Optimal Income Taxation with Asset Accumulation

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

April 13, 2011

Abstract

Several frictions restrict the government’s ability to tax capital income. First of all, it is very costly to monitor trades on international asset markets. Moreover, agents can resort to non-taxable low-return assets such as cash, gold or foreign currencies if taxes on capital income become too high. This paper shows that limitations to capital taxation have important consequences for the taxation of labor income. Using a dynamic moral hazard model of social insurance, we find that optimal labor income taxes typically become less progressive when capital taxation is restricted. We evaluate the effect quantitatively in a model calibrated to U.S. data and find evidence that restrictions to capital income taxation appear to be binding in the United States.

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

1 Introduction

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

The existing analyses of optimal income tax progressivity have largely focused on models where labor is the only source of income. In the present paper, we argue that the optimal shape of labor income taxes cannot be determined in isolation from the tax code on capital income. Specifically,
we show that when the government’s ability to tax capital income is limited, then optimal labor income taxes become less progressive.

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

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

A similar problem applies to transitional and developing countries, where foreign currency, gold, or other non-observable assets are often used for self-insurance. Motivated by these considerations, we contrast two stylized environments in this paper. In the first one, consumption and asset decisions are perfectly observable (and contractable) for the government. In the second environment, these choices are private information. We compare the constrained efficient allocations of the two scenarios. When absolute risk-aversion is convex, we find that optimal consumption in the scenario with hidden asset decisions moves in a less concave (or more convex) way with labor income. In this sense, the optimal allocation becomes less progressive in that scenario. This finding can be easily rephrased in terms of the progressivity of labor income taxes, since our model allows for a straightforward decentralization: optimal allocations can be implemented by letting agents pay nonlinear taxes on labor income and linear taxes on capital income (compare Gottardi and Pavoni 2010).\footnote{In the scenario with hidden assets, the tax rate on capital income is zero, of course.}

The decreased concavity of optimal consumption with respect to labor income then translates into marginal tax rates on labor income that increase less quickly.

We derive our results in a two-period model of social insurance. A continuum of ex-ante identical agents influence their labor incomes by exerting effort. Labor income realizations are not perfectly controllable, which creates a moral hazard problem. The social planner thus faces a trade-off between insuring agents against idiosyncratic income uncertainty on the one hand and the associated disincentive effects on the other hand. In addition, agents have access to a risk-free bond, which gives them a limited means for self-insurance. In this model, the planner wants to distort agents’ asset decisions, because the bond provides insurance against the realization of labor income and thereby reduces the incentives to exert effort.\footnote{See Diamond and Mirrlees (1978), Rogerson (1985), and Golosov, Kocherlakota and Tsyvinski (2003).}

Using the first-order approach (Abraham, Koehne and Pavoni 2010), we can switch from the observable asset case to the scenario with hidden asset accumulation by adding a Euler equation for...
the agent to the principal’s optimization problem. This constraint crucially changes the allocation of consumption across income states. Efficiency requires that for each state the costs of increasing the agent’s utility by a marginal unit equal the benefits of doing so. Due to the Euler equation, it becomes important how such changes in utility affect the agent’s marginal utility. One can show that a marginal increase of utility in a state with consumption \( c \) reduces the agent’s marginal utility in that state by \(-u''(c)/u'(c)\).\(^3\) This relaxes the Euler equation and thereby modifies how the gains of allocating utility vary in the cross-section. Obviously, the Euler equation affects the costs and benefits of allocating utility also by changing the shadow costs of the remaining constraints of the principal’s problem. However, we show that the former effect is key. If absolute risk-aversion is convex, we thus find that optimal consumption becomes a more convex function of labor income when asset accumulation is not observable. Put differently, marginal taxes on labor income become less progressive when capital income cannot be sufficiently taxed.

In a quantitative exercise, we estimate some of the key parameters of the model. We use consumption and income data from the PSID (Panel Study of Income Dynamics) as adapted by Blundell, Pistaferri and Preston (2008) and postulate that the data is generated by a tax system in which labor income taxes are set optimally given a capital income tax rate of 40\%.\(^4\) Using the implied parameters, we compute the optimal allocation when capital income taxation is unrestricted and compare it to the data. Under unrestricted capital taxation, the progressivity of the optimal allocation increases sizeably. Moreover, welfare increases by between 0.001\% and 13.23\% in consumption equivalent terms, depending on the coefficient of relative risk aversion. The required capital income tax rates are implausibly high, however, exceeding one hundred per cent for all specifications. This indicates that limitations to capital income taxation may be binding in the United States.

