DISCUSSION PAPER SERIES

IZA DP No. 8666

Human Capital and Optimal Redistribution

Winfried Koeniger
Julien Prat

November 2014

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Human Capital and Optimal Redistribution

Winfried Koeniger
University of St.Gallen (SEW-HSG), CEPR, CFS and IZA

Julien Prat
CNRS (CREST), CEPR, CESifo and IZA

Discussion Paper No. 8666
November 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 8666
November 2014

ABSTRACT

Human Capital and Optimal Redistribution*

We characterize optimal redistribution in a dynastic family model with human capital. We show how a government can improve the trade-off between equality and incentives by changing the amount of observable human capital. We provide an intuitive decomposition for the wedge between human-capital investment in the laissez faire and the social optimum. This wedge differs from the wedge for bequests because human capital carries risk: its returns depend on the non-diversifiable risk of children's ability. Thus, human capital investment is encouraged more than bequests in the social optimum if human capital is a bad hedge for consumption risk.

JEL Classification: E24, H21, I22, J24

Keywords: human capital, optimal taxation

Corresponding author:

Winfried Koeniger
Swiss Institute for Empirical Economic Research
University of St.Gallen
Varnbüelstr. 14
CH-9000 St.Gallen
Switzerland
E-mail: winfried.koeniger@unisg.ch

* We are grateful to Emmanuel Farhi and Iván Werning for making the code used in Farhi and Werning (2013) available, and to Carlo Zanella for excellent research assistance. We thank Robert Gary-Bobo, Marek Kapička, Etienne Lehmann, Jean-Baptiste Michau, Dominik Sachs, Josef Zweimüller as well as participants in various seminars for very helpful comments. Part of this research has been conducted while Winfried Koeniger was at Queen Mary, University of London; and Julien Prat was at the Institute for Economic Analysis (CSIC).

1 Introduction

Of all the factors shaping inequality, one of the most debated is the transmission of physical and human capital from parents to their offspring. As frequently argued, children from a privileged background get a head start that is difficult to reconcile with the provision of equal opportunity. Yet, eliminating inequality in inherited physical and human capital would be counterproductive since it removes the motivation of parents to provide their children with wealth and education. The optimal taxation of intergenerational transfers is therefore determined by the classic trade-off between insurance and incentives.

Mirrlees' (1971) seminal contribution on optimal income taxation provides a rigorous framework to analyze this trade-off. Mirrlees showed that asymmetric information about labor market productivity prevents full insurance because productive agents then would not find it optimal to reveal their true ability. We build on Mirrlees' insight, and the subsequent literature on optimal taxation, to analyze optimal redistribution in a model with altruistic dynasties. Each working-age generation of a dynasty decides how much labor effort to exert, how much to consume, to bequeath in terms of bonds and to invest into human capital of their offspring. Bequests and human capital are observable but the innate ability of each generation is private information. The ability of children is uncertain when parents make their decisions but may depend on parents' ability.

We show how taxes on labor income and bequests distort human capital investment. Thus, education and tax policies need to be jointly determined. Following the optimal taxation literature, we use the wedges between the laissez faire and the social optimum to characterize the implicit taxes or subsidies required to attain the social optimum.

We decompose the gross human capital wedge into three components. The first two components capture how the planner offsets the distortions at the intra- and intertemporal margin introduced by the wedges for labor effort and bequests. These wedges would otherwise induce suboptimal human capital investment. The third component can be interpreted as the net wedge since it isolates the part of the human capital wedge that does not result from optimal distortions of labor supply or bequests. We show that this net wedge is proportional to the effect that human capital has on the incentive-compatibility constraint. This makes explicit that the planner can use human capital investment to improve the trade-off between equality and incentives. Under the empirically plausible condition that the elasticity of substitution between human capital and innate ability is larger than $1/4$, more human capital reduces the disutility of labor and lowers the informational rents of productive families. Through this channel, human capital investments today allow the planner to mitigate the incentive problem and thus to achieve more equality in the future.

We establish that, at the social optimum, the net wedge for human capital exactly offsets the part of the gross human capital wedge that is related to the labor wedge. Thus, marginal perturbations of human capital at the social optimum do not affect incentive compatibility of the allocation so that the marginal cost of human capital investment equals the discounted value of its expected marginal returns. This allows us to derive a strikingly simple expression for the constrained-efficient wedge for human capital which shows that it is closely related, but not identical, to the wedge for bequests. The similarity is intuitive because parents can substitute physical with human capital when they transfer resources to their offspring. The two wedges are not identical, however, because the productivity of children is uncertain and parents cannot diversify this risk. The additional source of uncertainty associated with human capital discourages families' human capital investment as it provides a perverse hedge against consumption risk. It then follows that the planner should encourage human capital investment more than bequests.

We discuss the practical implications of these theoretical results, focussing on the case in which innate ability is uncorrelated across generations. We emphasize important features of taxes that allow to implement the socially optimal allocation and solve the model numerically. This allows us not only to confirm the insights discussed above but also to establish that the socially-optimal human capital investment into children should be decreasing in parents' ability. This striking result is explained by a wealth effect. In the constrained-efficient allocation without full insurance, children from a privileged background inherit larger bequests. The induced wealth effect reduces their labor supply so that it becomes relatively less efficient for the planner to invest into their human capital.

Related literature.—Our paper relates to the two large literatures on human capital and optimal taxation. For brevity we refer to the review of that literature in Stantcheva (2014a) and focus only on a number of recent contributions which allow us to put our main findings into context.

While the wedges for labor supply and bequests in our model correspond to previous findings in the literature (Farhi and Werning, 2013, Golosov et al. 2011, Kapička, 2013, Kocherlakota, 2010, Saez, 2001, and references therein), the wedge for human capital provides novel insights to the best of our knowledge. Compared with Farhi and Werning (2010), we find that the constrained efficient wedges for bequests and human capital are not identical since human capital investment carries more risk than bequests.

Our results on the human capital wedge in an intergenerational model relate to research on optimal redistribution and human capital accumulation over the life cycle. Findeisen and Sachs (2012) and Gary-Bobo and Trannoy (2014) analyze optimal student-loan contracts in asymmetric-information models with two periods. They show that the socially optimal allocation can be decentralized with student loans that have income-contingent repayment schedules. From a technical point of view, Findeisen and Sachs (2012) also use the generalized envelope condition derived by Kapička (2013) and Pavan et al. (2014) to characterize social optimality by the planner's first-order conditions.

Stantcheva (2014a) extends the analysis of Findeisen and Sachs (2012) to a multi-period setting in a more general model with training time and possibly unobservable human capital, building on the analyses of Kapička (2006) and Boháček and Kapička (2008). She proposes a decomposition of the human capital wedge that is similar, but not identical, to ours. In particular, Stantcheva (2014a) finds that human capital has a positive incentive effect if the elasticity of substitution between human capital and innate ability is larger than unity while we find that the elasticity only has to exceed $1/4$, for an empirically plausible value of 0.5 for the Frisch elasticity of labor supply. The explanation for this different finding is that we do not hold labor supply constant when deriving the effect of human capital on the incentive compatibility constraint. Formally, in our specification of the allocation problem, the planner chooses output while the planner in Stantcheva (2014a) chooses unobservable labor effort. In accordance with the revelation principle, the overall wedges do not depend on the choice of control variable but their decomposition and interpretation differs. Our decomposition makes transparent how to simplify the human capital wedge further, thereby highlighting its close relationship with the intertemporal wedge for bequests.

Besides these specific differences, the results in our model have a different interpretation because we are focusing on dynastic families. This relates our analysis to recent papers on optimal redistribution across generations. Gelber and Weinzierl (2014) analyze optimal taxation if the ability of future generations depends on the resources of the current generation. This is modelled by letting the probability of types directly depend on disposable income. Our model shares the feature that current resources may impact the earnings capacity of future generations but lets generations choose the amount of resources allocated to human capital accumulation. This allows us to analyze whether that choice is constrained efficient. Our assumptions of observable human capital (think of high-school or college degrees) and stochastic unobserved ability allow us to characterize the wedge for human capital when ability is not perfectly predictable across generations.

The complementary research by Erosa and Koreshkova (2007), Krueger and Ludwig (2013), Lee and Seshadri (2014) and Stantcheva (2014b) does not use the Mirrlees approach to analyze the effect of redistribution in models with human capital accumulation. Following the Ramsey approach, they specify parametric tax schedules and then analyze the welfare effects of changes in taxes.

Finally, our finding that the planner can improve the equality-efficiency trade-off over time by investing into human capital is akin to the economic mechanism in Koehne and Kuhn's (2014) model with habits or durable consumption. In their paper, the planner can exploit complementarities between durable and non-durable consumption choices over time to raise the marginal utility of non-durable consumption and thus the incentive to exert labor effort. Our paper emphasizes how human capital investment reduces the disutility of labor of future generations, makes the consumption of leisure less attractive and thus strengthens the power of incentives.

The rest of the paper is structured as follows. In Section 2 we describe the model set-up and solve the planner's problem. In Section 3 we derive the optimality conditions in the laissez faire and then characterize the wedges between the laissez faire and the social optimum. We discuss implementation of the constrained-efficient allocation in Section 4 and present the numerical solution for a calibrated version of the model in Section 5.

