identifying assumption: when the old group made their human capital effort decision they expected the old tax system to remain in place. In contrast, the young group expected the new tax system to be in place for all their life time. At this point, we compute the inflation-corrected discounted present value of both gross and net life-time income for all individuals at the age of 23. To make this comparable across age groups, we do two further adjustments. We adjust both for inflation and GDP growth using 1980 as the base year. We perform the growth adjustment to filter out any secular growth in wages, which is supposed to be independent of the choice of human capital effort.

In our model, agents are homogenous: given the exerted effort level, they all face the same income distribution. In the data, however, agents have different ability which can directly affect their life-time income. One of the attractive features of the NLSY79 data is that it contains a measure of ability, the AFQT score. It is also known that race may affect earnings through some form of discrimination. For these reason, we clean the data from variations due to race and ability.²⁹

Recall that taxes have become less progressive due to the tax reform. This can be seen graphically by comparing the implied tax rates for the two cohorts on Figure 5. It is clear from the picture that the younger cohort has faced significantly lower taxes for medium and high level of incomes. In our environment, this should have created increased incentives to exert effort. If this is the case, we should see a shift in the distribution of life-time income towards higher realizations for the second cohort. Figure 6 displays the two distributions. The distribution of life-time income of the young group is indeed shifted towards higher incomes and first-order stochastically dominates the distribution of life-time income of the

²⁷The results do not change in a significant way if we use stricter definitions of the two groups.

²⁸We summarize the main steps of the data processing for the NLSY79 data in Appendix D.

²⁹We provide details on the regressions in Appendix D.

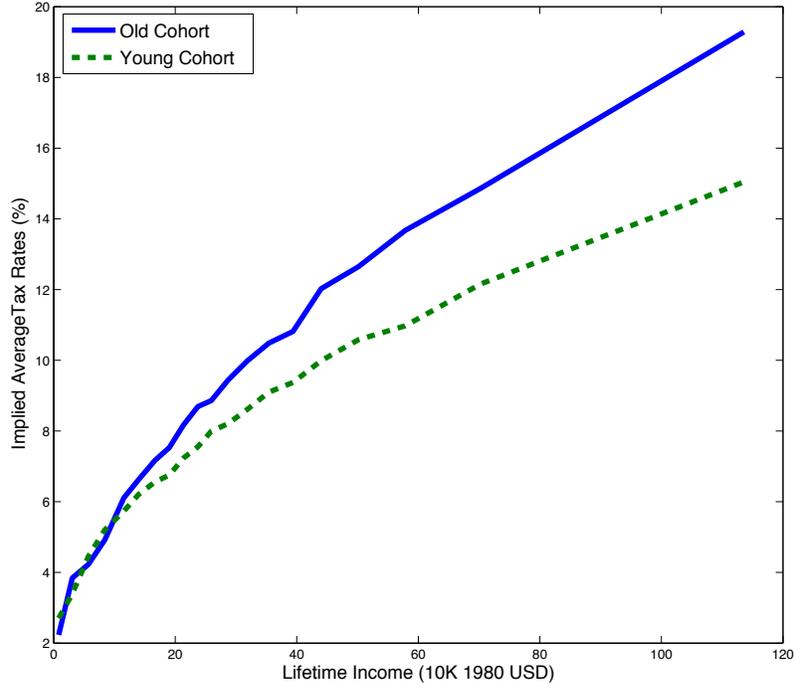


Figure 5: Implied average tax rates on life-time income for the two cohorts

old group.

For our estimation procedure, we define 20 income groups as the percentile groups of the distribution of gross life-time income of the old cohort. We use the (flat) probability density function of the old cohort together with a fitted probability density function for the young group as inputs for the estimation. The means of the 20 percentile groups of the old cohort provide the support for the discrete representation of life-time income distribution. This gives us the income vector $\{y_s\}_{s=1}^{20}$. Then we calculate for both groups the associated net life-time income levels as $\tilde{y}_s^j = \frac{\sum_i \tilde{y}_{i,s}^j y_s}{\sum_i y_{i,s}^j}$, where $\tilde{y}_{i,s}^j$ is the net life-time income of an agent i from group j (old or young) whose gross life-time income belongs to the s -th percentile group of the gross life-time income distribution of the old.