To the best of our knowledge, this is the first paper that examines how limits to capital taxation affect the optimal labor income tax code. Recent work on dynamic Mirrleesian economies studies optimal income taxation when capital taxation is unrestricted; see Golosov, Troshkin, and Tsyvinski (2009), and Farhi and Werning (2010). In that literature, the reason for capital taxation is similar to our model and stems from disincentive effects associated with the accumulation of wealth. While the Mirrlees (1971) framework focuses on redistribution in a population with heterogeneous skills, our approach highlights the social insurance (or ex-post redistribution) aspect of income taxation. In spirit, our model is therefore closer to the works by Varian (1980) and Eaton and Rosen (1980).

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

\(^4\)This rate is in line with U.S. effective tax rates on capital income calculated by Mendoza, Razin and Tesar (1994), and Domeij and Heathcote (2004).
An entirely different link between labor income and capital income taxation is explored by Conesa, Kitao and Krueger (2009). Using a life-cycle model with time-varying labor supply elasticities, they argue that capital income taxes and progressive labor income taxes are two alternative ways of mimicking age-dependent taxation. They then use numerical methods to determine the efficient relation between the two instruments. Interestingly, in the present environment we obtain a very different conclusion. While Conesa, Kitao and Krueger (2009) argue that capital income taxes and progressive labor income taxes are essentially substitutes, in our model they are complements. Laroque (2010) derives analytically a similar substitutability between labor income and capital income taxes, restricting labor taxation to be nonlinear but homogenous across age groups. In both these cases, the substitutability arises because exogenous taxes are in general imperfect instruments to perform redistribution. Since wealth is typically positively correlated to skill, it is optimal to compensate the imperfect distribution coming from labor income taxation with capital income taxation.

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

2 Model

Consider a benevolent social planner (the principal) whose objective is to maximize the welfare of the citizens. The (small open) economy consists of a continuum of ex-ante identical agents who live for two periods, \( t = 0, 1 \), and can influence their date-1 labor income realizations by exerting effort. The planner offers a tax/transfer system to insure them against idiosyncratic risk and provide them appropriate incentives for working hard. The planner’s budget must be (intertemporally) balanced.

Preferences The agent derives utility from consumption \( c_t \geq c \geq -\infty \) 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: \( u''_{c e}(c_t, e_t) \geq 0 \). This specification of preferences includes both the additively separable case, \( u(c, e) = u(c) - v(e) \), and the case with monetary costs of effort, \( u(c - v(e)) \), assuming \( v \) is strictly increasing and convex. The agent’s discount factor is denoted by \( \beta > 0 \).
Production and endowments  At date \( t = 0 \), the agent has a fixed endowment \( y_0 \). At date \( t = 1 \), the agent has a stochastic income \( y \in Y := [y_l, y_h] \). The realization of \( y \) is publicly observable, while the probability distribution over \( Y \) is affected by the agent’s unobservable effort level \( e_0 \) that is exerted at \( t = 0 \). The probability density of this distribution is given by the smooth function \( f(y, e_0) \). As in most of the the optimal contracting literature, we assume full support, that is \( f(y, e_0) > 0 \) for all \( y, e_0 \). There is no production or any other action at \( t \geq 2 \). Since utility is strictly decreasing in effort, the agent exerts effort \( e_1 = 0 \) at date 1. In what follows, we therefore use the notation \( u_1(c) := u(c, 0) \) for date-1 utility.

Allocations  An allocation \( (c, e_0) \) 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 income realization \( y \). An allocation \( (c_0, c(\cdot), e_0) \) is called feasible if it satisfies the planner’s budget constraint

\[
y_0 - c_0 + q \int_{y_l}^{y_h} (y - c(y)) f(y, e_0) dy - G \geq 0,
\]

