American Economic Journal: Macroeconomics 2019, 11(4): 271–309
https://doi.org/10.1257/mac.20160365

Optimal Taxation with Risky Human Capital†

By Marek Kapička and Julian Neira*

We study optimal tax policies in a life-cycle economy with permanent ability differences and risky human capital investments that have both an unobservable component, learning effort, and an observable component, schooling. The optimal policies balance redistribution across agents, insurance against human capital shocks, and incentives to learn and work. In the optimum, (i) high-ability agents face risky consumption while low-ability agents are insured; (ii) the optimal schooling subsidy is substantial but less than 100 percent; (iii) if utility is separable in labor and learning effort, the inverse labor wedge follows a random walk; and *(iv) if the utility is not separable then the "no distortion at the top" result does not apply. The welfare gains from switching to the optimal tax system are about 1 percent in annual consumption equivalents.* (*JEL* D15, H21, H24, I26, J24)

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. For example, Huggett, Ventura, and Yaron (2011) shows that such a model is able to account for key empirical features of the dynamics of earnings and consumption. We explore the implications of such a framework for optimal tax policy. We consider an economy where people are ex ante heterogeneous in their productive abilities and have two ways of investing in a risky human capital: learning effort, which takes time, and schooling, which costs money. At least since a pioneering study of Schultz (1961), the empirical literature has recognized that both types of costs constitute a significant

* Kapička: CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Politickych veznu 7, 111 21 Prague, Czech Republic (email: mkapicka@gmail.com); Neira: Department of Economics, University of Exeter Business School, Rennes Drive, Exeter, United Kingdom EX4 4PU (email: j.neira@exeter.ac.uk). A previous version of this paper was circulated under the title "Optimal Taxation in a Life-Cycle Economy with Endogenous Human Capital Formation." Richard Rogerson was coeditor for this article. We are grateful to two anonymous referees and 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, ENSAI-Rennes, CEPR, and other seminars and conferences. All remaining errors are our own responsibility. Kapička's research was supported by Purkyně Fellowship of the Czech Academy of Sciences, by GACR grant 13-29370S and by Marie-Curie's Integration grant No 631738. We acknowledge support from the Center for Scientific Computing from the CNSI, MRL: an NSF MRSEC (DMR-1121053) and NSF CNS-0960316.
† Go to https://doi.org/10.1257/mac.20160365 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

fraction of the total costs.[1] Yet, the previous dynamic mechanism design literature typically considers one cost at a time. The interaction between both types of investments, and its implications for the optimal tax policies, is a novel feature of this paper.

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 learning effort and human capital shocks are unobservable by the government. The interaction between both frictions is also a novel feature of our model. 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 the balance between learning and schooling; and provision of incentives to elicit high labor effort from agents with high human capital or ability.

There are several novel policy implications. First, the optimal labor taxes are contingent on human capital shocks in order to promote learning effort through a dispersion in second-period consumption. This is especially important for high-ability people, whose standard deviation of log-consumption is about 40 percent higher than in a calibrated economy. Learning effort is less important for low-ability people, so they are almost completely insured against human capital shocks. Overall, the bottom 90 percent of the population faces more consumption insurance under an optimal tax system than under the current tax system, while the top 10 percent of people face less consumption insurance. We show that the average tax rates for higher ability people are U-shaped in human capital shocks: they are high for low realizations of human capital to incentivize learning effort, but also high for high realizations of human capital to fund redistribution. The first rationale is missing for lower ability people, who face average labor taxes that are strictly increasing in human capital shocks. Marginal labor tax rates are, however, always decreasing in the human capital shocks to provide incentives for labor effort.

Schooling, in contrast to learning effort, is observable by the government in our model and can be directly subsidized. It has two advantages relative to learning effort: first, schooling decreases the agent's effective marginal cost of learning effort and alleviates the moral hazard problem. Second, it does not increase the required informational rent, while learning increases it whenever learning and labor effort are substitutes. For both of these reasons, schooling should be promoted relative to learning effort. The comparative advantage of schooling also implies that the fiscal externality—the expected increase in future tax revenue by the government—exceeds the marginal cost of higher schooling. This leads to a positive gross schooling wedge, which is in turn implemented as a positive gross schooling subsidy that decreases tax liabilities. We show, however, that due to the moral hazard considerations, the gross schooling subsidy is less than 100 percent, which contrasts, for example, with the results in Krueger and Ludwig (2016), who find a gross schooling subsidy of 170 percent. The calibrated version of the model shows that the gross schooling

[1] A recent Education at a Glance publication (OECD 2018) shows that the time cost in terms of foregone earnings is about 58 percent of the total private cost for the United States. This number is based on their tables 1.2, 5.1a, and 5.1b.

subsidy is substantial, between 86 to 100 percent of the schooling cost, depending on agents' ability.

Subtracting the fiscal externality from the gross schooling wedge defines a net schooling wedge, which measures how schooling is distorted relative to the frictionless economy. We identify two forces that determine the net schooling wedge. First, higher schooling subsidies increase the expected informational rent of agents, which hurts the government's ability to redistribute resources and contributes negatively. Second, higher schooling subsidies make it harder or easier to incentivize learning effort, depending on the complementarity between both inputs in the human capital production function. Theoretically, the sign of the net schooling wedge is ambiguous. In our quantitative exercise, however, the net schooling wedge is significantly negative, meaning that schooling is distorted downward relative to the first best. It is between -25 and -7 percent of the schooling cost and is the largest at both end points of the ability distribution. It is also substantially lower than in Stantcheva (2017), where the net wedge can be positive, negative, or zero, depending on the assumptions about complementarity between ability and human capital.

The interaction of a private information friction and a moral hazard friction produces other interesting 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. The reason is as follows. The optimal labor wedge decreases with consumption because of income effects and so co-moves negatively with second-period consumption. More precisely, we show that the inverse of the labor wedge is proportional to the inverse of the marginal utility of consumption. But the inverse of the marginal utility of consumption follows a random walk because the Inverse Euler Equation holds, and the labor wedge inherits this pattern. If the utility is not separable in labor and learning effort, we show that the well-known "no distortion at the top" result from the Mirrleesean literature does not apply and provide a novel argument for higher marginal income tax rates at the top of the distribution. 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 face a positive marginal tax. We show in a simple numerical exercise that this effect is quantitatively important for the top 10 percent of the ability distribution, and those agents face significantly higher labor wedge than in a separable case.

To investigate quantitatively the optimal tax policies and efficient allocations, we calibrate a two-period model to match a number of moments of the US economy (a "status quo" model). One of the key aspects of the exercise is the calibration of the relative importance of learning effort and schooling in the production of human capital. We assume a constant returns to scale Cobb-Douglas aggregator and find that the share of schooling is 0.87, much larger than the share of learning effort of 0.13. Despite its small share in the production function, the presence of learning effort and the corresponding moral hazard friction produce a substantial amount of consumption risk in the second period to incentivize effort, especially for high-ability agents. If the share of learning effort is one and schooling is eliminated, our economy reduces to a two-period version of Huggett, Ventura, and Yaron (2011).

The standard deviation of log-consumption increases still further, but the additional increase is rather small for most of the population.

We find significant welfare gains from implementing the efficient tax system. The unborn agent is indifferent between the efficient tax system and the status quo economy with an annual 0.96 percent higher consumption in every period and state of the world. Shutting the moral hazard friction yields an additional welfare gain of 0.21 percent. We compare the welfare gains of our model with those of a model with no schooling, as in Huggett, Ventura, and Yaron (2011). The welfare gains from implementing the efficient tax system in such a model are substantially lower, at 0.10 percent, coming from the fact that the efficient allocations require a higher increase in consumption risk. However, the welfare gains from shutting down the moral hazard constraint in this model are similar to the model with schooling, at about 0.21 percent.

Relationship to the Existing Literature.—We build on a large literature that looks at models with Ben-Porath (1967) technology for human capital formation. 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 and Storesletten, Telmer, and Yaron 2004) and the HIP (heterogeneous income profile) models (see, e.g., Lillard and Weiss 1979 and Guvenen 2007).[2] 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, Troshkin, and Tsyvinski (2016), 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 to study an environment of optimal taxation with human capital where both private information and moral hazard frictions are present.[3] In

[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.

[3] Ábrahám, Koehne, and Pavoni (2016) and Albanesi (2007) study the impact of moral hazard on optimal tax structures. Unlike our paper, they do not consider the interaction between moral hazard and private information. Gary-Bobo and Trannoy (2015) studies moral hazard and private information problems of student loans in a setting with two abilities and two shocks. They also find that repayments should be contingent on outcomes.

this respect, our paper is close to Shourideh (2012), 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 partially observable.[4] In contrast, da Costa and Maestri (2007); Jacobs, Schindler, and Yang (2012); and Stantcheva (2015) all study optimal taxation with risky Ben-Porath technology, but allow human capital investments to be fully observable.[5] We think that assuming that a part of human capital investments is unobservable 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.

In terms of human capital technology, our approach is complementary to Bovenberg and Jacobs (2005), Findeisen and Sachs (2016, 2017), Stantcheva (2017), and Koeniger and Prat (2018), who all assume that human capital investments are in terms of observable goods. Grochulski and Piskorski (2010), however, studies human capital investments in terms of unobservable goods. We consider human capital investments in terms of observable goods and unobservable effort and can therefore study the role of unobservable investments on shaping schooling subsidies.

Finally, our paper is also related to the Ramsey taxation literature that quantitatively studies optimal tax reforms in environments with endogenous human capital, such as Gorry and Oberfield (2012), Krueger and Ludwig (2016), and Peterman (2016). This literature is able to consider richer frameworks than ours by restricting taxes to specific functional forms.

I. The Model

Consider the following two-period life-cycle economy. Agents like to consume, dislike working and learning, and have preferences given by

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

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 allow the Frisch elasticity of labor $\gamma(\ell,e) = V_\ell(\ell,e)/(\ell V_{\ell\ell}(\ell,e))$ to depend on the learning effort, which is what some standard functional forms for V deliver. We, however, restrict the function V by assuming that, conditionally on the

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

[5] Best and Kleven (2013) and Makris and Pavan (2018) also assume that human capital investments are unobservable, but they do so in a model that features a learning-by-doing technology.

learning effort being zero, $\gamma(\ell, 0)$ is independent of ℓ. This delivers constant Frisch elasticity in the second period. We also assume that the cross-derivative $V_{e\ell}$ is nonnegative so that both uses of time are substitutes.

Agents can invest in human capital in two ways. One way is by exerting learning effort e, which is unobservable to the planner and incurs a utility cost. The second way to invest in human capital is by formal schooling investments s, which are observable to the planner and incur a resource cost $M(s)$, strictly increasing and differentiable in s. The 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 the agent's lifetime and is known to the agents at the beginning of the first period. The ability has continuous support $A = (\underline{a}, \bar{a})$, with $\bar{a}$ 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 $H = (\underline{h}, \bar{h})$, with $\bar{h}$ possibly infinite, and depends on an idiosyncratic human capital depreciation shock z_2, initial human capital h_1, learning effort e_1, and formal schooling s_1:

$$h_2 = \exp(z_2) f(e_1, s_1), \tag{3}$$

where the dependence on h_1 is kept implicit to reduce notation. The function f is strictly increasing, strictly concave, and differentiable in both arguments. We also assume that the cross-derivative f_{es} is nonnegative, so that both inputs complement each other. 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 schooling and denote it $p(h_2|f)$. The derivative of the density with respect to f, $p_f(h|f)$, exists, and we assume that the conditional 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.

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

Efficient Allocations.—The *information structure* is as follows: ability a, labor effort ℓ_1 and ℓ_2, learning effort e_1, and human capital shock z_2 are private information of the agent. Consumption c_1 and c_2, earnings y_1 and y_2, schooling s_1, 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 *allocation* σ consists of consumption allocation $\{c_1(a), c_2(a, h_2)\}$, earnings allocation $\{y_1(a), y_2(a, h_2)\}$, and the initial period schooling cost allocation $s_1(a)$. First-period allocation is conditional only on the ability report, while a second-period allocation is also conditional on the realization of human capital in the second period. The lifetime utility of an a-type agent who reports ability $\hat{a}$ and exerts effort e is $W(\hat{a}, e|a)$:

$$W(\hat{a}, e|a) \equiv U(c_1(\hat{a})) - V\left(\frac{y_1(\hat{a})}{a h_1}, e\right) + \beta \int_H \left[U(c_2(\hat{a}, h_2)) - V\left(\frac{y_2(\hat{a}, h_2)}{a h_2}, 0\right)\right] p(h_2 | f(e, s_1(\hat{a})))\, dh_2.$$

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 $\hat{a}$:

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

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:

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

The incentive-compatibility constraint reflects the assumption that, unlike schooling, effort is unobservable, and the deviating agent chooses whatever effort maximizes her utility. In order to reduce notational complexity, we define the utility-maximizing effort plan conditional on truth-telling by $e_1(a) \equiv \tilde{e}_1(a|a)$, and let $W(a) = W(a, e(a)|a)$ be the corresponding truth-teller's lifetime utility. We also write $f_1(a) = f(e_1(a), s_1(a))$ to denote the truth-teller's mean human capital at the beginning of period two.

