OPTIMAL TAXATION WITH RISKY HUMAN CAPITAL

Marek Kapička
Julian Neira

CERGE EI

Charles University
Center for Economic Research and Graduate Education
Academy of Sciences of the Czech Republic
Economics Institute

WORKING PAPER SERIES (ISSN 1211-3298)
Electronic Version

553

Working Paper Series **553**
(ISSN 1211-3298)

Optimal Taxation with Risky Human Capital

Marek Kapička
Julian Neira

CERGE-EI
Prague, November 2015

ISBN 978-80-7343-360-4 (Univerzita Karlova v Praze, Centrum pro ekonomický výzkum a doktorské studium)
ISBN 978-80-7344-358-0 (Národohospodářský ústav AV ČR, v. v. i.)

Optimal Taxation with Risky Human Capital*

Marek Kapička
U.C. Santa Barbara
and CERGE-EI†

Julian Neira
University of Exeter

November 25, 2015

Abstract

We study optimal tax policies in a life-cycle economy with risky human capital and permanent ability differences. The optimal policies balance redistribution across agents, insurance against human capital shocks, and incentives to learn and work. In the optimum, i) if utility is separable in labor and learning effort, the inverse labor wedge follows a random walk, ii) if the utility is not separable then the "no distortion at the top" result does not apply, and iii) quantitatively, high-ability agents face very risky consumption while low-ability agents are insured. The welfare gains from switching to an optimal tax system are large.

J.E.L. Codes: E6, H2
Keywords: optimal taxation, income taxation, human capital

*A previous version of this paper was circulated under the title *Optimal Taxation in a Life-Cycle Economy with Endogenous Human Capital Formation.* We are grateful to Laurence Ales, Javier Birchenall, Aspen Gorry, Finn Kydland, Ioana Marinescu, Gareth Myles, Peter Rupert, B. Ravikumar, Christian Siegel, Rish Singhania, Ali Shourideh, Aleh Tsyvinski, Andres Zambrano and Yuzhe Zhang for insightful comments, as well as to seminar participants at the 2012 NBER Summer Institute, SED, Midwest Macro, Stony Brook, the St. Louis Fed, the Chicago Fed, U. of Los Andes, LAEF and other seminars and conferences. All errors are our own.

†CERGE-EI, a joint workplace of Charles University in Prague and the Economics Institute of the Czech Academy of Sciences, Politickych veznu 7, 111 21 Prague, Czech Republic.

1 Introduction

Models of life-cycle economies with agents who have permanent differences in ability and face shocks to their human capital have been successful in understanding and quantifying the sources of inequality over the life cycle. Huggett, Ventura, and Yaron (2011) show that such a model is able to account for key empirical features of the dynamics of earnings and consumption. We show that this model is also a useful and tractable framework for studying optimal taxation.

We assume that the government's choices are limited by two frictions: a standard Mirrleesean private information friction where ability and *labor effort* are unobservable by the government, and a moral hazard friction where human capital investments (*learning effort*) and human capital shocks are also unobservable by the government. At the optimum, the government faces a nontrivial problem of balancing several competing objectives: redistribution of resources across agents of different abilities; insurance against human capital shocks; provision of incentives to accumulate human capital; and provision of incentives to elicit high labor effort from agents with high human capital or ability.

The interaction of a private information friction and a moral hazard friction produces several notable theoretical results. When the utility function is additively separable in labor and learning effort, we show that the inverse of the labor wedge follows a random walk, implying that the expected labor wedge increases with age (Proposition 5). This result is, to the best of our knowledge, novel and allows us to identify a new role of age-dependent taxes: they correct the "undesired" increase in the future labor effort implied by a positive savings wedge. This may seem surprising, because in a standard dynamic Mirleesean model, the role of a positive savings wedge is precisely to elicit a higher labor effort in the future by increasing the present value of future labor income. In our model, however, the positive savings wedge comes from the moral hazard friction, and its goal is to elicit a higher learning effort today. Hence, the optimal future labor wedge must increase to partially correct for this.

We also show that the well-known "no distortion at the top" result from the Mirrleesean literature does not apply if the utility function is not additively separable in labor effort and learning effort (Proposition 4). In our model, even the "top" agent needs incentives to increase his or her learning effort. If discouraging labor effort increases incentives to invest in

human capital, the 'top" agent will therefore face a positive marginal tax. Finally, we show that the marginal labor income tax rate is strictly decreasing in human capital realizations (Proposition 3). People with higher shock realizations are rewarded by higher consumption, and low labor income tax rates are needed in order to elicit the right quantity of labor supply.

We calibrate a two period model to match a number of moments of the U.S. economy (a "status quo" model) and investigate quantitatively the optimal tax policies and efficient allocations. We find that, at the optimum, only the top one percent of agents exert higher learning effort than in the status quo. This is achieved mainly by making their consumption very risky relative to the status quo economy. Specifically, high-ability agents face very low consumption after very low realizations of human capital shocks, and this gives them incentives to increase their learning effort. Interestingly, second period utility is also low for very *high* realizations of human capital shocks, where the provision of high labor effort becomes important. This conflicts with the government's desire to elicit high learning effort. For middle-ability agents, consumption risk is lower than for high-ability agents, but still higher than in the status quo economy. Middle-ability agents with high human capital realizations again provide high effort. This lowers their learning effort in the first period below the status quo economy. For low-ability agents, the desire to provide consumption insurance dominates, and they also choose a relatively low learning effort. Higher ability agents also face a higher savings wedge, because their incentives to self-insure are higher. Compared to the status quo, the top one percent face an increase in their capital income tax rate, while the optimal tax rate for the bottom 99 percent is lower than their current one.

The private information friction interacts with the moral hazard friction by limiting the amount of consumption risk that high-ability agents face: in its absence, high-ability agents would face even more consumption risk, and their learning effort would be higher. Interestingly, our results regarding the distribution of consumption risk stand in contrast to some of the previous literature with exogenous i.i.d. productivity shocks (e.g. Albanesi and Sleet, 2006) where high-ability agents typically have more consumption insurance.

We find large welfare gains from implementing the efficient tax system. The unborn agent is indifferent between the efficient tax system and the status quo economy with 4.5 percent higher consumption *in every period and state of the world.* We shut down each of the two

frictions and find that the private information friction is significantly more costly than the moral hazard friction.

1.1 Relationship to the existing literature

We build on a large literature that looks at models with Ben-Porath's (1967) technology for human capital formation, where investments in education are in terms of time, rather than in terms of physical resources. Properly parameterized life-cycle versions of such economies have been studied by Huggett, Ventura, and Yaron (2006) and Huggett, Ventura, and Yaron (2011) who are able to quantitatively account for the hump-shaped profile of average earnings and an increase in the earnings dispersion and skewness over the life-cycle. Moreover, the stochastic process for earnings generated by the model is consistent with both leading statistical models, the RIP (restricted income profile) models (see e.g. MaCurdy 1982, Storesletten, Telmer, and Yaron 2004) and the HIP (heterogeneous income profile) models (see e.g. Lillard and Weiss 1979, Guvenen 2007).[1] Finally, the Ben-Porath framework is also consistent with the increased dispersion in consumption over the life-cycle, as documented by Aguiar and Hurst (2013) or Primiceri and van Rens (2009). Our paper takes the economy with risky human capital and permanent ability differences as a starting point for the optimal taxation analysis.

On the normative side, our paper contributes to the growing literature that studies optimal taxation with endogenous human capital formation. The paper uses the Mirrlees approach (Mirrlees 1971, 1976) to optimal taxation. Recent dynamic extensions of the Mirrlees approach, including Golosov, Kocherlakota, and Tsyvinski (2003), Kocherlakota (2005), Farhi and Werning (2007), Albanesi and Sleet (2006), Werning (2007), Battaglini and Coate (2008), Farhi and Werning (2013) and Golosov, Tsyvinski, and Troshkin (2015), have mostly focused on cases when individual skills are exogenous. In contrast, this paper focuses on a case when individual skills are endogenous.

Our paper is the first one that studies an environment of optimal taxation with human capital where both private information and moral hazard frictions are present.[2] In this respect, our

[1] The difference between RIP and HIP models is that in HIP models people face heterogeneous life-cycle earning profiles, while in RIP models individuals face similar life-cycle earning profiles.

[2] Abraham, Koehne, and Pavoni (2014) and Albanesi (2007) study the impact of moral hazard on optimal tax

paper is close to Shourideh (2014), who studies an economy with unobservable risky physical, rather than human, capital investments and unobservable abilities and also considers the interplay between moral hazard and private information frictions. An important difference between hidden savings and hidden human capital investments is that hidden savings imply hidden consumption. Hidden consumption, in turn, implies that incentive compatibility constraints might be upward binding, potentially changing the nature of the redistributive problem.

We assume that human capital investments are risky and unobservable.[3] In contrast, da Costa and Maestri (2007), Jacobs, Schindler, and Yang (2012) and Stantcheva (2015a) all study optimal taxation with risky Ben-Porath technology, but allow human capital investments to be observable. This not only shuts down the moral hazard dimension of the problem, but also permits the government to use direct human capital subsidies conditional on learning effort. We think that unobservability of learning effort is a reasonable assumption to make. Even if the government could observe the number of hours that each individual spends by accumulating human capital in a formal setting, it is not obvious that this would be a good approximation of one's learning effort.

In term of human capital technology, our approach is complementary to Bovenberg and Jacobs (2005), Grochulski and Piskorski (2010), Findeisen and Sachs (2015a, 2015b) and Stantcheva (2015b), who all assume that human capital investments are in terms of goods (under various assumptions about observability of human capital investments) and thus study a different aspect of human capital formation. Our paper is also related to a recent research that quantitatively studies optimal tax reforms in environments with endogenous human capital such as Gorry and Oberfield (2012), Krueger and Ludwig (2013) and Peterman (2014). This literature is able to consider richer frameworks than ours by restricting taxes to specific functional forms.

---

structures. Unlike our paper, they do not consider the interaction between moral hazard and private information.

[3] Environments with riskless human capital have been previously studied by Diamond and Mirrlees (2002), Kapička (2006), Boháček and Kapička (2008) and Kapička (2015) and others.

