Optimal Taxation in a Life-Cycle Economy with Endogenous Human Capital Formation*

Marek Kapička
U.C. Santa Barbara

Julian Neira
U.C. Santa Barbara

November 22, 2012

Abstract

We study efficient allocations and optimal policies in a Mirrleesean life-cycle economy with risky human capital accumulation and permanent ability differences. We assume that ability, labor supply, learning effort and returns to human capital are all private information of the agents. We show that the "no distortion at the top" result from the Mirrleesean literature may not apply if discouraging labor supply increases incentives to invest in human capital. We also show that, under certain conditions, the inverse of the intratemporal wedge follows a random walk, implying that the average intratemporal wedge increases over time. This result is, to our knowledge, novel.

We calibrate a two-period economy and find several notable results. First, to elicit learning effort, it is efficient to make the consumption process risky for high-ability agents while insuring low-ability agents. Second, high-ability agents face the largest expected increase in the intratemporal wedge. Third, high-ability agents face a higher intertemporal wedge. These normative prescriptions differ significantly from the existing literature that abstracts from human capital. We also find large welfare gains for the U.S. from switching to an optimal tax system.

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

*We thank seminar participants at the NBER Summer Institute Macro Public Finance Session, the Workshop on Macroeconomic Applications of Dynamic Games and Contracts at Stony Brook, the Federal Reserve Bank of St. Louis, the Workshop on Financing Human Capital, Credit Constraints, and Market Frictions at the Chicago Fed, the XIII Latin-American Workshop in Economic Theory, the 2012 SED Meetings, and the Midwest Macro Meetings at Notre Dame for helpful comments and discussion.

1 Introduction

We explore the optimal tax structure and efficient allocations in a model where agents are heterogeneous in their ability to produce output, can invest in human capital to augment their productivity, and the rates of return to human capital evolve stochastically over the lifetime. We assume that the government's choices are limited by two frictions. The first one is a standard Mirrleesean private information friction where abilities and labor supply are unobservable by the government. The second one is a moral hazard friction where learning effort and realized rates of return to human capital are unobservable by the government as well. The model framework is similar to the one used by Huggett, Ventura, and Yaron (2011), who show that it is able to account for key empirical features of the dynamics of earnings and consumption. We show that with the informational structure described above, the model is also a useful and tractable framework for studying optimal taxation. In particular, due to the private information about abilities the optimum features permanent differences in earnings and consumption, but also consumption risk over the lifetime; consumption must vary with human capital realizations so as to provide agents with incentives to accumulate human capital.

We derive several theoretical results. We show that, when utility is additively separable in leisure and effort, *the inverse of the intratemporal (labor) wedge follows a random walk*. This result is, to the best of our knowledge, novel. It arises from two unique features of the model: the inverse Euler equation holds, and the agents receive no new information about abilities over the course of their lifetime. While the assumption of additive separability is restrictive, the result serves as a useful benchmark for the analysis of the dynamics of intratemporal wedges. The result immediately implies that the expected labor wedge is increasing with age. We also show that the well-known *"no distortion at the top" result from the Mirrleesean literature may not apply*; if discouraging labor supply increases incentives to invest in human capital, even the "top" agent will optimally face a positive marginal tax.

We calibrate a two-period model to the U.S. economy and show that consumption dispersion, as measured by the variance of log-consumption, is increasing in ability at the optimum. In other words, low-ability agents are relatively insured whereas high-ability agents face significant consumption risk. The reason is that high-ability agents must be provided the strongest incentives to accumulate human capital. These incentive are provided by making their second period consumption very sensitive to the realization of human capital shocks, i.e. by making their consumption very risky. This result stands in contrast to some of the previous literature with exogenous productivities (for example Albanesi and Sleet (2006)), where high-ability agents typically have more consumption insurance. The result implies that the intertemporal (savings) wedge increases in ability, which is again a conclusion that differs from e.g. Albanesi and Sleet (2006).

In order to get a sense of the magnitude of the normative prescriptions, we compare the efficient allocations to a benchmark decentralized economy with incomplete markets and U.S. tax system. In the benchmark, agents at the 97 percentile of the income distribution face average income tax rates of 23%. The model prescription raises average taxes for these agents to 63%. The efficient labor wedge for these agents is about 9% in the first period and 11% in the second period, whereas low-ability agents face a labor wedge of over 90%. The efficient savings wedge is close to zero for low-ability agents and it increases to 40% for high-ability agents. These normative prescriptions amount to substantial overall welfare gains for the U.S. The utilitarian representative agent is indifferent between the optimal tax system and the benchmark with 11% higher consumption *in every period and state of the world*. We find that, compared to the benchmark economy with U.S. taxes, the efficient allocation involves significant redistribution. Learning effort and output are reallocated from low-ability agents to high-ability agents. On average, the pre-tax labor earnings of the lowest ability drop by 81%, whereas those of the highest ability increase by 38%. Consumption, on the other hand, is reallocated from high-ability agents to low-ability agents. On average, the lowest ability agent increases his consumption by 96%, whereas the highest ability agent reduces his consumption by 31%. The outcome is that the social planner increases pre-tax labor earnings inequality while decreasing consumption inequality.

A key aspect that makes the optimal tax problem tractable is that we extend the method developed in Boháček and Kapička (2008) (for riskless observable human capital) and Kapička (2008) (for riskless unobservable human capital). Both papers show that with a first-order approach one can partially separate the redistributional dimension of the optimal tax problem, where the social planner redistributes resources across agents, and the dynamic dimension of the optimal tax problem, where the social planner chooses the optimal sequences of labor supply and learning effort. In addition, the dynamic dimension can be conveniently written recursively. Using recursive Lagrangean techniques of Marcet and Marimon (2009) we show that a similar decomposition is possible in this model. The result relies on the assumption that abilities are permanent (which is consistent with the model structure of Huggett, Ventura, and Yaron (2011)). The assumption that human capital is observable is also important for preserving tractability. It is, however, worth noting that due to unobservability of learning effort the model shares some features typically associated with models with unobserved human capital, namely that the incentives to accumulate human capital must be provided indirectly, through the income taxes.

Human capital in our model is both risky and observable. These assumptions are supported by a large literature which shows that displaced workers suffer a persistent decrease in wages. Jacobson, LaLonde, and Sullivan (1993) is a classic paper that argues that high-tenure workers separating from distressed firms suffer long-term wage losses. Neal (1995) and Parent (2000) provide evidence of industry-specific human capital whereas Poletaev and Robinson (2008) and Kambourov and Manovskii (2009) argue that human capital is specific to occupations. Following this literature, observability of human capital amounts to observability of a person's industry, firm, or occupation.

1.1 Relationship to the existing literature

Recent research on optimal taxation with private information followed the seminal contributions of Mirrlees (1971), Mirrlees (1976), and Mirrlees (1986), and extended them to dynamic economies. It has mostly focused on cases when the individual skills are exogenous (Golosov, Kocherlakota, and Tsyvinski (2003), Kocherlakota (2005), Albanesi and Sleet (2006), Battaglini and Coate (2008), Farhi and Werning (2005), Werning (2007)). A most complete life-cycle analysis is Golosov, Tsyvinski, and Troshkin (2010) and Farhi and Werning (2010) who analyze optimal taxation in an environment where individual skills are Markov (essentially a stripped-down version of the RIP model).

In contrast, this paper focuses on a case when individual skills are endogenous. A significant progress in this direction has been made by Grochulski and Piskorski (2010) who study a problem with unobservable risky human capital. However, investment in human capital in their model is only possible in the initial period and the dynamics in the remaining periods are technically similar to the above models with exogenous skills. Boháček and Kapička (2008) and Kapička (2008) study environments with riskless human capital in an infinite horizon setting. Boháček and Kapička (2008) assumes that human capital is observable, while Kapička (2008) assumes that it is unobservable.[1] While each of those models captures some important component of endogenous skill formation, neither of them is rich enough to fully capture the earnings and consumption dynamics observed in the data.

A related research has studied the impact of moral hazard on optimal tax structures. Abraham, Koehne, and Pavoni (2012) study optimal income taxation in a life-cycle framework with hidden effort. Our paper focuses on the interaction between hidden effort and private information. Albanesi (2007) studies a problem with observable risky physical capital, while Shourideh (2010) studies optimal allocation with entrepreneurial risk in a multiperiod setting.

One of the contributions of this paper is that it studies optimal taxation in a framework that is able to account for key features of the dynamics of the earnings and consumption that are observed in the data. As shown by Huggett, Ventura, and Yaron (2011), a properly parameterized life-cycle incomplete markets economy with risky human capital and heterogeneity in abilities is 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)).[2] Finally, the framework is also consistent with the increased dispersion in consumption over the life cycle, as documented by Aguiar and Hurst (2012) or Primiceri and van Rens (2009). Our paper takes the economy

[1] Kapička (2006) analyzes the optimal steady state allocations in a similar environment with unobservable human capital and a restriction that the government can only use current income taxes and agents cannot borrow or save. See also Diamond and Mirrlees (2002) who analyze unobservable human capital investments in a static framework.

[2] 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.

with risky human capital and heterogeneity in abilities as a starting point for the optimal taxation analysis.

2 The Model

Consider the following life-cycle economy. Agents live for $J > 1$ periods. They like to consume, dislike working and exerting learning effort, and have preferences given by

$$\mathbb{E}\sum_{j=1}^{J}\beta^{j-1}\left[U(c_j) - V(\ell_j, e_j)\right], \quad 0 < \beta < 1, \tag{1}$$

where j is age, c_j is consumption, ℓ_j is labor, and e_j 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.