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

$$\int_A \left[c_1(a) + M(s_1(a)) - y_1(a) + R^{-1} \int_H [c_2(a, h_2) - y_2(a, h_2)] p(h_2 | f_1(a))\, dh_2\right] q(a)\, da \leq 0, \tag{6}$$

where $q(a)$ is the probability distribution of abilities and R is the gross interest rate. The interest rate is taken as exogenous, and the government can both borrow and save at that rate.[6] 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:

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

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).

[6] A standard general equilibrium interpretation would be that there is an aggregate production function that is linear in capital.

First-Order Approach.—The first-order approach replaces the constraint (4) with a corresponding Euler equation in effort and (5) with an envelope condition. The Euler equation in effort can be written as

$$(8) \qquad \frac{V_e\big(\ell_1(a), e_1(a)\big)}{f_e\big(e_1(a), s_1(a)\big)} = \beta \int_H u_2(a, h_2)\, p_f\big(h_2 | f_1(a)\big)\, dh_2,$$

where $\ell_1 = y_1/ah_1$, $\ell_2 = y_2/ah_2$, and $u_2(a,h_2) = U(c_2(a,h_2)) - V(\ell_2(a,h_2),0)$. It asserts that, at the optimum, the marginal costs of learning effort must be equal to the expected marginal benefit of a learning effort, where both marginal costs and marginal benefits are expressed in terms of a unit increase in the mean human capital f. The expected marginal benefit comes from a more favorable distribution of future human capital shocks.

Let $W(\underline{a})$ denote the lifetime utility of the least able agent. The envelope condition states that the lifetime utility of an agent a must be the lifetime utility of the least able agent plus the informational rent from having ability level a:

$$(9) \quad W(a) = W(\underline{a}) + \int_{\underline{a}}^{a} \Big\{ V_\ell\big(\ell_1(\tilde{a}), e_1(\tilde{a})\big)\, \ell_1(\tilde{a}) + \beta \int_H V_\ell\big(\ell_2(\tilde{a}, h_2), 0\big)\, \ell_2(\tilde{a}, h_2) p\big(h_2 | f_1(\tilde{a})\big)\, dh_2 \Big\} \frac{d\tilde{a}}{\tilde{a}}.$$

Replacing the incentive constraint with the Euler equation 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).

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. We now show conditions ensuring that (8) and (9) 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 non-decreasing 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. Let $P(h|f) = \int_{\underline{h}}^{h} p(\tilde{h}|f)\, d\tilde{h}$. The Euler equation in effort (8) is sufficient if these conditions hold.

PROPOSITION 1. *Suppose that* $e^*(\hat{a}|a)$ *satisfies* (8), *and that* (*i*) $\int_{\underline{h}}^{h} P(\varepsilon|f)\, d\varepsilon$ *is non-increasing and convex in* f *for each* h, (*ii*) $\int_H hp(h|f)\, dh$ *is non-decreasing concave in* f, *and* (*iii*) $u_2(\hat{a}, h)$ *is non-decreasing and concave in* h. *Then* (4) *holds.*

The proof is omitted because it follows directly from Theorem 1 of Jewitt (1988). 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 earnings, schooling, and learning effort are all increasing in ability, one can recover the global incentive-compatibility constraint (5) from the envelope condition (9).

PROPOSITION 2: *Suppose that the allocation satisfies* (9). *If* (i) $e^*(\hat{a}|a)$, $s_1(\hat{a})$, $y_1(\hat{a})$, *and* $y_2(\hat{a},h_2)$ *are all non-decreasing in* $\hat{a}$ *for each* h_2 *and* (ii) $y_2(\hat{a},h_2)/h_2$ *is non-decreasing in* h_2 *for each* $\hat{a}$, *then the incentive-compatibility constraint* (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)

$$
\begin{aligned}
W(a) &- W(\hat{a}) \\
&= \int_{\hat{a}}^{a}\Bigg\{V_\ell\left(\frac{y_1(\tilde{a})}{\tilde{a}h_1},e_1^*(\tilde{a}|\tilde{a})\right)\frac{y_1(\tilde{a})}{\tilde{a}h_1} \\
&\qquad + \beta\int_H V_\ell\left(\frac{y_2(\tilde{a},h_2)}{\tilde{a}h_2},0\right)\frac{y_2(\tilde{a},h_2)}{\tilde{a}h_2}\,p\big(h_2|f\big(e_1^*(\tilde{a}|\tilde{a}),s_1(\tilde{a})\big)\big)\,dh_2\Bigg\}\frac{d\tilde{a}}{\tilde{a}} \\
&\geq \int_{\hat{a}}^{a}\Bigg\{V_\ell\left(\frac{y_1(\hat{\boldsymbol{a}})}{\tilde{a}h_1},e_1^*(\hat{\boldsymbol{a}}|\tilde{a})\right)\frac{y_1(\hat{\boldsymbol{a}})}{\tilde{a}h_1} \\
&\qquad + \beta\int_H V_\ell\left(\frac{y_2(\hat{\boldsymbol{a}},h_2)}{\tilde{a}h_2},0\right)\frac{y_2(\hat{\boldsymbol{a}},h_2)}{\tilde{a}h_2}\,p\big(h_2|f\big(e_1^*(\tilde{a}|\tilde{a}),s_1(\tilde{a})\big)\big)\,dh_2\Bigg\}\frac{d\tilde{a}}{\tilde{a}} \\
&\geq \int_{\hat{a}}^{a}\Bigg\{V_\ell\left(\frac{y_1(\hat{a})}{\tilde{a}h_1},e_1^*(\hat{a}|\tilde{a})\right)\frac{y_1(\hat{a})}{\tilde{a}h_1} \\
&\qquad + \beta\int_H V_\ell\left(\frac{y_2(\hat{a},h_2)}{\tilde{a}h_2},0\right)\frac{y_2(\hat{a},h_2)}{\tilde{a}h_2}\,p\big(h_2|f\big(e_1^*(\hat{\boldsymbol{a}}|\tilde{a}),s_1(\hat{\boldsymbol{a}})\big)\big)\,dh_2\Bigg\}\frac{d\tilde{a}}{\tilde{a}} \\
&= W\big(\hat{a},e_1^*(\hat{a}|a)|a\big) - W(\hat{a}).
\end{aligned}
$$

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 $y_2(\hat{a},h_2)/h_2$ 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, that s_1 increases in $\hat{a}$ and, again, that $e^*(\hat{a}|a)$ increases in $\hat{a}$. 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. ∎

The sufficiency conditions can be checked numerically by computing ex post the effort plan $e^*(\hat{a}|a)$ and verifying the monotonicity and concavity requirements.

These conditions are quite strict; however, they 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). Indeed, they are not satisfied in our quantitative exercise in Section IV, but the allocation is still incentive compatible, as we show in the online Appendix. In what follows, we will assume that the sufficiency conditions are satisfied, and the solution to the relaxed problem coincides with the efficient allocations.

II. Theoretical Implications

We will now characterize the properties of the efficient allocation and of the corresponding wedges. Let $\lambda \geq 0$, $\phi(a)\,q(a)$, and $\theta(a)\,q(a)$ be the Lagrange multipliers on the resource constraint (6), the Euler equation in effort (8), and the envelope condition (9).[7] The first-order conditions show that consumption in the second period depends on the realization of human capital shocks:

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

$$\frac{1}{U'\big(c_2(a,h_2)\big)} = \frac{1+\theta(a)+\phi(a)\dfrac{p_f\big(h_2|f_1(a)\big)}{p\big(h_2|f_1(a)\big)}}{\lambda}. \tag{10b}$$

Given the MLRP property, second-period consumption will be strictly increasing in the human capital shock if the Lagrange multiplier on the first-order condition in effort ϕ is strictly positive. The next proposition shows that this is indeed the case.

PROPOSITION 3: *If Assumption* 1 *holds*, *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 Euler equation 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

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

where $\Theta(a) \geq 0$ is the cumulative multiplier on the envelope condition (9). Equation (11) then implies that $\ell_2(a,h_2)$ is strictly increasing in h_2. By Assumption 1, 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 1 and equation (10b). ∎

[7] See Appendix A1 for the full Langrangean solution method.

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 marginal private 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. The moral hazard friction prevents that, and the social planner responds by making second-period consumption increasing in human capital.

A. *Effort versus Schooling*

Since they represent alternative ways of creating a more favorable human capital distribution in the second period, the choice of effort and schooling are closely related. In a first-best allocation, the private marginal cost of increasing f by schooling M'/f_s would be equated to the private marginal resource cost of increasing f by exerting effort, $V_e/(f_e U')$. In a second-best world, this equality no longer holds. Instead, the relationship between effort and schooling is determined by the following equation:[8]

$$\frac{M'(s_1)}{f_s} = \frac{V_e}{f_e}\frac{1}{U'(c_1)} + \frac{\Theta}{\lambda}\frac{V_{e\ell}\ell_1}{f_e} + \frac{\phi}{\lambda}\frac{1}{f_e^2}\left[V_{ee} + V_e\left(\frac{f_{es}}{f_s} - \frac{f_{ee}}{f_e}\right)\right]. \tag{12}$$

Equation (12) implies that $M'/f_s > V_e/(f_e U')$, and so schooling should be promoted relative to learning effort.[9] Higher effort is socially more costly than higher schooling in two aspects. First, unlike schooling, it increases the current informational rent if $V_{e\ell} > 0$. Second, higher effort increases the cost of satisfying the effort constraint (8) by increasing the marginal cost of increasing mean human capital f in terms of effort V_e/f_e. However, higher schooling decreases V_e/f_e if $f_{es} \geq 0$, which relaxes the effort constraint. For both reasons, it is optimal to increase private marginal cost of schooling above private marginal cost of effort.

The choice of the learning effort itself equates the marginal benefits from a higher expected return in the second period with its marginal costs. There are two sources of marginal costs, one due to the private information friction and one due to the moral hazard friction, as the first-order condition in effort shows[10]

$$\begin{aligned}\beta\int_H\left[y_2(h_2) - c_2(h_2)\right]p_f(h_2|f_1)\,dh_2 \\ = \frac{\Theta}{\lambda}\left[\frac{V_{e\ell}\ell_1}{f_e} + \beta\int_H V_\ell(\ell_2(h_2),0)\,\ell_2(h_2)\,p_f(h_2|f_1)\,dh_2\right] \\ + \frac{\phi}{\lambda}\left[\frac{V_{ee}f_e - V_e f_{ee}}{f_e^3} - \beta\int_H u_2(h_2)\,p_{ff}(h_2|f_1)\,dh_2\right],\end{aligned} \tag{13}$$

[8] We economize on notation by making the dependence on a implicit whenever possible.

[9] This follows from the fact that f_{es} and $V_{e\ell}$ are nonnegative, V_{ee} is strictly positive, and f_{ee} is strictly negative.

[10] Note that private marginal cost and private marginal benefits of higher effort do not show up in the optimality condition (13). They cancel out by virtue of the Euler equation in effort (8).

where $\Theta(a) = (aq(a))^{-1}\int_a^{\bar{a}} \theta(\tilde{a})\, q(\tilde{a})\, d\tilde{a} \geq 0$ is the cumulative Lagrange multiplier on the envelope condition. The left-hand side of equation (13) represents the marginal benefits of higher effort. The difference $y_2(h_2) - c_2(h_2)$ is just resources extracted from the agent in the second period if the agent gets a human capital shock h_2. The integral on the left-hand side then measures the expected increase in tax revenue (extracted resources) from an increase in mean human capital f, $dE[T_2]/df$. This value is called *fiscal externality* and is present even in the frictionless economy.[11]

The marginal social cost of higher effort is related to the two frictions in the economy. First, higher effort increases the informational rent that needs to be paid to the agent. If $V_{e\ell} > 0$, higher effort directly rises the required informational rent in the first period by increasing the disutility from working. Even if $V_{e\ell} = 0$, higher effort increases the expected informational rent tomorrow $E[V_\ell(\ell_2)\,\ell_2]$ by shifting the second-period distribution of the informational rent toward higher values. Private information friction thus tends to increase the cost of higher effort and reduces the optimal effort. Second, higher effort requires additional costly adjustments in the optimal allocation to ensure that the Euler equation (8) holds. Higher effort increases private marginal cost of effort, but decreases private marginal benefits of effort.[12] Since by the Euler equation (8) they must be equal, the social planner then must either increase schooling or dispersion in the second-period utility in order to restore equality between private marginal costs and private marginal benefits of higher effort. Since $\phi/\lambda > 0$ by Proposition 3, both of those actions create additional cost.