2 The Model

Consider the following life-cycle economy. Agents live for two periods. They like to consume, dislike working and exerting learning effort, and have preferences given by

$$U(c_1) - V(\ell_1, e_1) + \beta \mathbb{E}\left[U(c_2) - V(\ell_2, e_2)\right], \tag{1}$$

where $j \in \{1,2\}$ is age, $c_j \geq 0$ is consumption, $\ell_j \geq 0$ is labor effort, and $e_j \geq 0$ is learning effort. The function U is strictly increasing, strictly concave, and differentiable. The function V is strictly increasing, strictly convex, and differentiable in both arguments. We further restrict the function V by assuming that, conditionally on the learning effort being zero, the function V exhibits a constant Frisch elasticity of labor $\gamma(\ell, e) = \frac{V_\ell(\ell,e)}{\ell V_{\ell\ell}(\ell,e)}$:

Assumption 1. *The elasticity of labor effort $\gamma(\ell, 0)$ is independent of ℓ.*

An agent's earnings y_j are determined by the agent's ability a, current human capital h_j, and current labor effort ℓ_j:

$$y_j = a h_j \ell_j. \tag{2}$$

Ability is constant over an agent's lifetime and is known to the agents at the beginning of the first period. The ability has continuous support $A = (\underline{a}, \overline{a})$, with $\overline{a}$ possibly being infinite. All agents are born with the same initial human capital h_1. Human capital in the second period h_2 has continuous support $H = (\underline{h}, \overline{h})$, with $\overline{h}$ possibly infinite, and depends on an idiosyncratic human capital depreciation shock z_2, initial human capital h_1, and on learning effort e_1:

$$h_2 = \exp(z_2) F(h_1, e_1). \tag{3}$$

The function F is strictly increasing, strictly concave, and differentiable in both arguments. As is standard in the moral hazard literature, we transform the state-space representation of the problem to work directly with the distribution induced over h_2. To that end, we construct a probability density function of human capital in the second period conditional on first period effort and denote it $p(h_2|e_1)$. The derivative of the density with respect to effort $p_e(h|e)$ exists, and we assume that the conditional distribution of the second period human capital satisfies the Monotone Likelihood Ratio Property:

Assumption 2 (MLRP). $\frac{p_e(h|e)}{p(h|e)}$ *is strictly increasing in* h *for all* e.

The MLRP property has the usual interpretation that higher effort induces a more favorable distribution of human capital outcomes.

This economy is identical to Huggett, Ventura, and Yaron (2011), with two exceptions. First, this model includes leisure. That is essential for thinking about optimal taxation. Second, the ability a affects earnings directly, rather than indirectly through the human capital production function. That is irrelevant in the incomplete markets economy studied by Huggett, Ventura, and Yaron (2011) if the human capital production function takes the Ben-Porath form.[4] However, both formulations have different implications in a Mirrleesean economy with private information and observable human capital where it makes a difference whether h or ha is observed.[5] The formulation chosen in this paper has the advantage that it is entirely consistent with the existing Mirrleesean optimal taxation literature.

Our preferred interpretation of observable risky human capital is the observability of a person's industry, firm, or occupation. Human capital shocks then take the interpretation of shocks to industries, firms, or occupations. Support for this view comes from the literature which argues that human capital is occupation, firm, or industry-specific (see Jacobson, LaLonde, and Sullivan 1993, Neal 1995, Parent 2000, Poletaev and Robinson 2008, and Kambourov and Manovskii 2009).

3 Efficient Allocations

The *information structure* is as follows: ability a, labor effort ℓ_1 and ℓ_2, learning effort e_1 and e_2, and human capital shock z_2 are private information of the agent. Consumption c_1 and c_2, earnings y_1 and y_2, and human capital h_1 and h_2, are publicly observable. Agents report their ability level to the social planner in the first period. An agent's true ability is denoted by a whereas $\hat{a}$ denotes the ability report.

[4] To see that both formulations are isomorphic, let $F(h,e) = h + (eh)^\alpha$. Redefine human capital as follows: Let $\bar{h} = ha$ and $\bar{a} = a^{1-\alpha}$. Then the law of motion for human capital is $\bar{F}(h,e) = \bar{h} + \bar{a}(e\bar{h})^\alpha$, and the earnings are $y = \bar{h}\ell$, identical to the ones in Huggett, Ventura, and Yaron (2011).

[5] Both formulations are isomorphic if both h and a are either observable or unobservable. The first case is inconsistent with the Mirrleesean framework, while the second one would be extremely hard to solve in general.

An *allocation* (c,y) consists of consumption $c = \{c_1(\hat{a}), c_2(\hat{a},h_2)\}$ and earnings $y = \{y_1(\hat{a}), y_2(\hat{a},h_2)\}$. Consumption and earnings in the first period are conditional on ability report $\hat{a} \in A$. In the second period they are both conditional on the ability report in the first period and realization of human capital in the second period, $h_2 \in H$. Define the lifetime utility of an a-type agent who reports ability $\hat{a}$ and exerts effort e as $W(\hat{a},e|a)$, where

$$W(\hat{a},e|a) \equiv U(c_1(\hat{a})) - V\left(\frac{y_1(\hat{a})}{ah_1}, e\right) + \beta \int_H \left[U\left(c_2(\hat{a},h_2)\right) - V\left(\frac{y_2(\hat{a},h_2)}{ah_2}, 0\right)\right] p(h_2|e)dh_2.$$

Effort in the second period is trivially equal to zero. The first period effort $\tilde{e}_1(\hat{a}|a)$ maximizes the lifetime utility of an $a-$type agent who reports $\hat{a}$:

$$\tilde{e}_1(\hat{a}|a) \equiv \arg\max_{e \geq 0} W(\hat{a},e|a). \tag{4}$$

By the revelation principle, we restrict attention to the allocations that are *incentive compatible*, i.e. where the agent prefers to tell the truth about his or her ability:

$$W(a,\tilde{e}_1(a|a)|a) \geq W(\hat{a},\tilde{e}_1(\hat{a}|a)|a) \quad \forall a,\hat{a} \in A. \tag{5}$$

In order to reduce notational complexity we define the utility maximizing effort plan conditional on truthtelling by $e_1(a) \equiv \tilde{e}_1(a|a)$, and let $W(a) = W(a,e(a)|a)$ be the corresponding truthteller's lifetime utility.

An allocation is *feasible* if it satisfies the resource constraint

$$\int_A \left[c_1(a) - y_1(a) + R^{-1} \int_H \left[c_2(a,h_2) - y_2(a,h_2)\right] p(h_2|e_1(a))\, dh_2\right] q(a)\, da \leq 0, \tag{6}$$

where $q(a)$ is the probability distribution of abilities and R is the gross interest rate. For simplicity, we assume that $R = \beta^{-1}$.

The social welfare function is simply the expected utility of an agent who does not yet know his or her ability:

$$\mathscr{W} = \int_A W(a)q(a)\, da. \tag{7}$$

Definition 1. *An allocation is constrained efficient if it maximizes welfare (7) subject to the resource constraint (6) and the incentive compatibility constraint (5), where the learning effort is given by (4).*

First-Order Approach. The first-order approach replaces the constraints (4) and (5) with two conditions. The first one is the first-order condition in effort and says that, at the optimum, the marginal costs of learning effort must be equal to the expected marginal benefit of learning effort (given by the additional utility arising from the fact that the distribution of future human capital shocks is now more favorable):

$$V_e\left(\frac{y_1(a)}{ah_1}, e_1(a)\right) = \beta \int_H \left[U(c_2(a,h_2)) - V\left(\frac{y_2(a,h_2)}{ah_2}, 0\right)\right] p_e(h_2|e_1(a))\, dh_2. \quad (8)$$

The second condition is an envelope condition governing how the lifetime utility needs to vary with ability in order to deter the agent from misreporting his type. Let $W(\underline{a})$ denote the lifetime utility of the least able agent. The envelope condition is

$$W(a) = W(\underline{a}) + \int_{\underline{a}}^{a} \left\{ V_\ell\left(\frac{y_1(\tilde{a})}{\tilde{a}h_1}, e_1(\tilde{a})\right) \frac{y_1(\tilde{a})}{\tilde{a}h_1} + \beta \int_H V_\ell\left(\frac{y_2(\tilde{a},h_2)}{\tilde{a}h_2}, 0\right) \frac{y_2(\tilde{a},h_2)}{\tilde{a}h_2} p(h_2|e_1(\tilde{a}))\, dh_2 \right\} \frac{d\tilde{a}}{\tilde{a}}. \quad (9)$$

The envelope condition states that the variation in lifetime utility for an agent a is the lifetime utility of the least able agent plus the informational rent the agent obtains from having his or her given ability level. Replacing the incentive constraint with the first order condition in effort and the envelope condition leads to a *relaxed planning problem*:

Definition 2. *An allocation solves the relaxed planning problem if it maximizes welfare (7) subject to the resource constraint (6), the first-order condition in effort (8) and the envelope condition (9) .*

For now, we assume that the first-order approach is valid and the set of constrained efficient allocations are identical to the set of allocations that satisfy the relaxed planning problem. We examine its validity at the end of Section 3.1.

3.1 Theoretical Implications

We will now characterize the properties of the efficient allocation and of the labor and savings wedges. Let λ, $\phi(a)q(a)$ and $\theta(a)q(a)$ be the Lagrange multipliers on the resource constraint (6), the first order condition (8) and on the envelope condition (9).[6] The first-order conditions in consumption are

$$\frac{1}{U'(c_1(a))} = \frac{1+\theta(a)}{\lambda} \tag{10a}$$

$$\frac{1}{U'(c_2(a,h_2))} = \frac{1+\theta(a)+\phi(a)\frac{p_e(h_2|e_1(a))}{p(h_2|e_1(a))}}{\lambda}. \tag{10b}$$

The following proposition shows that if the elasticity of labor effort is constant in the second period and the monotone likelihood property hold, the Lagrange multiplier on the first-order condition in effort is positive and second period consumption is increasing in human capital realizations:

Proposition 1. *If Assumptions 1 and 2 hold, then $\phi(a)$ is strictly positive and $c_2(a,h_2)$ is strictly increasing in h_2.*

Proof. Consider a doubly relaxed problem where the first order condition in effort (8) is not imposed. Then $\phi(a) = 0$, which implies that $c_2(a,h_2)$ is independent of h_2 by (10b). The first-order condition in $\ell_2(a,h_2)$ is

$$\frac{\lambda a h_2}{V_\ell(\ell_2(a,h_2))} = 1+\theta(a)+\Theta(a)\left(\frac{1}{\gamma(\ell_2(a,h_2),0)}+1\right), \tag{11}$$

where $\Theta(a) \geq 0$ is the cumulative multiplier on the envelope condition (9). By Assumption 1, $\gamma(\ell,0)$ is independent of ℓ. Equation (11) then implies that $\ell_2(a,h_2)$ is strictly increasing in h_2. By Assumption 2, the right-hand side of (8) must then be strictly negative. Since the left-hand side of (8) is nonnegative, it is sufficient to require the right-hand side of (8) to be weakly greater than the left-hand side in order for (8) to hold as equality. The Kuhn-Tucker theorem then implies $\phi(a) > 0$. Strict monotonicity of c then follows from Assumption 2 and equation (10b). ■

A strictly positive Lagrange multiplier ϕ implies that the social planner would, in the absence of the effort constraint (8), increase private marginal costs of effort above the private

[6] See Appendix A for the full Langrangean solution method.

marginal benefits of effort. In fact, as the proof shows, the marginal benefits of higher effort would be negative: agents with higher human capital would see no consumption increase, but would work more. In the absence of a moral hazard friction, such an allocation could be supported by positive human capital subsidies (as in Jacobs, Schindler, and Yang 2012 and gross subsidy in Stantcheva 2015a). The moral hazard friction prevents that, and the social planner responds by making second period consumption increasing in human capital.