2 The model

Family dynasties are the decision units of our analysis. Each family is composed of parents and children in each generation and has a planning horizon equal to T. The family chooses the labor supply of the parents, as well as the bequests and education for the children. Preferences link generations in a time separable fashion. We make the common assumption that the per-period utility function $\mathbf{U}(c_t, l_t)$ is separable in consumption c_t and labor effort l_t:

$$\begin{aligned}
&[\mathbf{A1}]\text{: } \mathbf{U}(c_t, l_t) = u(c_t) - \mathbf{v}(l_t), \\
&\quad u(c_t) \in \mathcal{C}^2\left(\mathbb{R}^+\right) \text{ is increasing in } c_t \text{ and strictly concave,} \\
&\quad \mathbf{v}(l_t) \in \mathcal{C}^2\left(\mathbb{R}^+\right) \text{ is increasing in } l_t \text{ and strictly convex.}
\end{aligned}$$

As in the seminal paper of Mirrlees (1971), agents differ in their ability θ_t which cannot be observed by the planner. Both bequests b_t and human capital h_t are instead public knowledge. Output y_t is produced with technology $Y(h_t, l_t, \theta_t)$ which is increasing in its arguments and concave. We will use the production function to substitute l_t in the utility function and write $U(c_t, y_t, h_t, \theta_t)$ instead of $\mathbf{U}(c_t, l_t)$ or, with assumption **[A1]**, $v(y_t, \theta_t, h_t) = \mathbf{v}(l_t)$. Note that the planner cannot use observable output y_t to infer actual labor supply l_t because ability θ_t is stochastic and hidden.

In the spirit of Ben-Porath (1967), human capital in the next period h_{t+1} depends on the expenditure flow for education e_t and on the family background, which can be summarized by the stock of human capital of parents h_t.[1] The human capital production function $h_{t+1}(e_t, h_t)$ is increasing in its arguments and concave.[2]

The timing in the model is as follows. In any given period t, the family learns the parents' type θ_t and chooses to spend e_t on the children's human capital h_{t+1}, to supply parents' labor l_t, to consume c_t and thus bequeath b_{t+1}. We assume that abilities are uncorrelated across generations with types being drawn at the beginning of each period from a stationary distribution $F : \Theta \rightarrow [0, 1]$ over the fixed support $\Theta \equiv [\underline{\theta}, \overline{\theta}]$ with $\underline{\theta} > 0$. This assumption simplifies the analytic results without changing the main insights that human capital relaxes the incentive compatibility constraints and that the wedges of human capital and bequests are tightly related. We briefly discuss the extension of our model to persistent ability shocks in Section 3.2 and delegate the presentation of the results for this case to appendix A.3.

2.1 The planner's problem

According to the revelation principle, we can solve the planner's problem by focusing on a direct mechanism such that families truthfully report their types in each generation. Let $\theta^t \equiv \{\theta_0, \theta_1, ..., \theta_t\}$ denote the history of types within a given family. We do not impose any

[1] Human capital investment affects productivity in the next period (for the next generation) and not in the current period as in Stantcheva (2014a). This difference arises from the fact that Stantcheva analyzes human capital investment of individuals over the life cycle while we focus on human capital investment of parents into their children.

[2] We abstract from time use for human capital investments into children because the time effort exerted for human capital accumulation is plausibly as unobservable as is the time effort for production. Adding a second hidden action renders the analysis much less tractable because it requires ruling out joint deviations.

arbitrary restrictions on the allocation. In particular, we do not rule out history dependent allocations summarized by $\mathbf{x} \equiv \{x_t(\theta^t)\}_{t=0}^{T}$ where $x_t(\theta^t) \equiv \{c_t(\theta^t), h_{t+1}(\theta^t), y_t(\theta^t)\}$. The family's preferences over an allocation $\mathbf{x}$ are given by

$$\mathcal{U}(\mathbf{x}) \equiv \mathbb{E}_0 \left[\sum_{t=0}^{T} \beta^t \widetilde{U}\left(x_t\left(\theta^t\right), \theta_t\right) \right],$$

where $\mathbb{E}_0$ is the expectation operator conditional on information available at time 0 and β is the discount factor measuring the strength of the altruism towards future generations.

In general, families do not have to behave truthfully. They choose the *reporting strategy* $\mathbf{r} \equiv \{r_t(\theta^t)\}_{t=0}^{T}$ from the set $\mathcal{R}$ of feasible reports which maximizes their expected utility. Since types are private information, an allocation must be *incentive compatible*, i.e.,

$$\mathcal{U}(\mathbf{x}) \geq \mathcal{U}(\mathbf{x} \circ \mathbf{r}), \text{ for all } \mathbf{r} \in \mathcal{R}, \tag{1}$$

where $(\mathbf{x} \circ \mathbf{r})(\theta^t) \equiv \{x_t(r^t(\theta^t))\}_{t=0}^{T}$ is the allocation resulting from the reporting strategy $\mathbf{r}$ and history θ^t.

The planner discounts future utility with the factor q which equals the inverse interest factor.[3] As Farhi and Werning (2013), we abstract from feedbacks between choices of families due to equilibrium price effects so that the allocation problem can be analyzed separately for each family. Let $\mathcal{X}$ be the set of all feasible allocation. Cost minimization along the equilibrium path is achieved when an allocation solves the objective function

$$\min_{\mathbf{x} \in \mathcal{X}} \Pi(\mathbf{x}) \equiv \mathbb{E}_0 \left[\sum_{t=0}^{T} q^t \left[c_t\left(\theta^t\right) + e_t\left(\theta^t\right) - y_t\left(\theta^t\right) \right] \right],$$

subject to the incentive compatibility constraint (1), and to the promise keeping constraint $\mathcal{U}(\mathbf{x}) \geq \omega_0$ which ensures that the expected utility of truthful families is at least as high as the exogenously given level ω_0.

Recursive formulation.—Instead of directly solving the problem above, we apply two common modifications that simplify the analysis considerably. First, we write the planner's problem in recursive form. As shown by Abreu et al. (1990), when ability θ follows an i.i.d. process, we do not need to condition allocations on the entire history of reports but only on the realization of the equilibrium continuation value

$$\omega\left(\theta^t\right) \equiv \widetilde{U}\left(x_t\left(\theta^t\right), \theta_t\right) + \beta \int_{\Theta} \omega\left(\theta^t, \theta_{t+1}\right) dF\left(\theta_{t+1}\right) .$$

[3] We assume that the planner maximizes the welfare of the initial dynasty as in the infinite-horizon setting of Atkeson and Lucas (1992). See Farhi and Werning (2007, 2010) and Kocherlakota (2010), chapter 5, for analyses in which the planner may give additional weight to future generations. As shown in Farhi and Werning (2010), section IV.C, this generates a motive to subsidize education even when the effect of human capital on the labor supply of the next generation is ignored. We deliberately abstract from this effect to focus on the effect of human capital on incentives resulting from changes in labor supply.

At the beginning of each period, families compare the continuation value $\omega\left(\theta^t\right)$ of truthful reporting to those derived from arbitrary reporting strategies

$$\omega^{\mathbf{r}}\left(\theta^t\right) \equiv \widetilde{U}\left(x_t\left(r^t\left(\theta^t\right)\right), \theta_t\right) + \beta \int_\Theta \omega^{\mathbf{r}}\left(\theta^t, \theta_{t+1}\right) dF\left(\theta_{t+1}\right) \ .$$

Incentive compatibility is ensured when $\omega\left(\theta^t\right) \geq \omega^{\mathbf{r}}\left(\theta^t\right)$ for all θ^t and all $\mathbf{r} \in \mathcal{R}$.[4] Instead of considering all feasible reports, we focus on marginal deviations from the truth. In other words, we use a first-order approach. We replace the general incentive constraint by an envelope condition that is valid on the equilibrium path on which families truthfully reveal their types.[5] The recursive form of this relaxed planning problem reads[6]

$$\Gamma\left(V, h, t\right) = \min_{\{c,y,h',V'\}} \left\{ \int_\Theta \left[c\left(\theta\right) + g(h'(\theta), h) - y\left(\theta\right) + q\Gamma\left(V'\left(\theta\right), h'(\theta), t+1\right)\right] dF\left(\theta\right) \right\}$$

$$\text{s.t. } \omega\left(\theta\right) = U\left(c\left(\theta\right), y\left(\theta\right), \theta, h\right) + \beta V'\left(\theta\right), \tag{2}$$

$$V = \int_\Theta \omega\left(\theta\right) dF\left(\theta\right), \tag{3}$$

$$\frac{\partial \omega\left(\theta\right)}{\partial \theta} = \frac{\partial U\left(c, y, \theta, h\right)}{\partial \theta}, \tag{4}$$

where we have inverted the human capital accumulation function $h'(e,h)$ to substitute $e(\theta)$ with $g(h'(\theta), h)$. Note that costly human capital accumulation implies $\partial g(h', h)/\partial h' > 0$ and $\partial g(h', h)/\partial h < 0$ if costs are smaller for parents with more human capital.

The first constraint defines the continuation value $\omega\left(\theta\right)$ as the sum of the current and next period promised utilities, $U\left(\cdot\right)$ and $V'\left(\theta\right)$ respectively. Equation (3) is the promise-keeping constraint since it ensures that the expected value of the continuation utility is equal to the promised value V. The last equation is the local incentive-compatibility constraint captured by the envelope condition which is derived assuming that the first-order condition for truthful reporting is satisfied.[7] Condition (4) is necessary but not sufficient.