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

$$y_j = ah_j\ell_j \tag{2}$$

Ability is constant over an agent's lifetime and is known to the agents at the beginning of period 1. Ability and initial human capital h_1 are allowed to be correlated, and their joint distribution has density $q(a, h_1)$. The ability has a continuous support $A = (\underline{a}, \overline{a})$, with $\overline{a}$ possibly being infinite. Human capital in the first period, as well as in all other periods, has a continuous support $H = (\underline{h}, \overline{h})$, with $\overline{h}$ possibly infinite.

Human capital next period h_{j+1} depends on idiosyncratic human capital depreciation shock z_{j+1}, current human capital h_j, and on current learning effort e_j:

$$h_{j+1} = \exp(z_{j+1})F(h_j, e_j) \tag{3}$$

where the function F is strictly increasing, strictly concave, and differentiable in both arguments. The idiosyncratic human capital shock is serially uncorrelated, but its density can depend on age j. As is standard in the moral hazard literature, it is useful to transform the state-space representation of the problem to work directly with the distribution induced over h_j. To that end, we construct a probability density function of human capital in period $j+1$ conditional on $f_j = F(h_j, e_j)$, and denote it by $p_{j+1}(h_{j+1}|f_j)$.

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 (1967) form.[3] 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. The formulation chosen in this paper has the advantage that it is entirely consistent with the existing optimal taxation literature.

3 Optimal Taxation in a Two-Period Model

This section solves for optimal allocations in a two-period model. We also simplify the model by assuming that h_1 is the same for everyone in the economy. To reduce notation, the probability density of the second period human capital will be written directly as a function of the first period effort, $p(h_2|e)$.

3.1 Efficient Allocations

The *information structure* is as follows: ability a, labor supply ℓ_1, ℓ_2, learning effort e_1, and human capital shocks z_2 are private information of the agent. Consumption c_1, c_2, savings k_2, earnings y_1, y_2, and human capital h_1, 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.

An *allocation* $(\boldsymbol{c}, \boldsymbol{y})$ consists of consumption scheme $\boldsymbol{c} = \{c_1(\hat{a}), c_2(\hat{a}, h_2)\}$ and earnings scheme $\boldsymbol{y} = \{y_1(\hat{a}), y_2(\hat{a}, h_2)\}$. Consumption and earnings scheme in the first period are conditional on ability report $\hat{a} \in A$. In the second period they are both conditional on

[3] 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).

the ability report in the first period and realization of human capital in the second period, $h_2 \in H$.

Define lifetime utility of an a-type agent who reports ability $\hat{a}$ and exerts effort e as $W(\hat{a}, e|a)$,

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

Effort in the second period is trivially equal to zero. The first period effort $e_1^*(\hat{a}|a)$ maximizes the lifetime utility of an $a-$type that reports $\hat{a}$ and is given by

$$e_1^*(\hat{a}|a) \equiv \arg\max_e W(\hat{a}, e|a). \tag{4}$$

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 her ability:

$$W(a, e_1^*(a|a)|a) \geq W(\hat{a}, e_1^*(\hat{a}|a)|a) \quad \forall a, \hat{a} \in A \tag{5}$$

To reduce notational complexity we will define the utility maximizing effort plan conditional on truthtelling by $e_1(a) = e_1^*(a|a)$, and let $W(a) = W(a, e_1^*(a|a)|a)$ be the corresponding lifetime utility.

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

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

The social welfare function is simply the expected utility of someone who does not yet know his ability:

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

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).*

To reduce the complexity of the problem, we will assume that $R = \beta^{-1}$.

3.1.1 First-Order Approach

The first-order approach replaces the incentive constraint (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 (given by the disutility from spending an additional unit of time by 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):

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

The second one 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 $\tilde{a}$ stand for a dummy variable that distinguishes the variable of integration from the limit of integration, and let $W(\underline{a})$ denote the lifetime utility of the least able agent. The envelope condition is,

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

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 a given ability level. The least able agent has no informational rent.

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 return to the problem of validity of the first-order approach in Section 3.1.2.

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). We show in Appendix A that the planning problem can be written as a saddle point of the Lagrangean:

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

where

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

where $\Phi(a, h_2)$ is related to the multiplier on the the first-order condition in effort (8)

$$\Phi(a,h_2) = \phi(a)\frac{p_f(h_2|e_1(a))}{p(h_2|e_1(a))} \tag{10}$$

and $\Theta(a)$ is the cross-sectional cumulative of the multipliers on the envelope condition (9):

$$\Theta(a) = \frac{1}{aq(a)}\int_a^{\overline{a}} \theta(\tilde{a})q(\tilde{a})\, d\tilde{a}. \tag{11}$$

Both Φ and Θ have a direct economic interpretation: $\Phi(a)$ indicates how costly it is for the social planner to respect the first-order condition in effort. $\Theta(a)$ indicates how much the planner desires to redistribute resources across agents, and is a key in determining how much to distort labor supply of a $a-$type agent. Note also that the Lagrangean makes it easy to pinpoint the contribution of both frictions that are present in the model. In the absence of the moral hazard friction one would set $\Phi = \phi = 0$. In the absence of the private information friction one would set $\theta = \Theta = 0$. The dynamic program can therefore be easily adapted to "shut down" either one of those frictions to study its contribution to the optimal tax problem.

Appendix A also shows a recursive formulation of the Lagrangean that is useful for numerical simulations.

3.1.2 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 schooling choice, or because the envelope condition (9) fails to detect the utility maximizing report. We will now show the conditions for sufficiency of (8) and (9). For the following proposition, let $P(h|f) = \int_{\underline{h}}^{h} p(\tilde{h}|f)\, d\tilde{h}$.

Proposition 1. *Suppose that $e^*(\hat{a}|a)$ satisfies (8), and that*

i. $\int_{\underline{h}}^{h} P(\tilde{h}|f)\, d\tilde{h}$ *is nonincreasing and convex in f for each h.*

ii. $\int_H hp(h|f)\, dh$ *is nondecreasing concave in f.*

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 main difference from Jewitt (1988) 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 supply is increasing in h_2 sufficiently fast, the second period utility may decrease in h_2.

The following result shows that if both earning and effort is monotone in the report then the agent prefers to report the ability truthfully:

Proposition 2. *Suppose that the allocation satisfies (9), and that*

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.*

ii. $\frac{y_2(\hat{a},h_2)}{h_2}$ *is nondecreasing in h_2 for each $\hat{a}$.*

Then (5) holds.

Proof. The proof is similar to the proof of Mirrlees (1986) showing that, in a static environment, increasing income is sufficient for the first order approach to be valid. Suppose that an allocation satisfies (9). Assume that $\hat{a} < a$. Then (9) implies that (bold symbols indicate changes from the previous equation)

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

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 $\hat{a} > a$. $\square$

Taken together, Propositions 1 and 2 give a set of monotonicity conditions that ensure validity of the first order approach. They can be checked numerically by computing ex-post the schooling plan $e^*(\hat{a}|a)$ and verifying the monotonicity and concavity requirements. It is also worth mentioning that those 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).

3.2 Theoretical Implications

We will now characterize the properties of the efficient allocation, and also the properties of the intratemporal and intertemporal wedges. We will assume that the distribution of the second period human capital satisfies the Monotone Likelihood Ratio Property:

Assumption 1 (MLRP). $\frac{p_f(h|f)}{p(h|f)}$ *is strictly increasing in* h *for all* f.

We will also assume that the elasticity of labor supply is not increasing in labor supply, for a given effort.

Assumption 2. *The elasticity of labor supply* $\gamma(\ell, e) = \frac{V_\ell(\ell,e)}{\ell V_{\ell\ell}(\ell,e)}$ *is nonincreasing in* ℓ.

We first show that, under those conditions, the Lagrange multiplier on the first-order condition in effort is nonnegative:

Lemma 1. *Suppose that MLRP and Assumption 2 holds. Then* $\phi(a) > 0$.

Proof. The proof shows that the right-hand side of (12) is negative due to the fact that expected continuation utility decreases in future human capital realizations.

Consider a doubly relaxed problem where equation (8) holds as the following inequality:

$$\frac{V_e\left(\ell_1(a), e_1(a)\right)}{F_e\left(h_1, e_1(a)\right)} \geq \beta \int_H \left[U(c_2(a, h_2)) - V\left(\ell_2(a, h_2), 0\right)\right] \frac{p_f(h_2|e_1(a))}{p(h_2|e_1(a))}\, dh_2 \tag{12}$$

We will now prove that the constraint (12) is slack. In the absence of (12) one obtains that $c_2(a, h_2)$ is independent of h_2. The first-order condition in $\ell_2(a, h_2)$ is

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

If $\gamma(\ell, 0)$ is nonincreasing in ℓ then $\ell_2(a, h_2)$ is strictly increasing in h_2. Since Assumption 1 holds, the right-hand side of (12) is strictly negative. Since the left-hand side of (12) is nonnegative, (12) holds as a strict inequality. Hence, for the first order condition in effort constraint to hold, the opposite inequality must be imposed. From the Kuhn-Tucker theorem we have $\phi(a) > 0$. □

A strictly positive multiplier ϕ implies that the social planner would, in the absence of the constraint (8), increase private marginal costs of effort above the private marginal benefits of effort. In fact, as the proof shows, while the marginal costs would be positive, the marginal benefits of effort would be negative: people with higher human capital would see no consumption increase, but would be required to work more. The moral hazard friction prevents the social planner from achieving such allocations.