Taking the expectation of (10b) (and noting that $\int_H p_e\left(h_2|e_1(a)\right) dh_2 = 0$) implies immediately that, conditional on the ability type, the Inverse Euler equation holds:

$$\frac{1}{U'(c_1(a))} = \int_H \frac{1}{U'(c_2(a,h_2))} p\left(h_2|e_1(a)\right) dh_2 \quad \forall a \in A. \tag{12}$$

The Inverse Euler Equation would hold in the absence of the moral hazard friction as well. In that case, however, the second period consumption would be deterministic, conditional on ability.

Savings wedge. Define the *savings wedge* δ as the gap between the current marginal utility of consumption and the expected future marginal utility of consumption:

$$U'(c_1(a)) = \left(1-\delta(a)\right) \int_H U'(c_2(a,h_2)) p\left(h_2|e_1(a)\right) dh_2.$$

By Jensen's inequality, the Inverse Euler Equation (12) immediately implies the following:

Proposition 2. *The savings wedge $\delta(a)$ is strictly positive for each ability level a.*

Note, however, that a strictly positive savings wedge comes purely from the moral hazard friction on the model, and not from the private information friction. As we shall see later, this fact will have important consequences for the optimal structure of labor wedges.

Labor wedge. Similarly, define the *labor wedge* τ_j^ℓ as the gap between the marginal product of labor and the intratemporal marginal rate of substitution at each age:

$$ah_1\big(1-\tau_1^\ell(a)\big) = \frac{V_\ell\big(\ell_1(a),e_1(a)\big)}{U'(c_1(a))},$$
$$ah_2\big(1-\tau_2^\ell(a,h_2)\big) = \frac{V_\ell\big(\ell_2(a,h_2),e_2(a,h_2)\big)}{U'(c_2(a,h_2))}.$$

In the optimum, the labor wedges τ_1^ℓ and τ_2^ℓ satisfy

$$\frac{1}{U'(c_1(a))}\frac{\tau_1^\ell(a)}{1-\tau_1^\ell(a)} = \left(1+\frac{1}{\gamma(\ell_1(a),e_1(a))}\right)\Theta(a) + \frac{\phi(a)}{\lambda}\frac{V_{\ell e}\left(\ell_1(a),e_1(a)\right)}{V_\ell\left(\ell_1(a),e_1(a)\right)} \tag{14a}$$

$$\frac{1}{U'(c_2(a,h_2))}\frac{\tau_2^\ell(a,h_2)}{1-\tau_2^\ell(a,h_2)} = \left(1+\frac{1}{\gamma(\ell_2(a,h_2),0)}\right)\Theta(a), \tag{14b}$$

where $\Theta(a) = (aq(a))^{-1}\int_a^{\overline{a}}\theta(\tilde{a})q(\tilde{a})\,d\tilde{a}$ is the cumulative Lagrange multiplier on the envelope condition. In the following two propositions, we characterize the labor wedge. The proofs are omitted, since the results follow directly from the optimality conditions (14a) and (14b), and from Proposition 1.

Proposition 3. *If Assumptions 1 and 2 hold then $\tau_2^\ell(a,h_2)$ is strictly decreasing in h_2.*

The second period labor wedge is decreasing in the human capital shock because people with higher shock realizations are assigned higher consumption (Proposition 1), but for efficiency reasons they must be given enough incentives to supply labor. It is easy to see that if the support is unbounded and U satisfies the second Inada condition then the second period tax wedge converges to zero as h_2 converges to infinity. Those conditions will be satisfied, for example, if the distribution of h_2 is lognormal and the utility function U is of the CRRA form. To characterize the limits of the labor wedge in the ability dimension, we assume that $\Theta(a)$ converges to zero when a approaches its upper bound $\overline{a}$. This assumption is made in order to avoid obvious and well known cases when the top marginal tax rates do not converge to zero, and is satisfied whenever the ability distribution is bounded.

Proposition 4. *Suppose that $\Theta(a)$ converges to zero as a converges to $\overline{a}$. Then $\tau_1^\ell(a)$ con-*

verges to a positive (negative) value if $V_{\ell e}$ is positive (negative). In addition, $\tau_2^{\ell}(a,h_2)$ converges to zero for all $h_2 \in H$.

Thus, the "no distortion at the top" result from Mirrlees (1971) does not apply whenever the utility is not additively separable in labor and effort. Nonseparability allows the planner to change incentives to exert learning effort by changing first period labor effort. If $V_{\ell e} > 0$, discouraging labor effort in the first period increases incentives to exert learning effort, and it is optimal to do so, even for the "top" agent. This result differs from Kapička (2015) where human capital is unobservable but riskless. The absence of risk means that there is no scope for insurance against human capital risk. If the "top" agent faces a zero marginal tax, she will choose the efficient amount of learning effort, because she bears all the costs and benefits of the investment (the Lagrange multiplier ϕ is zero for the top agent, rather than being strictly positive). As a result, it is optimal to have a zero marginal tax on the "top" agent. Note also that this channel is absent in the second period where the "no distortion at the top" result applies.

Grochulski and Piskorski (2010) obtain a similar result, but their argument behind the nonzero tax at the top is different. In their model, the high-ability agents always face a negative marginal tax rate, because that helps to separate the truthtellers from deviators: deviators underinvest in human capital, have lower productivity, and are hurt by the negative marginal tax at the top more than truthtellers. This mechanism does not appear in our model because human capital realizations are observable. On the other hand, our mechanism is absent in Grochulski and Piskorski (2010), who do not allow for simultaneous labor effort and investment in human capital. [7]

If labor effort and learning effort are additively separable and labor effort has a constant elasticity then the inverse of the labor wedge follows a random walk:

Proposition 5. *If $\gamma(\ell,e)$ is a constant and $V_{\ell e} = 0$ then*

$$\frac{1}{\tau_1^{\ell}(a)} = \int_H \frac{1}{\tau_2^{\ell}(a,h_2)} p(h_2|e_1(a))\, dh_2.$$

[7] There are additional arguments for violation of the no distortion on the top result in the literature: Stiglitz (1982) obtains a negative tax on the top when skilled and unskilled labor are imperfect substitutes. Slavík and Yazici (2014) establish the same result when there is capital-skill complementarity. Those arguments rely on general equilibrium effects that are absent in our paper.

Proof. Since the right-hand sides of (14a) and (14b) are equal when $V_{\ell e} = 0$,

$$\frac{\tau_1^\ell(a)}{1-\tau_1^\ell(a)} \frac{1-\tau_2^\ell(a,h_2)}{\tau_2^\ell(a,h_2)} = \frac{U'(c_1(a))}{U'(c_2(a,h_2))}.$$

The result follows from using (12) and rearranging. ■

The result is due to several facts. First, a marginal tax revenue is proportional to $\tau_j^\ell/(1-\tau_j^\ell)$ for $j = 1,2$ (see e.g. Saez 2001). Second, since the ability shock is permanent, the social planner wants to keep the tax revenue valued at its marginal utility cost $1/U'(c_j)$ constant over time and state. Since $1/U'(c_j)$ follows a random walk, $(1-\tau_j^\ell)/\tau_j^\ell$ must follow a random walk as well. Jensen's inequality then implies that the average labor wedge is increasing over time:

Corollary 1.

$$\tau_1^\ell(a) < \int_H \tau_2^\ell(a,h_2) p\left(h_2|e_1(a)\right) dh_2. \tag{15}$$

To understand the intuition for the result, consider an individual who faces no risk, can borrow and save, his labor is taxed at rate τ_j^ℓ, and his savings are taxed at a rate τ^k.[8] The present value of the agent's earnings can be written as

$$w_1\ell_1(1-\tau_1^\ell) + \frac{w_2\ell_2}{1+r} \frac{(1+r)(1-\tau_2^\ell)}{1+r(1-\tau^k)}.$$

The savings tax affects the relative marginal labor tax in both periods by increasing the present value of future wages. A tax on savings thus decreases the effective tax rate on second period labor given by $\hat{\tau}_2^\ell = 1 - \frac{(1+r)(1-\tau_2^\ell)}{1+r(1-\tau^k)}$. By Corollary 1, the optimal tax on labor compensates for the implicit subsidy. Note that the tax on savings does *not* come from a dynamic private information friction, as in Golosov, Kocherlakota, and Tsyvinski (2003), Farhi and Werning (2013) and other dynamic Mirrleesean literature. In such case, there would be no need to compensate the savings tax by a higher future labor tax, because the point of the savings tax is precisely to elicit a higher future working effort (equivalently, relax future incentive constraints). In our model, the savings tax comes from the moral hazard part of the problem, and its purpose is to elicit higher learning effort today, and not higher labor supply tomorrow.

[8] In Section B we show how to implement the savings wedge as a one-to-one mapping to a regular tax on savings, τ^k.

In the absence of the moral hazard friction, the savings wedge is zero and it is optimal to have perfect tax smoothing, $\tau_1^\ell = \tau_2^\ell$. In its presence, it is optimal to "undo" the savings tax, at least partially, by a higher future labor tax.

Additively separable utility in labor effort and learning effort serves as a useful benchmark. If labor and effort are complements, the Frisch elasticity $\gamma(\ell_j, e_j)$ changes endogenously and the expected labor wedge may no longer be increasing. Estimates from Peterman (2014) suggest that the elasticity decreases over time in this scenario: it is higher when agents spend more time exerting effort, which happens at younger ages. This reinforces the increasing intertemporal profile of the labor wedge. On the other hand, $V_{\ell e} > 0$ increases the labor wedge in the first period. Since $V_{\ell e}$ is zero in the second period (because of zero effort), this weakens, or reverts, the increasing profile of the labor wedge.

Special Cases. It is instructive to explore the effect of each of the frictions on the wedges by shutting down the moral hazard and the private information individually. We do this by setting the Lagrange multipliers $\phi(a) = 0$ and $\theta(a) = 0$, respectively. If there is no moral hazard, the planner can then dictate learning effort directly. Thus, consumption no longer needs to vary with human capital realizations and is deterministic. Equation (12) implies that the savings wedge is zero. From Proposition 5, there is perfect tax smoothing across time and the labor wedge varies only with ability, as in a static Mirrlees model. With no private information, the planner can dictate labor effort directly. As a consequence, there is no need for the planner to induce labor effort through the labor wedge and $\tau_1^\ell = \tau_2^\ell = 0$. The savings wedge remains positive, as the moral hazard requires consumption uncertainty to induce optimal learning effort.

Implementation. We show in Appendix B that the efficient allocation can be implemented by a tax system consisting of income tax functions $T_1(y_1)$, $T_2(y_1, y_2, h_2)$ and a savings tax $\tau_k(s, y)$, where s are first period savings. While the first period income tax depends only on the current income, the second period income tax depends on the second period human capital realization, and can potentially depend on the history of incomes. In the optimum, the effective marginal income tax rate is equated to the labor wedge, and the marginal savings tax rate is equated to the savings wedge.