Constrained efficient allocations. Given the estimated parameters, we calculate the optimal allocation of consumption and effort for both hidden and observable asset accumulation. As before, for the second best, we assume that the planner can use capital income taxation without limitations and there are no hidden saving opportunities. For the third best, we assume that there are hidden storage opportunities, with an implied discount price

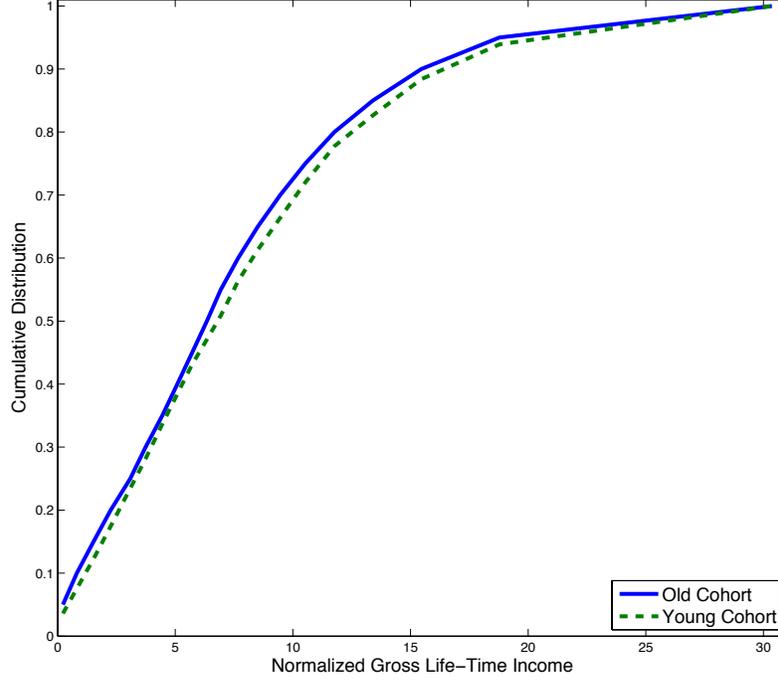


Figure 6: The estimated cumulative distribution of gross life-time income for the two cohorts. Note: Normalized income is defined as the annuized (one period is five years) value of life time income in 10,000\$ of 1980.

of $q^n = \tilde{q}$.

The optimal allocation in this environment is qualitatively very similar to the benchmark setup. Note that there is no uncertainty or any friction for periods $1, \dots, T$, hence consumption will evolve deterministically depending on the ratio q/β . This implies that we can express both the life-time utility and the present discounted value of total consumption as a function of c_1 . We can write the planner's problem as:³⁰

$$\max_{(c, e_0)} \frac{c_0^{1-\gamma} (1 - e_0)^{\gamma-\sigma}}{1 - \gamma} + \beta \Gamma_1 \sum_s \frac{c_s^{1-\gamma}}{1 - \gamma} p_s(e_0)$$

³⁰Here we have, for $i = 1, 2$, $\Gamma_i = \frac{1 - \hat{\Gamma}_i^T}{1 - \hat{\Gamma}_i}$ where $\hat{\Gamma}_1 = \beta \left(\frac{\beta}{q}\right)^{(1-\gamma)/\gamma}$ and $\hat{\Gamma}_2 = q \left(\frac{\beta}{q}\right)^{1/\gamma}$.

$$\begin{aligned}
\text{subject to} \quad & y_0 - c_0 + q \sum_s (y_s - c_s \Gamma_2) p_s(e_0) - G \geq 0 \\
& -\frac{\gamma - \sigma}{1 - \gamma} c_0^{1-\gamma} (1 - e_0)^{\gamma - \sigma - 1} + \rho \exp\{-\rho e_0\} \beta \Gamma_1 \sum_s \frac{c_s^{1-\gamma}}{1 - \gamma} \Delta \pi_s \geq 0 \\
& \tilde{q} c_0^{-\gamma} (1 - e_0)^{\gamma - \sigma} - \beta \sum_s c_s^{-\gamma} p_s(e_0) \geq 0.
\end{aligned}$$

Of course, the second best solves the same problem without imposing the last constraint.³¹

We are now ready to investigate the effect of asset observability on the optimal allocations. Similar to the previous application, we make these comparisons based on consumption (in period 1) and using the measure of progressivity introduced in the previous subsection. Figure 7 displays the optimal allocation of period-1 consumption under hidden and observable assets. We observe the same pattern as before: the allocation under hidden assets is clearly less concave, implying that the associated income taxes are less progressive. As before, we find that the difference in progressivity is shrinking in (life-time) income. However, in this application the difference in progressivity is of a smaller magnitude.³² In Table 2, we provide some quantitative measures for the comparison between these two allocations. Note that although the difference in progressivity is smaller, the welfare gains of asset observability and the capital wedge have a very similar magnitude as in the previous application.