B. *Wedges*

The efficient allocations can be characterized by the implied wedges between marginal costs and marginal benefits of the agents' choices. We have three types of wedges in this economy: a schooling wedge, which is a wedge between the marginal costs and benefits of schooling; a savings wedge, a wedge between marginal costs and benefits of saving; and a labor wedge, a wedge between marginal costs and benefits of working. We will now examine each one in turn. We will relegate the question of how the efficient allocation can be implemented in a market economy to the next section. For the interpretation of these wedged one should keep in mind that, as long as the optimal taxes are differentiable, the schooling wedge will correspond to the marginal schooling subsidy, the savings wedge will correspond to the marginal tax on savings, and labor wedges will correspond to the marginal taxes on labor income.

Schooling Wedge.—Unlike effort, schooling is observable and can be directly subsidized or taxed. We define the *gross schooling wedge* $\tau_1^{s,gross}$ as the percentage

[11] Bovenberg and Jacobs (2005) refers to the fiscal externality as the "Siamese twins" result.

[12] Properties of V and f imply that the first part of the last term, a change in private marginal cost $(V_{ee} f_e - V_e f_{ee})/f_e^3$, is positive. At the same time, if assumptions (i) and (iii) of Proposition 1 hold, the change in the private marginal benefits $\int_H u_2(a, h_2)\, p_{ff}(h_2|f)\, dh_2$ is negative. For this, see Conlon (2009).

difference between the marginal cost of schooling M' and the expected private marginal benefit of schooling:

$$(14)\quad \tau_1^{s,gross}(a) \equiv 1 - \frac{1}{M'(s_1(a))}\frac{\beta f_s(e_1(a), s_1(a))}{U'(c_1(a))}\int_H u_2(a,h_2)\, p_f(h_2|f_1(a))\, dh_2.$$

Assumption 1 implies that, for any non-decreasing function u, $\int_H u p_f \geq 0$. Since $u_2(a,h_2)$ is required to be non-decreasing by Proposition 1, the second term in equation (14) is positive. Thus, $\tau_1^s \leq 1$ and schooling cost are less than fully subsidized. The result is entirely due to the presence of moral hazard. In its absence, the second-period utility is decreasing in h_2 because people with higher human capital work more, but are not compensated in consumption. As a result, $\tau_1^s > 1$ and schooling subsidies would be more than 100 percent. Equation (12) implies that the optimal gross schooling wedge is

$$(15)\qquad \tau_1^{s,gross} = \frac{\Theta}{\lambda}\frac{V_{e\ell}\ell_1}{f_e} + \frac{\phi}{\lambda}\frac{1}{f_e^2}\left[V_{ee} + V_e\left(\frac{f_{es}}{f_s} - \frac{f_{ee}}{f_e}\right)\right].$$

That is, the gross schooling wedge is always positive, reflecting the desire to promote schooling relative to learning effort. One way to interpret this finding is that, in the optimum, the fiscal externality is always greater then the marginal social cost of schooling. They cannot be equal because the fiscal externality equals the marginal social cost of effort, the right-hand side of (13), and the marginal social cost of effort exceeds the marginal social cost of schooling, as explained in the discussion that follows equation (12). This creates a positive wedge between private marginal cost and private marginal benefits of schooling. To summarize, we have the following result.

PROPOSITION 4: $0 < \tau_1^{s,gross}(a) < 1$.

The economic literature often distinguishes between a gross schooling wedge and a net schooling wedge (Bovenberg and Jacobs 2005; Stantcheva 2015, 2017). The net schooling wedge corrects for the fiscal externality and measures distortion of schooling relative to a first-best allocation. We define the net schooling wedge $\tau_1^{s,net}$ as

$$\tau_1^{s,net} = \tau_1^{s,gross} - \frac{\beta f_s}{M'(s_1)}\int_H \left[y_2(h_2) - c_2(h_2)\right] p_f(h_2|f_1)\, dh_2.$$

The optimal net schooling wedge can be written, using (13) and (12),

$$(16)\qquad \frac{M'(s_1)}{f_s}\tau_1^{s,net} = -\beta\frac{\Theta}{\lambda}\int_H V_\ell(\ell_2(h_2), 0)\,\ell_2(h_2)\, p_f(h_2|f_1)\, dh_2$$
$$+ \frac{\phi}{\lambda}\left[\beta\int_H u_2(h_2)\, p_{ff}(h_2|f_1)\, dh_2 + \frac{V_e}{f_e^2}\frac{f_{es}}{f_s}\right].$$

The first term on the right-hand side reflects the role of private information. Just like in the case of effort, private information increases the cost because it increases the expected informational rent. The first term is then negative and decreases the net schooling wedge below zero. The second term reflects how higher schooling affects the Euler equation constraint. While the effect on the marginal private benefits of effort $\int_H u_2(a,h_2)\, p_{ff}(h_2|f)\, dh_2$ is negative, as discussed previously, the effect of higher schooling on marginal private costs of effort is ambiguous. It may decrease it if the complementarity between schooling and effort f_{es} is strong enough. If the effect is strong enough, higher schooling will relax the Euler equation constraint, and the second term will be positive. The net schooling wedge will, however, be unambiguously negative if $f_{es} = 0$.

The economic considerations behind the optimal net schooling wedge are quite different from Stantcheva (2017) and Koeniger and Prat (2018). In their models, the output production function (2) is generalized to allow for an arbitrary elasticity of substitution between agent's ability and human capital. Schooling then directly affects the current informational rent of the agent. The subsidy is positive if the elasticity of substitution is smaller than one, in which case schooling reduces the informational rent. Our result does not depend on the degree of complementarity between human capital and ability, but rather on the interaction between schooling, unobservable effort, and the expectation of the future informational rent. In addition, while in Stantcheva (2017) and Koeniger and Prat (2018) the optimal net schooling wedge is zero at both end points of the ability distribution when $\Theta = 0$, it is still nonzero in our framework, due to the moral hazard friction.

Savings Wedge.—Taking the expectation of (10b) (and noting that $\int_H p_f = 0$) implies immediately that, conditional on the ability type, the Inverse Euler equation holds:

$$\frac{1}{U'\big(c_1(a)\big)} = \int_H \frac{1}{U'\big(c_2(a,h_2)\big)}\, p\big(h_2|f_1(a)\big)\, dh_2 \quad \forall a \in A. \tag{17}$$

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.

Define the *savings wedge* δ in the usual way as the gap between the current marginal utility of consumption and the expected future marginal utility of consumption:

$$U'\big(c_1(a)\big) = \big(1 - \delta(a)\big) \int_H U'\big(c_2(a,h_2)\big)\, p\Big(h_2|f\big(e_1(a), s_1(a)\big)\Big)\, dh_2. \tag{18}$$

By Jensen's inequality, (17) immediately implies the following.

PROPOSITION 5: *The savings wedge $\delta(a)$ is strictly positive for each ability level a.*

The positive savings wedge comes purely from the moral hazard friction on the model, and not from the private information friction, as in Golosov, Kocherlakota, and Tsyvinski (2003), Farhi and Werning (2013), and other dynamic Mirrleesean literature. In these models, the point of the savings tax is to elicit a higher future labor 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. In the absence of the moral hazard friction, the savings wedge is zero.

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:

$$\tau_1^\ell(a) = 1 - \frac{1}{ah_1}\frac{V_\ell\big(\ell_1(a), e_1(a)\big)}{U'\big(c_1(a)\big)},$$

$$\tau_2^\ell(a,h_2) = 1 - \frac{1}{ah_2}\frac{V_\ell\big(\ell_2(a,h_2),0\big)}{U'\big(c_2(a,h_2)\big)}.$$

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

$$\frac{1}{U'(c_1)}\frac{\tau_1^\ell}{1-\tau_1^\ell} = \left(1 + \frac{1}{\gamma(\ell_1,e_1)}\right)\Theta + \frac{\phi}{\lambda}\frac{V_{\ell e}(\ell_1,e_1)}{V_\ell(\ell_1,e_1)} \tag{19a}$$

$$\frac{1}{U'\big(c_2(h_2)\big)}\frac{\tau_2^\ell(h_2)}{1-\tau_2^\ell(h_2)} = \left(1 + \frac{1}{\gamma}\right)\Theta. \tag{19b}$$

In the following two propositions, we characterize the labor wedge. The proofs are omitted, since the results follow directly from the optimality conditions (19a) and (19b) and from Proposition 3.

PROPOSITION 6: *If Assumption* 1 *holds, 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 3), 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 provide a sharper characterization of the limiting value, we assume that $\Theta(a)$ converges to zero. If it converges to a strictly positive limit, then the limiting labor wedges will be shifted upward by the limiting value.[13]

[13] This assumption is satisfied, for example, when the ability distribution is lognormal or bounded.

PROPOSITION 7: *Suppose that* $\Theta(a)$ *converges to zero as* a *converges to* $\bar{a}$. *Then* $\tau_1^\ell(a)$ *converges to a positive* (*negative*) *value if* $V_{\ell e}$ *is positive* (*negative*). *In addition,* $\tau_2^\ell(a, h_2)$ *decreases in* $h_2 \in H$.

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. Non-separability allows the planner to motivate learning effort by decreasing 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) obtains 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 truth-tellers from deviators: deviators underinvest in human capital, have lower productivity, and are hurt by the negative marginal tax at the top more than truth-tellers. This mechanism does not appear in our model because human capital realizations are observable. However, our mechanism is absent in Grochulski and Piskorski (2010), who do not allow for simultaneous labor effort and investment in human capital.[14]

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

$$\frac{1}{\tau_1^\ell(a)} = \int_H \frac{1}{\tau_2^\ell(a, h_2)} p\big(h_2 | f_1(a)\big)\, dh_2.$$

PROOF:

Since the right-hand sides of (19a) and (19b) are equal when $V_{\ell e} = 0$,

$$\frac{\tau_1^\ell(a)}{1 - \tau_1^\ell(a)} \frac{1 - \tau_2^\ell(a, h_2)}{\tau_2^\ell(a, h_2)} = \frac{U'\big(c_1(a)\big)}{U'\big(c_2(a, h_2)\big)}. \tag{20}$$

The result follows from using (17) and rearranging. ∎

[14] There are additional arguments for violation of the no-distortion-at-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) establishes the same result when there is capital-skill complementarity. Those arguments rely on general equilibrium effects that are absent in our paper.

Jensen's inequality then implies that the average labor wedge is increasing over time.

COROLLARY 1:

$$\tau_1^\ell(a) \ < \ \int_H \tau_2^\ell(a,h_2)\, p\big(h_2|f_1(a)\big)\, dh_2. \tag{21}$$

To understand the result, consider an agent with a high second-period human capital shock. Due to the moral hazard considerations, the agent receives high consumption (10b). However, high second-period consumption introduces a positive income effect. A positive income effect reduces the agent's hours worked, and it is optimal to respond by reducing the marginal labor tax rate or the labor wedge (Saez 2001). The opposite is true for agents with low human capital realizations. The labor wedge thus moves negatively with second-period consumption and positively with the marginal utility of second-period consumption. The inverse of the marginal utility of consumption follows a random walk, and the inverse of the labor wedge inherits this pattern.

The result in Proposition 8 holds because changes in the second-period consumption are generated by the moral hazard friction, which is distinct from the private information friction. In a standard dynamic optimal taxation model, e.g., in Golosov, Kocherlakota, and Tsyvinski (2003); Farhi and Werning (2013); or Golosov, Troshkin, and Tsyvinski (2016), the income effect affects the optimal labor wedge as well. However, in those models, current productivity shock determines not only the current consumption, but also the current Lagrange multiplier Θ. As a result, the optimal labor wedge is a product of two forces, and Proposition 8 no longer holds. One can see this mathematically by integrating equation (20) to obtain

$$\int_H \frac{\tau_2^\ell(a,h_2)}{1-\tau_2^\ell(a,h_2)}\frac{U'\big(c_1(a)\big)}{U'\big(c_2(a,h_2)\big)}\, p\big(h_2|f_1(a)\big)\, dh_2 \ = \ \frac{\tau_1^\ell(a)}{1-\tau_1^\ell(a)}$$

and comparing it to equation (13) in Farhi and Werning (2013) with $\rho \ = \ 1$. Their equation (13) includes an extra term on the right-hand side, which captures the effect of the additional private information in the current period. If that was the case in their model, current consumption would collapse to a value independent of the current shock; in our model, it still depends on h_2 due to moral hazard.