Validity of the First-Order Approach. The first-order approach might fail either because the first-order condition (8) fails to detect a utility-maximizing learning effort choice, or because the envelope condition (9) fails to detect the utility-maximizing report. In Appendix C, we show conditions that ensure that those two conditions are sufficient. Conditions for the sufficiency of (8) are similar to Jewitt (1988) (Theorem 1). The main difference is that it must be assumed that the second period utility is nondecreasing and concave in h_2. It cannot be inferred from the primitives because if labor effort is increasing in h_2 sufficiently fast, the second period utility may decrease in h_2. Conditions for sufficiency of (9) are relatively standard. Overall, the conditions for sufficiency are quite strict. In our quantitative exercise, they are not satisfied, but the allocation is still incentive compatible.

4 Quantitative Analysis

The benchmark model is the decentralized incomplete markets economy with observed U.S. capital and labor income tax rates, which we refer to as the "status quo". The status quo model is used to calibrate the initial human capital level and the parameters of the ability distribution. We then calculate the constrained efficient outcomes by replacing the status quo tax system for an efficient tax system, while keeping all other parameters of the status quo model unchanged.

4.1 Calibration

Parameters are set in two steps. First, standard parameters and those for which there are available estimates are set before solving the model. The remaining parameters are set to match moments from the data. Tables 1 and 2 summarize the calibration.

A model period is twenty years. The first period represents agents between 20 and 40 years of age, and the second period represents agents between 40 and 60 years of age.

Preferences. The instantaneous utility function for consumption is CRRA,

$$U(c) = \frac{c^{1-\rho}}{1-\rho}.$$

Table 1: Parameters Set Exogenously

Definition	Symbol	Value	Source/Target
CRRA parameter	ρ	1	Browning, Hansen, and Heckman (1999)
Frisch elasticity of labor	γ	0.5	Chetty, Guren, Manoli, and Weber (2011)
Elasticity of effort	ε	0.5	Same as Frisch elasticity
Discount factor	β	0.442	0.96 annual
Interest rate	r	119%	4% annual
H.C. technology	α	0.7	Browning, Hansen, and Heckman (1999)
Capital income tax rate	$\bar{\tau}^k$	45.88%	37% effective annual; McDaniel (2007)
Labor tax function	(ν_0, ν_1, ν_2)	$(0.182, 0.008, 1.496)$	Guner, Kaygusuz, and Ventura (2014)
Shock distribution	(μ_z, σ_z)	(-0.58, 0.496)	Huggett, Ventura, and Yaron (2011)

Table 2: Calibrated Parameters

Definition	Symbol	Value	Target Moment	Model	U.S. Data	Source
Initial human capital	h_1	0.680	$\bar{y}_1/\bar{y}_2$	0.868	0.868	HVY (2011)
St. dev. log-ability	σ_a	0.585	Earnings Gini	0.341	0.343	HVY (2011)
Mean log-ability	μ_a	0.170	Earnings Var.	0.390	0.390	HVY (2011)

The value of the parameter controlling intertemporal substitution and risk-aversion is set to $\rho = 1$, within the range of estimates surveyed by Browning, Hansen, and Heckman (1999). Preferences are additively separable in labor and effort with constant elasticities:

$$V(\ell, e) = \frac{\ell^{1+1/\gamma}}{1+1/\gamma} + \frac{e^{1+1/\varepsilon}}{1+1/\varepsilon}.$$

The Frisch elasticity of labor supply is set to $\gamma = 0.5$, consistent with micro estimates surveyed in Chetty, Guren, Manoli, and Weber (2011). The elasticity of learning effort is set to $\varepsilon = 0.5$, equal to the Frisch elasticity. Agents' discount factor is set to $\beta = (0.96)^{20} = 0.442$. Recall that we set $R = \beta^{-1}$.

Technology. The human capital production function is of the Ben-Porath form. Human capital in the second period h_2, depends on idiosyncratic human capital shock z_2, initial human capital h_1, and first period learning effort e_1:

$$F(h_1, e_1) = h_1 + (e_1 h_1)^{\alpha}.$$

The value of the human capital concavity parameter $\alpha = 0.7$ is the same used in Huggett, Ventura, and Yaron (2011), and is in the middle of the range of estimates surveyed by Browning, Hansen, and Heckman (1999).

The shock process is assumed to be i.i.d. and the shocks are drawn from a truncated normal distribution, $z \sim N(\mu_z, \sigma_z)$. The human capital shock process is estimated in Huggett, Ventura, and Yaron (2011). The Ben-Porath functional form implies that towards the end of the lifetime agents accumulate little human capital and the changes in human capital are mostly due to shocks. Thus, the parameters of the shock process can be approximated by assuming older workers in the data exert zero learning effort.[9] Huggett, Ventura, and Yaron (2011) estimate $\sigma_z^a = 0.111$ and $\mu_z^a = -0.029$ annually. We transform the shock process to its 20-year period equivalent, $\sigma_z = \sqrt{20(\sigma_z^a)^2} = 0.496$ and $\mu_z = 20\mu_z^a = -0.58$. These estimates imply that, in 20 years, a one-standard deviation shock moves wages by about 49.6 percent and human capital depreciates on average by 36.7 percent.

Status Quo Tax System. We approximate the U.S. tax system with a flat tax on capital income and a progressive tax on labor income. Tax revenues are redistributed lump-sum in the first period. The budget constraint for a given agent takes the form

$$\begin{aligned} c_1 + k_2 &= \big(1 + r(1-\bar{\tau}^k)\big)k_1 + ah_1\ell_1 - T(ah_1\ell_1) + LS_1 \\ c_2 &= \big(1 + r(1-\bar{\tau}^k)\big)k_2 + ah_2\ell_2 - T(ah_2\ell_2), \end{aligned}$$

where LS_1 is the lump-sum transfer in the first period. Without loss of generality, we set the lump-sum transfer in the second period to zero. Labor taxes take the Gouveia-Strauss form

$$T(y) = \nu_0[1 - (\nu_1 y^{\nu_2} + 1)^{-1/\nu_2}].$$

Guner, Kaygusuz, and Ventura (2014) estimate that the values of the Gouveia-Strauss function for individuals with no capital income are $\nu_0 = 0.182$, $\nu_1 = 0.008$, and $\nu_2 = 1.496$.

[9] Huggett, Ventura, and Yaron (2011) calculate wages from the Panel Study of Income Dynamics (PSID) for males between 55 and 65 years of age. Wages are total male labor earnings divided by total hours for the male head of household, using the Consumer Price Index to convert nominal wages to real wages. Then they estimate the parameters of the shock process from a log-wage difference regression. In our model, wages are unobservable but differences in log-wages are observable since abilities are permanent and are factored out.

These parameters imply progressive marginal tax rates starting at 8 percent and increasing to 24.5 percent. We obtain mean average tax rates for capital and consumption from McDaniel (2007).[10] We adjust labor and capital tax rates by the average consumption tax. This yields an effective annual capital tax rate of $\bar{\tau}_a^k = 37\%$, which we transform to a 20-year value.[11] The effective 20-year tax rate on capital income is $\bar{\tau}^k = 45.88\%$.

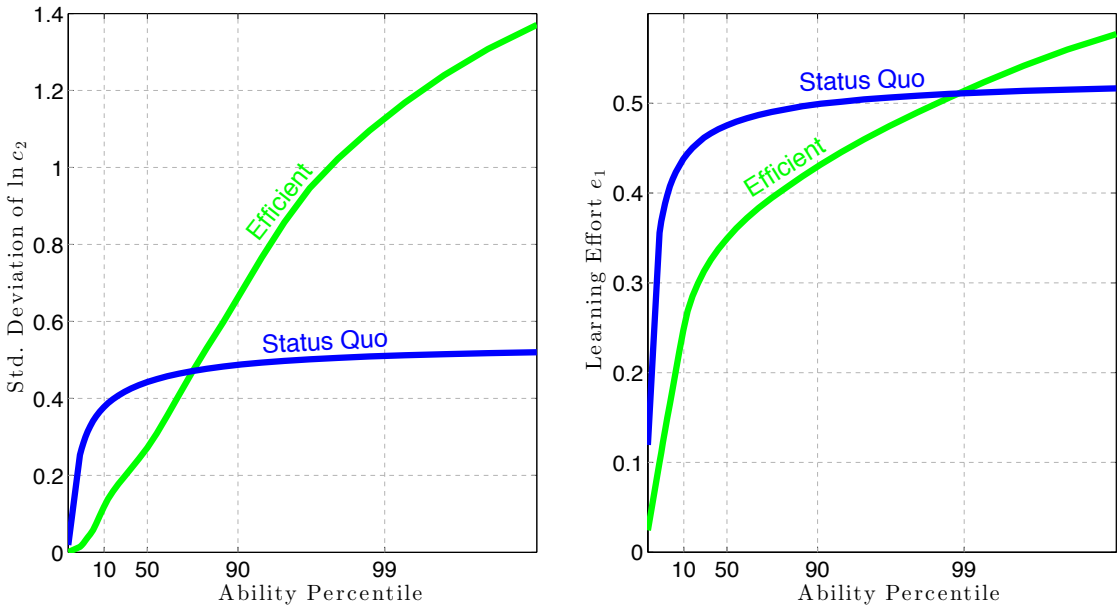
Initial Conditions. Following Huggett, Ventura, and Yaron (2006) and Huggett, Ventura, and Yaron (2011), we posit that the ability distribution is log-normally distributed, $q(a) = LN(\mu_a, \sigma_a^2)$. The initial human capital, h_1, is the same for all agents. We set μ_a, σ_a^2, and h_1 so that the equilibrium distribution of earnings matches data earnings moments. Huggett, Ventura, and Yaron (2011) estimate age profiles of mean earnings from the PSID 1969-2004 family files. We target three moments: The ratio of mean earnings of younger workers (ages 23 to 40) to mean earnings of older workers (ages 40 to 60), the earnings Gini coefficient, and the variance of earnings. Table 2 reports the results of the calibration. Parameters values $h_1 = 0.680$, $\sigma_a^2 = 0.585$, and $\mu_a = 0.170$ best approximate the model to the data targets.

4.2 Findings

Insurance, Redistribution and Incentives. Proposition 1 shows that a risky second period consumption is necessary to provide optimal incentives to accumulate human capital. Figure 1a shows that the standard deviation of log-consumption in the second period increases dramatically with ability. Relative to the status quo economy, the efficient tax system yields lower consumption risk for low-ability and medium-ability agents but significantly higher consumption risk for high-ability agents (about 99th percentile). Figure 1b shows the corresponding effort elicited from the agents in both the efficient and in the status quo allocation. For agents above the 99th percentile the efficient allocation elicits higher learning effort. However, for agents below the 99th percentile the efficient effort is lower than in the status quo economy. In particular, the agents between the 77th percentile and the 99th percentile experience both an increase in the variability of consumption and a decrease in effort. Higher

[10] We use tax rates for the 1969-2004 period for compatibility with the PSID sample used to calculate the human capital shocks.

[11] The effective 20-year capital income tax rate is the solution to $(1 + r_a(1 - \bar{\tau}_a^k))^{20} = 1 + r(1 - \bar{\tau}^k)$.