Table 2: Quantitative measures of progressivity, welfare loss and capital wedge (risk aversion=3)

Average progressivity ($-c''(y)/c'(y)$)	
Perfect K tax (observable assets)	0.173
Limited K tax (hidden assets)	0.154
Welfare loss (%)	1.478
Perfect capital wedge (%)	
$\tau^K = 1 - q/\tilde{q}$	48.29

³¹Note that our estimation procedure above delivers the implied value of government consumption as the discounted expected (average) difference between gross and net life-time income of the young cohort: $G = q \sum_{s=1}^N p_s(e^2) (y_s - \tilde{y}_s)$.

³²The estimated likelihood ratio function for this application exhibits the same qualitative pattern as that displayed in Figure 4, but the effect of effort on the curvature of the likelihood ratio in y is smaller.

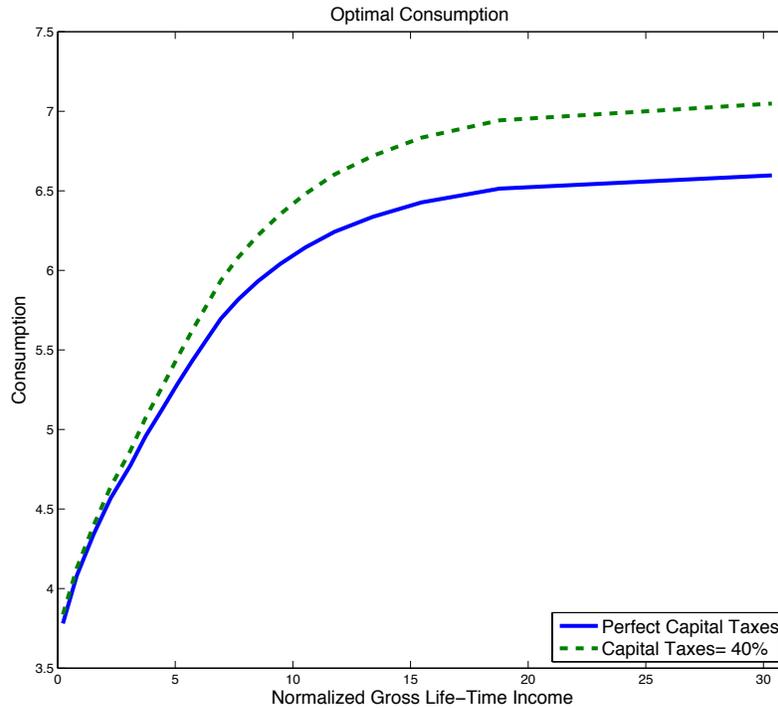


Figure 7: Optimal period-1 consumption with perfect and limited capital taxation

4.3 Summary of the quantitative results

We have provided two simple quantitative applications to explore the effect of asset observability on the progressivity of optimal taxes. On the one hand, these two applications were different in terms of the identification strategy, the data we used and in terms of their economic interpretation. On the other hand, they used similar (and standard) functional forms. Despite all these differences, we got very similar qualitative results. (i) Hidden assets will lead to an optimal tax scheme which is considerably less progressive. (ii) The welfare costs of hidden assets are significant. (iii) The model with observable assets requires implausibly high taxes on capital (or capital income).

We also note that in both applications (and under both asset observability scenarios) the optimal income taxes are significantly more progressive than actual taxes. This observation probably indicates that actual governments face additional constraints in addition to moral hazard and unobserved assets. Potential candidates for these additional frictions could be income tax evasion and political economy constraints. The potential interaction of these frictions with moral hazard and hidden assets is left for future research.