The following proposition characterizes the consumption allocation. It's second part draws on the result of Lemma 1.

Proposition 3. *Consumption $\boldsymbol{c}$ satisfies the Inverse Euler Equation:*

$$\frac{1}{U'(c_1(a))} = \int_H \frac{1}{U'(c_2(a,h_2))} p\left(h_2 | F(h_1, e_1(a))\right) dh_2 \quad \forall a \in A \tag{13}$$

If MLRP holds then $c_2(a, h_2)$ is also strictly increasing in h_2.

Proof. The first-order conditions in consumption are

$$\frac{1}{U'(c_1(a))} = \frac{1+\theta(a)}{\lambda} \tag{14}$$

$$\frac{1}{U'(c_2(a,h_2))} = \frac{1+\theta(a)+\Phi(a,h_2)}{\lambda} \tag{15}$$

Take the expectation of (15) and note that $\int_H p_f\left(h_2|F(h_1, e_1(a))\right) dh_2 = 0$. If MLRP holds then $\phi > 0$ by Lemma 1, and so $\Phi(a, h_2)$ is strictly increasing in h_2. The result then follows from (15). □

The first part of the proposition is relatively standard. It 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. The second part of the proposition shows that with moral hazard this is no longer the case, provided that MLRP holds. Conditional on ability, there is a dispersion in consumption in the second period.

3.2.1 Wedges

Define the *intratemporal wedge* τ_j as the gap between the marginal product of labor and the intratemporal marginal rate of substitution. Similarly, define the *intertemporal wedge* δ as the wedge between the current marginal utility of consumption and the expected marginal utility of consumption tomorrow:

$$ah_j(1-\tau_j(a,h_j)) = \frac{V_\ell\big(\ell_j(a,h^j),e_j(a,h^j)\big)}{U'(c_j(a,h^j))}, \quad \text{for } j=1,2$$
$$U'(c_1(a)) = (1-\delta(a))\int_H U'(c_2(a,h_2))p\left(h_2|F(h_1,e_1(a))\right)dh_2.$$

Proposition 3 immediately implies that $\delta(a)$ is strictly positive for each ability level a. The first-order conditions in labor imply that the intratemporal wedge τ satisfies

$$\frac{1}{U'(c_1(a))}\frac{\tau_1(a)}{1-\tau_1(a)} = \left(1+\frac{1}{\gamma(\ell_1(a),e_1(a))}\right)\Theta(a) + \frac{\phi(a)}{\lambda}\frac{V_{\ell e}\left(\ell_1(a),e_1(a)\right)}{V_\ell\left(\ell_1(a),e_1(a)\right)}\frac{1}{F_e\left(h_j,e_1(a)\right)}$$
$$\frac{1}{U'(c_2(a,h_2))}\frac{\tau_2(a,h_2)}{1-\tau_2(a,h_2)} = \left(1+\frac{1}{\gamma(\ell_2(a,h_2),0)}\right)\Theta(a).$$

In what follows, we characterize the limiting intratemporal wedge and its dependence on realized human capital shock.

Proposition 4.

i. Suppose that $\lim_{a\to\overline{a}}\Theta(a)=0$*. Then*

$$\lim_{a\to\overline{a}}\tau_1(a) \lesseqgtr 0 \text{ if } V_{\ell e} \lesseqgtr 0$$
$$\lim_{a\to\overline{a}}\tau_2(a,h_2) = 0 \quad \forall h_2\in H.$$

ii. Suppose that MLRP and Assumption 2 holds. Then $\tau_2(a,h_2)$ *is strictly decreasing in* h_2*.*

The proof follows directly from the first order conditions in labor, from Lemma 1, and from Proposition 3. The implication of the first part is that the "no distortion at the top" result from Mirrlees (1971) does not apply whenever the utility is not additively separable in labor and effort. The intuition is that nonseparability gives the planner the option to change incentives to exert effort by changing first period labor supply. If $V_{\ell e} > 0$, discouraging labor supply in the first period increases incentives to exert effort. Hence it is optimal to do so, even for the top agent. This channel is absent in the second period where the "no distortion at the top" result applies. The second part of the proposition shows that, under the stated conditions, the second period wedge is decreasing in the human capital shock. It is easy to see that if the support is unbounded and $U'(c_2(a, h_2))$ converges to infinity as h_2 converges to infinity, then the second period tax wedge converges to zero in h_2 as well. 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.

Grochulski and Piskorski (2010) also obtain that the "no distortion at the top" does not apply, but their argument 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 supply and investment in human capital. Note also that the result is different from Kapička (2008) 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.[4]

If labor supply and learning effort are additively separable and labor supply has a constant elasticity then we have the following sharp characterization of the intratemporal wedges:

[4]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 (2012) establish the same result when there is capital-skill complementarity. Those arguments rely on general equilibrium effects that are absent in our paper.

Proposition 5. *Suppose that γ is a constant and that $V_{\ell e} = 0$. Then*

$$\frac{1}{\tau_1(a)} = \int_H \frac{1}{\tau_2(a,h_2)} p(h_2|h_1, e_1(a))\, dh_2.$$

Proof. The expression for wedges reduces to

$$\frac{1}{U'(c_2(a,h_2))} \frac{\tau_2(a,h_2)}{1-\tau_2(a,h_2)} = \left(1 + \frac{1}{\gamma\big(\ell_2(a,h_2),0\big)}\right) \frac{\Theta(a)}{\lambda}.$$

To prove the second part, note that if γ is a constant and $V_{\ell e} = 0$ then the intratemporal wedges satisfy

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

Since (13) holds,

$$\frac{1-\tau_1(a)}{\tau_1(a)} = \int_H \frac{1-\tau_2(a,h_2)}{\tau_2(a,h_2)} p(h_2|h_1, e_1(a))\, dh_2.$$

Rearranging, the result follows. $\square$

The result is due to several facts. First, the tax revenue of an $a-$type agent is proportional to $\frac{\tau_j(a,h_j)}{1-\tau_j(a,h_j)}$ for $j = 1, 2$ (Saez (2001)). Second, if the assumptions of Proposition 5 hold then (since the ability shock is permanent) the social planner wants to keep the tax revenue valued at the utility cost $\frac{1}{U'(c_j(a,h_j))}$ constant over time and state. Hence the expression $\frac{1}{U'(c_j(a,h_j))} \frac{\tau_j(a,h_j)}{1-\tau_j(a,h_j)}$ is constant over time and state. Since $\frac{1}{U'(c_j(a,h_j))}$ follows a random walk, the result follows. Jensen's inequality then implies that the average intratemporal wedge is increasing over time:

Corollary 3.

$$\tau_1(a) < \int_H \tau_2(a,h_2) p\left(h_2|F(h_1, e_1(a))\right)\, dh_2. \tag{16}$$

4 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.

In the first step, define the tax on savings as follows. Enlarge the direct mechanism by allowing the agent to save and modify the consumption allocation:

$$c_1 + m(x) \leq c_1(\hat{a}) \tag{17}$$

$$c_2 \leq c_2(\hat{a}, h_2) + x \quad \forall h_2, \tag{18}$$

where x are after-interest savings, and $m(x)$ represents a transformation of the nonlinear tax of savings. We also set $m(0) = 0$. That is, an agent who follows the allocation chosen by the planner pays no tax. The value $m(x)$ is the amount by which current consumption must be reduced in order to increase future consumption by x. As such, it can be easily transformed to a more usual tax on interest income $\tau^k(s)$, where an agent reducing current consumption by s increases future consumption by $\left[1 + r\left(1 - \tau^k(s)\right)\right] s$. Let

$$\begin{aligned}\hat{W}(x; m|a) = \max_{\hat{a}} \Bigg\{ & U(c_1(\hat{a}) - m(x)) - V\left(\frac{y_1(\hat{a})}{ah_1}, e\right) \\ & + \beta \int_H \left[U(c_2(\hat{a}, h_2) + x) - V\left(\frac{y_2(\hat{a}, h_2)}{ah_2}, 0\right)\right] p(h_2|e) dh_2 \Bigg\}\end{aligned}$$

be the lifetime utility from the utility maximizing report, conditional on savings being x. Now define, for each ability level a, a function $m^*(\cdot, a)$ to be such that the agent is *indifferent* among all the savings levels:

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

Differentiating the function M^* and evaluating at $x = 0$, one obtains

$$m^*_x(0, a) = \frac{1}{1 - \delta}$$

$$m^*_a(0, a) = 0$$

That is, the derivative with respect to the savings is equal to the inverse of the intratemporal wedge, and the derivative with respect to one's type is always zero, when evaluated at zero savings.[5] The second derivative follows simply from the fact that $m^*(0, a) = 0$ for all a.