Additively separable utility in labor effort and learning effort serves as a useful benchmark. If $V_{\ell e} \ > \ 0$, there are two additional forces. First, the first-period labor wedge increases in order to motivate higher learning effort. This weakens, or reverts, the increasing profile of the labor wedge. Second, the Frisch elasticity $\gamma(\ell_j, e_j)$ changes endogenously. Estimates from Peterman (2016) 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, however, reinforces the increasing intertemporal 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 dictate learning effort directly and the Euler equation (8) no longer holds. Let τ_1^e be the effort wedge:

$$\frac{V_e(e_1,\ell_1)}{f_e}(1 - \tau_1^e) = \beta \int_H u_2(h_2)\, p_f(h_2|f)\, dh_2.$$

The following proposition shows that the gross schooling wedge is now equal to the effort wedge.

PROPOSITION 9: *If effort is observable, the gross schooling wedge* $\tau^{s,gross}$ *is given by*

$$\tau^e = \tau^{s,gross}.$$

Without moral hazard, consumption also no longer needs to vary with human capital realizations and is deterministic. Equation (17) implies that the savings wedge is zero. From Proposition 8, 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.

III. A Market Economy

We now investigate a decentralized economy with taxes and schooling subsidies and provide a connection between the efficient allocations and a tax system that implements it. We consider the following tax and subsidy system $\mathcal{T} = (T_1, T_2, X)$. It consists of a tax on labor income $T_1(y_1, M_1(s_1))$ in the first period and $T_2(y_1, y_2, h_2)$ in the second period and a tax on savings $X(k_1)$. All taxes are nonlinear. The income tax in the first period is a joint function of schooling expenditures M and first-period income.[15] Income tax in the second period exhibits history dependence. It is also stochastic because it depends explicitly on the realization of the human capital shock. Finally, the tax on savings depends only on the level of savings.[16]

Given a tax and subsidy system $\mathcal{T}$, the agents in a decentralized economy face budget constraints:

(22a) $$c_1 + M(s_1) + k_1 \le y_1 - T_1(y_1, M(s_1)),$$

(22b) $$c_2(h_2) \le X(k_1) + y_2(h_2) - T_2(y_1, y_2(h_2), h_2),$$

[15] We write the income tax to be a function of $M(s)$ rather than a function of the schooling investment s, to make it easier to express marginal schooling subsidies as a fraction of marginal schooling cost.

[16] The savings tax is defined as a function of the level of savings, but it can easily be transformed to a more usual tax on interest income $\tau^k(k_1)$ by $X(k_1) = [1 + r(1 - \tau^k(k_1))]\, k_1$.

where labor earnings are given by (2). The solution to this *market problem* is given by (σ, k_1), where the allocation $\sigma(a)$ for agent a and his savings $k_1(a)$ maximize the expected utility (1) subject to the budget constraints (22). The choice of effort level is left implicit.

Implementation.—The functional forms of $\mathcal{T}$ are not chosen arbitrarily. They are needed in order to implement the efficient allocations in a decentralized economy. We say that a tax and subsidy system $\mathcal{T}$ implements an allocation σ if each agent selects zero savings and the allocation assigned to him given $\mathcal{T}$, i.e., if $(\sigma(a), 0)$ solves the agent's problem in a market economy.

The implementation relies on the taxation principle by Hammond (1979) and Rochet (1985), where the taxes are such that the incentive-compatible allocations are just budget feasible by the agent and are set low enough for other choices to prevent joint deviations, either in first and second-period income or in first-period income and schooling investment. The need to prevent joint deviations explains why both the second-period income tax and schooling subsidy depend on the first-period income. In addition, we follow Werning (2011) in designing the nonlinear savings tax in a way that eliminates all the gains from joint deviations. The details of how the tax on savings is constructed is in Appendix A2, which also contains the proof of the following main result of this section.

PROPOSITION 10: *If an allocation σ satisfies the incentive constraint* (5)*,t hen there exists a tax system $\mathcal{T}$ such that $X(0) = 0$ for all y, and $(\sigma, 0)$ solves the market problem. Conversely, let $\mathcal{T}$ be a tax system such that $X(0) = 0$ and suppose that $(\sigma, 0)$ solves the market problem. Then the allocation σ is incentive compatible.*

If the first-period tax function $T_1(y_1, M)$ decreases in M, then higher schooling expenditures decrease tax liability. That is, schooling expenditures are partially deductible. If the tax system is differentiable, then it is easy to show that $\partial T_1(y_1(a), M_1(a))/\partial M = -\tau_1^{s,gross}(a)$. A positive gross schooling wedge thus measures the degree of deductibility of schooling expenditures. Proposition 4 then implies that, in the optimum, schooling expenditures will be deductible, but not fully. How much deductibility will be optimal will be investigated in the next section.

Note also that the second-period tax has a special feature: it is stochastic because it depends directly on the realization of the second-period human capital shock. Such a feature is needed in order to provide the right incentives to exert a first-period effort, as one can see from the Euler equation (8) and from the second-period labor wedges (19b). We will again investigate this feature quantitatively in the next section.

The implementation adopted in this paper is not unique. One possible alternative is a system where the agents receive a loan for their educational investment and partially repay it in the second period. This is a subsidy system adopted in Stantcheva (2017). In our setting, the tax system would by represented by a time invariant income tax $T(y)$ (not uniquely defined and possibly set to zero), loan schedule

$L_1(M(s_1))$, a second-period repayment schedule $D(L_1, y_1, y_2, h_2)$, and a saving tax X and would lead to budget constraints of the following form:

$$c_1 + M(s_1) - L\big(M(s_1)\big) + k_1 \leq y_1 - T(y_1),$$

$$c_2(h_2) \leq X(k_1) + y_2(h_2) - T(y_2) - D\big(L_1(s_1), y_1(h_2), y_2(h_2), h_2\big).$$

Our tax system prevents joint deviations by making the first-period and second-period income tax non-separable; the alternative tax system achieves the same result by making the repayment schedule D non-separable. Given that we work with a two period model, a model period represents about 20 years of one's life. This alternative tax system would therefore have agents repay the loan only after 20 years.

IV. Quantitative Analysis

The benchmark model is the decentralized incomplete markets economy under the current US tax system, which we refer to as the "status quo." The status quo model is used to calibrate the parameters of the ability distribution and human capital accumulation technology. We then calculate the constrained-efficient outcomes by solving the relaxed planning problem while keeping all other parameters of the status quo model unchanged and verifying the validity of the relaxed problem.

A. *Calibration*

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.

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.

Preferences.—The instantaneous utility function for consumption is CRRA:

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

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:

$$V(\ell, e) = \frac{\ell^{1+1/\gamma}}{1+1/\gamma} + \frac{e^{1+1/\epsilon}}{1+1/\epsilon}. \tag{23}$$

The Frisch elasticity of labor supply is set to $\gamma = 0.5$, consistent with micro-estimates surveyed in Chetty et al. (2011). The elasticity of learning effort is set to $\epsilon = 0.5$, equal to the Frisch elasticity. The interest rate is set to the standard

Table 1—Assigned Parameters

Definition	Symbol	Value	Source/Target
CRRA parameter	ρ	1	Browning, Hansen, and Heckman (1999)
Frisch elasticity of labor	γ	0.5	Chetty et al. (2011)
Elasticity of effort	ϵ	0.5	Same as Frisch elasticity
Discount factor	β	0.457	4% annual interest rate
H.C. technology	α	0.7	Browning, Hansen, and Heckman (1999)
Capital income tax rate	τ^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)
Schooling subsidy	τ^s	3.5%	Stantcheva (2017)
Shock distribution	(μ_z, σ_z)	(−0.58, 0.496)	Huggett, Ventura, and Yaron (2011)

Table 2—Calibrated Parameters

Definition	Symbol	Value	Target moment	Value	Source
Effort share	η	0.13	Wage premium	1.8	Heathcote, Perri, and Violante (2010)
Initial human capital	h_1	0.146	Mean earnings ratio	0.868	Huggett, Ventura, and Yaron (2011)
SD log-ability	σ_a	0.468	Earnings Gini	0.343	Huggett, Ventura, and Yaron (2011)
Mean log-ability	μ_a	1.621	Earnings variance	0.390	Huggett, Ventura, and Yaron (2011)
Pareto tail, ability	λ_a	7.441	Mean-to-median earn. young	1.335	Huggett, Ventura, and Yaron (2011)
Linear cost	b_ℓ	0.269	Schooling cost ratio	6%	Stantcheva (2017)
Adjustment cost	b_a	0.160	Mean hours ratio	0.988	Rupert and Zanella (2015)

annual value of 4 percent, which implies a 20-year interest rate of 119 percent. Agents' discount factor is set to $R = \beta^{-1}$, so that $\beta = 0.4566$. This implies an annual discount rate of 0.962.

Technology.—The human capital production function is of the Ben-Porath form:

$$f(e_1, s_1) = h_1 + \left(e_1^{\eta} s_1^{1-\eta} h_1\right)^{\alpha}.$$

We posit that learning effort and schooling aggregate in a constant returns fashion. The share of learning effort in human capital production is determined by the parameter η. Setting η equal to one shuts down schooling, as in the Huggett, Ventura, and Yaron (2011) economy. 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).[17]

[17] While there is no clear way of mapping the concavity of a human capital production function in a one-period model to the concavity in a multi-period model, our results are robust to reasonable variations around the choice of $\alpha = 0.7$.

The monetary costs of schooling consist of a linear and a quadratic term:

$$M(s_1) = b_\ell s_1 + b_a s_1^2.$$

We target three moments to calibrate the three human capital parameters η, b_ℓ, and b_a for which we do not have direct estimates. The learning effort share η targets the college wage premium of 1.8, in line with the US average for men since the 1990s (see Heathcote, Perri, and Violante 2010). As there is no discrete college choice in the model, we need to redefine model moments appropriately. Following Stantcheva (2017), we set this value to match the wage premium of the top 42.7 percent of agents, ranked by educational expenses, to the bottom 57.3 percent of agents.[18] We set b_ℓ to match the ratio of monetary costs of schooling to lifetime earnings. Stantcheva (2017) calculates this value to be about 6 percent after college. Finally, b_a is set to match the ratio of average labor hours of 20- to 40-year-olds to the average labor hours of 40- to 60-year-olds. This ratio is 0.988 in PSID data reported in Rupert and Zanella (2015). The parameters that best match these moments are $\eta = 0.1314$, $b_\ell = 0.2694$, and $b_a = 0.1600$.

The shock process is assumed to be independently and identically distributed across agents. The shocks are drawn from a truncated normal distribution, $z_2 \sim N(\mu_z, \sigma_z)$. The human capital shock process is estimated in Huggett, Ventura, and Yaron (2011). The Ben-Porath functional form implies that toward 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 do not invest in human capital.[19] Huggett, Ventura, and Yaron (2011) estimates annual values of $\sigma_z^{ann} = 0.111$ and $\mu_z^{ann} = -0.029$. We transform the shock process to its 20-year period equivalent, $\sigma_z = \sqrt{20(\sigma_z^{ann})^2} = 0.496$ and $\mu_z = 20\mu_z^{ann} = -0.58$. 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 US tax system with a flat tax on capital income, a progressive tax on labor income and a linear schooling subsidy. Labor income taxes take the Gouveia-Strauss form

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

Without loss of generality, we set the lump-sum transfer *LS* to zero in the second period, but allow them to be nonzero in the first period in order to redistribute tax revenues back to the agents. Guner, Kaygusuz, and Ventura (2014) estimates that $\nu_0 = 0.182$, $\nu_1 = 0.008$, and $\nu_2 = 1.496$. These parameters imply progressive marginal tax rates

[18] Autor, Katz, and Krueger (1998) finds that the top 42.7 percent of the population have a full-time college equivalent education. We extrapolate these numbers to formal schooling after college.