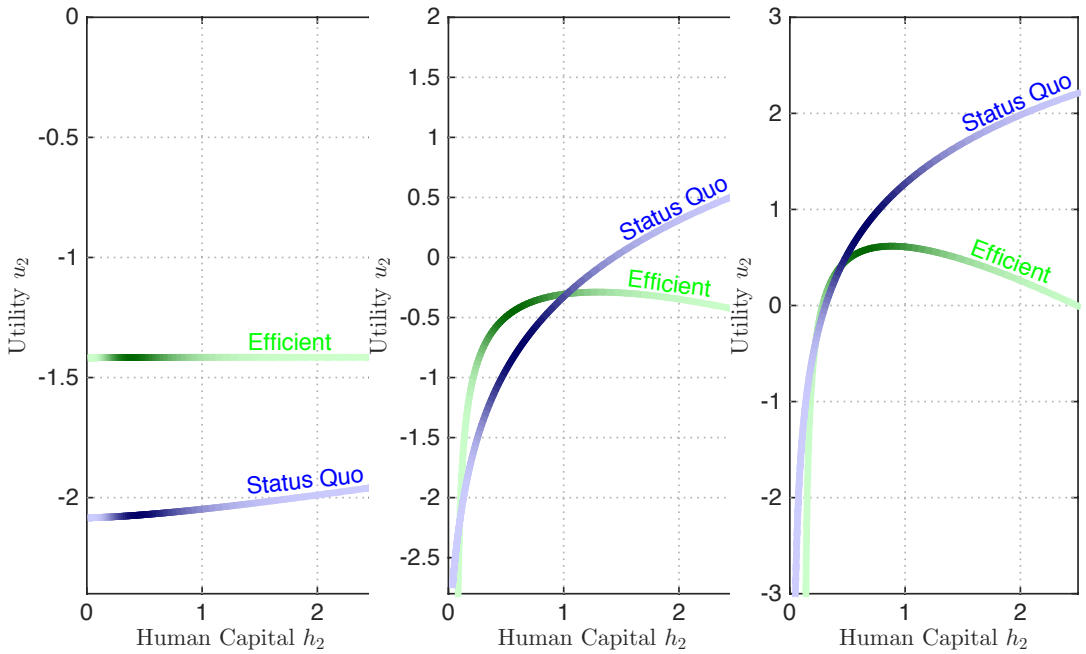


(a) Standard deviation of log-consumption in second period

(b) Effort

Figure 1: Consumption Risk and Effort

dispersion of consumption does not seem to motivate them to increase effort.



(a) Low Ability (0%)

(b) High Ability (80%)

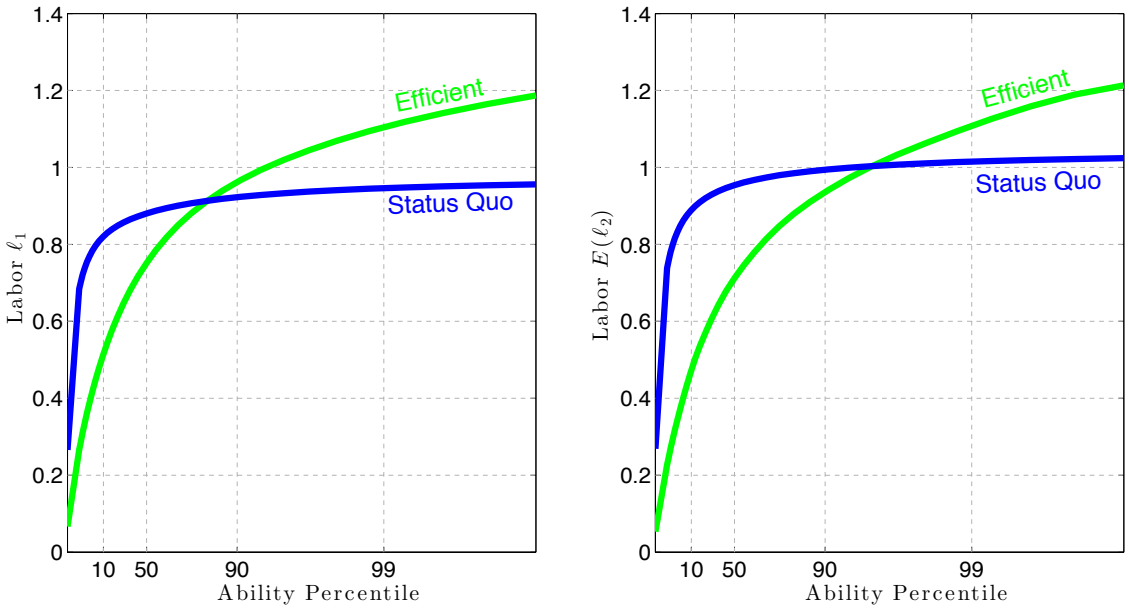
(c) Very High Ability (99.9%)

Figure 2: Second Period Utility Profiles. Darker shades represent higher probabilities.

To understand this surprising result, one needs to realize that, for each type, the efficient distribution of second period consumption and utilities reflects the conflicting objectives of providing incentives to accumulate human capital, providing consumption insurance, and eliciting efficient labor effort in the second period. Figure 2 shows the distribution of second period utilities for three ability levels (darker segments of the lines represent higher probability realizations). Second period utilities are, in general, hump-shaped in the human capital shock: the agents face a downside risk both at the left-tail and in the right-tail of the distribution. Consider the picture in the middle, which shows second period utilities for an agent with higher dispersion of consumption and lower effort (between 77th and 99th percentile). For the lowest realizations of h_2 the efficient allocation provides a very low second period utility, significantly lower than the status quo allocation. This increases the incentives to accumulate human capital and is also responsible for the high standard deviation of second period log-consumption. For the highest realizations of h_2 the efficient allocation is decreasing in the shock. It is efficient to have the agents exert high labor effort in the second period, and consumption insurance limits the rewards for doing so. Finally, for the likeliest realizations of h_2 in the middle range, the desire to insure against those shocks dominates other considerations, and the second period utility is relatively flat. Both the efficiency consideration at the top and insurance consideration in the middle of the distribution decrease the learning effort, which is then mainly elicited by the first effect, essentially a threat of very low consumption in the unlikely case of a low shock realization.[12] For the agent in the middle picture, the last two effects dominate and their learning effort is lower than in the status quo economy. For agents at the bottom of the ability distribution (pitured in the left picture) the insurance objective clearly dominates, their second period utility is almost flat, and their learning effort is very low. At the top of the ability distribution (pictured on the right), on the other hand, the motivations to provide higher learning effort and higher labor effort for high shocks outweigh the insurance motive.

Figure 2 also shows that low-ability agents enjoy higher utility in the second period, while high-ability agents face the largest decrease in the second period utility. Relative to the

[12] A consequence of concave utility is that utility is more responsive at lower levels of consumption, so it is less costly for the planner to incentivize learning effort by punishing low realizations than by rewarding good realizations. Relative to the status quo, the efficient second period consumption function is significantly more concave.



(a) First period.

(b) Second period (expected values).

Figure 3: Labor, by ability percentiles.

status quo, the consumption distribution thus becomes more equal within each period. Low-ability agents enjoy, in expectation, higher consumption than in the status quo economy, and resources are redistributed away from high-ability agents. Labor effort, shown in Figure 3, moves in the opposite direction. Relative to the status quo, the labor earnings distribution (expected labor earnings in the second period) becomes more unequal, as it is optimal to concentrate labor earnings at the top of the distribution.

Overall, the moral hazard friction increases consumption dispersion in the second period, while the private information friction tends to equalize consumption across agents. On the other hand, both the moral hazard and private information friction tend to increase labor effort dispersion. Numerical simulations also show that the private information friction tends to limit the amount of risk faced by high-ability agents. In its absence, the hump-shaped profiles in Figure 2 are significantly more pronounced: higher ability agents face more left-tail downside risk, but they also face more right-tail downside risk of having lower utility after receiving high shocks.

Savings wedge. Figure 4 illustrates that the savings wedge is strictly increasing in ability. Since high-ability agents face the riskiest second period consumption, they have the highest incentive to self-insure through savings. In order to discourage that, the social planner imposes a higher savings tax on them. This result stands in contrast to the previous literature that features exogenous i.i.d. shocks (e.g. Albanesi and Sleet 2006), where high-ability agents typically face less consumption risk, and a lower savings wedge than low-ability agents.

For the highest ability agents, the savings wedge is 34 percent. This translates into an efficient tax on annual capital income τ_a^k of 49.3 percent.[13] Thus, for the highest ability agents the efficient tax on capital income is about 12.3 percent higher than the current U.S. tax rate. However, for the 99th percentile the savings wedge is 27.3 percent, which translates into an efficient tax on annual capital income τ_a^k of 37 percent, almost exactly identical to the current U.S. tax rate. Thus, while the top one percent would optimally face an increase in the capital income tax rate, the optimal tax rate for the bottom 99 percent is lower than the current one. An individual with median abilities would, for example, face a capital income tax rate of only 5.3 percent.

Labor wedge. The labor wedge in the first period and the expected labor wedge in the second period are shown in Figure 5a. Several features stand out. First, since high-ability agents are faced with the highest savings tax, they also need the highest increase in the labor wedge. Hence the difference between the first period labor wedge and the second period expected labor wedge is highest for high-ability agents (but small, at about 2 percent increase). Second, the labor wedge in the first period decreases with abilities and converges to zero. This is due to the assumption of lognormally distributed abilities, as in the static Mirrlees case, and the fact that the utility function is additively separable in labor supply and effort. The second period labor wedge is shown in Figure 5b. The labor wedge is very high for low human capital realizations and then decreases with human capital, as predicted in Proposition 3. The decrease is most rapid for higher ability levels, reflecting the fact that higher ability agents face a more risky consumption profile.

[13] We use the relationship $1-\delta = \beta(1+r(1-\tau^k))$ to recover the tax on capital earnings over 20 years τ^k, and then convert it to an annual value τ_a^k.