In the second step consider a tax system consisting of income tax income functions $\boldsymbol{T} = (T_1(y_1), T_2(y_1, y_2, h_2))$ and a savings tax $M(x, y)$ satisfying $M(0, y) = 0$. 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:

$$c_1 + M(x, y_1) \leq y_1 - T_1(y_1) \tag{19}$$

$$c_2 \leq x + y_2 - T_2(y_1, y_2, h_2) \quad \forall h_2. \tag{20}$$

A consumer of a given ability a maximizes the expected utility

$$U(c_1) - V\left(\frac{y_1}{ah_1}, e\right) + \beta \int_H \left[U(c_2(h_2)) - V\left(\frac{y_2(h_2)}{ah_2}, 0\right)\right] p(h_2|e) dh_2 \tag{21}$$

subject to the budget constraints (19) and (20). The solution to this *market problem* for all abilities is given by $(\tilde{\boldsymbol{c}}, \tilde{\boldsymbol{y}}, x)$, where $(\tilde{\boldsymbol{c}}, \tilde{\boldsymbol{y}})$ is an allocation and $x(a)$ are savings. We prove the following version of the taxation principle (see Hammond (1979)):

Proposition 6. *If an allocation $(\boldsymbol{c}, \boldsymbol{y})$ satisfies the incentive constraint (5) then there exists a tax system $(\boldsymbol{T}, M)$ such that $M(0, y) = 0$ for all y, and $(\boldsymbol{c}, \boldsymbol{y}, 0)$ solves the market problem.*

Conversely, let $(\boldsymbol{T}, M)$ be a tax system such and $(\boldsymbol{c}, \boldsymbol{y}, x)$ solves the market problem. Then the allocation $(\boldsymbol{c}, \boldsymbol{y})$ is incentive compatible.

Proof. Suppose that an allocation $(\boldsymbol{c}, \boldsymbol{y})$ satisfies the incentive constraint (5). Define the tax

[5] If one converts m^* back to a nonlinear tax on savings $\tau^k(s)$, then $\tau^k(0) = \frac{\delta(1+r)}{r}$.

functions to be such that they satisfy

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

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

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

By the definition of m^*,

$$W(a|a) = \tilde{W}(0, a|a) \geq \max_x \tilde{W}(x, \hat{a}|a) \geq \tilde{W}(0, \hat{a}|a) = W(\hat{a}|a) \quad \forall \hat{a} \in A.$$

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

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

It follows from the Proposition that one can take the efficient allocation and find a tax system $(\boldsymbol{T}^*, M^*)$ that implements the efficient allocation. It is easy to show that the marginal income taxes evaluated at the optimal allocation are equal to their respective intratemporal wedges τ_1 and τ_2, and that the marginal tax on savings is equal to the inverse of the intertemporal wedge $\frac{1}{1-\delta}$.

5 Quantitative Analysis

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

5.1 Calibration

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

Table 1: Parameters Set Exogenously

Definition	Symbol	Value	Source/Target
Number of periods	J	2	20 years per period
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	1.19	0.04 annual
HC technology	α	0.7	Browning, Hansen, and Heckman (1999)
Capital income tax rate	$\bar{\tau}_k$	0.37	McDaniel (2007)
Labor income tax rate	$\bar{\tau}_\ell$	0.26	McDaniel (2007)
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	0.580	y_1/y_2	0.868	0.868	HVY (2011)
St. dev. log-ability	σ_a	0.620	Earnings Gini	0.346	0.345	HVY (2011)

We set $J = 2$ periods. A model period is 20 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,

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

as in Huggett, Ventura, and Yaron (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,

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

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 $\epsilon = 0.5$, equal to the Frisch elasticity. Agents' discount factor is set to $\beta = (0.96)^{20} = 0.442$.

Technology

The human capital production function

$$F(h, e) = h + (eh)^{\alpha}$$

is of the Ben-Porath form. The value of the parameter $\alpha = 0.7$ is the same used in Huggett, Ventura, and Yaron (2011) and 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. Hence, we can approximate the parameters from the shock process by assuming older workers in the data exert zero learning effort.

The parameters of the shock distribution are calibrated as follows. Wages are calculated from the Panel Study of Income Dynamics (PSID) for males between 55 and 65 years of age.[6] Huggett, Ventura, and Yaron (2011) estimate the parameters of the shock process from a log-wage difference regression. It is not immediately clear that we can use the same log-wage difference regression since labor income is also a product of ability in our model. However, since abilities are permanent, log-wage differences cancel out abilities from the regression. In particular, consider a multiperiod version of our model where each period is a year. Also assume that there is zero investment in human capital from date t to date $t+n$. Difference of log-wages yields,

$$\begin{aligned}
w_{t+n} &= ah_{t+n} = a\exp(z_{t+n})F(h_{t+n-1},0) = a\prod_{i=1}^{n}\exp(z_{t+i})h_t \\
\ln w_{t+n} &= \ln a + \sum_{t+i}^{n} z_{t+i} + \ln h_t \\
\Delta \ln w_{t+n} &= \ln w_{t+n} - \ln w_{t+n-1} + \zeta_{t+n} - \zeta_t = \sum_{t+i}^{n} z_{t+i} + \zeta_{t+n} - \zeta_t
\end{aligned}$$

where $\zeta_{t+n} - \zeta_t$ are measurement error differences. The last equation is the regression equation. It shows that differences in log-wages are solely attributed to shocks and measurement error. Also notice that the the model implies that wages are not observable. However, differences in log-wages are observable since abilities do not enter the equation.

Let superscript a denote values at an annual frequency, as opposed to a 20-year frequency. Huggett, Ventura, and Yaron (2011) estimate $\sigma_z^a = 0.111$ and $\mu_z^a = -0.029$. 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

[6] Huggett, Ventura, and Yaron (2011) calculate real wages as total male labor earnings divided by total hours for male head of household, using the Consumer Price Index to convert nominal wages to real wages.

by about 49.6% and human capital depreciates on average 36.69%.[7]

Benchmark Tax System

We approximate the U.S. tax system with a flat tax on capital income and a flat tax on labor income,

$$T(k_j, y_j, h_j) = (1 - \bar{\tau}_k) r k_j + (1 - \bar{\tau}_l) y_j \quad \text{for } j = 1, 2.$$

We calculate the values of the tax rates and interest rate in three steps. First, we obtain the mean average tax rates from McDaniel (2007) for the 1969-2004 period. Second, we normalize tax rates by the average consumption tax. This is enough to obtain the value of the labor income tax rate of $\bar{\tau}_l = 0.26$ and the annual capital income tax rate τ_k^a. Third, the effective 20-year capital income tax rate and interest rate are the solution to the following two equations:

$$(1 + r^a(1 - \tau_k^a))^{20} = 1 + r(1 - \bar{\tau}_k) \tag{22}$$

$$(1 + r^a)^{20} = 1 + r \tag{23}$$

The annual interest rate $r^a = 0.04$ is set to the historical risk-free rate of return in the U.S. The effective 20-year tax rate on capital income is $\bar{\tau}_k = 0.37$. The effective 20-year interest rate is $r = 1.19$. Tax revenues are transferred lump-sum back to the agents.

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.

[7] Since shocks enter multiplicative, the 20-year shock process is distributed $\exp(z) = LN(20\mu_z^a, 20(\sigma_z^a)^2)$, or $LN(-0.58, 0.2464)$. As an intuitive check, the average depreciation rate of human capital in Huggett, Ventura, and Yaron (2011) is $1 - \exp(\mu_z^a + \frac{(\sigma_z^a)^2}{2}) = 2.26\%$ per year. This implies a 20-year depreciation rate of $1 - (1 - .0226)^{20} = 36.69\%$. The shock process has to be such that $63.3\% = \exp(\mu_z + \frac{\sigma_z^2}{2})$, which is consistent with our numbers.

We target two moments: The ratio of mean earnings of younger workers (ages 23 to 40) to mean earnings of older workers (ages 40 to 60) and the earnings Gini coefficient for all age groups. Table 2 displays the results. Parameters values $\mu_a = -0.125$, $\sigma_a^2 = 0.68$, and $h_1 = 0.59$ best approximate the model to the data targets.

5.2 Findings

We focus on four main findings. First, it is efficient to provide consumption insurance for low ability agents and increase consumption risk for high ability agents relative to the benchmark. Second, the increase in the intratemporal wedge is the highest for high ability agents. Third, the intertemporal wedge increases with ability. Fourth, implementing an optimal tax system brings large welfare gains for the overall economy. However, welfare gains are distributed unevenly across the economy. High ability agents lose a significant amount of welfare.

Incentives for High Ability, Insurance for Low Ability

Figure 1 shows the variance of log-consumption in the second period for the benchmark economy and the constrained efficient economy.

In the benchmark, variance of second period log-consumption is slightly increasing in ability. As seen in Figure 2a, higher ability agents exert more learning effort. Consequently high ability agents face a higher expected human capital as compared to low ability agents, as seen in Figure 2b.