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

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

\[
\max_{(c, e_0)} u(c_0, e_0) + \beta \int_{y_l}^{y_h} u_1(c(y)) f(y, e_0) dy
\]

subject to \( c_0 \geq c, c(y) \geq c, e_0 \geq 0 \), the planner’s budget constraint

\[
y_0 - c_0 + q \int_{y_l}^{y_h} (y - c(y)) f(y, e_0) dy - G \geq 0
\]

and the incentive compatibility constraint for effort

\[
e_0 \in \arg \max_e u(c_0, e) + \beta \int_{y_l}^{y_h} u_1(c(y)) f(y, e) dy.
\]

2.1 Decentralization and the first-order approach

Any second best allocation can be generated as an equilibrium outcome of a competitive environment where agents exert effort and trade bonds facing an appropriate tax system. To simplify the analysis, we assume throughout this paper that the first-order approach (FOA) is valid. This enables us to characterize the agent’s choice of effort \( e_0 \) and bond trades \( b_0 \) based on the associated
first-order conditions. Sufficient conditions for the validity of the FOA in this setup are given in Abraham, Koehne, and Pavoni (2010). Specifically, the FOA is valid if the agent has nonincreasing absolute risk aversion and the cumulative distribution function of income is log-convex in effort.\(^5\)

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

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

\[
\begin{align*}
c_0 &= y_0 + \tau_0, \\
c(y) &= y + \tau(y), \quad y \in Y, \\
(e_0, 0) &\in \arg \max_{(e,b)} u(y_0 + \tau_0 - \tilde{q}b, e_0) + \beta \int_y y u_1(y + \tau(y) + b)f(y, e_0) dy. \quad (5)
\end{align*}
\]

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

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

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

\[
\begin{align*}
u'_c(y_0 + \tau_0, e_0) + \beta \int_y y u_1(y + \tau(y))f_c(y, e_0) dy &\geq 0, \quad (6) \\
\tilde{q}u'_c(y_0 + \tau_0, e_0) - \beta \int_y y u'_1(y + \tau(y))f(y, e_0) dy &\geq 0. \quad (7)
\end{align*}
\]

The second condition determines the agent’s asset decision based exclusively on consumption levels and the price \(\tilde{q}\). This means that the planner can essentially ignore the problem of joint deviations when taxing asset trades.

\(^5\)As argued by Abraham, Koehne, and Pavoni (2010), both conditions have quite a broad empirical support. First, virtually all estimations for \(u\) reveal NIARA; see Guiso and Paiella (2008) for example. The condition on the distribution function essentially restricts the agent’s Frisch elasticity of labor supply. This restriction is satisfied as long as the Frisch elasticity is smaller than unity. In fact, most empirical studies find values for this elasticity between 0 and 0.5; see Domeij and Floden (2006), for instance.
Notice that we have normalized asset holdings to \( b_0 = 0 \) in the above proposition. This is without loss of generality, since there is an indeterminacy between \( 0 \) 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 \( \varepsilon \). This indeterminacy is of course not surprising, because the timing of tax collection is irrelevant by the Ricardian equivalence result.

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

\[
\frac{\lambda}{u'_c(c_0, e_0)} = 1 + \mu \frac{u''_c(c_0, e_0)}{u'_c(c_0, e_0)},
\]

(8)

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

(9)

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

Finally, we note that a tax on the bond price is equivalent to a tax rate \( t \) on the bond return (constant across agents) given by:

\[
\left(1 + \left(\frac{1}{\tilde{q}} - 1\right)(1 - t)\right)^{-1} = \tilde{q}.
\]

### 2.2 Frictions and third best allocations

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

Notice that, even though we focus on a particular decentralization mechanism in this paper, the above problem is general. Decentralizations that allow asset taxes to depend on the agent’s period-1 income realization (Kocherlakota 2005), for instance, can generate zero asset taxes on average, but generally require high tax rates for a sizable part of the population.\(^7\)

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

---

\(^{6}\)A sufficient condition for interiority is, for example, \( u'_c(c_0, e_0) = 0 \) for all \( c > c \) in combination with the Inada condition \( \lim_{c \to c} u''_c(c, 0) = \infty \).

\(^{7}\)For example, assuming additively separable preferences and CRRA consumption utility, the tax rate on asset holdings in such a decentralization would be \( 1 - \frac{q}{\tilde{q}} \left( \frac{c(y)}{c_0} \right)^\gamma \). For incentive reasons, \( c(y) \) tends to be significantly below \( c_0 \) for a range of income levels \( y \), which results in tax rates on assets close to 1 at those income levels. In other words, almost their entire wealth (not just asset income) would be taxed away for those agents.
consider the polar case where the bond price cannot be taxed at all, so that agents have access to the same technology for intertemporal transfers as the planner. Our results extend easily to the case where taxation of the bond price is possible, but restricted by an upper bound below the level of $\bar{q}$ required for Proposition 1.