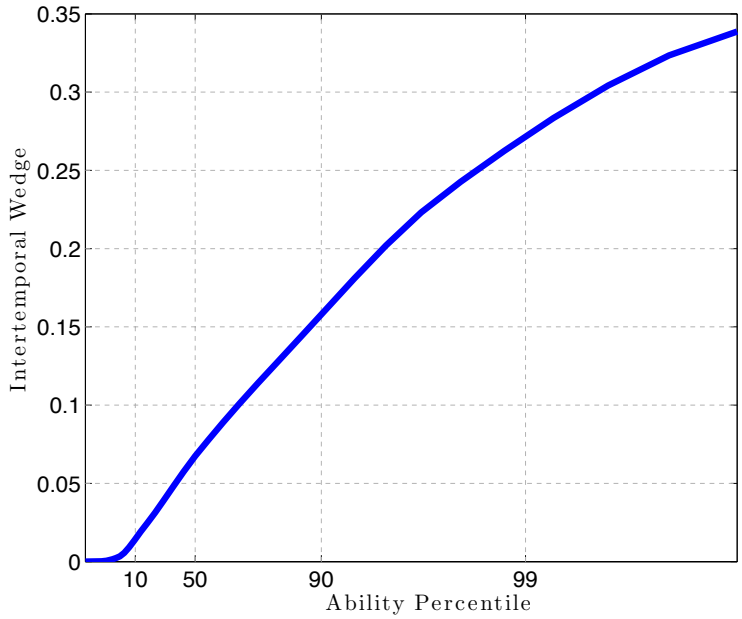
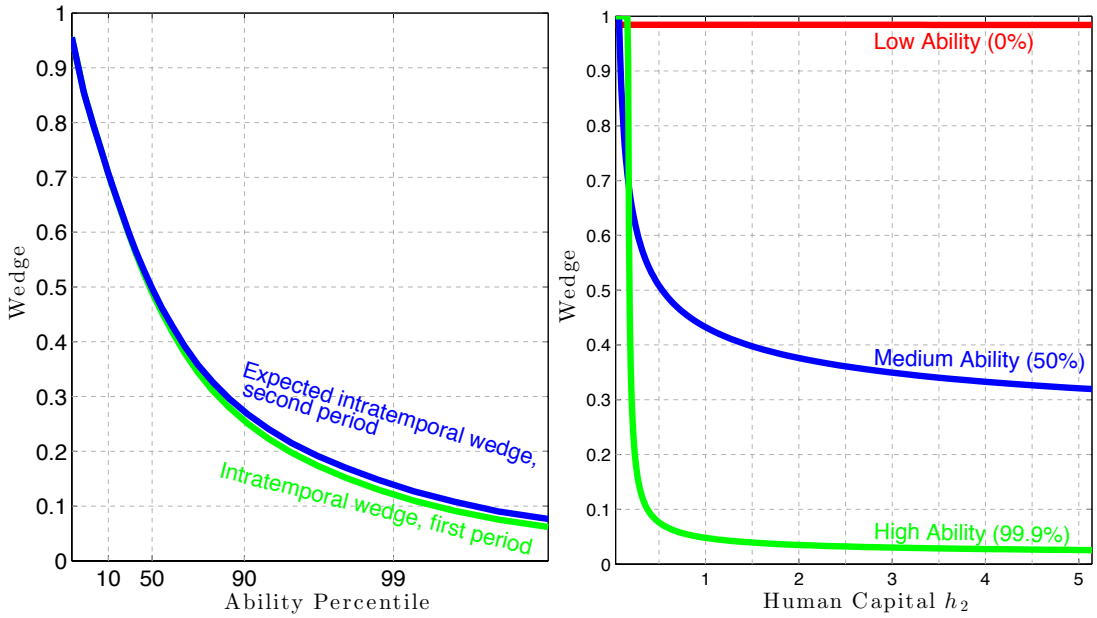


Figure 4: Savings Wedge



(a) (b)

Figure 5: Labor Wedge

Welfare. We now compute the welfare gains relative to the status quo. The overall welfare gain is defined as the percentage increase in period consumption that would make an agent

who does not yet know her type indifferent between the status quo allocation and the constrained efficient allocation, keeping labor and effort unchanged. Specifically, the welfare gain is the η that solves,

$$W\left((1+\eta)c_1^{SQ},(1+\eta)c_2^{SQ},\ell_1^{SQ},\ell_2^{SQ},e_1^{SQ}\right)=W^{CE}.$$

where the SQ denote status quo allocations and CE denotes constrained efficient allocations. We find that the welfare gains of switching to an optimal tax system are equivalent to a $\eta = 4.5\%$ increase in consumption in every period and state of the world.

Table 3: Welfare Gains

	Welfare Gains
Status Quo	0%
Constrained Efficient	4.5%
No moral hazard	6.4%
No private information	26.8%
No frictions	32.7%

We report the welfare gains in Table 3. Shutting down the moral hazard friction completely (by setting the Lagrange multiplier $\phi(a) = 0$) yields welfare gains of 6.4 percent over the status quo. Shutting down the private information friction completely (by setting the Lagrange multiplier $\theta(a) = 0$) yields a much larger gain in welfare of 26.8 percent. Shutting down both constraints completely yields a first-best welfare gain of 32.7 percent.

The distribution of welfare changes across types is illustrated in Figure 6. The large welfare gains accrue at the bottom of the ability distribution. In contrast, the top abilities lose a substantial amount of welfare compared to the status quo economy. However, it is worth noting that welfare is still increasing with ability in the constrained efficient economy in order to prevent agents from misreporting their ability level (as expressed by the envelope condition (9)).

Verifying the validity of the First Order Approach. Since second period utility is not monotonically increasing in human capital realizations, as shown in Figure 2, the sufficient conditions in Proposition 7 are not satisfied. We therefore verified directly that the expected

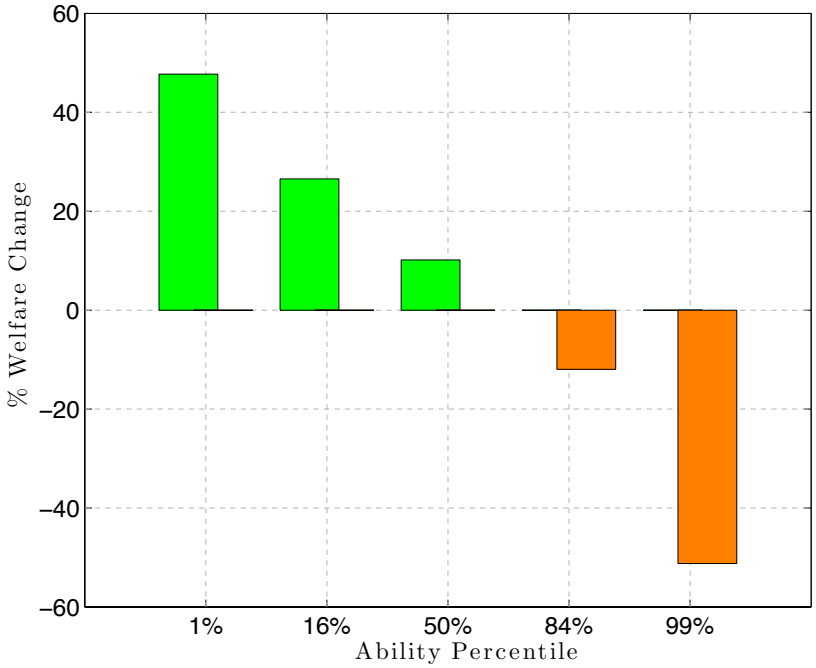


Figure 6: Welfare change of the unborn agent, from the status quo to the constrained efficient economy, by ability percentiles.

lifetime utility is concave in the effort. Figure 7 shows that this is indeed the case, and that the first order approach is valid.

5 Conclusion

This paper addresses two questions: What is the optimal tax structure when there is endogenous human capital accumulation? What are the welfare gains for the U.S. from switching to an optimal tax system? We answer these normative questions in a human capital framework that has been successful at positive analysis. The prominent features of the framework are permanent ability differences, Ben-Porath human capital accumulation technology, and risky returns to investments in human capital. The model is sufficiently rich to be useful for policy analysis, and we show that it is also tractable enough for the normative analysis.

The interaction of private information and moral hazard frictions is novel in the optimal taxation with human capital literature, and produces several prominent results. First, the labor wedge has a sharp characterization. When utility is separable in labor and learning efforts,

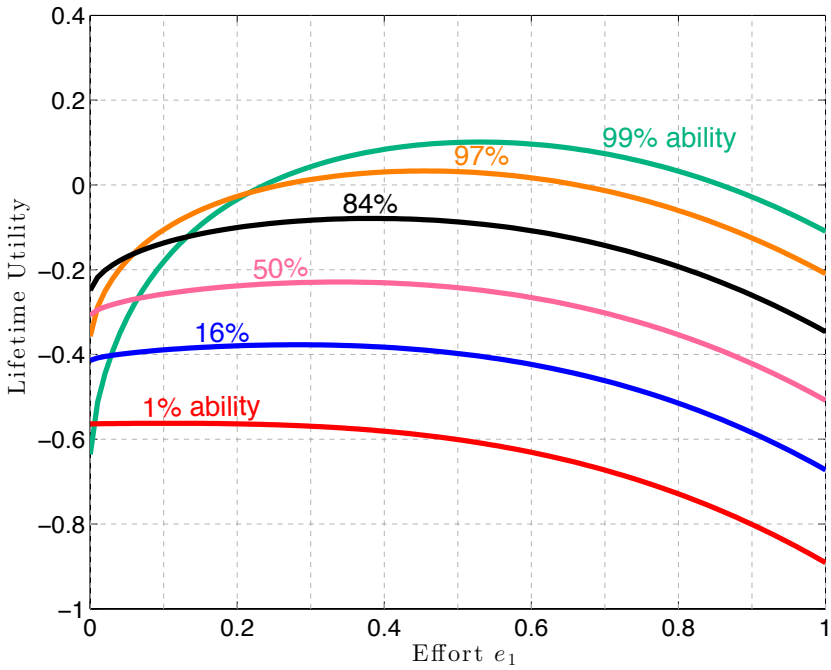


Figure 7: Verification of concavity of effort decision

the average intratemporal wedge is increasing over an agent's lifetime in order to correct for undesired distortions from a positive savings wedge. Second, the "no distortion at the top" result from Mirrlees (1971) might not apply when taxing labor encourages human capital accumulation. Quantitatively, we find that it is efficient to design the tax system so that higher ability people face higher consumption risk in order to provide them with incentives to accumulate human capital. Consequently, higher ability people have the highest incentives to self-insure through physical savings, and the optimal system implements a savings tax that is increasing in ability. Overall, the optimal system yields high welfare gains.

We have assumed that human capital investments are completely unobservable. While this is more likely the case for older workers who have completed their formal education, human capital investments early in life have some observable components, such as schooling hours. It would be interesting to explore how incentives to elicit unobservable human capital investments interact with schooling subsidies and other incentives to elicit observable human capital investments.