In the constrained efficient economy, the social planner finds it optimal to increase the variance of log-consumption with ability. In contrast to the benchmark economy, low-ability agents are more insured across realizations of human capital shocks, whereas high-ability agents face a significantly higher amount of uncertainty. The effort profile shown in Figure 2a provides insight as to what the social planner accomplishes through this variance profile. Increasing the uncertainty of second period consumption for high agents (as compared to the benchmark) makes these agents exert higher learning effort as a form of self-insurance. Figure 2b shows, in turn, that high-ability agents face a more favorable distribution of human capital shocks compared to other agents and also as compared to the benchmark. At the bottom end of the distribution, the social planner finds it optimal to provide consumption

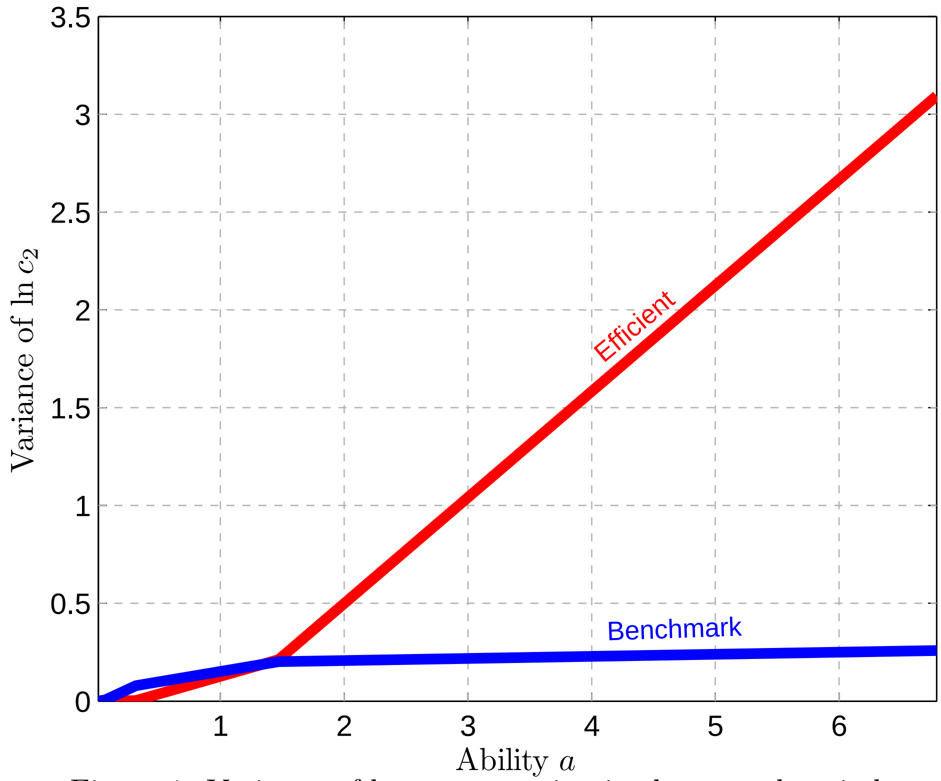


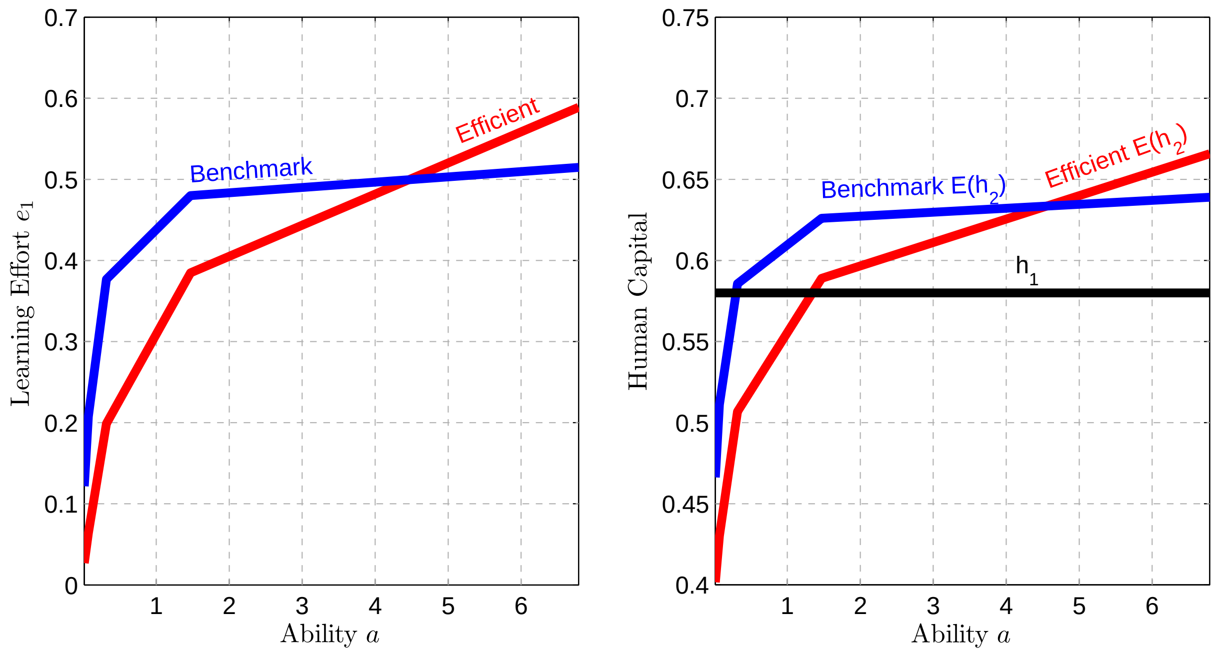
Figure 1: Variance of log-consumption in the second period.

insurance. Low ability agents, in turn, exert little effort as they see little necessity to accumulate human capital. Consequently, high ability agents increase their expected human capital in the constrained efficient scenario by more than what they were increasing in the benchmark.

Wedges

Figure 3a shows the intratemporal wedge in the second period as a function of human capital. The wedges are shown for three selected ability levels, low, medium and high. In all cases the intratemporal wedge is very high for low human capital realizations and then decreases with human capital. Recall from Propositions 3 and 4 that the intratemporal wedge decreases with human capital so that consumption increases with human capital realizations. The decrease is most rapid for higher ability levels.

The intratemporal wedge in the first period and the expected intratemporal wedge in the second period are shown in Figure 3b. The figure shows that the expected intratemporal



(a) Effort

(b) Expected human capital in the second period and human capital first period.

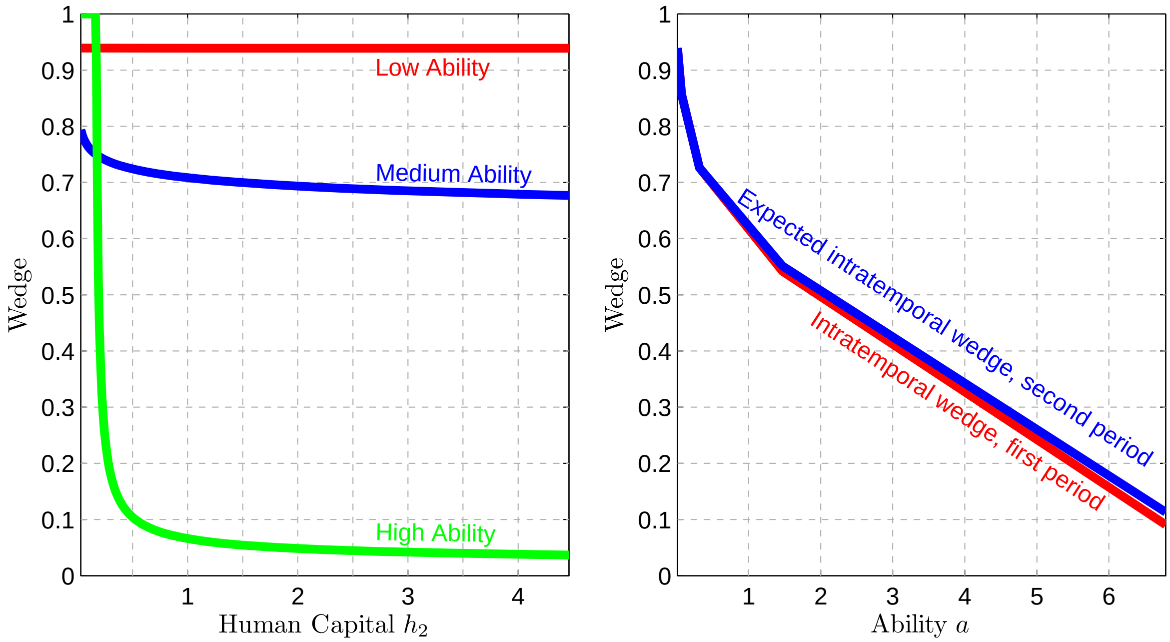
Figure 2: Effort and Human Capital

wedge is higher than the intratemporal wedge in the first period. This confirms the second theoretical finding in Proposition 4. The difference is most pronounced for higher ability agents, where the agents expect second period intratemporal wedge to be about 2% higher than in the first period. Overall, the intratemporal wedge in period 1 decreases with abilities and converges to zero due to the fact that abilities are lognormally distributed.

The intertemporal wedge is shown in Figure 4. The figure shows that the intertemporal wedge is strictly increasing in ability. This result stands in contrast to the previous literature (e.g. Albanesi and Sleet (2006)). The planner discourages savings of high ability agents to increase the incentives of high ability agents to self-insure through human capital accumulation.

Consumption Profiles and Redistribution

As stated in Proposition 3, the social planner provides higher consumption in good states of the world and low consumption in bad states. However, Figure 5 shows that, in the



(a) (b)

Figure 3: Intratemporal Wedge

constrained efficient economy low ability agents receive higher consumption in *all* states of the world relative to the benchmark. As abilities increase, the constrained efficient consumption allocations become flatter relative to the benchmark. The highest ability level faces the largest decrease in consumption relative to the benchmark, with close to zero consumption for low realizations of human capital. Another aspect that stands out is the concavity of the consumption profile for high ability agents. This concavity indicates that it is efficient to punish low realizations more than reward good realizations.

Figures 6 and 7 show that the distribution of consumption and earnings move in opposite directions. Relative to the benchmark, the labor earnings distribution becomes more unequal while the consumption distribution becomes more equal within each period. In the constrained efficient economy, it is optimal to concentrate labor earnings - equivalent to output in the model - at the top of the distribution and enact a tax system that redistributes earnings enough to reduce consumption inequality compared to the benchmark. However, overall labor earnings, and consequently output, decreases in the economy. Redistribution comes at the expense of lower output.