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

$$\max_{(c,e_0)} u(c_0, e_0) + \beta \int_y u_1(c(y)) f(y, e_0) \, dy$$

subject to $c_0 \geq c$, $c(y) \geq c$, $e_0 \geq 0$, the planner’s budget constraint

$$y_0 - c_0 + q \int_y (y - c(y)) f(y, e_0) \, dy - G \geq 0$$

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

$$u_1'(c_0, e_0) + \beta \int_y u_1(c(y)) f_e(y, e_0) \, dy \geq 0,$$

$$qu_1'(c_0, e_0) - \beta \int_y u_1'(c(y)) f(y, e_0) \, dy \geq 0.$$  

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

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

$$\tau_0 = c_0 - y_0,$$

$$\tau(y) = c(y) - y, \quad y \in Y.$$  

If agents face this tax system and have access to the bond market at price $q$, the resulting allocation will obviously be $(c_0, c(\cdot), e_0)$.

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

$$\frac{\lambda}{u_1'(c_0, e_0)} = 1 + \mu \frac{u''_e(c_0, e_0)}{u_1'(c_0, e_0)} + \bar{q} \frac{u''_e(c_0, e_0)}{u_1'(c_0, e_0)},$$

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

where $\lambda$, $\mu$ and $\xi$ are the (nonnegative) Lagrange multipliers associated with the budget constraint (11), the first-order condition for effort (12), and the Euler equation (13), respectively.
Proposition 2 Suppose that the FOA is valid and let \((c_0, c(\cdot), e_0)\) be a third best allocation that is interior. Then equations (14) and (15) characterizing the consumption scheme are satisfied with \(\xi > 0\).

Proof. From the Kuhn-Tucker theorem we have \(\xi \geq 0\). If \(\xi > 0\), we are done. If \(\xi = 0\), then the first-order conditions of the Lagrangian read

\[
\frac{\lambda}{u'_c(c_0, e_0)} = 1 + \frac{u''_{ce}(c_0, e_0)}{u'_c(c_0, e_0)},
\]

\[
\frac{\lambda q}{\beta u'_1(c(y))} = 1 + \frac{f_e(y, e_0)}{f(y, e_0)}, \quad y \in [y, \overline{y}].
\]

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

\[
\int_y^{\overline{y}} \frac{\lambda q}{\beta u'_1(c(y))} f(y, e_0) dy = 1.
\]

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

\[
\frac{\lambda}{u'_c(c_0, e_0)} \geq 1 = \int_y^{\overline{y}} \frac{\lambda q}{\beta u'_1(c(y))} f(y, e_0) dy \geq \frac{\lambda q}{\beta \int_y^{\overline{y}} u'_1(c(y)) f(y, e_0) dy},
\]

where the last inequality follows from Jensen’s inequality. This inequality is in fact strict, since the agent cannot be fully insured when effort is interior. Hence, we conclude

\[
\lambda \beta \int_y^{\overline{y}} u'_1(c(y)) f(y, e_0) dy > \lambda q u'_c(c_0, e_0),
\]

which is incompatible with the agent’s Euler equation (13). This shows that \(\xi\) cannot be zero.

Q.E.D.

Comparing the characterization of third best consumption schemes, (14), (15), to the characterization of second best consumption schemes, (8), (9), we notice that the difference between the two environments is closely related to the effect of the agent’s Euler equation (13) and the associated Lagrange multiplier \(\xi\). We discuss the implications of this finding in detail in the next section.

3 Absolute Progressivity and Linear Likelihoods

We are interested in the shape of second best and third best consumption schemes \(c(y)\). Clearly, this shape is closely related to the curvature of labor income taxes in the associated decentralizations.

Definition 1 We say that an allocation \((c_0, c(\cdot), e_0)\) is progressive if \(c'(y)\) is decreasing in \(y\). We call the allocation regressive if \(c'(y)\) is increasing in \(y\).
Recall that \( \tau(y) = c(y) - y \) denotes the agent’s transfer, hence \(-\tau(y)\) represents the labor income tax. Definition 1 implies that whenever a consumption scheme is progressive (regressive), we have a tax system with increasing (decreasing) marginal taxes \(-\tau'(y)\) on labor income supporting it.