Appendix A A Lagrangean Solution Method

Let λ, $\phi(a)q(a)$ and $\theta(a)q(a)$ be the Lagrange multipliers on the resource constraint (6), the first order condition (8) and on the envelope condition (9). The planning problem can be written as a saddle point of the Lagrangean

$$\max_{c,y,e} \min_{\lambda,\theta,\phi} \mathscr{L},$$

where

$$\begin{aligned}
\mathscr{L} = \int_A \Bigg\{ & \Big(1+\theta(a)\Big)W(a) - \theta(a)W(\underline{a}) \\
& - \lambda \left[c_1(a) - y_1(a) + \beta \int_H \left[c_2(a,h_2) - y_2(a,h_2) \right] p(h_2|e_1(a))\, dh_2 \right] \\
& - \theta(a) \int_{\underline{a}}^{a} \left[V_\ell \left(\frac{y_1(\tilde{a})}{\tilde{a}h_1}, e_1(\tilde{a}) \right) \frac{y_1(\tilde{a})}{\tilde{a}h_1} + \beta \int_H V_\ell \left(\frac{y_2(\tilde{a},h_2)}{\tilde{a}h_2}, 0 \right) \frac{y_2(\tilde{a},h_2)}{\tilde{a}h_2} p(h_2|e_1(\tilde{a}))\, dh_2 \right] \frac{d\tilde{a}}{\tilde{a}} \\
& - \phi(a) \left[V_e \left(\frac{y_1(a)}{ah_1}, e_1(a) \right) - \beta \int_H \left[U(c_2(a,h_2)) - V \left(\frac{y_2(a,h_2)}{ah_2}, 0 \right) \right] p_e(h_2|e_1(a))\, dh_2 \right] \Bigg\} q(a)\, da
\end{aligned}$$

The first-order condition in $W(\underline{a})$ implies $\int_A \theta(a)q(a)\, da = 0$. Using this, integrating by parts and rearranging terms, one obtains

$$\begin{aligned}
\mathscr{L} = \int_A \Bigg\{ & \Big(1+\theta(a)\Big)W(a) - \lambda \left[c_1(a) - y_1(a) + \beta \int_H \left[c_2(a,h_2) - y_2(a,h_2) \right] p(h_2|e_1(a))\, dh_2 \right] \\
& - \Theta(a) \left[V_\ell \left(\frac{y_1(a)}{ah_1}, e_1(a) \right) \frac{y_1(a)}{ah_1} + \beta \int_H V_\ell \left(\frac{y_2(a,h_2)}{ah_2}, 0 \right) \frac{y_2(a,h_2)}{ah_2} p(h_2|e_1(a))\, dh_2 \right] \\
& - \phi(a) \left[V_e \left(\frac{y_1(a)}{ah_1}, e_1(a) \right) + \beta \int_H \left[U(c_2(a,h_2)) - V \left(\frac{y_2(a,h_2)}{ah_2}, 0 \right) \right] \frac{p_e(h_2|e_1(a))}{p(h_2|e_1(a))} p(h_2|e_1(a))\, dh_2 \right] \Bigg\} q(a)\, da
\end{aligned}$$

where $\Theta(a) = (aq(a))^{-1} \int_a^{\overline{a}} \theta(\tilde{a})q(\tilde{a})\, d\tilde{a}$ is the cross-sectional cumulative of the Lagrange multipliers on the envelope condition.

Appendix B Implementation

In this section we decentralize the efficient allocations through a tax system. We describe the tax system in two steps. In the first one, we follow Werning (2011) to augment the direct mechanism and allow the agents to borrow and save, but design the savings tax in such a way that the agents choose not to do so. In the second step, we design an indirect tax mechanism that implements the efficient allocation.

Step 1. In the first step, define the tax on savings as follows. Suppose that an a-type agent reports $\hat{a}$. Enlarge the direct mechanism by allowing the agent to borrow and save. Let s be pre-tax savings, and $x(s)$ be second period after-tax savings satisfying $x(0) = 0$. The agent's budget constraints are

$$c_1 + s \leq c_1(\hat{a}) \tag{A-1}$$

$$c_2 \leq c_2(\hat{a}, h_2) + x(s) \quad \forall h_2. \tag{A-2}$$

$x(s)$ can be easily transformed to a more usual tax on interest income $\tau^k(s)$ by $x(s) = \left[1 + r\left(1 - \tau^k(s)\right)\right] s$. Let the lifetime utility from the utility-maximizing report, conditional on savings being s be

$$\hat{W}(s; x|a) = \max_{\hat{a}} \left\{ U(c_1(\hat{a}) - s) - V\left(\frac{y_1(\hat{a})}{a h_1}, e\right) + \beta \int_H \left[U(c_2(\hat{a}, h_2) + x(s)) - V\left(\frac{y_2(\hat{a}, h_2)}{a h_2}, 0\right)\right] p(h_2|e) dh_2 \right\}.$$

For each ability level a, define now a function $x^*(\cdot, a)$ to be a function x such that the agent is *indifferent* among all the savings levels:

$$\hat{W}(s; x^*(\cdot, a)|a) = W(a) \quad \forall s.$$

Differentiating the function x^* and evaluating at $s = 0$, one obtains $x^*_s(0, a) = 1 - \delta(a)$ and $x^*_a(0, a) = 0$. That is, the derivative with respect to the savings is equal to the inverse of the

savings wedge, and the derivative with respect to one's type is always zero, when evaluated at zero savings.[14]

Step 2. In the second step consider a tax system consisting of income tax functions $T_1(y_1)$, $T_2(y_1,y_2,h_2)$ and a savings tax $X(s,y_1)$ satisfying $X(0,y_1)=0$. While the first period income tax depends only on the current income, the second period income tax depends on the second period human capital realization, and can potentially depend on the history of incomes. The agent faces the following budget constraints:

$$c_1+s\leq y_1-T_1(y_1) \tag{A-3}$$

$$c_2\leq X(s,y_1)+y_2-T_2(y_1,y_2,h_2)\quad \forall h_2. \tag{A-4}$$

A consumer with ability a maximizes the expected utility

$$U(c_1)-V\left(\frac{y_1}{ah_1},e\right)+\beta\int_H\left[U(c_2(h_2))-V\left(\frac{y_2(h_2)}{ah_2},0\right)\right]p(h_2|e)dh_2 \tag{A-5}$$

subject to the budget constraints (A-3) and (A-4). The solution to this *market problem* for all abilities is given by $(\tilde{c},\tilde{y},s)$, where $(\tilde{c},\tilde{y})$ is an allocation and $s(a)$ are savings. We prove the following version of the taxation principle (see Hammond (1979)):

Proposition 6. *If an allocation (c,y) satisfies the incentive constraint (5) then there exists a tax system (T_1,T_2,X) such that $X(0,y)=0$ for all y, and $(c,y,0)$ solves the market problem. Conversely, let (T_1,T_2,X) be a tax system such and $(c,y,0)$ solves the market problem. Then the allocation (c,y) is incentive compatible.*

Proof. Suppose that an allocation (c,y) satisfies the incentive constraint (5). Define the tax functions (T_1,T_2,X) to be such that they satisfy

$$\begin{aligned}T_1(y_1(a))&=c_1(a)-y_1(a)\\ T_2(y_1(a),y_2(a,h_2),h_2)&=c_2(a,h_2)-y_2(a,h_2)\\ X(s,y_1(a))&=x^*(s,a).\end{aligned}$$

[14]The second derivative follows simply from the fact that $x^*(0,a)=0$ for all a.

For other values in the domain set the taxes T_1 and T_2 high enough so that no agent chooses such values. Let

$$\tilde{W}(s,\hat{a}|a) = U\left(c_1(\hat{a}) - s\right) - V\left(\frac{y_1(\hat{a})}{ah_1}, e\right) + \beta \int_H \left[U\left(c_2(\hat{a},h_2) + x^*(s)\right) - V\left(\frac{y_2(\hat{a},h_2)}{ah_2}, 0\right)\right] p(h_2|e)\, dh_2.$$

Note that $\hat{W}(s;x^*|a) = \max_{\hat{a}} \tilde{W}(s,\hat{a}|a)$. By the definition of x^*,

$$W(a|a) = \tilde{W}(0,a|a) \geq \max_s \tilde{W}(s,\hat{a}|a) \geq \tilde{W}(0,\hat{a}|a) = W(\hat{a}|a) \quad \forall \hat{a} \in A,$$

Choosing $(c,y,0)$ yields lifetime utility $W(a|a)$. Any other choice yields $\tilde{W}(s,\hat{a}|a)$ or lower. Hence $(c,y,0)$ solves the market problem.

Conversely, take any tax system (T_1,T_2,M), and let (c,y,s) be the solution to the market problem. Then a type a agent prefers $(c(a),y(a),0)$ to $(c(\hat{a}),y(\hat{a}),0)$. The allocation (c,y) is thus incentive compatible. ■

Appendix C Validity of the First-Order Approach

Proposition 7 shows the conditions for sufficiency of (8) and Proposition 8 does the same for (9). In what follows, let $P(h|e) = \int_{\underline{h}}^{h} p(\tilde{h}|e)\, d\tilde{h}$.

Proposition 7. *Suppose that $e^*(\hat{a}|a)$ satisfies (8), and that (i) $\int_{\underline{h}}^{h} P(\tilde{h}|e)\, d\tilde{h}$ is nonincreasing and convex in e for each h, (ii) $\int_H hp(h|e)\, dh$ is nondecreasing concave in e, and (iii) $U(c_2(\hat{a},h)) - V\left(\frac{y_2(\hat{a},h)}{ah}, 0\right)$ is nondecreasing and concave in h. Then (4) holds.*

The proof is omitted because it follows directly from Jewitt (1988) (Theorem 1). It shows that under the conditions of the proposition the objective function is strictly concave in e, implying sufficiency of the first-order conditions.

The next proposition shows that if both earnings and learning effort are monotone in the report, one can recover the global incentive compatibility constraint (5) from the envelope condition (9):

Proposition 8. *Suppose that the allocation satisfies (9). If (i) $e^*(\hat{a}|a)$, $y_1(\hat{a})$ and $y_2(\hat{a},h_2)$ are all nondecreasing in $\hat{a}$ for each h_2, and (ii) $\frac{y_2(\hat{a},h_2)}{h_2}$ is nondecreasing in h_2 for each $\hat{a}$, then (5) holds.*

Proof. Suppose that an allocation satisfies (9). Assume that $\hat{a} < a$. Then (9) implies that (bold symbols indicate changes from the previous equation)

$$
\begin{aligned}
& W(a) - W(\hat{a}) \\
&= \int_{\hat{a}}^{a} \left\{ V_\ell \left(\frac{y_1(\tilde{a})}{\tilde{a}h_1}, e_1^*(\tilde{a}|\tilde{a}) \right) \frac{y_1(\tilde{a})}{\tilde{a}h_1} + \beta \int_H V_\ell \left(\frac{y_2(\tilde{a},h_2)}{\tilde{a}h_2}, 0 \right) \frac{y_2(\tilde{a},h_2)}{\tilde{a}h_2} p(h_2|e_1^*(\tilde{a}|\tilde{a}))\, dh_2 \right\} \frac{d\tilde{a}}{\tilde{a}} \\
&\geq \int_{\hat{a}}^{a} \left\{ V_\ell \left(\frac{y_1(\hat{a})}{\tilde{a}h_1}, e_1^*(\hat{a}|\tilde{a}) \right) \frac{y_1(\hat{a})}{\tilde{a}h_1} + \beta \int_H V_\ell \left(\frac{y_2(\hat{a},h_2)}{\tilde{a}h_2}, 0 \right) \frac{y_2(\hat{a},h_2)}{\tilde{a}h_2} p(h_2|e_1^*(\tilde{a}|\tilde{a}))\, dh_2 \right\} \frac{d\tilde{a}}{\tilde{a}} \\
&\geq \int_{\hat{a}}^{a} \left\{ V_\ell \left(\frac{y_1(\hat{a})}{\tilde{a}h_1}, e_1^*(\hat{a}|\tilde{a}) \right) \frac{y_1(\hat{a})}{\tilde{a}h_1} + \beta \int_H V_\ell \left(\frac{y_2(\hat{a},h_2)}{\tilde{a}h_2}, 0 \right) \frac{y_2(\hat{a},h_2)}{\tilde{a}h_2} p(h_2|e_1^*(\hat{a}|\tilde{a}))\, dh_2 \right\} \frac{d\tilde{a}}{\tilde{a}} \\
&= W(\hat{a}, e_1^*(\hat{a}|a)|a) - W(\hat{a}).
\end{aligned}
$$

The first equality applies (9). The first inequality follows from the assumption that $e^*(\hat{a}|a)$, $y_1(\hat{a})$ and $y_2(\hat{a},h_2)$ are all increasing in $\hat{a}$. The second inequality follows from the fact that $\frac{y_2(\hat{a},h_2)}{h_2}$ increases in h_2 for all $\hat{a}$, that the distribution p is such that, for any increasing function $f(h)$, $\int_H f(h)p(h|e)\,dh$ increases in e, and that $e^*(\hat{a}|a)$ increases in $\hat{a}$ again. Finally, the last equality follows from the fundamental theorem of calculus. The proof is similar for $\hat{a} > a$. Therefore, global incentive compatibility (5) holds. ■

Taken together, Propositions 7 and 8 give a set of monotonicity conditions that ensure validity of the first order approach. They can be checked numerically by computing ex-post the effort plan $e^*(\hat{a}|a)$ and verifying the monotonicity and concavity requirements, which we do in Section 4. These conditions are sufficient, but not necessary. If they fail, one may still be able to verify incentive compatibility by checking directly the conditions (4) and (5).