[19] Huggett, Ventura, and Yaron (2011) calculates 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.

starting at 8 percent and increasing to 24.5 percent. We obtain mean average tax rates for capital and consumption from McDaniel (2007).[20] We adjust labor and capital tax rates by the average consumption tax. This yields an effective annual capital tax rate of $\tau^k_{ann} = 37$ percent, which we transform to a 20-year value.[21] The effective 20-year tax rate on capital income is $\tau^k = 45.88$ percent. Schooling subsidies are calculated by Stantcheva (2017) at 35 percent annually for two years. This translates to a subsidy of $\tau^s_1 = 3.5$ percent in a 20-year period.

Initial Conditions.—We assume that the ability distribution is Pareto-lognormal, $q(a) \sim PLN(\mu_a, \sigma_a, \lambda_a)$. The initial human capital, h_1, is the same for all agents. We set μ_a, σ_a, λ_a, and h_1 so that the equilibrium distribution of earnings matches data earnings moments. Huggett, Ventura, and Yaron (2011) estimates age profiles of mean earnings from the PSID 1969–2004 family files. We target four 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, the mean-to-median earning ratio of young workers (ages 20 to 40), and the variance of earnings. Table 2 reports the results of the calibration. Parameter values $h_1 = 0.1456$, $\sigma_a = 0.4675$, $\mu_a = 1.6211$, and $\lambda_a = 7.4413$ best approximate the model to the data targets.

B. *Findings*

Insurance, Redistribution, and Incentives.—Proposition 3 shows that a risky second-period consumption is necessary to provide optimal incentives to accumulate human capital. Figure 1, panel A shows the standard deviation of log-consumption in the second period as a function of 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 (those above the ninetieth percentile). Figure 1, panels B and C show the corresponding effort and schooling elicited from the agents in both the efficient and in the status quo allocation. Schooling is significantly higher in the efficient allocation for all agents above the fiftieth percentile, whereas learning effort is lower for all agents. The private information and moral hazard costs associated with learning effort makes effort more costly to elicit than schooling. The outcome is that it is optimal to substitute learning effort for more schooling from high-ability agents. Figure 1, panels B and C show that the current US tax system is inefficiently skewed toward learning effort and away from schooling investment.

Why is it that higher consumption risk does not motivate top agents to increase their learning effort? To understand this 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.

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

[21] The effective 20-year capital income tax rate is the solution to $(1 + r_{ann}(1 - \tau^k_{ann}))^{20} = 1 + r(1 - \tau^k)$.

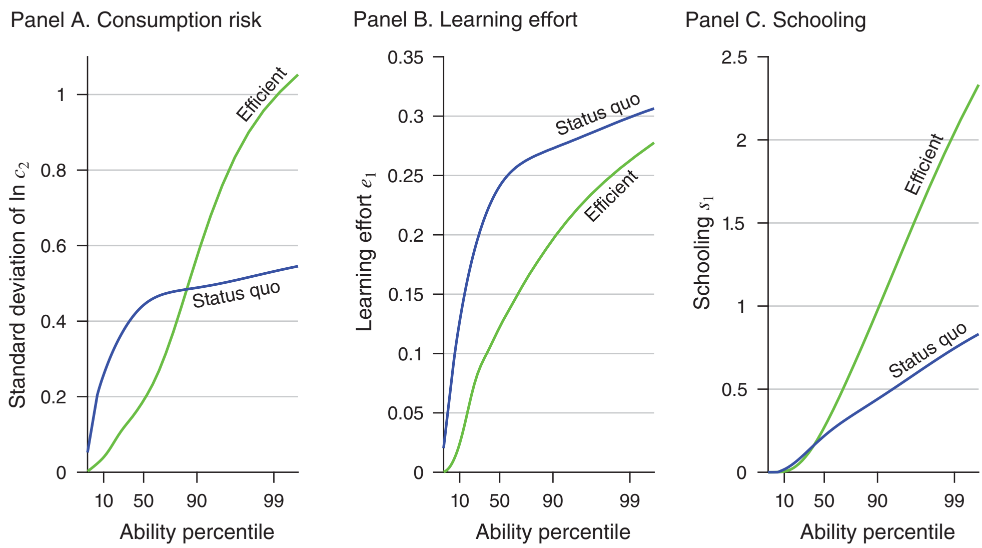

Figure 1

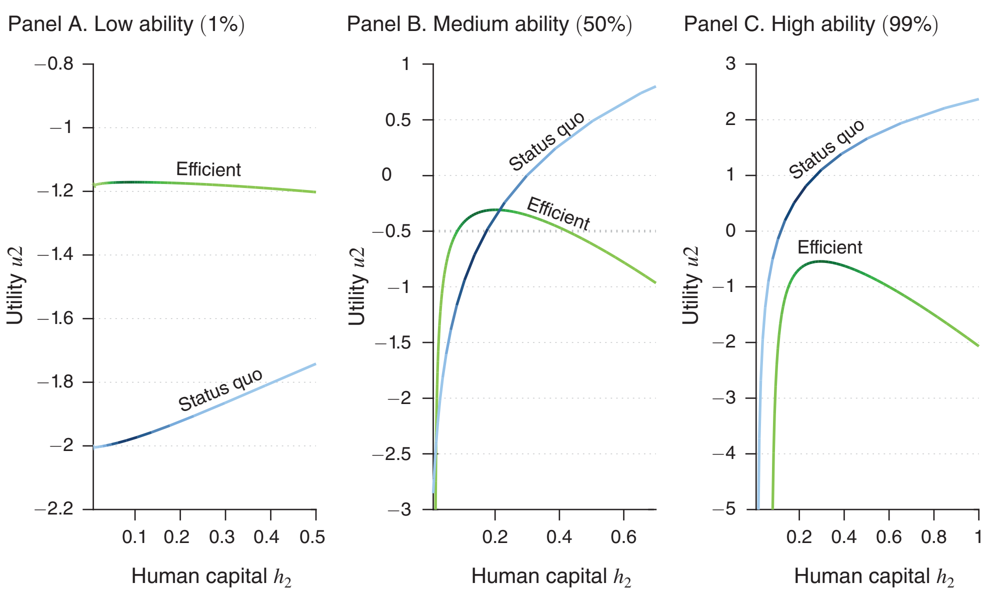

Figure 2. Second-Period Utility Profiles

Note: Darker shades represent higher probabilities.

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 panel C, which shows second-period utilities for a high-ability agent. 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. The efficiency consideration at the top decreases 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.[22] For the agents in panels B and C, both of those effects are strong, producing a significantly hump-shaped second-period utility profile. For agents at the bottom of the ability distribution (shown in panel A), both effects are mild, and their second-period utility is almost flat.

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 status quo, the consumption distribution thus becomes more equal within each period. Labor effort 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.

Labor Wedge and Labor Taxes.—The labor wedge in the first period, the expected labor wedge in the second period, and the marginal labor tax rates in the status quo are shown in Figure 3, panel A. The expected second-period labor wedge exceeds the first-period labor wedge as expected from Corollary 1, although the difference is very small, at less than 1 percent. The labor wedge in the first period decreases with abilities. In contrast, the status quo marginal labor tax rates are increasing in ability and bounded between 8 and 24 percent. The second-period labor wedge is shown in Figure 3, panel B. The labor wedge decreases with human capital realizations, as predicted in Proposition 6. The decrease is most rapid for higher ability levels, reflecting the fact that higher ability agents face a riskier consumption profile.

Whereas the labor wedge paints a picture of a regressive optimum tax system, the average labor tax rate paints a different picture. Figure 4, panel A shows that, in expectation, average labor tax rates are increasing in ability, consistent with the redistribution objective of the planner.[23] Low to middle abilities receive a net subsidy that is funded, in part, by positive average labor taxes from agents with high abilities. Average labor tax rates in the status quo also increase with ability, but are never negative and the dispersion is much smaller. Figure 4, panel B shows how average labor taxes vary with human capital realizations for different abilities. There are three interesting features in this figure. First, for any given human capital realization, higher abilities pay a higher average labor tax rate. This is, again, arising from the redistribution objective of the planner. Second, average labor taxes are high for

[22] 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.

[23] The average labor tax rate is calculated as $(y_1 - c_1 - M(s_1)(1 - \tau_1^{s,gross}))/y_1$ in the first period and $\int_H T_2/y_2 p_2(h_2|f_1)\, dh_2$ in the second period.

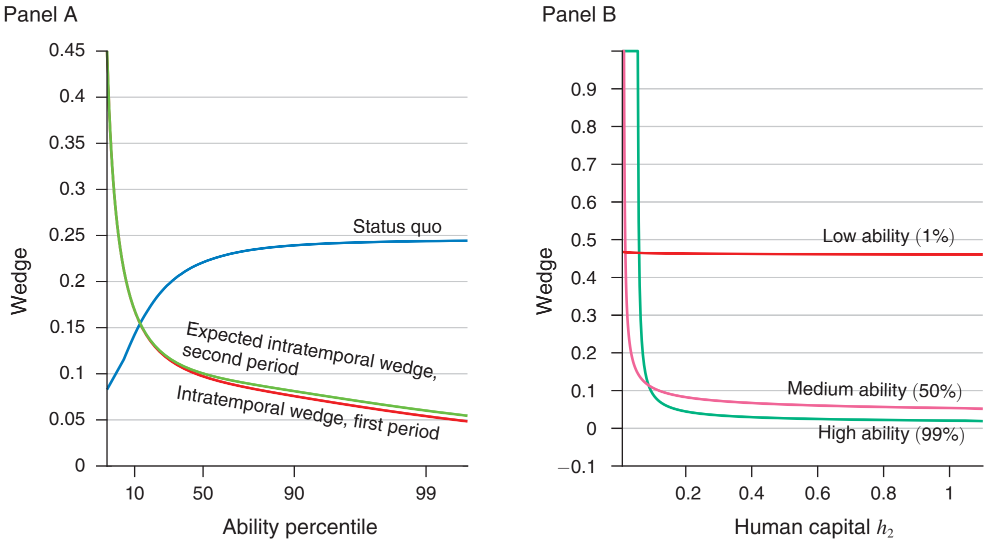

Figure 3. Labor Wedge

Notes: In panel A, the labor wedge is plotted by the ability percentile. In panel B, the labor wedge is plotted by human capital realizations, second period.

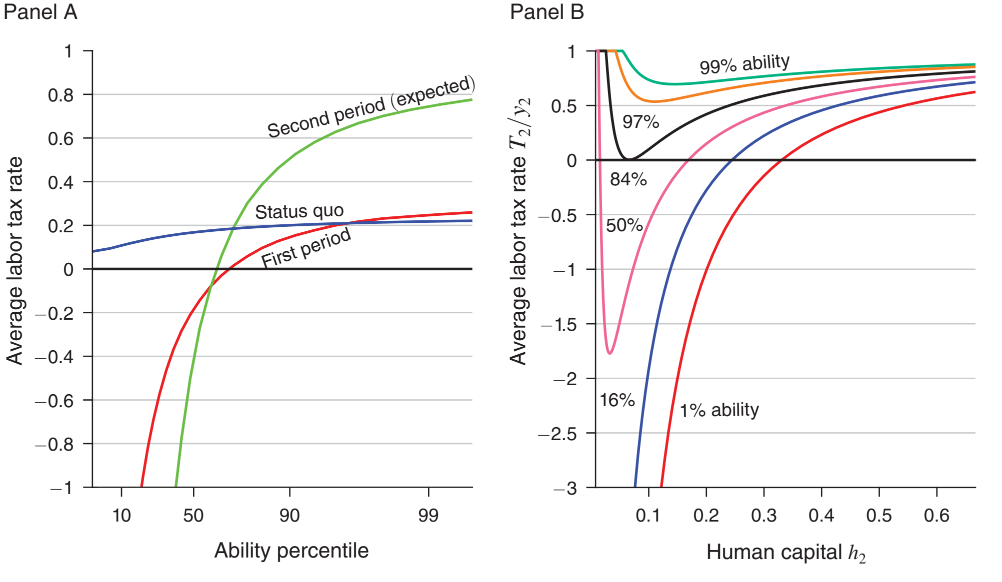

Figure 4. Average Labor Tax Rate

Notes: In panel A, the labor tax rate is plotted by the ability percentile. In panel B, the labor tax rate is plotted by human capital realizations, second period.

high realizations of human capital and start dropping as human capital realizations drop. This redistributes resources from the lucky agents with high human capital realizations to others. Third, for low realizations of human capital, two possibilities

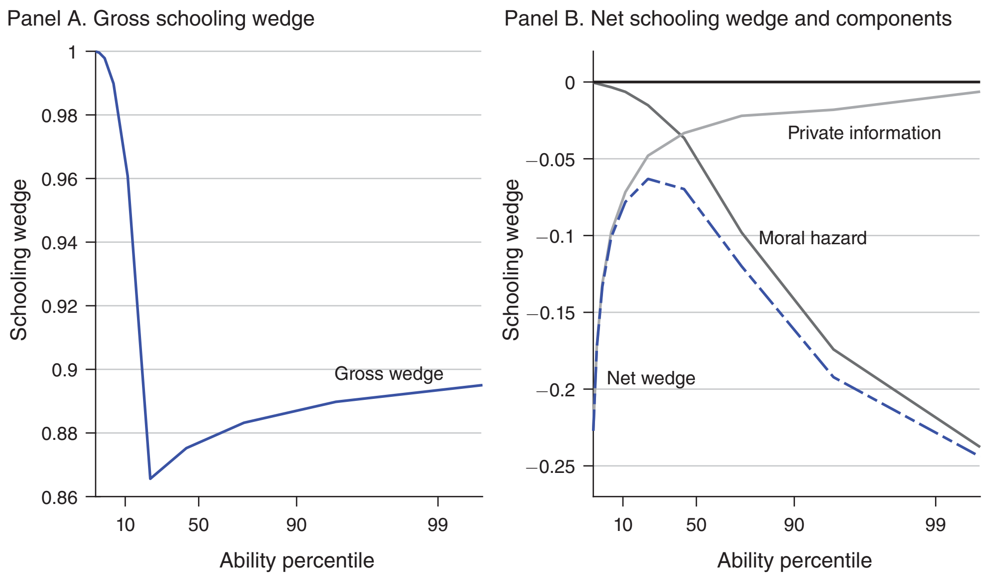

FIGURE 5. SCHOOLING WEDGE

arise. High-ability agents see their average labor taxes again increase. The threat of high taxes provides incentives for higher effort in the first period. For low-ability agents, however, the insurance aspect dominates, there is no punishment for low human capital realizations, and average taxes further decrease.