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

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

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

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

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

\[
g(c) := \frac{\lambda q}{\beta u'_{1}(c)} - \xi a(c).
\]

By concavity of \( u \), \( \frac{1}{u'_{1}(c)} \) is always increasing. Therefore, if \( \frac{1}{u'_{1}(c)} \) is convex and \( \xi = 0 \) (or \( \xi > 0 \) and \( a(\cdot) \) decreasing and concave), then \( g(\cdot) \) is increasing and convex. Given the validity of the FOA, equation (9) (or equation (15), respectively) shows that second best (third best) consumption schemes are characterized as follows:

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

where, by assumption, the right-hand side is a positive affine transformation of a concave function. By applying the inverse function of \( g(\cdot) \) to both sides, we see that \( c(\cdot) \) is concave since it is an increasing and concave transformation of a concave function. **Q.E.D.**
Note that in the previous proposition, since the function \( g \) is increasing, consumption is increasing as long as the likelihood ratio function \( l(y, e) \) is increasing in \( y \).

Proposition 3 implies that CARA utilities with concave likelihood ratios lead to progressive schemes, both in the second best and the third best.\(^8\) In the second best, progressive schemes are also induced by concave likelihood ratios and CRRA utilities with \( \sigma \geq 1 \), since \( \frac{1}{u_1(c)} = c^\sigma \) is convex in this case. For logarithmic utility with linear likelihood ratios we obtain second best schemes that are proportional, since \( \frac{1}{u_1(c)} = c \) is both concave and convex. Interestingly, third best schemes are regressive in this case (since absolute risk aversion \( a(c) = \frac{1}{c} \) is convex).\(^9\)

This particular finding sheds light on a more general pattern under convex absolute risk aversion: when capital income can be fully taxed (second best), the allocation has a ‘more concave’ relationship between labor income and consumption. In other words, capital income taxation calls for more progressivity in the labor income tax system. The next result formalizes this insight.

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

**Proof.** Given validity of the FOA, by equations (9) and (15) the consumption schemes \( c(y) \) and \( \hat{c}(y) \) are characterized as follows:

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

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

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

Finally, notice that convexity of \( g_\lambda \) is equivalent to concavity of \( c \), since \( l(y, e) \) is linear in \( y \). \( \text{Q.E.D.} \)

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

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

\(^8\)Other cases where progressivity/regressivity does not differ between second best and third best are when \( a \) has the same shape as \( \frac{1}{c^2} \) (quadratic utility) and when \( a \) is linear (and hence increasing).

\(^9\)More precisely, consumption is characterized by \( \frac{\lambda}{\beta} c(y) - \frac{\lambda}{\beta} \frac{1}{e(y)} = 1 + \mu l(y, e) \) in this case. Since the left-hand side is concave in \( c \) and the right-hand side is linear in \( y \), the consumption scheme \( c(y) \) must be convex in \( y \).
This expression equates the discounted present value (normalized by \( f(y,e_0) \)) of the costs and benefits of increasing the agent’s utility by one unit in state \( y \). The increase in utility costs the planner \( \frac{q}{\beta u_1'(c(y))} \) units in consumption terms. Multiplied by the shadow price of resources \( \lambda \), we obtain the left-hand side of the above expression. In terms of benefits, first of all, since the agent’s utility is increased by one unit, there is a return of 1. Furthermore, increasing the agent’s utility also relaxes the incentive constraint for effort, generating a return of \( \mu \frac{f_e(y,e_0)}{f(y,e_0)} \). Finally, by increasing \( u_1(c(y)) \) the planner alleviates the saving motive of the agent. This gain, measured by \( \xi a(c(y)) \), depends crucially on the multiplier \( \xi \) of the agent’s Euler equation. When capital income can be fully taxed (second best), we have \( \xi = 0 \) and this gain vanishes. Intuitively, by lowering the net return of the bond, the planner is able to circumvent the first-order incentive constraint for assets. However, when capital income taxation is ruled out (third best), this constraint is binding and we have \( \xi > 0 \). Under convex absolute risk aversion, the term \( \xi a(c(y)) \) is convex. This implies that, ceteris paribus, the benefits of increasing the agent’s utility change in a more convex way with labor income. As a consequence, in the third best the agent’s utility must also change in a more convex way with labor income, hence consumption becomes more convex in \( y \) in this case.