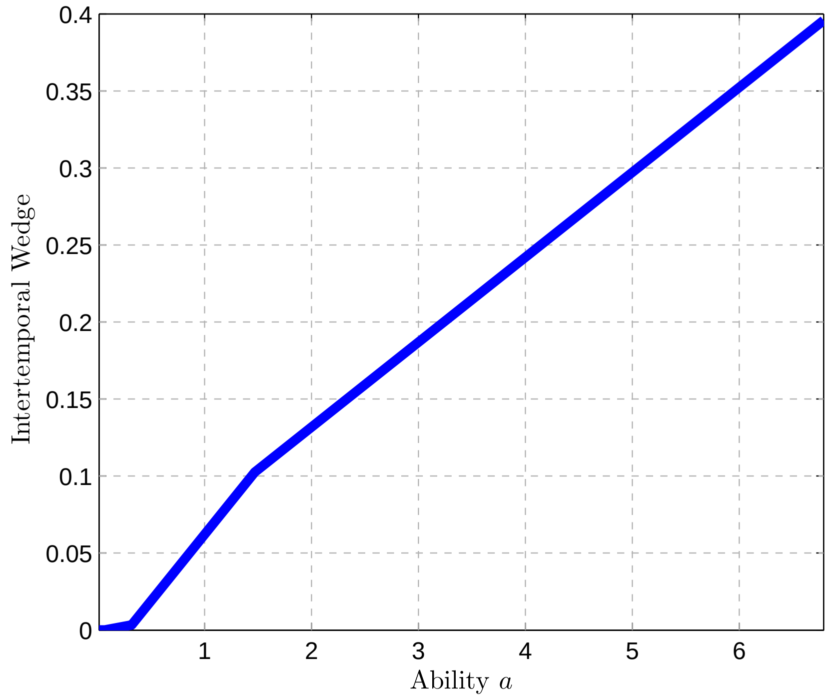
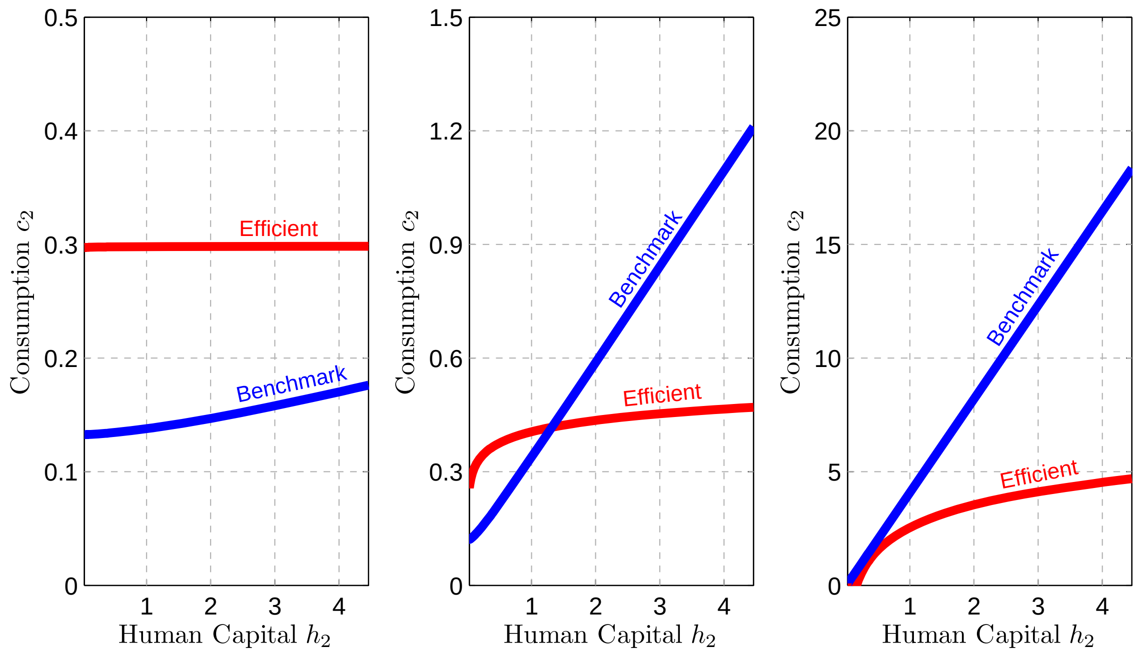


Figure 4: Intertemporal Wedge

These results are indicative that a significant amount of welfare gains are due to redistribution. Consumption is reallocated from low marginal utility agents (high-ability) to high marginal utility agents (low-ability). Effort (and labor) is reallocated from low marginal product agents (low-ability) to high marginal product agents (high-ability).

In order to get a sense of the amount of redistribution that happens in the constrained efficient economy, we plot consumption as a percentage of labor earnings for each ability type. The results, plotted in Figure 8, reveal that there is a substantive amount of redistribution in the efficient scenario. The lowest ability agent in the economy, the zero percentile agent, consumes between 45,000%-76,000% more than he earns. In contrast, the agent in the 97 percentile of the ability distribution consumes only 43 to 34 percent of what he earns, indicating overall earnings taxes of 57 to 64 percent.



(a) Low Ability (b) Medium Ability (c) High Ability

Figure 5: Second Period Consumption Profiles

Welfare

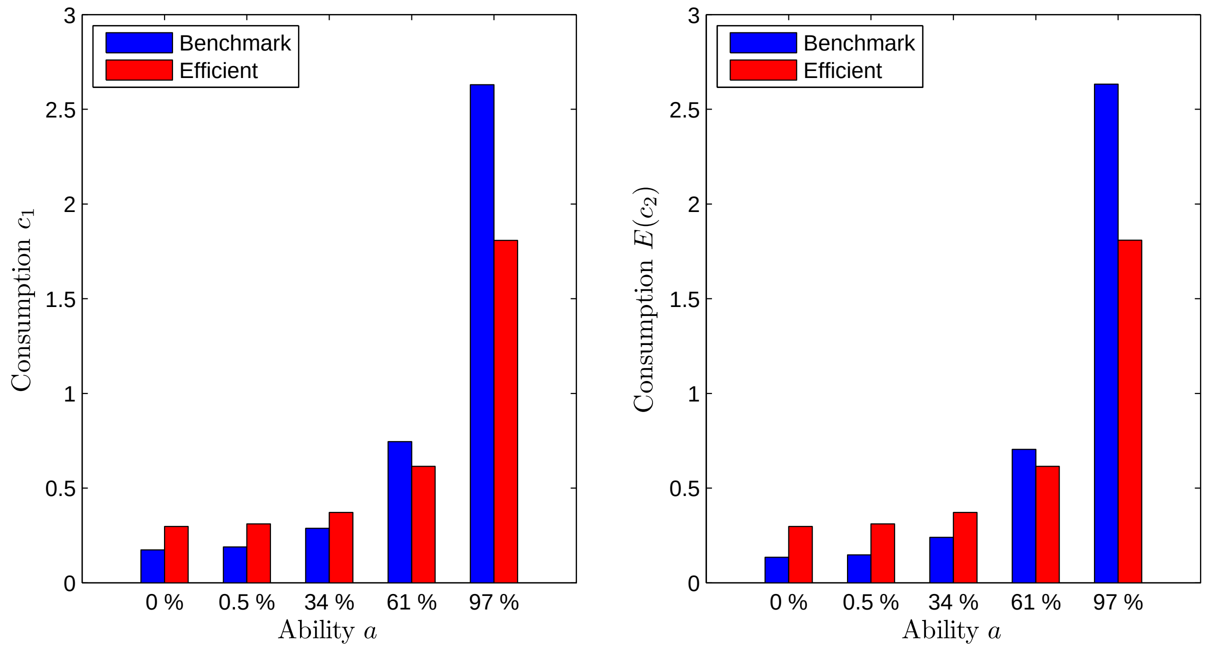
In this section we explore the welfare gains for the benchmark economy from switching to the efficient tax system. We are interested in the overall welfare gain and in the welfare change for each ability type.

The overall welfare gain is defined as the percentage increase in period consumption that would make the representative agent indifferent between the benchmark allocation and the constrained efficient allocation, keeping labor and effort unchanged. Specifically, the welfare gain is the η that solves,

$$W((1+\eta)c_1^B, (1+\eta)c_2^B, l_1^B, l_2^B, e_1^B) = W^{CE}.$$

where the B superscript denote benchmark allocations. We find that the welfare gains of switching to an optimal tax system are equivalent to a $\eta = 11\%$ increase in consumption every period and state of the world.

The change of welfare across types is illustrated in Figure 9. The large welfare gains accrue



(a) First period.