5 Concluding remarks

This paper analyzed how limitations 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 perfectly controlled by the planner. In terms of our decentralization, this implies that taxes on labor income become less progressive when limitations to capital income taxation are binding. We complemented our theoretical results with a quantitative analysis based on two different estimation strategies. The first one used individual level U.S. data on consumption and income. The second one relied on a U.S. tax reform that affected the progressivity of labor income taxes.

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 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).³³

³³In 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, we obtain the formula for third-best allocations as:

$$\frac{q\lambda}{\beta u'(c(y))} = 1 + \mu l(y; e) + \xi a(c(y)) - \frac{d\phi(y)}{\beta f(y; e)}, \quad (22)$$

where the multiplier associated to the intensive-margin incentive-constraint $\phi(y)$ is related to the Spence-Mirrlees condition and the labor supply distortion, and it satisfies $\phi(\underline{y}) = \phi(\bar{y}) = 0$. The comparison between the case with restricted and unrestricted capital taxation amounts again to considering the cases with $\xi > 0$ and

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] AGUIAR, M. AND E. HURST (2007): "Measuring Trends In Leisure: The Allocation Of Time Over Five Decades", *Quarterly Journal Of Economics*, 122(3), 969-1006.
- [3] BLUNDELL R, PISTAFERRI L., PRESTON I. (2008): "Consumption inequality and partial insurance." *American Economic Review*, 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. MIRRLI (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," *Quarterly Journal of Economics*, 95(2), 357–64.
- [11] FARHI, E., AND I. WERNING (2013): "Insurance and Taxation over the Life Cycle," *Review of Economic Studies*, 80(2), 596-635.
- [12] GAYLE G.-L., AND R. MILLER (2009): "Has Moral Hazard Become a More Important Factor in Managerial Compensation?" *American Economic Review* , 99(5), 1740–1769.
- [13] GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKI (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70(3), 569–587.

$\xi = 0$ respectively. Although the forces at play are the same as above, an analytic analysis with an intensive margin (and private information on y) is complicated by the fact that both λ, μ , and the schedule $\phi(\cdot)$ change.

- [14] GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (2013): "Redistribution and Social Insurance," *mimeo*.
- [15] GOLOSOV, M., AND A. TSYVINSKI (2007): "Optimal Taxation with Endogenous Insurance Markets," *Quarterly Journal of Economics*, 122(2), 487–534.
- [16] GOUVEIA, M. AND R. STRAUSS (1994): "Effective Federal Individual Income Tax Functions: An Exploratory Empirical Analysis," *National Tax Journal*, 47(2), 317–339.
- [17] GROSSMAN, S. J., AND O. D. HART (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51(1), 7–45.
- [18] GUIISO, L., AND M. PAIELLA (2008): "Risk Aversion, Wealth, and Background Risk," *Journal of the European Economic Association*, 6(6), 1109–1150.
- [19] GUNER, N., R. KAYGUSUZ AND G. VENTURA (2014): "Income Taxation of US Households: Facts and Parametric Estimates," *Review of Economic Dynamics*, forthcoming, <http://dx.doi.org/10.1016/j.red.2014.01.003>.
- [20] HOLMSTROM, B. (1979): "Moral hazard and observability," *Bell Journal of Economics*, 10(1), 74–91.
- [21] KOCHERLAKOTA, N. R. (2005): "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica*, 73(5), 1587–1621.
- [22] LAROQUE, G. (2011): "On Income and Wealth Taxation in A Life-Cycle Model with Extensive Labour Supply," *Economic Journal*, 121(554), F144–F161.
- [23] 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.
- [24] MIRRLEES, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(114), 175–208.
- [25] ROGERSON, W. (1985), "Repeated Moral Hazard," *Econometrica*, 53: 69–76.
- [26] SADKA, E. (1976): "On Progressive Income Taxation," *American Economic Review*, 66(5), 931–35.
- [27] SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–29.
- [28] VARIAN, H. R. (1980): "Redistributive taxation as social insurance," *Journal of Public Economics*, 14(1), 49–68.