A closely related intuition for equation (15) can be obtained by rewriting it as follows:

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

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

4 General Results on Progressivity

Since at least Holmstrom (1979), it is well known that consumption patterns under moral hazard are crucially influenced by the shape of the likelihood ratio function \( l(\cdot,e) \). Stated in more negative terms, one can always find functions \( l(\cdot,e) \) so that the shape of consumption is almost arbitrary.\footnote{Of course, if the increase in consumption is done in a state with a negative likelihood ratio, this represents a cost since the incentive constraint is in fact tightened.}
To make the impact of capital income taxation on the shape of optimal consumption easier to observe, we have therefore normalized the curvature of the likelihood ratio by assuming linearity in Proposition 4.

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

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

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

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

**Proof.** Recall that we have

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

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

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

Conversely, suppose that \(c\) is a convex transformation of \(l\). Using (18), we see that \(g_\lambda\) is then concave. Convexity of \(a(c)\) implies that \(\hat{g}_{\lambda, \xi}\) is then also concave, which shows that \(\hat{c}\) is a convex transformation of \(l\). Q.E.D.

The previous result obviously generates a sense in which \(c\) is ‘more progressive’ than \(\hat{c}\). Note that this finding generalizes Proposition 4 to arbitrary shapes of the likelihood ratio function \(l\). As a drawback, we can rank the curvature of \(c\) and \(\hat{c}\) only when, for example, \(c\) is a concave transformation of \(l\). We will now reduce the set of possible utility functions to facilitate such comparisons.
Let us consider the class of HARA (or linear risk tolerance) utility functions, namely
\[ u_1(c) = \rho \left( \eta + \frac{c}{1-\gamma} \right)^{1-\gamma} \]
with \( \rho > 0 \) and \( \eta + \frac{c}{1-\gamma} > 0 \).

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

**Lemma** Given a strictly increasing, differentiable function \( u_1 : \mathbb{R} \rightarrow \mathbb{R} \), consider the two functions defined as follows:
\[ g_\lambda (c) := \frac{\lambda q}{\beta u'_1(c)}, \]
\[ \hat{g}_{\lambda, \xi} (c) := \frac{\lambda q}{\beta u'_1(c)} - \hat{\xi} a(c). \]

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

**Proof.** If \( u \) belongs to the HARA class, we obtain
\[ \hat{g}_{\lambda, \xi}(c) = \frac{\hat{\lambda}}{\lambda} g_\lambda(c) - \hat{\xi} a(c) = \frac{\hat{\lambda}}{\lambda} g_\lambda(c) - \hat{\xi} \lambda^{1-\gamma} \kappa (g_\lambda(c))^{-\frac{1}{\gamma}}, \text{ with } \kappa = \left[ \frac{\gamma q}{\beta \rho (1-\gamma)} \right]^\frac{1}{\gamma} > 0. \]

In other words, we have
\[ \hat{g}_{\lambda, \xi}(c) = h(g_\lambda(c)), \text{ where } h(g) = \frac{\hat{\lambda}}{\lambda} g - \hat{\xi} \lambda^{\frac{1}{\gamma}} \kappa g^{-\frac{1}{\gamma}}. \]

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

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

Recall that second best and third best consumption schemes are characterized as follows:
\[ g_\lambda (c(y)) = 1 + \mu l(y, e_0), \]
\[ \hat{g}_{\lambda, \xi} (\hat{c}(y)) = 1 + \hat{\mu} l(y, e_0). \]

For logarithmic utility, \( g_\lambda \) is linear. The lemma therefore has the following consequence.