[4] Note that this imposes incentive compatibility for all $\theta^t \in \Theta^t$. Thus we now require truth telling to be optimal after any history of shocks, whereas the incentive constraint (1) only requires truth telling to be ex-ante optimal. But the difference is immaterial to our analysis because the two notions can only differ on a set of measure zero histories. In other words, allocations that are ex-ante incentive compatible are also ex-post incentive compatible almost everywhere.

[5] To shorten the exposition, we do not explicitly derive the recursive formulation from first principles. We refer readers interested in the validity of the first-order approach to Kapička (2013) for an in-depth discussion of the intermediate steps and technical issues.

[6] To simplify the notation, we only keep a time index for the value function, otherwise we drop the indexes and use a prime $'$ to denote the next period.

[7] Totally differentiating the continuation value of a truthful family yields

$$\frac{\partial \omega\left(\theta\right)}{\partial \theta} = \left.\frac{\partial U\left(c\left(r\right), y\left(r\right), \theta, h\right)}{\partial \theta}\right|_{r=\theta} + \left.\frac{\partial U\left(c\left(r\right), y\left(r\right), \theta, h\right)}{\partial r}\right|_{r=\theta} + \beta \left.\frac{\partial V'\left(r\right)}{\partial r}\right|_{r=\theta} .$$

The local optimality condition is equivalent to (4) because the sum of the last two terms on the right hand side equals zero when the first-order condition for truthful reporting is satisfied.

However, ability θ is i.i.d. and preferences satisfy the single-crossing condition: under assumption **[A1],** $\partial^2 U(\cdot)/\partial\theta\partial y = -\partial^2 v(\cdot)/\partial\theta\partial y > 0$. Thus, the first-order approach is valid when the allocation is monotone in ability, a requirement that can easily be verified ex post.[8]

2.2 Optimality conditions

In the first best environment without information asymmetries, $\mu(\theta) = 0$ for all θ and agents are fully insured against changes in ability. Consumption remains constant across families and is therefore separated from production. With information asymmetries instead, the planner faces an insurance-incentive trade-off whose optimal resolution is determined by the following conditions.

Proposition 1 *If [**A1**] holds, the first-order conditions of the planner problem are*

$$V'(\theta) : \left[-\frac{\beta}{\frac{\partial u(c(\theta))}{\partial c(\theta)}} + q\lambda'(\theta) \right] f(\theta) = 0, \tag{5}$$

$$\begin{aligned} h'(\theta) : \frac{\partial g(h'(\theta), h)}{\partial h'(\theta)} &+ q \int_{\Theta} \left(\frac{\frac{\partial v(y'(\theta'),\theta',h'(\theta))}{\partial h'(\theta)}}{\frac{\partial u(c'(\theta'))}{\partial c'(\theta')}} + \frac{\partial g(h''(\theta'), h'(\theta))}{\partial h'(\theta)} \right) dF(\theta') \\ &- q \int_{\Theta} \mu'(\theta') \frac{\partial^2 v(y'(\theta'), \theta', h'(\theta))}{\partial\theta'\partial h'(\theta)} d\theta' = 0, \end{aligned} \tag{6}$$

$$y(\theta) : \left[\frac{\frac{\partial v(y(\theta),\theta,h)}{\partial y(\theta)}}{\frac{\partial u(c(\theta))}{\partial c(\theta)}} - 1 \right] f(\theta) - \frac{\partial^2 v(y(\theta), \theta, h)}{\partial\theta\partial y(\theta)} \mu(\theta) = 0, \tag{7}$$

with

$$\mu(\theta) = \int_{\underline{\theta}}^{\theta} \left[\lambda - \frac{1}{\partial u(c(x))/\partial c(x)} \right] dF(x), \ \textit{and} \ \lim_{\theta\to\underline{\theta}} \mu(\theta) = \lim_{\theta\to\overline{\theta}} \mu(\theta) = 0. \tag{8}$$

Consumption and Output.—Equation (5) implies that the reciprocal Euler equation continues to hold in our model with human capital. To see why, note that evaluating the law of motion (8) of the costate variable at the upper bound of the ability distribution yields

$$\lambda - \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial c(\theta)}{\partial\omega(\theta)} dF(\theta) = \mu(\overline{\theta}) = 0.$$

[8] See example 1 in Battaglini and Lamba (2014) with discrete types as in any numerical approximation. For continuous ability types and persistent shocks to ability, see Kapička (2013) and Pavan et al. (2014), or the discussion in Farhi and Werning (2013).

Using that $\partial c(\theta)/\partial\omega(\theta) = [\partial u(c(\theta))/\partial c(\theta)]^{-1}$ and leading this equation one period ahead, we find that $\lambda'(\theta) = \mathbb{E}\left[[\partial u(c'(\theta))/\partial c'(\theta)]^{-1}\right]$. Thus the *reciprocal Euler equation*

$$\frac{1}{\frac{\partial u(c(\theta))}{\partial c(\theta)}} = \frac{q}{\beta}\mathbb{E}\left[\frac{1}{\frac{\partial u(c'(\theta'))}{\partial c'(\theta')}}\right]$$

holds and the inverse of the marginal utility of consumption follows a martingale when $q = \beta$.

The condition for optimal production (7) is analogous to the optimality condition in the standard Mirrlees problem. Thus, we postpone its analysis to the next section where we characterize the constrained efficient wedges.

Human capital.—Turning our attention to education, let us repeat the optimality condition

$$\frac{\partial g(h',h)}{\partial h'} = -q\int_{\Theta}\left(\frac{\frac{\partial v(y'(\theta'),\theta',h'(\theta))}{\partial h'(\theta)}}{\frac{\partial u(c'(\theta'))}{\partial c'(\theta')}} + \frac{\partial g(h'',h')}{\partial h'}\right)dF(\theta') + q\int_{\Theta}\mu'(\theta')\frac{\partial^2 v(y',\theta',h')}{\partial\theta'\partial h'}d\theta'. \tag{9}$$

The marginal cost of human capital investment on the left hand side are equated to the marginal benefit. The latter is made of three components. Firstly, human capital lowers the disutility of labor to produce a given quantity of output. This allows the planner to spend less on consumption and still provide the family with the same continuation value.[9] Secondly, when education costs vary with the family background so that $\partial g(h'',h')/\partial h' < 0$, more human capital investment reduces the cost of accumulating human capital for the next generation. Thirdly, the second integral on the right hand side captures how human capital affects the incentive compatibility constraint. This term is central to our analysis so that we elaborate on it.

In the absence of informational frictions, families are perfectly insured against transitory shocks to ability so that $\partial\omega(\theta)/\partial\theta = 0$. With hidden ability types instead, information revelation is profitable solely if

$$\frac{\partial\omega(\theta)}{\partial\theta} = \frac{\partial U(c,y,\theta,h)}{\partial\theta} = -\frac{\partial v(y,\theta,h)}{\partial\theta} > 0,$$

where the inequality follows under the assumption that higher ability reduces the disutility of effort, i.e., $\partial v(\cdot)/\partial\theta < 0$. Incentive compatibility prevents full insurance: children with more able parents enjoy higher lifetime utilities. An increase in the slope $|\partial v(\cdot)/\partial\theta|$ of the disutility term widens the gap separating the constrained-efficient allocation from the first best. Hence, the cross-derivative $\partial^2 v(\cdot)/(\partial\theta\partial h)$ measures the effect that human capital has on the incentive compatibility constraint: if $\partial^2 v(\cdot)/(\partial\theta\partial h) > 0$, more human capital reduces informational rents and mitigates the incentive problem.

[9] As shown in the proof of Proposition 1 in appendix A.1, this benefit for the planner is captured by $-\frac{\partial v\left(y'(\theta'),\theta',h'(\theta)\right)}{\partial h'(\theta)}/\frac{\partial u\left(c'(\theta')\right)}{\partial c'(\theta')} > 0$.

These gains can be translated into consumption units through multiplication by the costate variable $\mu'(\theta')$ which measures the marginal cost of violating the incentive constraint. The resulting products in (9) are integrated over all potential realizations of θ' because neither the planner nor the family know the value of θ' when the human-capital investment is made.[10]

The sign of the cross derivative $\partial^2 v(\cdot)/(\partial\theta\partial h)$ is determined by: (i) the Frisch elasticity of labor supply and, (ii) the degree of complementarity between human capital and ability. Both are captured by a single parameter if we assume that the disutility of labor and the production function for output have the following functional forms.

Corollary 1 *Assume that*

$$\textbf{[A1']:}\quad \mathbf{U}(c,l) = u(c) - \mathbf{v}(l)\,,\ \textit{where } \mathbf{v}(l) = \zeta l^{\alpha},\ \textit{with } \zeta > 0 \textit{ and } \alpha > 1,$$

$$\textbf{[A2]:}\quad Y(h,l,\theta) = A(\theta,h)\,l,\quad \textit{with } A(\theta,h) = \left[\xi\theta^{\chi} + (1-\xi)\,h^{\chi}\right]^{1/\chi},\ \chi \in (-\infty, 1] \textit{ and } \xi \in (0,1)\,.$$

Then $\partial^2 v(y,\theta,h)/(\partial\theta\partial h) \geq 0$ *if and only if* $\chi \geq -\alpha$.

The Frisch elasticity of labor supply is $1/(\alpha-1)$ and the degree of complementarity is measured by the parameter χ. If the production function is Cobb Douglas, $\chi = 0$. Hence, negative χ imply more complementarity between ability and human capital than in the Cobb-Douglas case. Corollary 1 shows that informational rents are *decreasing* in human capital when the sign of $\chi + \alpha$ is positive: that is when the parameter α, which is inversely related to the Frisch elasticity of labor supply, is greater than the degree of complementarity χ between ability and human capital.