Appendices

A Proofs omitted from the main text

Proof of Proposition 2. Fix q^n . From the Kuhn-Tucker theorem we have $\xi \geq 0$. If $\xi > 0$, we are done. If $\xi = 0$, 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, because the agent cannot be fully insured when effort is interior. Since we have $\lambda > 0$ from the previous condition, we conclude

$$\beta \int_{\underline{y}}^{\bar{y}} u'_1(c(y)) f(y, e_0) dy > q u'_c(c_0, e_0). \quad (23)$$

Clearly, exactly the same allocation delivering condition (23) is obtainable for all q^n by ignoring the agent's Euler equation. If we now define $\bar{q} > q$ such that

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

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

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

$$g(c) := \frac{\lambda q}{\beta u'_1(c)} - \xi a(c).$$

By concavity of u , $\frac{1}{u'_1(\cdot)}$ is increasing. Therefore, if $\frac{1}{u'_1(\cdot)}$ is convex and $\xi = 0$ (or $\xi > 0$ and $a(\cdot)$ decreasing and concave), then $g(\cdot)$ is increasing and convex. Given the validity of the FOA, equation (8) (or equation (13), respectively) shows that second best (third best) consumption schemes are characterized as follows:

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

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.**

Proof of Proposition 6. The first-order conditions for consumption imply

$$g^{sb}(c^{sb}(y)) = 1 + \mu^{sb} l(y, e_0^{sb}), \quad (24)$$

$$g^{tb}(c^{tb}(y)) = 1 + \mu^{tb} l(y, e_0^{tb}), \quad (25)$$

where the functions g^{sb} and g^{tb} are defined as in (14) and (15), respectively. First, suppose that c^{tb} is a concave transformation of $l(y, e)$. Since the right-hand side of (25) is a positive affine transformation of $l(y, e)$, this is equivalent to the condition that g^{tb} is convex. Now, since $a(c)$ is convex by assumption, convexity of g^{tb} is sufficient (but not necessary) for $g^{sb}(c) = (g^{tb}(c) + \zeta^{tb} a(c)) \lambda^{sb} / \lambda^{tb}$ being convex as well. Finally, using (24) we note that g^{sb} is convex if and only if c^{sb} is a concave transformation of $l(y, e)$.

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

Proof of Lemma 1. Simple algebra shows

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

If u belongs to the HARA class, we obtain

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

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

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

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

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

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

Proof of Proposition 8. 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:

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

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

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

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

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

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 $c_0^{j*} = \kappa y_0^j$. By non-satiation of preferences, t^j is given by

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

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

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

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

This implies

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

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

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

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

B Estimation of the labor income taxation model

Given the fixed parameters, the first group of remaining parameters of the model are the effort technology parameter 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 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:

$$p_s(e_0^*) = \exp(-\rho e_0^*) \pi_s^l + (1 - \exp(-\rho e_0^*)) \pi_s^h \text{ for } s = 1, \dots, N - 1, \quad (26)$$

$$\frac{q}{\beta} \lambda^*(\varepsilon_s^*)^\gamma = 1 + \mu^* \rho \frac{\exp(-\rho e_0^*) (\pi_s^h - \pi_s^l)}{p_s(e_0^*)} + \xi^* \frac{\gamma}{\varepsilon_s^*} \text{ for } s = 1, \dots, N, \quad (27)$$

where (27) 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 the parameters σ 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. Given risk aversion γ , the parameter α determines a value for σ due to the restriction $1 - \gamma = \alpha(1 - \sigma)$. Also notice that, given $p_s(e_0^*) = 1/20$ and

$\varepsilon_s^* = \hat{c}(\eta_s)$ for all s , if we sum equation (27) across income levels using weights as $p_s(e_0^*) = 1/20$ we obtain

$$\frac{q}{\beta} \lambda^* \frac{1}{20} \sum_{s=1}^{20} \hat{c}(\eta_s)^\gamma = 1 + \frac{\xi^* \gamma}{20} \sum_{s=1}^{20} \frac{1}{\hat{c}(\eta_s)}. \quad (28)$$

Consequently, the data implies a further restriction between the parameters and endogenous variables $(\beta, \sigma, \lambda^*, \xi^*)$, 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):

$$\tilde{q} (1 - e_0^*)^{\gamma - \sigma} = \beta \sum_{s=1}^N p_s(e_0^*) (\varepsilon_s^*)^{-\gamma}. \quad (29)$$

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

$$\frac{\gamma - \sigma}{1 - \gamma} \frac{(1 - e_0^*)^{\gamma - \sigma}}{1 - e_0^*} = \beta \rho \exp(-\rho e_0^*) \sum_{s=1}^N (\pi_s^h - \pi_s^l) \frac{(\varepsilon_s^*)^{1 - \gamma}}{1 - \gamma}, \quad (30)$$