Schooling Wedge.—The gross schooling wedge is shown in Figure 5, panel A. As stated in Proposition 4, the schooling subsidy is positive for all agents. It is substantial, with values ranging from 86.5 percent to 100 percent. That is, a substantial fraction of schooling expenditures is, on the margin, deductible. The optimal schooling policy starts with almost full deductibility for the lowest ability agents, with deductibility initially decreasing with ability until about the sixteenth ability percentile, and then increasing for the remainder of the population. The abrupt change in the shape of the schooling wedge corresponds to that same change in Figure 4, panel B. As abilities decrease, there is an ability threshold between the sixteenth and the fiftieth percentile such that incentives for learning effort are more costly than the benefits. For those above the threshold, a higher schooling subsidy corresponds to higher learning effort incentives. For those below the threshold, redistribution dominates so that lower abilities receive a larger subsidy.

The net schooling wedge, shown in Figure 5, panel B, is the sum of a private information component and a moral hazard component, as laid out in equation (16). While the sign of the moral hazard component is theoretically ambiguous, it turns out to be negative for all ability levels, and moreover, it decreases with ability. High-ability agents not only exert more effort, but the benefits from higher effort are smaller due to redistributional concerns, see the right panel of Figure 2. For both reasons, the Euler equation constraint becomes harder to satisfy, which depresses the wedge. In contrast, the private information component is the highest for low-ability agents, since the private information constraint is most binding for them. Summing together, the net schooling wedge is negative everywhere and hump-shaped. For both high and low-ability agents, it is substantial at about −25 percent.

It is interesting to compare the schooling wedges to those found in other studies. For example, the gross schooling subsidy in Stantcheva (2017) is between −7 and 19 percent, depending on the elasticity of substitution between human capital and ability. We obtain higher gross schooling subsidies in part because that paper obtains lower marginal labor tax rates and therefore a lower fiscal externality and because the gross schooling subsidy is pushed down by the impact of schooling on contemporaneous human capital. This contemporaneous effect is absent in our model. The average gross schooling wedge in Findeisen and Sachs (2016) is 111 percent, ranging from around 275 percent for the lowest abilities to −21 percent for middle abilities, increasing again to 38 percent for highest abilities.[24] In contrast, our moral hazard constraint bounds the schooling wedge below one, and we obtain a smaller dispersion of values. As for the Ramsey taxation literature, Krueger and Ludwig (2016) finds optimal gross schooling subsidies of 170 percent in a life-cycle Ramsey taxation framework, arising mostly from a dominant fiscal externality force. This number is similar to the gross schooling subsidy we get in the first-best economy, where the only force driving the schooling wedge is the fiscal externality. The gross schooling wedge in our first-best economy is around 150 percent. As for the net schooling wedge, the only quantitative estimate available comes from Stantcheva (2017). That paper obtains values between around −2 percent and 5 percent. We obtain much larger negative values in Figure 5, panel B, due in part to the downward moral hazard force.

C. *HVY Economy*

It is interesting to compare our benchmark economy to one without monetary costs of schooling, as in Huggett, Ventura, and Yaron (2011). We do this by setting the learning effort share parameter η to one and recalibrating the model.[25] Figure 6, panel A shows that the agents, especially higher ability agents, now face a substantially higher consumption risk. Schooling, by reducing the role of effort in generating human capital in the second period, thus decreases the required consumption risk. Figure 6, panel B shows that without schooling, the optimal effort schedule is much steeper and even crosses the status quo at the top. Schooling reduces the dependence of human capital on learning effort and also gives the planner an extra tool to incentivize this effort. Therefore, the planner is less reliant on risk and learning effort altogether.

[24] Findeisen and Sachs (2016) calculates tuition at \$11,100 per year, at 4 percent interest rate for four years, which comes to a total schooling cost of \$47,136. Given an absolute subsidy of \$52,557, the subsidy is 52,557/47,136 = 111 percent. We thank Dominik Sachs for facilitating these calculations over email correspondence.

[25] Values of $h_1 = 0.6193$, $\sigma_a = 0.5779$, $\mu_a = 0.0802$, and $\lambda_a = 6.3702$ best fit the model to mean earnings ratio, earnings Gini, earnings variance, and mean-to-median earnings moments.

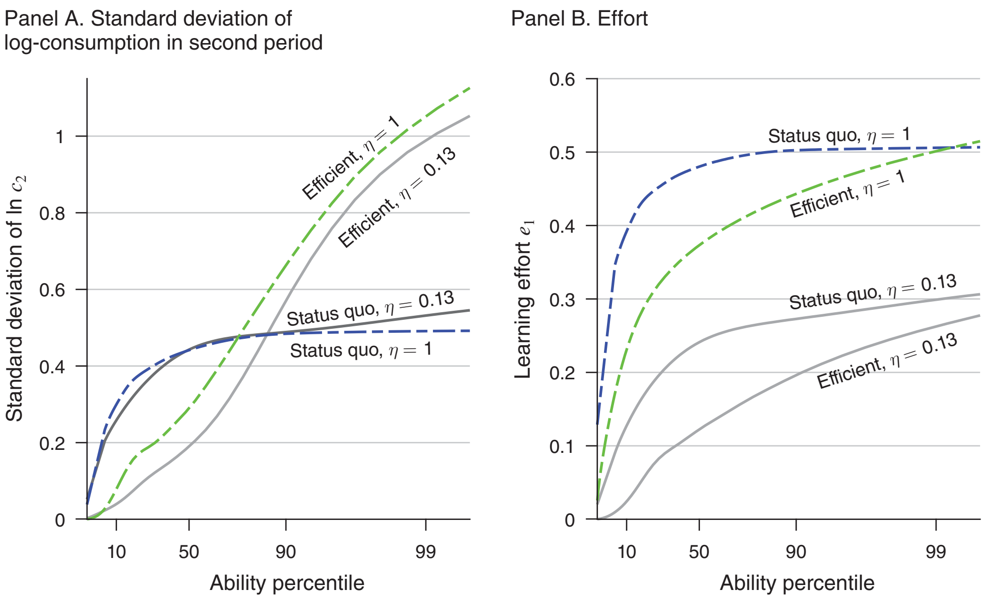

Figure 6. Consumption Risk and Effort with and without Schooling

D. *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 ζ that solves

$$W\Big((1+\zeta)c_1^{SQ},(1+\zeta)c_2^{SQ},\ell_1^{SQ},\ell_2^{SQ},e_1^{SQ},s_1^{SQ}\Big) \;=\; W^{CE},$$

where SQ denotes 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 0.96 percent annual increase in consumption in every period and state of the world.[26]

We report the welfare gains in Table 3. The left column reports welfare gains for the benchmark economy with η equal to 0.13. Shutting down the moral hazard friction completely yields welfare gains of 1.17 percent over the status quo. Shutting down the private information friction completely yields a welfare gain of 1.71 percent. Shutting down both constraints completely yields a first-best welfare gain of 1.87 percent.

The column on the right reports welfare gains for the case without schooling by setting η equal to one. The welfare gains from the constrained-efficient allocations are 0.10 percent, significantly lower than in the benchmark. In this sense, observable

[26] Annualized welfare gains are calculated as $\zeta_{ann} \;=\; \exp\Big((1+\beta)\ln(1+\zeta)\frac{1-\beta_{ann}}{1-\beta_{ann}^{41}}\Big) - 1.$

Table 3—Annualized Welfare Gains

	Benchmark $\eta = 0.13$	HVY economy $\eta = 1$
Status quo	0%	0%
Constrained efficient	0.96%	0.10%
No moral hazard	1.17%	0.31%
No private information	1.71%	1.56%
No frictions	1.87%	1.77%

schooling provides a powerful tool for a planner. Additional gains from shutting down the moral hazard constraint, however, are similar to the benchmark. In both cases, shutting down the moral hazard constraint yields additional welfare gains of about 0.20 percent.

The distribution of welfare changes across types is illustrated in Figure 7. 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.

E. *Non-separability between Labor and Effort*

The optimal schooling subsidy in (15) suggests that the cross-derivative $V_{e\ell}$ plays a critical role in determining its value. To investigate this, we numerically simulate a simple example with two alternative functional forms for V. The first one is additively separable in ℓ and e and takes the form given in (23). The second one is non-separable and takes the form

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

To make the comparison between both functional forms easier, we now assume that the utility is linear in consumption in the first period and, furthermore, that the social planner solves a Rawlsian problem of maximizing the expected utility of the least able agent $\underline{a}$ subject to the resource constraint. That way, the cumulative Lagrange multiplier on the envelope condition is simply equal to $\Theta \;=\; (1 - Q)/(aq)$ and so is identical for both utility functions. The parameters of the model are as calibrated in the previous section. That includes the assumption that Q is Pareto-lognormal.

Figure 8, panel A shows the gross schooling wedge for both functional forms. The common feature of both wedges is that they are relatively high, with 78–100 percent of schooling cost being subsidized. But the patters are different, and the gross subsidy is less progressive under non-separability. If the function V is separable, the subsidy had a pronounced U-shaped pattern. It is decreasing for the bottom 15 percent of the population and then increases for most of the remaining population, almost to the original level. In contrast, if the function V is non-separable, schooling subsidies are higher but decreasing for the bottom half of the population and do not increase significantly afterward. The reason why the subsidies are higher for the

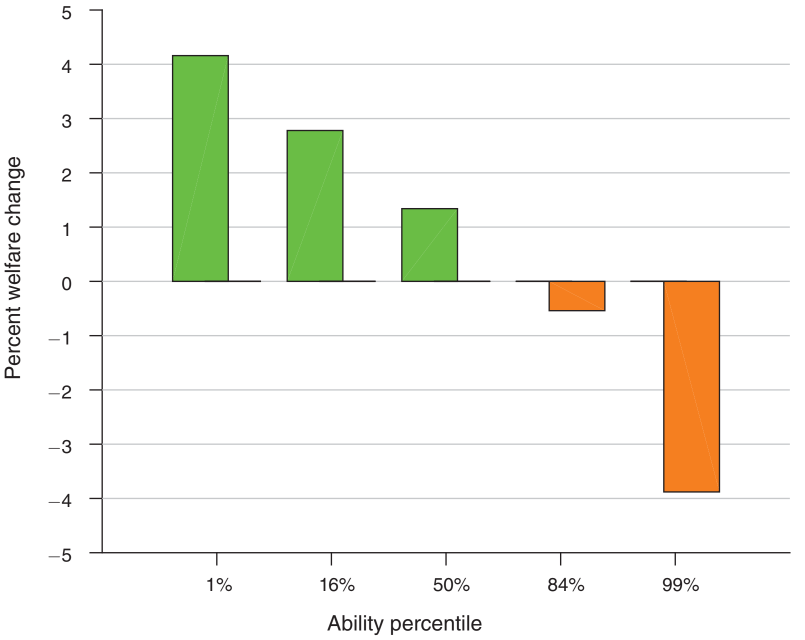

Figure 7. Welfare Change by Ability

Note: This figure shows the welfare change of the unborn agent from the status quo to the constrained efficient economy.