References

Abraham, A., S. Koehne, and N. Pavoni (2014). Optimal income taxation with asset accumulation. Working paper. 4

Aguiar, M. and E. Hurst (2013). Deconstructing life-cycle expenditure. *Journal of Political Economy 121*(3), 437 – 492. 4

Albanesi, S. (2007). Optimal taxation of entrepreneurial capital with private information. Working paper, Columbia University. 4

Albanesi, S. and C. Sleet (2006). Dynamic optimal taxation with private information. *The Review of Economic Studies 73*(1), 1–30. 3, 4, 23

Battaglini, M. and S. Coate (2008). Pareto efficient income taxation with stochastic abilities. *Journal of Public Economics 92*(3-4), 844–868. 4

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy 75*(4), pp. 352–365. 4

Boháček, R. and M. Kapička (2008). Optimal human capital policies. *Journal of Monetary Economics 55*, 1–16. 5

Bovenberg, A. L. and B. Jacobs (2005). Redistribution and education subsidies are Siamese twins. *Journal of Public Economics 89*(11). 5

Browning, M., L. P. Hansen, and J. J. Heckman (1999). Micro data and general equilibrium models. In M. Woodford and J. Taylor (Eds.), *Handbook of Macroeconomics, Volume 1, Chapter 8*. North Holland. 17, 18

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011). Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *American Economic Review 101*(3), 471–75. 17

da Costa, C. E. and L. J. Maestri (2007). The risk properties of human capital and the design of government policies. *European Economic Review 51*(3), 695 – 713. 5

Diamond, P. and J. Mirrlees (2002). Optimal taxation and the Le Chatelier principle. Working paper, MIT. 5

Farhi, E. and I. Werning (2007). Inequality and social discounting. *Journal of Political Economy 115*(1), 365–402. 4

Farhi, E. and I. Werning (2013). Insurance and taxation over the life cycle. *The Review of Economic Studies 80*(2), 596–635. 4, 14

Findeisen, S. and D. Sachs (2015a). Designing efficient college and tax policies. Working paper, CESifo Working Paper Series No. 5435. 5

Findeisen, S. and D. Sachs (2015b). Education and optimal dynamic taxation: The role of income-contingent student loans. Working paper, University of Mannheim. 5

Golosov, M., N. R. Kocherlakota, and A. Tsyvinski (2003). Optimal indirect and capital taxation. *The Review of Economic Studies 70*, 569–587. 4, 14

Golosov, M., A. Tsyvinski, and M. Troshkin (2015). Redistribution and social insurance. Working paper, Yale University. 4

Gorry, A. and E. Oberfield (2012). Optimal taxation over the life cycle. *Review of Economic Dynamics 15*(4), 551–572. 5

Grochulski, B. and T. Piskorski (2010). Risky human capital and deferred capital income taxation. *Journal of Economic Theory 145*(3), 908 – 943. 5, 13

Guner, N., R. Kaygusuz, and G. Ventura (2014, October). Income taxation of U.S. households: Facts and parametric estimates. *Review of Economic Dynamics 17*(4), 559–581. 17, 18

Guvenen, F. (2007). Learning your earning: Are labor income shocks really very persistent? *American Economic Review 97*(3), 687–712. 4

Hammond, P. J. (1979). Straightforward individual incentive compatibility in large economies. *The Review of Economic Studies 46*, 263–282. 30

Huggett, M., G. Ventura, and A. Yaron (2006). Human capital and earnings distribution dynamics. *Journal of Monetary Economics 53*, 265–290. 4, 19

Huggett, M., G. Ventura, and A. Yaron (2011). Sources of lifetime inequality. *American Economic Review 101*(7), 2923–54. 2, 4, 7, 17, 18, 19

Jacobs, B., D. Schindler, and H. Yang (2012). Optimal taxation of risky human capital. *The Scandinavian Journal of Economics 114*(3), 908–931. 5, 11

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993). Earnings losses of displaced workers. *American Economic Review 83*(4), 685–709. 7

Jewitt, I. (1988). Justifying the first-order approach to principal-agent problems. *Econometrica 56*(5), pp. 1177–1190. 16, 31

Kambourov, G. and I. Manovskii (2009). Occupational specificity of human capital. *International Economic Review 50*(1), 63–115. 7

Kapička, M. (2006). Optimal taxation and human capital accumulation. *Review of Economic Dynamics 9*, 612–639. 5

Kapička, M. (2015). Optimal Mirrleesean taxation in a Ben-Porath economy. *American Economic Journal: Macroeconomics 7*(2), 219–48. 5, 13

Kocherlakota, N. R. (2005). Zero expected wealth taxes: A Mirrless approach to dynamic optimal taxation. *Econometrica 73*, 1587–1621. 4

Krueger, D. and A. Ludwig (2013, January). Optimal progressive taxation and education subsidies in a model of endogenous human capital formation. *American Economic Review P&P 103*, 496–501. 5

Lillard, L. A. and Y. Weiss (1979). Components of variation in panel earnings data: American scientists 1960-70. *Econometrica 47*, 437–453. 4

MaCurdy, T. (1982). The use of time-series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics, 18*, 83–114. 4

McDaniel, C. (2007). Average tax rates on consumption, investment, labor and capital in the OECD 1950-2003. Working paper, Arizona State University. 17, 19

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies 38*, 175–208. 4, 13, 27

Mirrlees, J. A. (1976). Optimal tax theory: A synthesis. *Journal of Public Economics 6*, 327–358. 4

Neal, D. (1995). Industry-specific human capital: Evidence from displaced workers. *Journal of Labor Economics 13*(4), pp. 653–677. 7

Parent, D. (2000). Industry-specific capital and the wage profile: Evidence from the National Longitudinal Survey of Youth and the Panel Study of Income Dynamics. *Journal of Labor Economics 18*(2), pp. 306–323. 7

Peterman, W. B. (2014). The effect of endogenous human capital accumulation on optimal taxation. Working Paper 2012-03, Board of Governors of the Federal Reserve System. 5, 15

Poletaev, M. and C. Robinson (2008). Human capital specificity: Evidence from the dictionary of occupational titles and displaced worker surveys, 1984-2000. *Journal of Labor Economics*. 7

Primiceri, G. E. and T. van Rens (2009). Heterogeneous life-cycle profiles, income risk and consumption inequality. *Journal of Monetary Economics 56*, 20–39. 4

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *The Review of Economic Studies 68*, 205–229. 14

Shourideh, A. (2014). Optimal taxation of wealthy individuals. Working paper, Wharton School, University of Pennsylvania. 5

Slavík, C. and H. Yazici (2014). Machines, buildings, and optimal dynamic taxes. *Journal of Monetary Economics 66*(0), 47 – 61. 13

Stantcheva, S. (2015a). Learning and (or) doing: Human capital investments and optimal taxation. Working paper, Harvard. 5, 11

Stantcheva, S. (2015b). Optimal taxation and human capital policies over the life cycle. Working paper, Harvard. 5

Stiglitz, J. E. (1982). Self-selection and Pareto efficient taxation. *Journal of Public Economics 1*(2), 213 – 240. 13

Storesletten, K., C. Telmer, and A. Yaron (2004). Consumption and risk sharing over the life cycle. *Journal of Monetary Economics 51*, 609–633. 4

Werning, I. (2007). Optimal fiscal policy with redistribution. *Quarterly Journal of Economics 22*(3), 925967. 4

Werning, I. (2011). Nonlinear capital taxation. Working paper, MIT. 29

Abstrakt

V článku studujeme optimální daňové politiky v ekonomice s životním cyklem, rizikovým lidským kapitálem a trvalými rozdíly ve schopnostech jedinců. Optimální politiky se snaží dosáhnou několika cílů zároveň: redistribuovat zdroje mezi agenty, poskytnout pojištění proti šokům do lidského kapitálu, a motivovat k práci a vzdělávání. V optimu platí, že i) pokud je užitková funkce separabilní mezi prací a investicemi do lidského kapitálu, inverze daňového klínu sleduje náhodnou procházku, ii) pokud užitková funkce separabilní není tak i nejschopnější agent čelí distorzím (známý výsledek „no distortion at the top“ neplatí) a iii) kvantitativní výsledky ukazují že agenti s vyššími schopnostmi čelí rizikové spotřebě zatímco agenti s nižšími schopnostmi jsou téměř úplně pojištěni. Zisky z adopce optimálního daňového systému jsou velké.

Working Paper Series
ISSN 1211-3298
Registration No. (Ministry of Culture): E 19443

Individual researchers, as well as the on-line and printed versions of the CERGE-EI Working Papers (including their dissemination) were supported from institutional support RVO 67985998 from Economics Institute of the ASCR, v. v. i.

Specific research support and/or other grants the researchers/publications benefited from are acknowledged at the beginning of the Paper.

Published by
Charles University in Prague, Center for Economic Research and Graduate Education (CERGE) and
Economics Institute of the CAS, v. v. i. (EI)
CERGE-EI, Politických vězňů 7, 111 21 Prague 1, tel.: +420 224 005 153, Czech Republic.
Printed by CERGE-EI, Prague
Subscription: CERGE-EI homepage: http://www.cerge-ei.cz

Phone: + 420 224 005 153
Email: office@cerge-ei.cz
Web: http://www.cerge-ei.cz

Editor: Jan Zápal

The paper is available online at http://www.cerge-ei.cz/publications/working_papers/.

ISBN 978-80-7343-360-4 (Univerzita Karlova v Praze, Centrum pro ekonomický výzkum a doktorské studium)
ISBN 978-80-7344-358-0 (Národohospodářský ústav AV ČR, v. v. i.)

CERGE-EI
P.O.BOX 882
Politických vězňů 7
111 21 Praha 1
Czech Republic
http://www.cerge-ei.cz