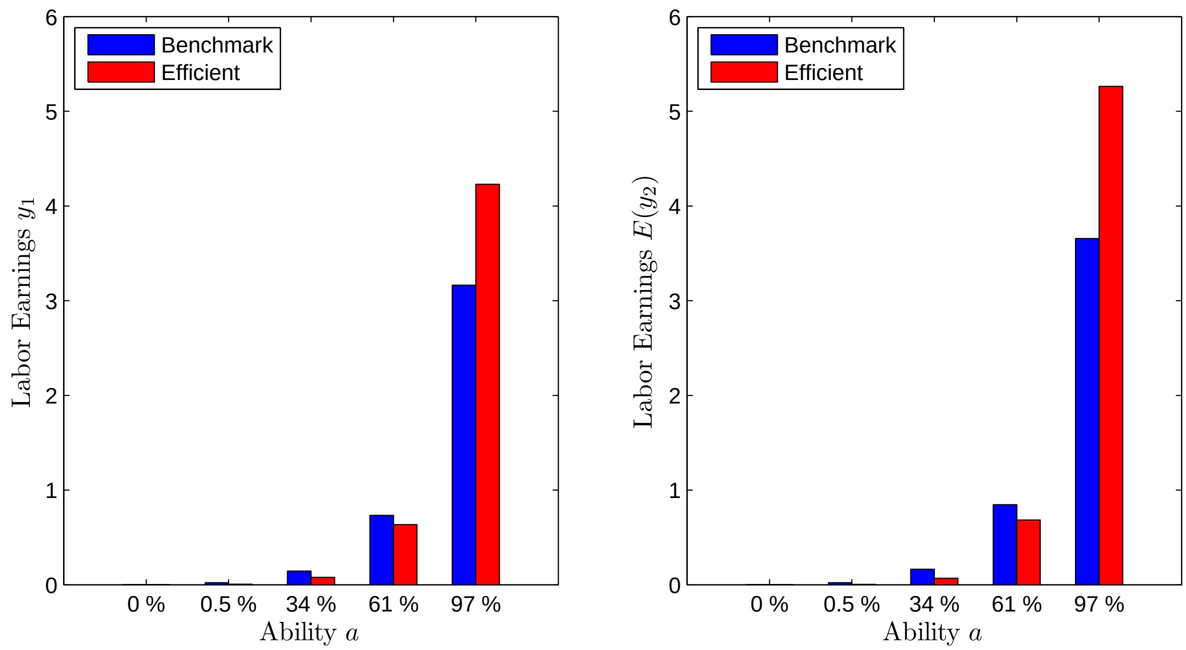
(b) Second period.

Figure 6: Distribution of Consumption, by ability percentiles.

at the bottom of the ability distribution. In contrast, the top abilities lose a substantial amount of welfare compared to the benchmark economy. However, it is worth noting that welfare is still increasing with ability in the constrained efficient economy. Monotonicity of welfare on ability is a direct consequence of restricting allocation to those that are locally incentive compatible, so that the envelope condition (9) holds. The implication is that, by construction, welfare is monotonically increasing in ability.

6 Conclusion

This paper addresses two questions: What are the features of an optimal tax system? What are the welfare gains for the U.S. from switching to an optimal tax system? We answer these questions in a Mirrleesian life-cycle economy in which agents are heterogeneous in their ability to produce output, can invest in human capital to augment their productivity, and the rates of return to human capital evolve stochastically. The model is sufficiently rich to be useful for policy analysis, and we show that it is also tractable enough for the normative



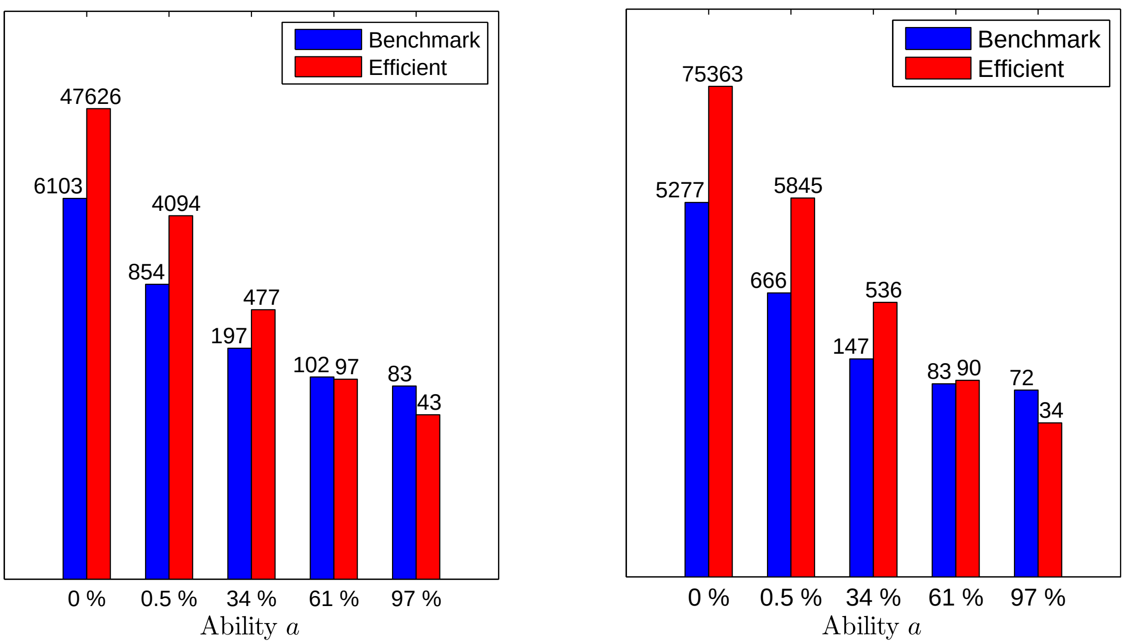
(a) First period. (b) Second period.

Figure 7: Distribution of Earnings, by ability percentiles.

analysis.

We highlight five main findings. First, the "no distortion at the top" result from Mirrlees (1971) might not apply when taxing labor encourages human capital accumulation. Second, it is efficient to design the tax system so that consumption risk is increasing in ability. This finding has the interpretation of focusing incentives of exerting effort in high ability agents while providing insurance for low ability agents. Third, under certain conditions, the average intratemporal wedge is increasing over an agent's lifetime. Fourth, the intertemporal wedge is increasing in ability. Finally, there are large welfare gains from switching to an optimal tax system in the benchmark.

There are reasons to believe that welfare gains might not be as large in practice. First, separability of labor and effort in the utility function implies that the social planner is able to incentivize high-ability agents to supply high labor and exert high effort. In reality, it is likely that there are complementarities in labor and effort. Second, approximating the U.S. tax function with flat taxes results in a highly unequal distribution of consumption in the benchmark. Replacing flat taxes with a more realistic progressive taxation benchmark



(a) First period.

(b) Second period.

Figure 8: Consumption as a % of labor earnings, by ability percentiles

will likely drive down the welfare gains. Third, a two-period model does not leave room for self-insurance through precautionary savings. Consumption in the second period is simply a residual. Extending the model to more than two periods might allow agents to better self-insure through savings. Fourth, ability is perfectly correlated with labor earnings in the two-period model. This implies that the agents with high marginal utility of consumption have low marginal product of labor. Adding dispersion in initial human capital or extending the model to more than two-periods will affect this correlation and might reduce the welfare gains of redistribution.

We are currently working on extending the model to more than two periods in order to obtain more precise policy prescriptions. So far, we have been able to show that at least two of the main results from the two-period model extend to a finite number of periods: There is a positive distortion at the top with labor and learning effort complementarities, and the intratemporal wedge is increasing over the lifetime.

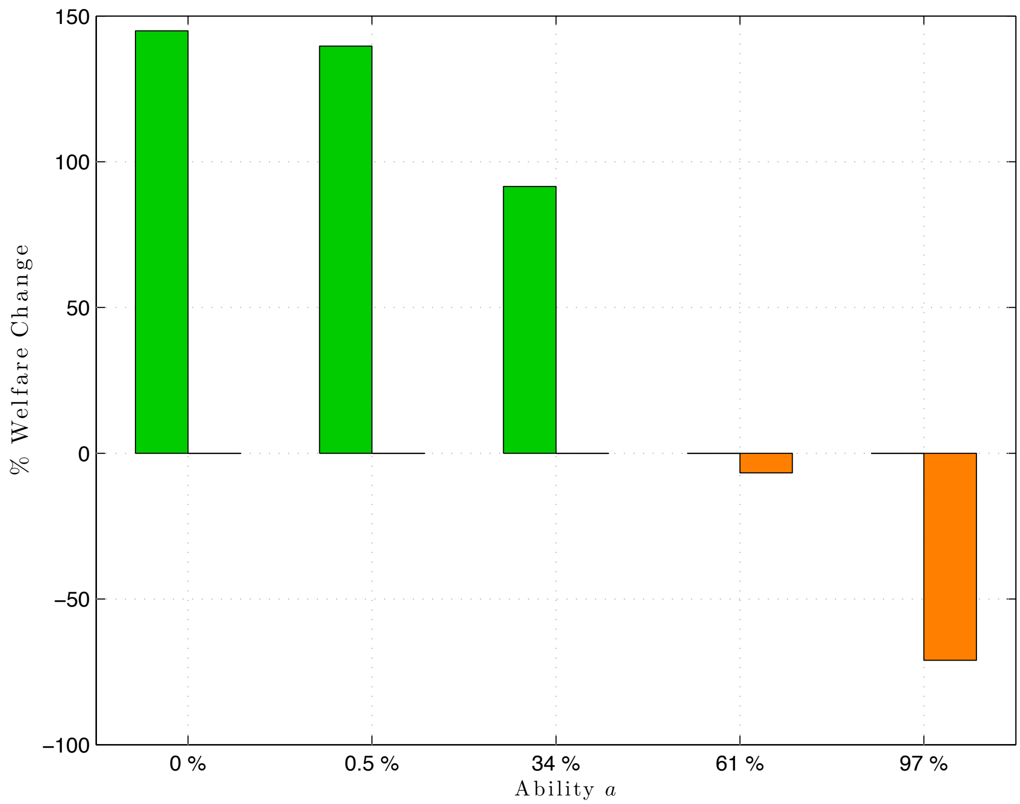


Figure 9: Welfare change from the benchmark to the constrained efficient economy, by ability percentiles.