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

$$\frac{\lambda^*}{(1 - e_0^*)^{\gamma - \sigma}} = 1 - \xi^* \tilde{q} \gamma - \mu^* \frac{\gamma - \sigma}{(1 - e_0^*)}, \quad (31)$$

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

$$q \lambda^* \sum_s p'_s(e_0^*) (\eta_s - \varepsilon_s^*) + \mu^* \left(\beta \sum_s p''_s(e_0^*) \frac{(\varepsilon_s^*)^{1 - \gamma}}{1 - \gamma} - \frac{(\gamma - \sigma)(\gamma - \sigma - 1)}{1 - \gamma} (1 - e_0^*)^{\gamma - \sigma - 2} \right) + \xi^* \left(-\beta \sum_i p'_s(e_0^*) \varepsilon_s^{-\gamma} - \tilde{q} (\gamma - \sigma) (1 - e_0^*)^{\gamma - \sigma - 1} \right) = 0. \quad (32)$$

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

$$G^* = q \left(\sum_i \gamma y_0^i \right) \sum_{s=1}^N p_s(e_0^*) (\eta_s - \varepsilon_s^*). \quad (33)$$

Here we have used $y_0 - c_0^* = 0$, the unit root process of income and Proposition 8.

C Robustness exercise for the labor income taxation model: nondurable consumption data

As another robustness check, we examined how the results would change if we use nondurable consumption as our measure of consumption. As we have seen in Figure 1, the main difference between the two consumption measures is that nondurable consumption is less dispersed (the average slope is significantly lower). Table 3 contains the average measures of progressivity, capital wedges and the welfare losses of limited capital taxa-

tion for the benchmark risk aversion case. With nondurable consumption, we again have a significant increase in progressivity when we impose perfect capital taxes. This once more implies a sizeable welfare gain and an implausibly high tax rate on capital.

Table 3: Different consumption measures

Risk aversion = 3	nondurable	total expenditure
Average measure of progressivity ($-c''(y)/c'(y)$)		
Perfect K tax (endog. weights)	0.849	0.963
Perfect K tax (equal weights)	0.853	0.978
Limited K tax (40% on K income)	0.687	0.644
Welfare loss from limited capital tax (%)		
	0.434	1.309
Perfect capital wedge (%)		
$\tau^K = 1 - q/\tilde{q}$	27.03	39.69

Hence, we can conclude that the following three main points of our analysis are robust to different measures of consumption (and to different levels of risk aversion). (i) Limited (as opposed to perfect) capital taxation leads to less progressive optimal income taxes. (ii) There are significant welfare losses due to limited capital taxation. (iii) The perfect capital taxes are implausibly high.

D Estimation of the human capital model

Given the preset parameters, we estimate the remaining parameters by matching moments. The set of parameters to be estimated is represented by the following two vectors: $\Theta := (\rho, \beta, \sigma, y_0)$ and $\{\pi_s^l, \pi_s^h\}_{s=1}^S$. First of all, as we explain in the main text, given e_1, e_2 , and ρ , we can recover the baseline probability vectors $\{\pi_s^l, \pi_s^h\}_{s=1}^S$ using (20) and (21) by matching the life-time income distribution of the two cohorts. This is what we do: we match the two life-time income distributions exactly. However, it is easy to see that conditions (20) and (21) do not suffice to guarantee that the values for π_s^l, π_s^h represent probabilities. This will be true only for certain values of ρ, e_1 , and e_2 . We hence look for the values of the parameters Θ such that:

- The baseline probability vectors $\{\pi_s^l, \pi_s^h\}_{s=1}^S$ match exactly the lifetime income distributions of both cohorts and satisfy the following set of inequalities: $0 \leq \pi_s^h \leq 1$ and $0 \leq \pi_s^l \leq 1$ for all s .
- Neither cohort saves in period zero, that is $c_0^j = y_0$ for both cohorts.³⁴
- The sum of squares of the following five moment restrictions is minimized:
 - 1 & 2) The two effort first-order conditions for $j = 1, 2$:

$$0 = \frac{\gamma - \sigma}{1 - \gamma} c_0^{1-\gamma} (1 - e_j)^{\gamma - \sigma - 1} - \beta \rho \exp(-\rho e_j) \sum_{s=1}^N (\pi_s^h - \pi_s^l) V(\tilde{y}_s^j + b_j) \quad (34)$$

³⁴Note that this implies that neither the value $V(\tilde{y}_s^j + b_j)$ nor the associated consumption plan depend on the agent's choice variables (effort, saving), but only on data of life-time after-tax income.