This result is illustrated in Figure 1 which plots the disutility $v(\cdot)$ as a function of supplied labor. The vertical lines display the values of $l = \hat{y}/A(\theta,h)$ resulting from different combinations of θ and h that are consistent with a given level of output $\hat{y}$. The differences reported on the vertical axis measure the values of $\Delta v(h_i) \equiv v(\hat{y},\theta_2,h_i) - v(\hat{y},\theta_1,h_i) < 0$, with $\theta_2 > \theta_1$. The derivative of the disutility of labor with respect to ability is given by $\partial v(\hat{y},\theta_2,h_i)/\partial\theta = \lim_{\theta_1\to\theta_2} \Delta v(h_i)/(\theta_2-\theta_1)$ and the cross derivative $\partial^2 v(\hat{y},\theta_2,h_2)/(\partial\theta\partial h)$ is of the same sign as $\Delta v(h_2) - \Delta v(h_1)$ when both $\theta_1 \to \theta_2$ and $h_1 \to h_2$. Note that $\Delta v(h_1) < \Delta v(h_2) < 0$ in Figure 1 so that $\Delta v(h_2) - \Delta v(h_1) > 0$ for $h_2 > h_1$, illustrating the result in Corollary 1.

We now discuss the intuition for this result. Note that an increase in h shifts the vertical lines to the left in Figure 1: more human capital means that a given unit of output $\hat{y}$ can be produced with less labor. Since the disutility of labor is convex, it is less sensitive to changes in l for smaller l. The size of this effect is proportional to the convexity of $v(\cdot)$ and thus increasing in the elasticity parameter α.

Although human capital lowers the informational rents by reducing the amount of labor effort required to produce a given unit of output, this effect may be offset by the fact that human capital also affects the sensitivity of labor supply with respect to changes in

[10]Note, by defintion (8), the costate variable $\mu'(\theta')$ also captures the probability weight for each type.

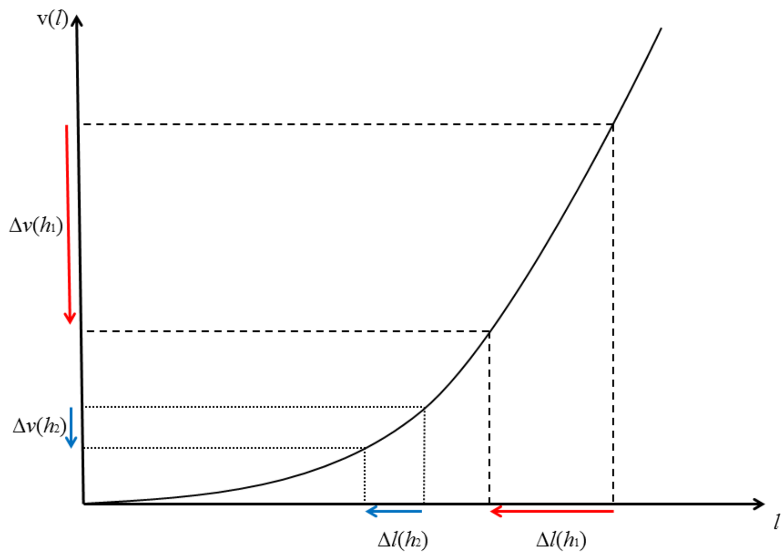

Figure 1: Example for $\partial^2 v(\hat{y}, \theta, h) / (\partial\theta\partial h) > 0$.

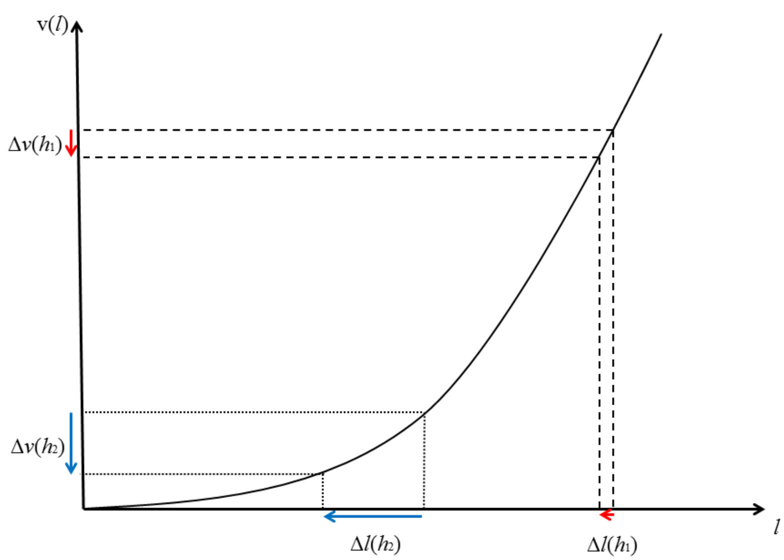

Figure 2: Example for $\partial^2 v(\hat{y}, \theta, h) / (\partial\theta\partial h) < 0$.

ability θ. If human capital and ability are strongly complementary for labor productivity, with more human capital h_2 much less labor is needed to produce a given output $\hat{y}$ if ability increases from θ_1 to θ_2. As shown in Figure 2, this means that the interval $\Delta l(h_i) \equiv l(\hat{y}, \theta_2, h_i) - l(\hat{y}, \theta_1, h_i)$ reported on the horizontal axis may be much larger at h_2 than at h_1, for $h_2 > h_1$. The figure shows how a sufficiently large difference between $\Delta l(h_2)$ and $\Delta l(h_1)$ can offset the labor supply effect discussed in the previous paragraph. Thus, $\partial^2 v(y, \theta, h)/(\partial\theta\partial h) > 0$ only if human capital is not too complementary to innate ability. Then a marginal increase in human capital relaxes the incentive compatibility constraint.

3 The wedges

We now compare the optimality conditions in the laissez faire to those for the constrained-efficient allocation derived in the previous section. The wedges between these conditions characterize the implicit taxes or subsidies which are necessary to attain the social optimum.

In the laissez faire each family solves the maximization problem

$$\begin{aligned} W(\theta, b, h, t) = \max_{\{b', h', l\}} & \left\{ \mathbf{U}(c, l) + \beta \int_\Theta W(\theta', b', h', t+1)\, dF(\theta') \right\} \\ \text{s.t. } b' &= (1+r)b - c - e + y, \\ y &= Y(h, \theta, l), \\ h' &= h'(e, h) \text{ so that } e = g(h', h), \end{aligned}$$

where b is the bequest. Below we extend this problem by introducing a borrowing constraint that restricts assets, and thus bequests, of families to be non-negative. For clarity we first derive results abstracting from such a constraint.

Proposition 2 *The laissez faire is characterized by the following first-order conditions for bequests, human capital and labor supply:*

$$\begin{aligned} \frac{\partial \mathbf{U}(c, l)}{\partial c} &= \beta(1+r)\mathbb{E}\left[\frac{\partial \mathbf{U}(c', l')}{\partial c'}\right], \\ \frac{\partial g(h', h)}{\partial h'}\frac{\partial \mathbf{U}(c, l)}{\partial c} &= \beta\mathbb{E}\left[\left(\frac{\partial y'}{\partial h'} - \frac{\partial g(h'', h')}{\partial h'}\right)\frac{\partial \mathbf{U}(c', l')}{\partial c'}\right], \\ -\frac{\partial \mathbf{U}(c, l)}{\partial l} &= \frac{\partial y}{\partial l}\frac{\partial \mathbf{U}(c, l)}{\partial c}. \end{aligned}$$

We assume preferences and technologies for production and human capital accumulation such that the conditions in Proposition 2 are necessary and sufficient.[11] Then the

[11] Note that human capital is chosen for the next generation (current human capital is a state variable) and thus does not imply a direct intratemporal substitution effect for the labor supply of the current generation. This timing assumption, which is plausible in our setting with families who invest into the education of their children, avoids the potential non-concavities discussed in Bovenberg and Jacobs (2005), Section 2.2.

results of Propositions 1 and 2 can be combined to derive interpretable conditions for the wedges between the choices in the laissez faire and the constrained-efficient allocation of the planner. We start with the following definition.

Definition 1 *The wedges for bequests* τ_b*, labor supply* τ_l *and human capital* τ_h *are*

$$\tau_b\left(\theta^t\right) \equiv 1 - \frac{q}{\beta} \frac{\partial u\left(c\right)/\partial c}{\mathbb{E}\left[\partial u\left(c'\right)/\partial c'\right]}, \tag{10}$$

$$\tau_l\left(\theta^t\right) \equiv 1 - \frac{\partial v(y,\theta,h)/\partial y}{\partial u\left(c\right)/\partial c}, \tag{11}$$

$$\tau_h\left(\theta^t\right) \equiv \frac{\beta}{\frac{\partial g(h',h)}{\partial h'}} \int_\Theta \left[\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} \left(\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right)\right] dF\left(\theta'\right) - 1. \tag{12}$$

Wedges are defined as the deviations from the laissez faire. In general, the wedges depend on the whole history of shocks since the allocation $\{c, h', y\}$ is a function of θ^t which we suppressed in the notation for convenience. In the following we denote the wedges as $\tau_j \equiv \tau_j\left(\theta^t\right)$, and the corresponding leads and lags of the wedges as $\tau_j' \equiv \tau_j'\left(\theta^{t+1}\right)$ and $\tau_{j-} \equiv \tau_{j-}\left(\theta^{t-1}\right)$, $j = b, l, h$. The wedges have a useful interpretation: constrained efficiency requires that the planner discourages (encourages) bequests, labor supply or human capital, respectively, if the optimality conditions which characterize the social optimum imply that $\tau_j > 0$ ($\tau_j < 0$), $j = b, h, l$.