**Corollary** Suppose \( u_1 \) is logarithmic. Then \( c \) is a concave transformation of \( \hat{c} \).
Proof. By the previous lemma, there exists a concave function \( \tilde{h} \) such that \( c \) and \( \hat{c} \) are related as follows:

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

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

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

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

This concept generalizes the standard notion of convexity. It is easy to see that a function \( f \) is convex if and only if it is \( G \)-convex for any increasing affine function \( G \). Moreover, it can be shown that if \( G \) is concave and \( f \) is \( G \)-convex then \( f \) must be convex, but the converse is false.\(^{11}\)

**Proposition 6** Assume \( u_1 \) belongs to the HARA class with \( \gamma \geq -1 \). Then \( c \) is \( g_\lambda \)-convex (\( g_\lambda \)-concave) if and only if \( \hat{c} \) is \( \hat{g}_\lambda \)-convex (\( \hat{g}_\lambda \)-concave).\(^ {12}\)

**Proof.** Recall that consumption is determined as follows:

\[
\begin{align*}
g_\lambda (c(y)) &= 1 + \mu l(y, e_0), \\
\hat{g}_\lambda \hat{c}(y) &= 1 + \hat{\mu} l(y, e_0).
\end{align*}
\]

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

\[
\frac{1}{\mu} \left( g_\lambda (c(y)) - 1 \right) = \frac{1}{\hat{\mu}} \left( \hat{g}_\lambda \hat{c}(y) - 1 \right).
\]

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

**Corollary** Assume \( u_1 \) belongs to the HARA class with \( \gamma \geq -1 \). If \( \hat{c} \) is \( g_\lambda \)-concave then \( c \) is \( g_\lambda \)-concave. Conversely, if \( c \) is \( g_\lambda \)-convex then \( \hat{c} \) is \( g_\lambda \)-convex.

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

\(^{12}\)In fact, this statement is not only true for concavity and convexity, but more generally for any property defined with respect to the transformations \( g_\lambda \) and \( \hat{g}_\lambda \).
Proof. Let \( \hat{c} \) be \( g_\lambda \)-concave. By the Lemma, we have \( \hat{g}_\lambda \hat{c} = h \circ g_\lambda \) for some increasing and concave function \( h \). Hence, when \( \hat{c} \) is \( g_\lambda \)-concave, then \( \hat{c} \) must also be \( \hat{g}_\lambda \hat{c} \)-concave. Now Proposition 6 implies that \( c \) is \( g_\lambda \)-concave.

To verify the second statement, let \( c \) be \( g_\lambda \)-convex. From Proposition 6, we see that \( \hat{c} \) is \( \hat{g}_\lambda \hat{c} \)-convex, i.e., \( \hat{g}_\lambda \hat{c} \circ \hat{c} \) is convex. By the Lemma, we have \( \hat{g}_\lambda \hat{c} = h \circ g_\lambda \) for some increasing and concave function \( h \). Since the inverse of \( h \) must be convex, we conclude that \( g_\lambda \circ \hat{c} = h^{-1} \circ \hat{g}_\lambda \hat{c} \) is convex. Q.E.D.

The corollary shows that whenever \( \hat{c} \) satisfies the \( g_\lambda \)-concavity property, then \( c \) satisfies this property. In this sense, we note again that \( c \) is ‘more progressive’ than \( \hat{c} \).

5 Quantitative Analysis

6 Concluding Remarks

This paper analyzed how restrictions to capital taxation change the optimal tax code on labor income. Assuming preferences with convex absolute risk aversion, we found that optimal consumption moves in a more convex way with labor income when asset accumulation cannot be controlled by the planner. In terms of our decentralization, this implies that marginal taxes on labor income become less progressive when restrictions to capital income taxation are binding. We complemented our theoretical results with a quantitative analysis based on individual level U.S. data on consumption and income.

The model we presented here is one of action moral hazard, similar to Varian (1980) and Eaton and Rosen (1980). This is mainly done for tractability. Although a more common interpretation of this model is that of insurance, we believe that it conveys a number of general principles for optimal taxation that also apply to models of ex-ante redistribution. Of course, the quantitative analysis might change in this case. A recent example of a quantitative study of optimal income taxation under observable assets (in two periods) is Golosov, Troshkin, and Tsyvinski (2009). In their model, however, capital taxation is unrestricted and thus the government can implement the second best. This creates a possible direction for further research that can be addressed by including ex-ante (unobservable) heterogeneity in our model.\(^{13}\)

\(^{13}\)The case with observable heterogeneity can be handled quite easily and none of our results would change.
References


Proof of Proposition 7.

The Pareto weights imply transfers across individuals such that we can solve the problem in two stages. Stage one, transfers across individuals. Stage two, separate optimal contracts. Clearly, the transfers must be such that the individual $i$ receives $c^*_0 = \gamma y^*_0$. We will chose the Pareto weights according the statement in the proposition, and construct the solution.