References

Abraham, A., S. Koehne, and N. Pavoni (2012, May). Optimal income taxation with asset accumulation. MPRA Paper 38629, University Library of Munich, Germany. 5

Aguiar, M. and E. Hurst (2012). Deconstructing lifecycle expenditure. Working Paper 13893, NBER. 5

Albanesi, S. (2007). Optimal taxation of entrepreneurial capital with private information. Working paper, Columbia University. 5

Albanesi, S. and C. Sleet (2006). Dynamic optimal taxation with private information. *The Review of Economic Studies 73*(1), 1–30. 3, 4, 27

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. 7

Boháček, R. and M. Kapička (2008). Optimal human capital policies. *Journal of Monetary Economics 55*, 1–16. 3, 5, 39

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. 21, 22

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011, May). 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. 21, 22

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

Farhi, E. and I. Werning (2005). Inequality and social discounting. *Journal of Political Economy 115*(1), 365–402. 4

Farhi, E. and I. Werning (2010). Insurance and taxation over the life-cycle. Working paper, MIT. 4

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

Golosov, M., A. Tsyvinski, and M. Troshkin (2010). Optimal dynamic taxes. Working paper, Yale University. 4

Grochulski, B. and T. Piskorski (2010). Risky human capital and deferred capital income taxation. *Journal of Economic Theory 145*(3), 908 – 943. 4, 16

Guvenen, F. (2007). Learning your earning: Are labor income shocks really very persistent? *American Economic Review 97*(3), 687–712. 5

Hammond, P. J. (1979). Straightforward individual incentive compatibility in large economies. *The Review of Economic Studies 46*, 263–282. 19

Huggett, M., G. Ventura, and A. Yaron (2006). Human capital and earnings distribution dynamics. *Journal of Monetary Economics 53*, 265–290. 24

Huggett, M., G. Ventura, and A. Yaron (2011). Sources of lifetime inequality. *American Economic Review 101*(7), 2923–54. 2, 4, 5, 7, 21, 22, 23, 24

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

Jewitt, I. (1988). Justifying the first-order approach to principal-agent problems. *Econometrica 56*(5), pp. 1177–1190. 11

Kambourov, G. and I. Manovskii (2009, 02). Occupational specificity of human capital. *International Economic Review 50*(1), 63–115. 4

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

Kapička, M. (2008). The dynamics of optimal taxation when human capital is endogenous. Working paper, UC Santa Barbara. 3, 5, 16, 39

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

Lillard, L. A. and Y. Weiss (1979). Components of variation in panel earnings data: American scientists 1960-70. *Econometrica 47*, 437–453. 5

MaCurdy, T. (1982). The use of time-series processes to model the er- ror structure of earnings in a longitudinal data analysis. *Journal of Econometrics, 18*, 83–114. 5

Marcet, A. and R. Marimon (2009). Recursive contracts. *Unpublished manuscript*. 4, 39

McDaniel, C. (2007). Average tax rates on consumption, investment, labor and capital in the oecd 1950-2003. Technical report, Arizona State University, mimeo. 21, 24

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies 38*, 175–208. 4, 15, 32

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

Mirrlees, J. A. (1986). The theory of optimal taxation. In K. J. Arrow and M. D. Intriligator (Eds.), *Handbook of Mathematical Economics, vol. III, Chapter 24*. Elsevier. 4, 12

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

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. 4

Poletaev, M. and C. Robinson (2008, 07). Human capital specificity: Evidence from the dictionary of occupational titles and displaced worker surveys, 1984-2000. *Journal of Labor Economics 26*(3), 387–420. 4

Primiceri, G. E. and T. van Rens (2009). Heterogeneous life-cycle profiles, income risk and consumption inequality. *Journal of Monetary Economics 56*, 20–39. 5

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *The Review of Economic Studies 68*, 205–229. 17

Shourideh, A. (2010). Optimal taxation of entrepreneurial income: A mirrleesian approach to capital accumulation. Working paper, University of Minnesota. 5

Slavík, C. and H. Yazici (2012). Machines, buildings, and optimal dynamic taxes. Working paper, Goethe University in Frankfurt. 16

Stiglitz, J. E. (1982). Self-selection and pareto efficient taxation. *Journal of Public Economics 1*(2), 213 – 240. 16

Storesletten, K., C. Telmer, and A. Yaron (2004). Consumption and risk sharing over the life cycle. *Journal of Monetary Economics 51*, 609–633. 5

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. 17

Appendix A A Lagrangean Solution Method

Recall that λ, $\phi(a)q(a)$ and $\theta(a)q(a)$ are the Lagrange multipliers on the resource constraint (6), the first order condition (8) and on the envelope condition (9). After combining lifetime utility $W(a)$ from the objective function and the envelope condition, the Lagrangean is

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

The first-order condition in $W(\underline{a})$ implies

$$\int_A \theta(a)q(a)\, da = 0. \tag{A-1}$$

Integrating the Lagrangean by parts, using (A-1), and rearranging terms, one obtains

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

where $\Theta(a)$ is the cross-sectional cumulative of the Lagrange multipliers on the envelope condition:

$$\Theta(a) = \frac{1}{aq(a)} \int_a^{\bar{a}} \theta(\tilde{a})q(\tilde{a})\, d\tilde{a}. \tag{A-2}$$

Finally, take the expectation of the first order condition in effort and divide it by $F_e\left(h_1, e_1(a)\right)$. Rearranging the terms involving ϕ, one obtains

$$\mathcal{L} = \int_A \Bigg\{ \Big(1+\theta(a)\Big) W(a) - \lambda \left[c_1(a) - y_1(a) + \beta \int_H \left[c_2(a,h_2) - y_2(a,h_2)\right] p(h_2|e_1(a))\, dh_2 \right]$$
$$- \Theta(a) \left[V_\ell \left(\frac{y_1(a)}{ah_1}, e_1(a) \right) \frac{y_1(a)}{ah_1} + \beta \int_H V_\ell \left(\frac{y_2(a,h_2)}{ah_2}, 0 \right) \frac{y_2(a,h_2)}{ah_2} p(h_2|e_1(a))\, dh_2 \right]$$
$$- \phi(a) \frac{V_e\left(\frac{y_1(a)}{ah_1}, e_1(a)\right)}{F_e\left(h_1, e_1(a)\right)} + \beta \int_H \Phi(a,h_2) \left[U(c_2(a,h_2)) - V\left(\frac{y_2(a,h_2)}{ah_2}, 0 \right) \right] p(h_2|e_1(a))\, dh_2 \Bigg\} q(a)\, da$$

where

$$\Phi(a,h_2) = \phi(a) \frac{p_f(h_2|e_1(a))}{p(h_2|e_1(a))} \tag{A-3}$$

The expression for Φ is simplified by the property

$$p_e(h_2)(h_2|e_1(a)) = p_f(h_2|e_1(a)) F_e\left(h_1, e_1(a)\right).$$

A Recursive Formulation

The above problem can be solved using the following three-step procedure adapted from Boháček and Kapička (2008) and Kapička (2008). First, fix a Lagrange multiplier λ. Second, fix a function $\theta(a, h_1)$ and compute the cumulative Lagrange multiplier $\Theta(a, h_1)$ using (A-2). Conditional on those values, the problem has a recursive representation that uses the recursive Lagrangean method of Marcet and Marimon (2009). Denote Φ to be the costate variable corresponding to the cumulative Lagrange multiplier on the first-order condition (8) in effort. Let $\Omega_j(h, \Phi; a, h_1)$ be the value of having human capital h and a costate variable Φ at the beginning of period j for an agent with ability a and initial human capital h_1. It can be shown that Ω_j is given by

$$\Omega_j(h,\Phi;a,h_1) = \max_{c,l,e} \min_{\phi} \Bigg\{ \left[1+\theta(a,h_1)+\Phi\right]\left[U(c) - V\left(\ell,e\right)\right] - \lambda\left(c - ahl\right) - \phi \frac{V_e\left(\ell,e\right)}{F_e\left(h,e\right)}$$
$$- V_\ell\left(\ell,e\right) l \Theta(a,h_1) + \beta \int_H \Omega_{j+1}\left[h', \Phi'(h'); a, h_1\right] p_{j+1}(h'|F(h,e))\, dh' \Bigg\} \tag{A-4}$$

where the law of motion for Φ is

$$\Phi'(h') = \Phi + \phi \frac{p_{f,j+1}\left(h'|F(h,e)\right)}{p_{j+1}\left(h'|F(h,e)\right)}.$$

The dynamic program is initiated at $(h, \Phi) = (h_1, 0)$, and is terminated with $\Omega_{J+1} = 0$.

Once the problem is computed, one updates the θ function by using a period 1 first order condition in consumption. This generates a new function $T\theta$:

$$T\theta(a, h_1) = \frac{\lambda}{U'\left(c_1(a, h_1)\right)} - 1$$

The iterations proceed until T converges. Finally, one iterates on the Lagrange multiplier λ until the resource constraint (6) clears.

The dynamic program (A-4) is at the heart of this computational procedure. It has four state variables, but it is worth noting that only two of them (h and Φ) are changing over time; the other two (a and h_1) are constant. Also, h_1 enters the problem only through θ and Θ, which will somewhat simplify the computational procedure. Furthermore, the optimization problem itself is relatively simple, as it features no constraints (apart from nonnegativity constraints on c, ℓ and e) and only four variables. It is therefore expected that, despite its complexities, it will be feasible to solve the problem numerically.