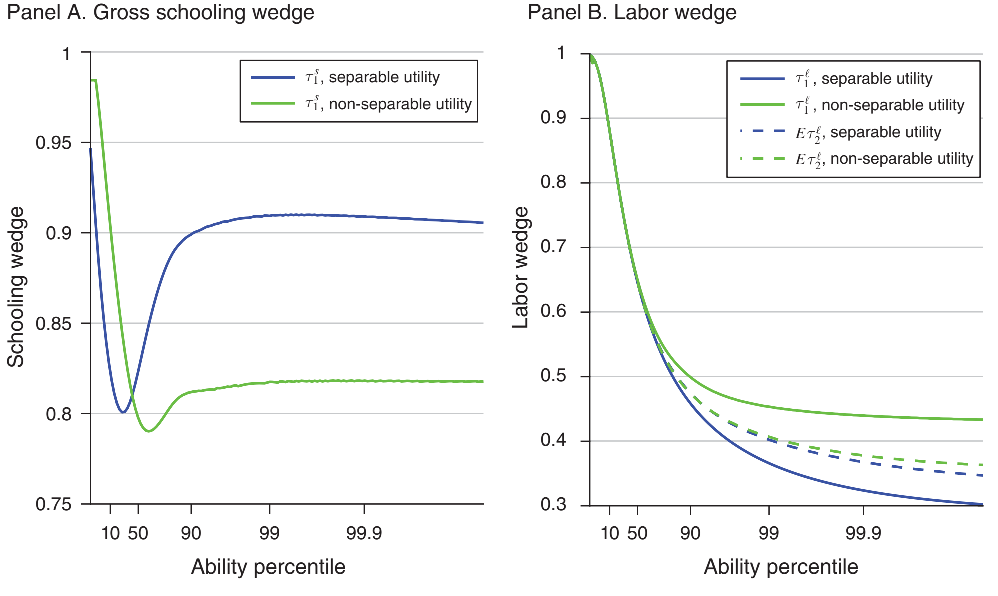

Figure 8. Rawlsian Example

bottom half comes from the first term in equation (15). With abilities being essentially lognormally distributed on the left, Θ is rather high for low abilities. High learning effort is then very costly to the social planner, and the planner responds by providing incentives for formal schooling through very high subsidies. However, most of the difference for the top half of the population comes from the second term in equation (15). Non-separability mechanically increases marginal cost of both first-period labor and effort. In the optimum, they are both lower, and the need for subsidies is diminished. Decomposing the gross wedge into the fiscal externality component, private information component, and moral hazard component, one finds that most of the difference comes from the moral hazard component. The fiscal externality term is very similar in both economies.

Figure 8, panel B shows the first-period labor wedge and the expected second-period labor wedge for the two utility specifications. The solid blue line represents the first-period labor wedge for the separable case, while the solid green line represents the non-separable case. For the bottom 90 percent of the ability levels, both specifications yield very similar labor wedges. However, for the top 10 percent, the differences are substantial. At 99.9th percentile the labor wedge for the non-separable case is about 12 percentage points higher than in the separable case and decreases at a slower rate. Non-separability of the utility function thus has potentially important consequences for the top marginal tax rates, both in terms of levels and in terms of their progressivity.

V. Conclusion

We study the interaction between observable and unobservable human capital investment and their role in shaping optimal tax policy in a life-cycle framework with risky human capital, permanent ability differences, and learning-or-doing human capital accumulation technology. The interaction of private information and moral hazard frictions is novel in the optimal taxation with human capital literature and produces several prominent results. The planner uses consumption risk to incentivize learning effort. The labor wedge is increasing over time, and the "no distortion at the top" might not apply when taxing labor encourages human capital accumulation. The gross schooling wedge is strictly positive but bounded at one, while the net schooling wedge is negative. Overall, the optimal system yields high welfare gains.

We conclude by several observations that relate the optimal tax system to the existing tax policies and discuss how to implement the optimum. We find that the current US tax system favors inefficiently high learning at the expense of inefficiently low schooling investments. We have shown that these subsidies can be implemented by two ways. First, they can be implemented by making the educational expenses partially tax deductible, similar to the American Opportunity Tax Credit or the Lifetime Learning Credit in the United States. Our result indicates that it would be optimal to extend the American Opportunity Tax Credit significantly. Its subsidies are currently capped at US$2,500, while the optimal subsidies are close to 100 percent for the bottom 10 percent of the population, and then decrease to 86–89 percent for the rest of the population (see Figure 5, panel B). A different way to implement the subsidy is via educational loans with income-contingent repayment schedules, similar to the United Kingdom's Tuition

Fee and Maintenance Loans. Under the UK system, students receive an educational loan from the government, and repayments are collected as a tax on the amount earned above an income threshold after graduation.

In order to reduce the complexity of the model, we relied on the assumption that human capital realizations are observable. One important implication is that the optimal tax system is contingent not only on income but on human capital realizations. To approximate that, one could condition taxes and repayments on observable characteristics that are correlated with human capital, for example, by making them occupation specific of dependent on one's type of degree obtained. Making human capital realizations unobservable would have the attractive feature, from a policy perspective, that the planner cannot condition taxes on them. It is, however, not clear that the first-order approach holds if one relaxes this assumption. Exploring the implications of unobservable human capital realizations on optimal tax policy is a desirable avenue for future research.

Appendix

A1. *A Lagrangean Solution Method*

Let λ, $\phi(a)\,q(a)$, and $\theta(a)\,q(a)$ be the Lagrange multipliers on the resource constraint, the Euler equation in effort and on the envelope condition. The planning problem can be written as a saddle point of the Lagrangean:

$$\max_{c,y,e,s}\ \min_{\lambda,\theta,\phi} \int_A \mathcal{L}(a)\, q(a)\, da,$$

where

$$\begin{aligned}
\mathcal{L}(a) \;=\; & \big(1+\theta(a)\big)W(a) - \theta(a)W(\underline{a}) \\
& - \lambda\Big[c_1(a) + M(s_1) - y_1(a) \\
& \qquad + \beta \int_H \big[c_2(a,h_2) - ah_2\,\ell_2(a,h_2)\big]\, p\big(h_2|f_1(a)\big)\, dh_2\Big] \\
& - \theta(a) \int_{\underline{a}}^{a} \Big[V_\ell\big(\ell_1(\tilde{a}), e_1(\tilde{a})\big)\,\ell_1(\tilde{a}) \\
& \qquad + \beta \int_H V_\ell\big(\ell_2(\tilde{a},h_2),0\big)\,\ell_2(\tilde{a},h_2) p\big(h_2|f_1(a)\big)\, dh_2\Big]\frac{d\tilde{a}}{\tilde{a}} \\
& - \phi(a)\Bigg[\frac{V_e\big(\ell_1(a), e_1(a)\big)}{f_e\big(e_1(a), s_1(a)\big)} \\
& \qquad - \beta \int_H \big[U\big(c_2(a,h_2)\big) - V\big(\ell_2(a,h_2),0\big)\big]\, p_f\big(h_2|f_1(a)\big)\, dh_2\Bigg]
\end{aligned}$$

and

$$W(a) = U\big(c_1(a)\big) - V\big(\ell_1(a), e_1(a)\big) + \beta\int_H \Big[U\big(c_2(a,h_2)\big) - V\big(\ell_2(a,h_2),0\big)\Big]\, p\Big(h_2 | f\big(e_1(a), s_1(a)\big)\Big)\, dh_2.$$

The first-order condition in $W(\underline{a})$ implies $\int_A \theta(a)\, q(a)\, da = 0$. Let $\Theta(a) = \big(\int_a^{\bar{a}} \theta(\tilde{a}) q(\tilde{a})\, d\tilde{a}\big)/(aq(a))$. Integrating by parts and rearranging terms, one obtains, conditional on a,

$$\begin{aligned}
\text{(A-1)} \quad \mathcal{L} = {} & (1+\theta)\big(U(c_1) - V(\ell_1, e_1)\big) - \lambda\big[c_1 + M(s_1) - ah_1\ell_1\big] \\
& - \Theta V_\ell(\ell_1, e_1)\, \ell_1 - \phi \frac{V_e(\ell_1, e_1)}{f_e(e_1, s_1)} \\
& + \beta\int_H \Big[(1+\theta)\big[U\big(c_2(h_2)\big) - V\big(\ell_2(h_2),0\big)\big] \\
& \qquad - \lambda\big[c_2(h_2) - a\,h_2\,\ell_2(h_2)\big]\Big] p\big(h_2 | f(e_1, s_1)\big)\, dh_2 \\
& - \beta\Theta\int_H V_\ell\big(\ell_2(h_2),0\big)\, \ell_2(h_2)\, p\big(h_2 | f(e_1, s_1)\big)\, dh_2 \\
& + \phi\int_H \big[U\big(c_2(h_2)\big) - V\big(\ell_2(h_2),0\big)\big]\, p_f\big(h_2 | f(e_1, s_1)\big)\, dh_2.
\end{aligned}$$

A2. *Implementation*

We will show how to decentralize the efficient allocations through a tax system. We describe the tax system in two steps. In the first step, we 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 (Werning 2011). Armed with the savings tax function, we prove Proposition 10, a version of a taxation principle, in the second step.

Constructing the Tax Function.—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 and to change their schooling investment. Let k_1 be pretax savings and $x(k_1)$ be second-period after-tax savings satisfying $x(0) = 0$. The agent's choice sets are

$$\text{(A-2)} \qquad c_1 + k \leq c_1(\hat{a}),$$

$$\text{(A-3)} \qquad c_2 \leq c_2(\hat{a}, h_2) + x(k) \quad \forall h_2.$$

Clearly, choosing $k = 0$ is feasible and yields consumption $c_1(\hat{a})$ and $c_2(\hat{a}, h_2)$. Let the lifetime utility from the utility-maximizing report, conditional on savings and schooling investment being k, be

$$\hat{W}(k; x|a) = \max_{\hat{a},e}\left\{ U\left[c_1(\hat{a}) - k\right] - V\left(\frac{y_1(\hat{a})}{a h_1}, e\right) \right.$$
$$+ \beta \int_H \left[U\left(c_2(\hat{a}, h_2) + x(k)\right) - V\left(\frac{y_2(\hat{a}, h_2)}{a h_2}, 0\right)\right]$$
$$\left. \times\, p\left(h_2 | f\left(e, s_1(\hat{a})\right)\right) dh_2 \right\}.$$

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

$$\hat{W}\left[k; x^*(\cdot, a)|a\right] = W(a) \quad \forall k. \tag{A-4}$$

We further take the supremum over a to define a tax function $x^{**}(k_1) = \sup_{a \in A} x^*(k_1, a)$. It follows from (A-4) that

$$W(a) \geq \hat{W}\left(k; x^{**}|a\right). \tag{A-5}$$

PROOF OF PROPOSITION 10:

(i) Necessity: suppose that σ satisfies the incentive constraint (5). Let $\Delta_1 \subseteq R^2$ contain all pairs of $(y_1(a), M(s_1(a)))$ for $a \in A$ and $\Delta_2(h_2) \subseteq R^2$ contain all pairs of $(y_1(a), y_2(a, h_2))$ for $a \in A$. Define $\mathcal{T} = (T_1, T_2, X)$ by

$$T_1\left(y_1(a), M\left(s_1(a)\right)\right) = c_1(a) - y_1(a) - M\left(s_1(a)\right),$$

$$T_2\left(y_1(a), y_2(a, h_2), h_2\right) = c_2(a, h_2) - y_2(a, h_2),$$

$$X(k_1) = x^{**}(k_1).$$

For values of $(y_1, M_1) \in R^2 \backslash \Delta_1$, set the taxes T_1 high enough so that no agent chooses such values. Similarly, for all pairs $(y_1, y_2(h_2)) \in R^2 \backslash \Delta_2(h_2)$, set T_2 high enough to ensure that no one makes such a choice.