Bequest and labor wedges.—Reinserting the optimality conditions into the definition of the wedges allows us to derive their expressions in the constrained efficient allocation.

Proposition 3 *Under assumption **[A1]**, the first-order conditions of the planner's problem imply that the constrained efficient wedges for bequests* τ_b^* *and for labor* τ_l^* *are given by*

$$\tau_b^* = 1 - \frac{1}{\mathbb{E}\left[\frac{1}{\frac{\partial u(c')}{\partial c'}}\right] \mathbb{E}\left[\frac{\partial u(c')}{\partial c'}\right]}, \tag{13}$$

$$\tau_l^* = -\frac{\partial^2 v(y,\theta,h)}{\partial\theta\partial y} \frac{\mu\left(\theta\right)}{f\left(\theta\right)}. \tag{14}$$

By Jensen's inequality, we obtain the standard result that the wedge for bequests $\tau_b^* > 0$. The planner reduces intergenerational transfers to discourage double deviations in which parents leave bequests and their children shirk. The expression for the labor wedge τ_l^* is also standard. Since ability increases productivity, $\partial^2 v(y,\theta,h)/(\partial\theta\partial y) < 0$, and $\tau_l^* > 0$ whenever the costate variable $\mu\left(\theta\right)$ is positive. The intuition is that an additional unit of required output tightens the incentive compatibility constraint, increases the information rents and thus allows for less redistribution. Families do not internalize this effect when choosing their optimal labor supply. Corollary 2 below shows that the labor wedge in our model is analogous to the wedge in Mirrlees (1971).[12]

Corollary 2 *Under assumption [**A1'**] and [**A2**]*

$$\frac{\tau_l^*}{1-\tau_l^*} = \alpha \frac{\xi \theta^{\chi}}{A^{\chi}} \frac{\partial u(c)/\partial c}{\theta f(\theta)} \int_{\underline{\theta}}^{\theta} \left[\lambda - \frac{1}{\frac{\partial u(c(x))}{\partial c(x)}} \right] dF(x),$$

where $\alpha = \varepsilon^{-1} + 1$ and ε denotes the Frisch elasticity of labor supply.

Human capital wedge.—Our contribution consists in deriving an explicit decomposition for the optimal human capital wedge τ_h^*.

Proposition 4 *Under assumption [**A1**], the constrained efficient human capital wedge τ_h^* can be decomposed as*

$$\tau_h^* = \Delta_l + \Delta_b + \Delta_i, \tag{15}$$

with

$$\begin{aligned} \Delta_l &\equiv \frac{q}{\frac{\partial g(h',h)}{\partial h'}} \mathbb{E}\left[\frac{\partial y'}{\partial h'} \tau_l^{*\prime} \right], \\ \Delta_b &\equiv \frac{q}{\frac{\partial g(h',h)}{\partial h'}} \mathbb{E}\left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} \right] \frac{\tau_b^*}{1-\tau_b^*} \\ &\quad + \frac{\beta}{\frac{\partial g(h',h)}{\partial h'} \frac{\partial u(c)}{\partial c}} \mathrm{Cov}\left(\frac{\partial u(c')}{\partial c'}, \frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} \right), \qquad (16) \\ \Delta_i &\equiv -\frac{q}{\frac{\partial g(h',h)}{\partial h'}} \mathrm{Cov}\left(\frac{1}{\frac{\partial u(c')}{\partial c'}}, \frac{\partial v(y',\theta',h')}{\partial h'} \right). \end{aligned}$$

The first component Δ_l relates the human capital wedge τ_h^* to expectations about the labor wedge $\tau_l^{*\prime}$. These expectations are weighed by the marginal product of human capital. To see why τ_h^* positively depends on future labor wedges $\tau_l^{*\prime}$, note that the first-order condition for human capital in the social optimum (6) and the definition of the

[12] Compared with Mirrlees (1971), the multiplier λ is in the numerator since the shadow price λ is in units of marginal utils and not of public funds of the planner. Furthermore, $\lim_{\theta \to \underline{\theta}} \mu(\theta) = 0$ and $\lim_{\theta \to \overline{\theta}} \mu(\theta) = 0$ imply that

$$\int_{\underline{\theta}}^{\theta} \left[\lambda - \frac{1}{\frac{\partial u(c(x))}{\partial c(x)}} \right] dF(x) = \int_{\theta}^{\overline{\theta}} \left[\frac{1}{\frac{\partial u(c(x))}{\partial c(x)}} - \lambda \right] dF(x).$$

labor wedge (11) imply that[13]

$$\frac{\partial g(h',h)}{\partial h'} = q\int_{\Theta}\left(\left(1-\tau_l^{*\prime}\right)\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right)dF\left(\theta'\right) + q\int_{\Theta}\mu'\left(\theta'\right)\frac{\partial^2 v(y',\theta',h')}{\partial\theta'\partial h'}d\theta'. \quad (17)$$

Since

$$\mathbb{E}\left[\left(1-\tau_l^{*\prime}\right)\frac{\partial y'}{\partial h'}\right] = \left(1-\mathbb{E}\left[\tau_l^{*\prime}\right]\right)\mathbb{E}\left[\frac{\partial y'}{\partial h'}\right] - \mathrm{Cov}\left(\tau_l^{*\prime},\frac{\partial y'}{\partial h'}\right),$$

equation (17) shows that the benefits from human capital accumulation for the social planner are smaller when the labor wedges $\tau_l^{*\prime}$ in the next generation are on average larger. If children are expected to supply more labor in the laissez faire than is socially optimal ($\mathbb{E}\left[\tau_l^{*\prime}\right] > 0$), then the planner wants parents to invest less into human capital so as to reduce the children's incentives to work.[14] This effect is dampened if the covariance between the wedge $\tau_l^{*\prime}$ and the marginal product of human capital $\partial y'/\partial h'$ is positive, thus reducing the riskiness of human capital investment.

The second component Δ_b relates the wedge for human capital to the wedge for bequests τ_b^*. The first term in Δ_b is of the same sign as τ_b^*.[15] The equality

$$q\frac{\tau_b^*}{1-\tau_b^*} = \mathbb{E}\left[\beta\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} - q\right]$$

makes explicit that the size of $q\tau_b^*/(1-\tau_b^*)$ depends on the difference between the stochastic discount factor of the family and the planner. The wedge required to discourage bequests implies that the stochastic discount factor of the family $\beta\frac{\partial u(c')}{\partial c'}/\frac{\partial u(c)}{\partial c}$ is expected to be higher than the planner's discount factor q. In order to correct the distortion that families expect to discount the returns to human capital investment less than the planner, the planner has to render human capital accumulation less attractive. Otherwise families would invest too much into human capital as an alternative way of transferring utility from the current to the future generation.

Bequests and human capital are not perfect substitutes, however, because the return to human capital depends on future ability and is thus risky. The risk adjustment is captured by the second term in (16) which depends on the covariance between the return to human capital and the marginal utility of consumption. Since both the return to human capital and consumption of the next generation are likely to increase with ability θ', we expect the covariance to be negative. The planner needs to discourage human-capital investment relatively less than bequests because human capital provides a bad

[13] To derive equation (17), we have used results from the proof of Remark 1 in appendix A.2 showing that

$$\frac{\partial v\left(y',\theta',h'\right)}{\partial h'} = -\frac{\partial v\left(y',\theta',h'\right)}{\partial y'}\frac{\partial y'}{\partial h'}.$$

[14] In Section 4 we explain why this result is not at odds with Bovenberg and Jacobs (2005) who show that human capital should be subsidized to offset the distortions of labor income taxation.

[15] To see this, notice that (i) the return to human capital, $\partial y'/\partial h' - \partial g(h'',h')/\partial h'$, is positive; and (ii) equation (13) implies that the constrained-efficient wedge for bequests $\tau_b^* \in (0,1)$.

hedge against consumption risk. Hence, human capital accumulation is less attractive for families.

The components Δ_l and Δ_b show that the planner counteracts the distortions in the choice of human capital induced by the intra- and intertemporal wedges τ_l^* and τ_b^*. Filtering out these effects, we are left with the component Δ_i. Following the terminology of Stantcheva (2014a), Δ_i can therefore be interpreted as the *net* human capital wedge. We now explain that Δ_i captures the effect of human capital accumulation on the incentive-compatibility constraint.

The incentive wedge Δ_i for human capital.—In order to understand the economic effect captured by Δ_i, it is useful to rewrite it as[16]

$$\Delta_i = -\frac{q}{\frac{\partial g(h',h)}{\partial h'}} \int_{\Theta} \mu'(\theta') \frac{\partial^2 v(y',\theta',h')}{\partial\theta'\partial h'} d\theta' \ . \tag{18}$$

As explained in Section 2.2, the integral on the right hand side measures by how much a marginal increase in human capital is expected to relax the incentive-compability constraint for the next generation. This benefit is ignored by the families because they take the allocation as given, and thus, ignore the impact that their investments have on the feasibility of the allocation. As families do not internalize the effect of human capital on the informational rents, there is a wedge between the optimal choice of families and that of the planner. Whether this wedge is positive or negative depends on the elasticity of labor supply and complementarity between human capital and ability.