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

$$V(\pi^i) = \max_{c_h, c_s, c_b} \pi^i \left[ \left( c^*_h \right)^{\gamma} \left( v \left( T - c^*_0 \right) \right)^{1 - \gamma} \right]^{1 - \gamma} + \pi^k \beta \sum_s p_s \left( e^*_b \right)^{\gamma} \left[ \left( c^*_s \right)^{\gamma} \left( v \left( T \right) \right)^{1 - \gamma} \right]^{1 - \gamma}$$
for an appropriate choice of \( c_i \). Now, consider the following 'normalized' problem:

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

s.t.

\[
\frac{1}{\gamma} - 1 + q \sum_s p_s(e_0) \left( \frac{\eta_s}{\gamma} - \varepsilon_s \right) - \tau^* \geq 0;
\]

\[
-(1-\alpha) (1-\sigma) v'(T - e_0) \frac{(v(T - e_0))^{1-\alpha}}{1-\sigma} 
\leq \beta \sum_s p'_s(e_0) \left( \frac{(\varepsilon_s)^\alpha (v(T))^{1-\alpha}}{1-\sigma} \right);
\]

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

Now, consider the following 'normalized' problem:

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

s.t.

\[
\frac{1}{\gamma} - 1 + q \sum_s p_s(e_0) \left( \frac{\eta_s}{\gamma} - \varepsilon_s \right) - \tau^* \geq 0;
\]

\[
-(1-\alpha) (1-\sigma) v'(T - e_0) \frac{(v(T - e_0))^{1-\alpha}}{1-\sigma} 
\leq \beta \sum_s p'_s(e_0) \left( \frac{(\varepsilon_s)^\alpha (v(T))^{1-\alpha}}{1-\sigma} \right);
\]

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

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

---

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

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

then for an appropriate choice of \( \kappa \) that solves the Euler equation, such contract solves the incentive constraints as well, since

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

Existence is hence guaranteed by continuity.
Now, consider the objective function for the guessed Pareto weights as in the proposition:

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

s.t.

\[
y_0 - c_0 + q \sum_s p_s(e_0) [y_0 - c_0 e_s] - \gamma y_0 \geq 0;
\]

\[
- (1 - \alpha) (1 - \sigma) \frac{v'(T - e_0)}{v(T - e_0)} \left[ \frac{(v(T) - e_0)^{1-\alpha}}{1 - \sigma} \right]^{1-\sigma} = \beta \sum_s p_s'(e_0) \left[ \frac{(e_s^\alpha (v(T))^{1-\alpha})}{1 - \sigma} \right];
\]

\[
\tilde{q} \left[ (v(T) - e_0)^{1-\alpha} \right]^{1-\sigma} = \beta \sum_s p_s(e_0) \frac{1}{\tilde{e}_s} \left[ (e_s^\alpha (v(T))^{1-\alpha}) \right]^{1-\sigma}.
\]

Note that since the IC constraints are homogeneous of degree zero in \(c_0\), they are indetical to those in the normalized problem. The same argument we made above guarantees that this problem has at least one solution: let \((c_0^*(y_0), e_0^*(y_0), \{\tilde{e}_s^*(y_0)\})_s\) such a solution.\(^{15}\)

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

Given the required allocation, we solve the budget constraint by choosing the value for \(G\) as a residual. More precisely, given \(\tilde{e}_0^*\), we have \(\tau^* := \frac{1}{\gamma} - 1 + q \sum_s p_s(\tilde{e}_0^*) \left[ \frac{d_s}{\gamma} - \tilde{e}_s \right]\), hence \(G^* = \tau^* \sum \gamma y_0^*\). Q.E.D.

\(^{15}\)A necessary and sufficient condition for the optimality of \(c_0^*\) is

\[
\frac{\gamma y_0}{(\gamma y_0)^{1-\sigma} \alpha} \left\{ [(\tilde{e}_0^*)^\alpha (v(T) - e_0)^{1-\alpha}]^{1-\sigma} + \beta \sum_s p_s(e_0) [(\tilde{e}_0^*)^\alpha (v(T))^{1-\alpha}]^{1-\sigma} \right\} = \lambda \left[ 1 + q \sum_s p_s(e_0) \tilde{e}_s^* \right]
\]

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

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