The tax system thus provides sufficient penalties for all choices y_1, s_1, and $y_2(h_2)$ that do not mimic the allocation of some other agent $\hat{a}$, $y_1(\hat{a})$, $s_1(\hat{a})$, and $y_2(\hat{a}, h_2)$, and one only needs to check that an agent with ability a

prefers $\sigma_{ys}(a)$ to $\sigma_{ys}(\hat{a})$ for any other $\hat{a}$, where $\sigma_{ys} = \{y_1(a), s_1(a), y_2(a, h_2)\}$ is a subset of σ. Let

$$\tilde{W}(k_1, \hat{a}|a) = \max_e \Bigg\{ U\big(c_1(\hat{a}) - k_1\big) - V\bigg(\frac{y_1(\hat{a})}{ah_1}, e\bigg) + \beta \int_H \bigg[U\big(c_2(\hat{a}, h_2) + x^{**}(k_1)\big) - V\bigg(\frac{y_2(\hat{a}, h_2)}{ah_2}, 0\bigg)\bigg] \times p\big(h_2 | f\big(e, s_1(\hat{a})\big)\big)\, dh_2 \Bigg\}$$

be the utility of an agent a if the agent chooses savings k_1 and reports $\hat{a}$. Note that $\hat{W}(k_1; x^{**}|a) = \max_{\hat{a}} \tilde{W}(k_1, \hat{a}|a) \geq \tilde{W}(k_1, \hat{a}|a)$. But then

$$\Omega\big(0, \sigma_{ys}(a)|a\big) = \tilde{W}\big(0, a|a\big) \geq \max_k \hat{W}\big(k; x^{**}|a\big) \geq \tilde{W}\big(0, \hat{a}|a\big) = \Omega\big(0, \sigma_{ys}(\hat{a})|a\big) \quad \forall \hat{a} \in A,$$

where the first inequality follows from (A-5), and the second inequality follows from the fact that $k = 0$ and $\hat{a}$ may not be a utility-maximizing choice. Both equalities follow from the budget constraints of the agent under the tax system as defined. Hence, $\{\sigma(a), 0\}$ solves the market problem for agent a.

(ii) Sufficiency: define, for any hours and schooling choice $\sigma_{ys} = \{y_1, y_2, s_1\}$ and for a given tax and subsidy system (T_1, T_2, X), the utility from such a choice in a market economy as

$$\Omega\big(k_1, \sigma_{ys}|a\big) = \max_e \Bigg\{ U\big[y_1 - T_1\big(y_1, M(s_1)\big) - M(s_1) - k_1\big] - V\bigg(\frac{y_1}{a h_1}, e\bigg) + \beta \int_H \bigg[U\big[X(k_1, y_1) + y_2 - T_2(y_1, y_2, h_2)\big] - V\bigg(\frac{y_2(h_2)}{a h_2}, 0\bigg)\bigg] \times p\big(h_2 | f(e, s_1)\big)\, dh_2 \Bigg\}.$$

Suppose that $\sigma(a)$ and $k_1 = 0$ solves the tax problem for agent a. Then

$$W\big(a, \tilde{e}_1(a|a)|a\big) = \Omega\big(0, \sigma_{ys}(a)|a\big) \geq \max_{k_1} \Omega\big(k_1, \sigma_{ys}(\hat{a})x|a\big) \geq \Omega\big(0, \sigma(\hat{a})|a\big) = W\big(\hat{a}, \tilde{e}_1(\hat{a}|a)|a\big).$$

The first and last equality follows from the budget constraints and the fact that $\tilde{e}_1(\hat{a}|a)$ maximizes utility given that agent a chooses allocation $\sigma(\hat{a})$. The first inequality follows from the fact that $\sigma_{ys}(a), 0$ maximizes agent a's utility. The second inequality follows from the fact that choosing zero savings cannot dominate choosing the best savings. ∎

REFERENCES

Ábrahám, Árpád, Sebastian Koehne, and Nicola Pavoni. 2016. "Optimal Income Taxation When Asset Taxation Is Limited." *Journal of Public Economics* 136: 14–29.

Aguiar, Mark, and Erik Hurst. 2013. "Deconstructing Life-Cycle Expenditure." *Journal of Political Economy* 121 (3): 437–92.

Albanesi, Stefania. 2007. "Optimal Taxation of Entrepreneurial Capital with Private Information." Columbia University Department of Economics Discussion Paper 0607-11.

Albanesi, Stefania, and Christopher Sleet. 2006. "Dynamic Optimal Taxation with Private Information." *Review of Economic Studies* 73 (1): 1–30.

Autor, David H., Lawrence F. Katz, and Alan B. Krueger. 1998. "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics* 113 (4): 1169–1213.

Battaglini, Marco, and Stephen Coate. 2008. "Pareto Efficient Income Taxation with Stochastic Abilities." *Journal of Public Economics* 92 (3–4): 844–68.

Ben-Porath, Yoram. 1967. "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy* 75 (4): 352–65.

Best, Michael Carlos, and Henrik Jacobsen Kleven. 2013. "Optimal Income Taxation with Career Effects of Work Effort." https://www.henrikkleven.com/uploads/3/7/3/1/37310663/best-kleven_careers_feb2013.pdf.

Bohacek, Radim, and Marek Kapicka. 2008. "Optimal Human Capital Policies." *Journal of Monetary Economics* 55 (1): 1–16.

Bovenberg, A. Lans, and Bas Jacobs. 2005. "Redistribution and Education Subsidies Are Siamese Twins." *Journal of Public Economics* 89 (11–12): 2005–35.

Browning, Martin, Lars Peter Hansen, and James J. Heckman. 1999. "Micro Data and General Equilibrium Models." In *Handbook of Macroeconomics*, Vol. 1, Part A., edited by John B. Taylor and Michael Woodford, 543–633. Amsterdam: North-Holland.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea 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.

Conlon, John R. 2009. "Supplement to 'Two New Conditions Supporting the First-Order Approach to Multisignal Principal-Agent Problems.'" *Econometrica* 77 (1): 249–78.

da Costa, Carlos E., and Lucas J. Maestri. 2007. "The Risk Properties of Human Capital and the Design of Government Policies." *European Economic Review* 51 (3): 695–713.

Diamond, Peter A., and James A. Mirrlees. 2002. "Optimal Taxation and the Le Chatelier Principle." MIT Department of Economics Working Paper 02–30.

Farhi, Emmanuel, and Iván Werning. 2007. "Inequality and Social Discounting." *Journal of Political Economy* 115 (3): 365–402.

Farhi, Emmanuel, and Iván Werning. 2013. "Insurance and Taxation over the Life Cycle." *Review of Economic Studies* 80 (2): 596–635.

Findeisen, Sebastian, and Dominik Sachs. 2016. "Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans." *Journal of Public Economics* 138: 1–21.

Findeisen, Sebastian, and Dominik Sachs. 2017. "Optimal Need-Based Financial Aid." www.dominiksachs.com/downloads/need_based.pdf.

Gary-Bobo, Robert J., and Alain Trannoy. 2015. "Optimal Student Loans and Graduate Tax under Moral Hazard and Adverse Selection." *RAND Journal of Economics* 46 (3): 546–76.

Golosov, Mikhail, Narayana Kocherlakota, and Aleh Tsyvinski. 2003. "Optimal Indirect and Capital Taxation." *Review of Economic Studies* 70: 569–87.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski. 2016. "Redistribution and Social Insurance." *American Economic Review* 106 (2): 359–86.

Gorry, Aspen, and Ezra Oberfield. 2012. "Optimal Taxation over the Life Cycle." *Review of Economic Dynamics* 15 (4): 551–72.

Grochulski, Borys, and Tomasz Piskorski. 2010. "Risky Human Capital and Deferred Capital Income Taxation." *Journal of Economic Theory* 145 (3): 908–43.

Guner, Nezih, Remzi Kaygusuz, and Gustavo Ventura. 2014. "Income Taxation of U.S. Households: Facts and Parametric Estimates." *Review of Economic Dynamics* 17 (4): 559–81.

Guvenen, Fatih. 2007. "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?" *American Economic Review* 97 (3): 687–712.

Hammond, Peter J. 1979. "Straightforward Individual Incentive Compatibility in Large Economies." *Review of Economic Studies* 46 (2): 263–82.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante. 2010. "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967–2006." *Review of Economic Dynamics* 13 (1): 15–51.

Huggett, Mark, Gustavo Ventura, and Amir Yaron. 2006. "Human Capital and Earnings Distribution Dynamics." *Journal of Monetary Economics* 53 (2): 265–90.

Huggett, Mark, Gustavo Ventura, and Amir Yaron. 2011. "Sources of Lifetime Inequality." *American Economic Review* 101 (7): 2923–54.

Jacobs, Bas, Dirk Schindler, and Hongyan Yang. 2012. "Optimal Taxation of Risky Human Capital." *Scandinavian Journal of Economics* 114 (3): 908–31.

Jewitt, Ian. 1988. "Justifying the First-Order Approach to Principal-Agent Problems." *Econometrica* 56 (5): 1177–90.

Kapička, Marek. 2006. "Optimal Taxation with Human Capital Accumulation and Limited Record Keeping." *Review of Economic Dynamics* 9 (4): 612–39.

Kapička, Marek. 2015. "Optimal Mirrleesean Taxation in a Ben-Porath Economy." *American Economic Journal: Macroeconomics* 7 (2): 219–48.

Kapička, Marek, and Julian Neira. 2019. "Optimal Taxation with Risky Human Capital: Dataset." *American Economic Journal: Macroeconomics*. https://doi.org/10.1257/mac.20160365.

Kocherlakota, Narayana R. 2005. "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation." *Econometrica* 73 (5): 1587–1621.

Koeniger, Winfried, and Julien Prat. 2018. "Human Capital and Optimal Redistribution." *Review of Economic Dynamics* 27: 1–26.

Krueger, Dirk, and Alexander Ludwig. 2016. "On the Optimal Provision of Social Insurance: Progressive Taxation versus Education Subsidies in General Equilibrium." *Journal of Monetary Economics* 77: 72–98.

Lillard, Lee A., and Yoram Weiss. 1979. "Components of Variation in Panel Earnings Data: American Scientists 1960–70." *Econometrica* 47 (2): 437–54.

MaCurdy, Thomas E. 1982. "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis." *Journal of Econometrics* 18 (1): 83–114.

Makris, Miltos, and Alessandro Pavan. 2018. "Taxation under Learning-by-Doing." faculty.wcas.northwestern.edu/~apa522/LBD.pdf.

McDaniel, Cara. 2007. "Average Tax Rates on Consumption, Investment, Labor and Capital in the OECD 1950–2003." paulklein.ca/newsite/teaching/mcdaniel_tax_2007.pdf.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38 (2): 175–208.

Mirrlees, James A. 1976. "Optimal Tax Theory: A Synthesis." *Journal of Public Economics* 6 (4): 327–58.

OECD. 2018. *Education at a Glance 2018: OECD Indicators*. Paris: OECD.

Peterman, William. 2016. "The Effect of Endogenous Human Capital Accumulation on Optimal Taxation." *Review of Economic Dynamics* 21: 46–71.

Primiceri, Giorgio E., and Thijs van Rens. 2009. "Heterogeneous Life-Cycle Profiles, Income Risk and Consumption Inequality." *Journal of Monetary Economics* 56 (1): 20–39.

Rochet, J.C. 1985. "The Taxation Principle and Multi-time Hamilton-Jacobi Equations." *Journal of Mathematical Economics* 14 (2): 113–28.

Rupert, Peter, and Giulio Zanella. 2015. "Revisiting Wage, Earnings, and Hours Profiles." *Journal of Monetary Economics* 72: 114–30.

Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies* 68 (1): 205–29.

Schultz, Theodore W. 1961. "Investment in Human Capital." *American Economic Review* 51 (1): 1–17.

Shourideh, Ali. 2014. "Optimal Taxation of Wealthy Individuals." http://www.shourideh.com/uploads/3/6/8/3/3683044/captaxent.pdf.

Slavík, Ctirad, and Hakki Yazici. 2014. "Machines, Buildings, and Optimal Dynamic Taxes." *Journal of Monetary Economics* 66: 47–61.

Stantcheva, Stefanie. 2015. "Learning and (or) Doing: Human Capital Investments and Optimal Taxation." NBER Working Paper 21381.

Stantcheva, Stefanie. 2017. "Optimal Taxation and Human Capital Policies over the Life Cycle." *Journal of Political Economy* 125 (6): 1931–90.

Stiglitz, Joseph E. 1982. "Self-Selection and Pareto Efficient Taxation." *Journal of Public Economics* 17 (2): 213–40.

Storesletten, Kjetil, Christopher I. Telmer, and Amir Yaron. 2004. "Consumption and Risk Sharing over the Life Cycle." *Journal of Monetary Economics* 51 (3): 609–33.

Werning, Iván. 2007. "Optimal Fiscal Policy with Redistribution." *Quarterly Journal of Economics* 122 (3): 925–67.

Werning, Iván. 2011. "Nonlinear Capital Taxation." Unpublished.