Corollary 3 *Under assumptions [**A1'**] and [**A2**], $\partial^2 v(y',\theta',h')/(\partial\theta'\partial h') > 0$ if and only if $\chi > -\alpha$. Then $\Delta_i < 0$, showing that the planner has a motive to increase human capital accumulation in order to relax the incentive compatibility constraint.*

Estimates for the Frisch elasticity of 0.5 documented in Chetty (2012) imply $\alpha = \varepsilon^{-1} + 1 = 3$. While the evidence in Cunha et al. (2006), Table 3, shows some complementarity between human capital and ability, $\chi \geq -3$, or an elasticity of substitution between human capital and ability larger than $1/4$, seems empirically plausible.

The importance of the wedge for bequests for the human capital wedge.— Proposition 4 shows that the constrained efficient wedge for human capital can be decomposed into three terms. We now prove that the two components related to the labor wedge and incentives, Δ_l and Δ_i, offset each other at the social optimum since then

$$\frac{\partial g(h',h)}{\partial h'} = q\mathbb{E}\left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right].$$

Thus,

$$\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\mathbb{E}\left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right] = 1$$

[16]See the proof of Proposition 3 for a derivation of equation (18).

and $\tau_h^* = \Delta_b$ implies that the wedge for human capital is tightly linked to the wedge for bequests.

Proposition 5 *At the social optimum, $\Delta_l = -\Delta_i$ whenever investment in human capital is positive. It follows that the constrained efficient wedge for human capital*

$$\tau_h^* = \frac{\tau_b^*}{1-\tau_b^*} + \frac{\beta}{\frac{\partial g(h',h)}{\partial h'}\frac{\partial u(c)}{\partial c}} \text{Cov}\left(\frac{\partial u\left(c'\right)}{\partial c'}, \frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right) .$$

Proposition 5 shows that the motives that lead the planner to distort human capital investment and bequests are similar. This should not be surprising since both forms of capital transfer resources from one generation to the next. There is, however, a significant difference between the two type of intergenerational transfers: parents cannot diversify the risk associated with their children's abilities, and so human capital investment carries more risk than bequests. This is why the constrained efficient wedge for human capital will be lower when $\text{Cov}\left(\frac{\partial u(c')}{\partial c'}, \frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right) < 0$, that is when human capital is a bad hedge against consumption risk.

3.1 Comparison with the literature

We find that socially optimal distortions of human capital investment are tightly related to the optimal distortions of bequests. Compared with Farhi and Werning (2010), Proposition 5, both distortions are not equal in our model because ability is stochastic so that the returns to human capital are more uncertain than the return to bequests. Although we allow for labor supply choices of the next generation compared with Farhi and Werning (2010), we find that the labor wedge does not matter for the human capital wedge at the social optimum: the effect of the labor wedge on the human capital wedge is exactly balanced by the incentive wedge for human capital.

Our finding that human capital alleviates incentive constraints for empirically plausible parameters differs from Stantcheva (2014a) since we make explicit the effect of human capital accumulation on incentive compatibility through changes in labor supply. As already discussed in the introduction, the difference is that we specify the planner's problem with output as control and not labor effort. Thus, in the planner's optimality condition for human capital there is an additional term which captures how human-capital investment affects incentives to produce a given level of output by changing labor supply. This is made explicit in the proof of Remark 1, Appendix A.1, when taking the derivative of $\partial v/\partial\theta$ with respect to h to obtain the key cross derivative $\partial^2 v/\left(\partial\theta\partial h\right)$. The sign of this cross derivative determines how human capital affects incentives and we find that human capital mitigates the incentive problem if $\chi > -\alpha$, instead of $\chi > 0$ as in Stantcheva (2014a).

Findeisen and Sachs (2012) find a negative incentive effect of human capital, opposite to our paper. The reason is a different assumption about how human capital affects productivity. In Findeisen and Sachs (2012), more human capital and higher innate ability both favorably shift the distribution function of labor market productivity but do not enter as inputs in the production technology. Thus, human capital does not alter the incentive problem as in our model by changing the amount of labor supply needed to produce a given unit of output. Since Findeisen and Sachs (2012) assume that more human capital reinforces the effect of innate ability on the distribution function of productivity, human capital increases the informational rents of high-ability types. Thus, the incentive compatibility constraint tightens and it is optimal to tax human capital investment ceteris paribus.

In our paper instead, we assume a standard production technology in which labor productivity depends on human capital and innate ability with an aggregator function that exhibits a constant elasticity of substitution. This technology implies that, for plausible degrees of complementarity between innate ability and human capital, the disutility of effort to produce a given output decreases *less* in innate ability if human capital is higher. Then, more human capital reduces the effort cost for all agents to produce a given output, and this effect is stronger for agents with low innate ability. It follows that more human capital alleviates the incentive problem so that the planner has a motive to subsidize human capital.

3.2 Extensions

Liquidity constraints.—One may argue that parents cannot require children to make transfers to them and that children within a family cannot take on debt obligations. In our model this corresponds to the constraint $b' \geq 0$, i.e., bequests cannot be negative. We characterize in Appendix A.2 how the possibility of a binding liquidity constraint affects the wedges. A binding liquidity constraint implies a lower labor wedge ceteris paribus because the planner encourages labor effort to generate income and alleviate the constraint. The wedges for bequests and human capital become larger instead to offset that a binding constraint increases resources of the future generation.

Persistent types.—Our main results extend to a model where, instead of being independent, types are persistent across generations. Adding this feature captures the genetic transmission of characteristics from parents to their offspring. As shown in Appendix A.3, where we let the density $f(\cdot)$ from which children's abilities are drawn vary with the type of the parents, human capital mitigates the incentive problem more when the probability distribution has the monotone likelihood ratio property.[17] This common assumption implies that the planner is more likely to observe higher future output for dynasties that have high current ability (for example, Rogerson, 1985). Furthermore, we show that an analogous expression to that in Proposition 5 holds for the constrained-efficient human capital wedge. This demonstrates that the intertemporal wedge is crucial for socially optimal human capital investment also in models with persistent ability shocks.

[17] A probability density function satisfies the monotone likelihood ratio property when $[\partial f(\theta'|\theta)/\partial\theta]/f(\theta'|\theta)$ is increasing in θ'.

4 Implementation of the constrained-efficient allocation

An important question is how the socially optimal allocation of the planner's problem can be implemented in a decentralized economy. One possibility is to rely on loans for human capital accumulation with payments that are contingent on the history of loans and income. The socially optimal allocation is then implemented if these history-dependent loan repayments are combined with taxes on labor income and bequests that condition only on current income and current bequests, respectively.

Analogous to results in Albanesi and Sleet (2006), with i.i.d. ability θ, the history dependence of the tax system becomes much simpler since the history can be summarized by two state variables: bequests and human capital (see also Stantcheva, 2014a). It is then possible to implement the constrained efficient allocation, for example, either with means-tested grants that depend on labor income y, human capital investment h' and condition on the initial state variables b and h; or with loans for human capital accumulation featuring repayment schedules that depend on y, h', condition on b and h and are complemented with labor income taxes which only depend on current income y.

Existing tax and subsidy systems for student loans in continental Europe and Anglo-Saxon countries contain elements which resemble such tax schedules. The conditioning on bequests and human capital roughly corresponds to grants or repayment schedules for student loans that condition on parents' permanent income (which is highly correlated with human capital) and parents' wealth (which is correlated with bequests).[18] For concreteness, let us now illustrate how marginal taxes relate to the respective wedges. Assume the tax schedule $T(b, h, y, h')$ so that agents solve the maximization problem

$$\begin{aligned} W(\theta, b, h, t) = \max_{\{b', h', l\}} & \left\{ \mathbf{U}(c, l) + \beta \int_{\Theta} W(\theta', b', h', t+1)\, dF(\theta') \right\} \\ \text{s.t. } b' &= (1+r)b - c - g(h', h) + y - T(b, h, y, h'), \\ y &= Y(h, \theta, l), \\ h' &= h'(e, h) \text{ so that } e = g(h', h). \end{aligned}$$

The first-order condition for labor supply and the definition of the labor wedge (11) imply that the labor wedge equals the marginal income tax: $\tau_l = \partial T(\cdot)/\partial y$. The first-order condition for human capital

$$\begin{aligned} \left(\frac{\partial g(h', h)}{\partial h'} + \frac{\partial T(\cdot)}{\partial h'} \right) \frac{\partial \mathbf{U}(c, l)}{\partial c} = \beta \int_{\Theta} & \left[\frac{\partial y'}{\partial h'} \left(1 - \frac{\partial T'(\cdot)}{\partial y'} \right) \frac{\partial \mathbf{U}(c', l')}{\partial c'} \right] dF(\theta') \\ - \beta \int_{\Theta} & \left[\left(\frac{\partial g(h'', h')}{\partial h'} + \frac{\partial T'(\cdot)}{\partial h'} \right) \frac{\partial \mathbf{U}(c', l')}{\partial c'} \right] dF(\theta') \end{aligned}$$

[18] An interesting question for further research is how large the persistence of ability across generations has to be so that the simple tax and subsidy schedules observed in reality imply sizable deviations from the social optimum and thus substantial welfare losses.

and the definition of the wedge for human capital (12) imply that

$$\frac{\partial T(\cdot)}{\partial h'} = \frac{\partial g(h',h)}{\partial h'}\tau_h - \beta \int_\Theta \left[\left(\frac{\partial y'}{\partial h'} \frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'} \right) \frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} \right] dF\left(\theta'\right). \tag{19}$$

As pointed out by Stantcheva (2014a), a positive wedge for human capital does not necessarily imply a positive current marginal tax on human capital accumulation in a dynamic model. The second term on the right-hand side shows that this also depends on how human capital changes taxes in the next period and how these changes are correlated with the marginal utility of consumption.