3 & 4) The Euler equation satisfied with equality for the first, and the Euler equation satisfied possibly with inequality for the second cohort:³⁵

$$0 = c_0^{-\gamma} (1 - e_1)^{\gamma-\sigma} - \frac{\beta}{\bar{q}} \sum_{s=1}^N p_s(e_1) c_{s,1}^{-\gamma} \quad (35)$$

$$0 = \min \left\{ 0, c_0^{-\gamma} (1 - e_2)^{\gamma-\sigma} - \frac{\beta}{\bar{q}} \sum_{s=1}^N p_s(e_2) c_{s,2}^{-\gamma} \right\} \quad (36)$$

5) The distance between effort of 16–22 year old individuals in the data (American Time Use Surveys 1985; see below for further details) and arithmetic average effort in the model.

Next, we summarize the main steps of the data processing for the NLSY79 data. Given that the sample period is characterized by a large increase in female labor force participation and a large reduction of the male-female wage gap (which our model abstracts from), we focus on males.

The main step of the procedure is the computation of (realized) individual life-time income in the sample. This required some imputations of annual income for some years for some individuals. First, the survey was conducted every year until 1994, but then only biannually between 1996 and 2010, so we need to impute annual income for the missing years. For these cases, we have used linear interpolations. Second, the individuals in the sample were between 45 and 53 years in 2010, so we need to impute their income for the remaining years before retirement. Due to time discounting, income in the period close to retirement has only a modest quantitative effect on life-time incomes and human capital investments. We assumed that earnings remain at the same level after the last available data point until the agent reaches age 65. Finally, there is some missing income data; here we used linear interpolation of available income data.

As we explain in the main text, from the annual income data for every agent for every year (when the agent was alive), we calculate the discounted present value of gross and net life-time income using the sequence of the nominal interest rates to discount future values. The net life-time income has been computed using the ‘hypothetically expected’ net annual income by applying the tax functions estimated by Gouveia and Strauss (1994). In particular, we used the 1981 tax parameters for the old cohort and the 1983 parameters for the young cohort. We have thus calculated the inflation-corrected discounted present value of both gross and net life-time income for all individuals at the age of 23. To make this comparable across age groups, we adjusted both for inflation and GDP growth using 1980 as the base year. Moreover, we performed the growth adjustment to filter out any secular growth in wages.

One of the attractive features of the NLSY79 data is that it contains a measure of ability, the AFQT score. We have run regressions of the following form:

$$\log(y_i) = \alpha + \beta' X_i + \varepsilon_i \quad \text{and} \quad \log(\tilde{y}_i) = \nu + \tilde{\beta}' X_i + u_i, \quad (37)$$

³⁵This will effectively imply that the first cohort is just indifferent between saving and borrowing but the second cohort is borrowing constrained in equilibrium. Given the properties of the tax reform, it is easy to see indeed that the desire to save is always higher for the first cohort.

where y_i (\tilde{y}_i) is the gross (net) life-time income of individual i . The independent variables include the AFQT score and dummies for blacks and hispanics. Then the cleaned measures of lifetime income are obtained as

$$Y_i = \exp(\alpha + \beta' \bar{X} + \hat{\varepsilon}_i) \quad \text{and} \quad \tilde{Y}_i = \exp(\nu + \tilde{\beta}' \bar{X} + \hat{u}_i),$$

where $\hat{\varepsilon}_i$ and \hat{u}_i are the predicted residuals of the regressions in (37) and \bar{X} contains the mean values of the independent variables.

Our estimation also uses a data moment on the level of effort exerted by young agents. The NLSY79 does not provide any direct information on the effort devoted to human capital investment. For this reason, we used actual time use data from the American Time Use Surveys to calculate average effort. We use core market work time (this variable excludes work related but not human capital enhancing time use such as commuting and meal breaks) and education as the measure of human capital effort for the 16–22 age group. The time use survey closest to the tax reform was conducted in 1985 and we used this survey for our data target. The average share of time spent on core market work and education in the 1985 survey equals 0.21.