Equation (19) allows us to relate our results further to Bovenberg and Jacobs (2005) who show that human capital should be subsidized if taxation of labor income distorts the decision to accumulate human capital. We recover the analogon of this result in our model: if, as in Bovenberg and Jacobs (2005), human capital accumulation is socially optimal in the laissez faire without tax distortions, then $\tau_h = 0$ and equation (19) reads

$$\begin{aligned} \frac{\partial T(\cdot)}{\partial h'} = & -\beta \mathbb{E}\left[\frac{\partial y'}{\partial h'} \frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'} \right] \mathbb{E}\left[\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} \right] \\ & - \frac{\beta}{\frac{\partial u(c)}{\partial c}} \mathrm{Cov}\left(\frac{\partial y'}{\partial h'} \frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'}, \frac{\partial u\left(c'\right)}{\partial c'} \right). \end{aligned}$$

Thus, the current marginal tax on human capital accumulation $\partial T(\cdot)/\partial h'$ is negatively related to the expected tax change for the next generation, $\mathbb{E}\left[\frac{\partial y'}{\partial h'} \frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'} \right]$, resulting from an additional marginal unit of human capital. Compared with Bovenberg and Jacobs (2005), this tax change does not only consist of the additional marginal income tax but also of the change of taxes due to the higher human capital stock of the next generation. Moreover, the returns to human capital are uncertain in our model so that it matters whether the tax changes reduce consumption risk. Human capital accumulation should then be subsidized if $\mathbb{E}\left[\frac{\partial y'}{\partial h'} \frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'} \right] > 0$ and the future tax changes caused by human capital accumulation do not reduce consumption risk too much (i.e., the covariance is not too negative).

This is not the whole story, however, since $\tau_h^* = \Delta_b \neq 0$ at the social optimum, as shown in Proposition 5. This is why we have an additional term in equation (19). As derived in Appendix A.4, equation (19) implies

$$\begin{aligned} \frac{\partial T(\cdot)}{\partial h'} = & \frac{\partial g(h',h)}{\partial h'} \frac{\tau_b^*}{1-\tau_b^*} \\ & - q\mathbb{E}\left[\frac{\partial y'}{\partial h'} \frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'} \right] \\ & + \frac{\beta}{\frac{\partial u(c)}{\partial c}} \mathrm{Cov}\left(\frac{\partial u\left(c'\right)}{\partial c'}, \left(1 - \frac{\partial T'(\cdot)}{\partial y'}\right) \frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} - \frac{\partial T'(\cdot)}{\partial h'} \right). \end{aligned}$$

The equation shows that human capital is subsidized more ($\partial T(\cdot)/\partial h' < 0$ is more negative) if the wedge for bequests τ_b is smaller, or if human capital investment increases the tax burden $\mathbb{E}\left[\frac{\partial y'}{\partial h'}\frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'}\right]$, or if the after-tax return to human capital is a bad hedge for consumption risk (in which case the covariance of the last term is negative).

Concerning the implementation of the wedge for bequests τ_b, the first-order condition with respect to b' implies

$$\frac{\partial u(c)}{\partial c} = \beta \mathbb{E}\left[\left(1 + r - \frac{\partial T'(\cdot)}{\partial b'}\right)\frac{\partial u(c')}{\partial c'}\right].$$

The wedge for bequests τ_b generally has to be implemented by taxes that ensure that the wedge for bequests also holds *ex post*. Then, the Euler equation of families holds for each consumption level at the reported values of stochastic ability (Kocherlakota, 2010). Otherwise families may find it optimal to deviate from the social optimum by bequeathing and letting their children exert little labor effort. Stantcheva (2014a), Proposition 5, shows that the constrained efficient allocation can also be implemented with a tax system in which savings taxes condition only on current wealth and make it optimal for families to leave no bequests.

5 Numerical analysis

We now uncover further interesting features of the allocation and wedges by solving the model numerically. In doing so, we check that the solution of the relaxed problem, based on the first-order approach, is indeed incentive compatible. We start by discussing how we calibrate the model so that the quantitative implications of the simulations are comparable to U.S. data.

5.1 Calibration

Utility function.—We set the length of a period to 30 years to approximate the time until labor-market entry of a new-born generation and the length of the labor-market career. For the standard assumption of an annual discount rate of 4%, this implies that $\beta = 0.308$. We assume $q = \beta$ to abstract from intergenerational redistribution motives arising from differences in the planner's and households' discount factors (see, for example, Farhi and Werning, 2010). We specify the utility function as $\mathbf{U}(c, l) = \ln(c) - l^{\alpha}/\alpha$, which satisfies the parametric assumption [**A1'**] made above. Based on estimates for the Frisch elasticity of 0.5 documented in Chetty (2012), we obtain that $\alpha = \varepsilon^{-1} + 1 = 3$.

Production technology.—We assume that labor productivity is Cobb-Douglas so that $A(\theta, h) = \theta^{\xi} h^{1-\xi}$. From a practical standpoint, the assumption of Cobb-Douglas productivity has the advantage that, under the assumption of competitive labor markets, wages $w(\theta, h)$ are log-linear in human capital and unobserved ability:

$$\ln w(\theta, h) = \ln A(\theta, h) = (1 - \xi)\ln h + \xi \ln \theta. \tag{20}$$

Our model thus predicts that differences in unobserved ability θ generate the residual wage dispersion which remains in the data after regressing log-wages on years of schooling (where years of schooling S correspond to $\ln h$ in our model). We assume that θ is drawn from a log-normal distribution with mean 1 and standard deviation $\sqrt{0.2}/\xi$, based on estimates by Heathcote et al. (2008, 2010). They show that the variance of residual log-wages among U.S. workers due to persistent shocks has been equal to 0.2 in 2005.[19] We use the variance resulting from persistent shocks because θ is fully persistent in our model during a generation's labor-market career and transitory shocks (at least partially) wash out.

In order to calibrate the parameter ξ of the production function, we use the large body of empirical evidence on Mincerian wage regressions. As surveyed by Card (1999), the literature shows that the marginal returns of an additional year of schooling are remarkably consistent across studies and close to 0.1. Since years of schooling S correspond to $\ln h$ in our model, equation (20) then implies $1 - \xi = 0.1$ and $\xi = 0.9$.

Education costs.—The interpretation of $\ln h$ as years of schooling S allows us to use data on educational expenditure to determine parameters of the cost function $g(h', h)$. For simplicity, we abstract from the effect of family background h on the cost of human capital accumulation so that $\partial g(h', h)/\partial h = 0$ for the flexible but parsimonious cost function $g(h', h) = \kappa\, [h'^{\varsigma} - 1]$. Since years of schooling $S = \ln(h') = 0$ if $h' = 1$, this function ensures that it is costless to provide children with 0 years of non-compulsory education. Non-compulsory education in the data corresponds to additional years of schooling starting from the first year of upper-secondary education, i.e., grade 9 in the U.S.

We now use data on the costs of upper-secondary and tertiary education to calibrate the parameters ς and κ. The parameter ς is identified by the cost of tertiary education relative to upper-secondary education whereas κ is identified by the level of upper-secondary education costs. For the assumed functional form, the ratio of cumulative costs for tertiary education to the cumulative cost for upper-secondary education is equal to $(\exp(S_2)^{\varsigma} - 1) \,/\, (\exp(S_1)^{\varsigma} - 1)$. Using actual expenditures reported in OECD (2011), we find that $\varsigma = 0.214$.[20]

The parameter κ is calibrated to match the actual cost of the first year of upper-secondary education. We thus have to relate the monetary costs observed in the data to units of the model. We make the empirically plausible assumption that the median worker of those workers *without* any non-compulsory education does not receive, or leave, any significant bequests, so that she is approximately a hand-to-mouth consumer. The lifetime income of such a worker in the laissez-faire economy is then equal to 1 which we use as

[19] See panel C of Figure 3 in Heathcote et al. (2008).

[20] Annual expenditure per year in the U.S. amounts to \$12,690 for upper-secondary education and to \$29,910 for tertiary education (Tables B.1.2 and B.1.6 in OECD, 2011). Hence, the cumulative costs for $S_1 = 4$ years of upper-secondary education is \$50,760. The cumulative cost for $S_2 = 8$, with additional four years of tertiary education, is \$50,760 + \$119,640 = \$170,400. Thus, the cost ratio is 3.357, which for $\varsigma = 0.214$ equals $(\exp(8)^{\varsigma} - 1) \,/\, (\exp(4)^{\varsigma} - 1)$.

numéraire.[21] According to census data, the mean annual earnings of high-school dropouts have been equal to \$20, 241 in 2010.[22] By comparison, the annual expenditure per year for upper-secondary students was \$12, 690. Computing the cost-income ratio, we find that the cost of an additional year of upper-secondary education amounts to 62.6% of annual income or, given our 30-year period, to 2.08% of lifetime income of the median worker without non-compulsory education. It follows that $\kappa = 0.0208/\left[\exp(1)^{0.214} - 1\right] = 0.087$.

TABLE 1: Calibration

Parameters	Model	Target	Source
Utility function			
$\beta = q = 0.308$	Discount rate	Annualized 4%	Standard
$\alpha = 3$	$v(l) = l^{\alpha}/\alpha$	Frisch Elasticity 1/2	Chetty (2012)
Production technology			
$\xi = 0.9$	$y/l = \theta^{\xi} h^{1-\xi}$	Returns to education 10%	Card (1999)
$\sigma = \sqrt{0.2}/\xi$	$\log \theta' \sim \mathcal{N}(-\sigma^2/2, \sigma^2)$	Variance residual wages	Heathcote *et al.* (2008)
Education cost			
$\varsigma = .214$	Cost function:	Costs for tertiary/upper-	OECD (2011)
$\kappa = .087$	$g(h', h) = \kappa\left[h'^{\varsigma} - 1\right]$	secondary education	

The algorithm follows Farhi and Werning (2013) closely. We initialize the level of human capital and the promised value so that the planner breaks even when we account for the cost of compulsory education: $\Gamma(V_{\text{initial}}, h_{\text{initial}}, T) = 0$ and the level of human capital h_{initial} corresponding to high-school graduation ($S = 4$). To facilitate interpretation, we report the results for a dynasty which exists for two generations.

5.2 Results

Human capital investment and parent's ability.—Figure 3 shows that optimal investment into children's years of schooling decreases in the ability of parents. This result may be surprising but is a natural consequence of the asymmetric information problem. In the first best, human capital investment would be constant in ability given that ability shocks are i.i.d. With asymmetric information, the planner's insurance of the current generation is constrained by incentive compatibility. This requires that the planner promises families with currently high ability additional utility for their children. The planner achieves

[21] For a hand-to-mouth consumer without bequests, $c = y$. The optimal labor supply for such a consumer in the laissez-faire economy solves $l^*(\theta, h) \equiv \arg\max\{\ln(A(\theta, h)\, l) - v(l)\}$, so that $l^*(\theta, h) = (A(\theta, h)/c)^{\frac{1}{\alpha-1}} = \theta^{\frac{\xi}{\alpha-1}} h^{\frac{1-\xi}{\alpha-1}} c^{\frac{1}{1-\alpha}}$. Evaluating this solution for the median worker with $S = \exp(h) = 0$, one gets $l^*(1,1) = c^{\frac{1}{1-\alpha}} = y(1,1)^{\frac{1}{1-\alpha}}$ given hand-to-mouth behavior. Since $y^*(\theta, h) = A(\theta, h)\, l^*(\theta, h)$, the income of the average worker without any non-compulsory education is $y^*(1,1) = 1$.

[22] See Table 232 in the Statistical Abstract of the United States 2012 published by the U.S. Census Bureau. For data sources see also http://www.census.gov/population/www/socdemo/educ-attn.html)

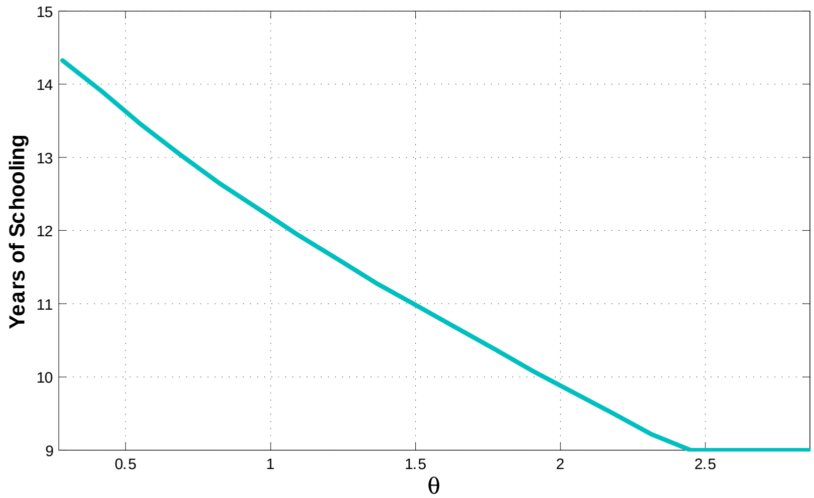

Figure 3: Human capital as a function of θ

this by giving children of high ability parents more consumption and by letting them produce less output, thus reducing their disutility of labor. The smaller labor effort of children of high-ability parents in turn makes it less attractive for the planner to invest into their human capital: in Figure 3, children of very able parents only receive nine years of compulsory schooling. Human capital investment into hard-working children of low-ability parents is instead a much more efficient way to transfer resources intertemporally. Figure 3 shows that these children receive more than five years of schooling on top of the compulsory level, roughly corresponding to a college degree.

The result is striking that children of low-ability parents receive more schooling in the social optimum. We expect that this result would be modified if ability is persistent across generations. Although there is no conclusive empirical evidence whether there actually is such persistence, it is worthwhile to briefly discuss its implications for optimal human capital investment. In the first best, human capital would then be increasing in parent's ability. It is thus not clear whether children of high-ability parents optimally receive less education in an economy with asymmetric information and persistent ability. The basic insight seems robust, however, that asymmetric information reduces the slope of human capital investment in parent's ability. An interesting question for future research is how high the persistence in ability has to be across generations to make human capital investment increase in parent's ability and whether this persistence of ability across generations is at all plausible.

The human capital wedge.— We have shown in Proposition 4 that the human capital wedge $\tau_h^* = \Delta_l + \Delta_b + \Delta_i$ can be decomposed into three components which are plotted in Figure 4. The figure illustrates the result of Proposition 5 that $\Delta_l = \Delta_i$ at the social optimum so that the human capital wedge is tightly related to the wedge for bequests.

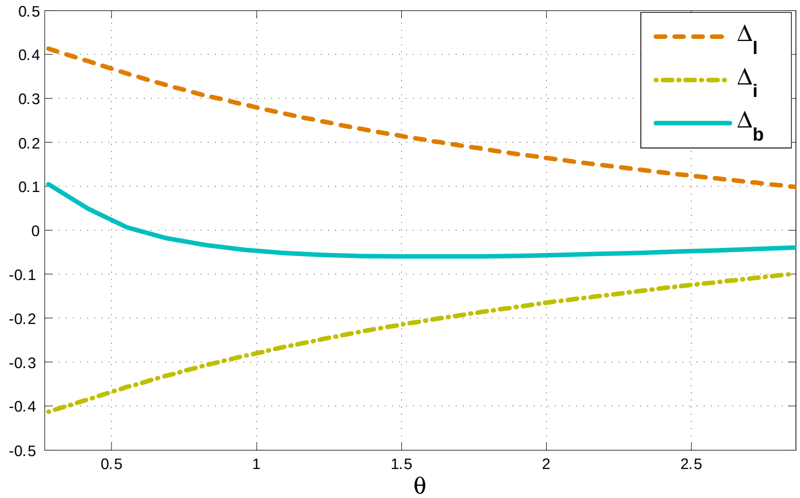

Figure 4: Decomposition of the human capital wedge

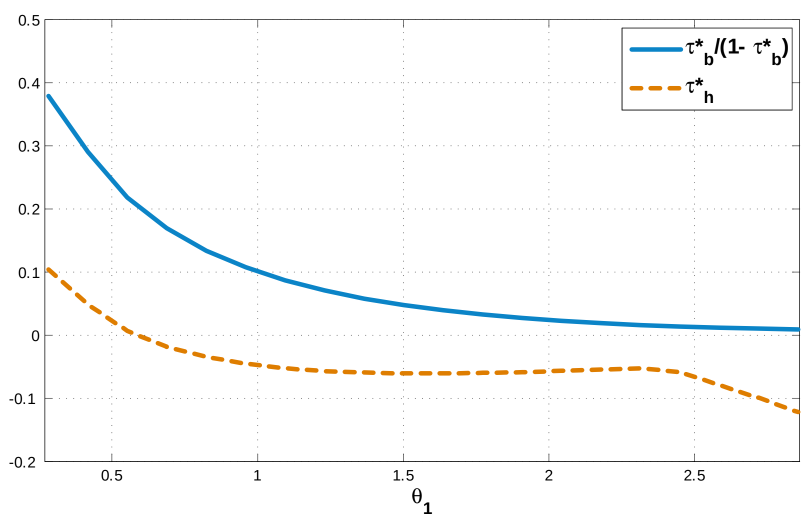

Figure 5: The relationship between the wedges for human capital and bequests

Figure 5 shows that $\tau_h^* = \Delta_b$ at interior social optima and that additional implicit subsidies for human capital are necessary for very high ability types whose human capital investment is at the corner of nine compulsory schooling years (see also Figure 3).

Figure 5 further shows that the wedges for bequests and human capital are tightly related but not identical. As is well known, the wedge for bequests is regressive (decreasing in ability) because the planner wants to discourage families with bequests to shirk and report a low type. As discussed in Proposition 5, the wedge for human capital differs from the wedge for bequests because human capital carries risk. Figure 5 shows that the risky human capital is a bad hedge for consumption risk so that $\tau_h^* < \tau_b^*/(1 - \tau_b^*)$.

Since parents cannot diversify the risk associated with their children's ability, the sign and slope of the human capital wedge in Figure 5 differs from the wedge for bequests. The implicit tax on bequests τ_b^* is always positive whereas the implicit tax on human capital τ_h^* is mostly negative, implying that human capital should be subsidized for all but the lowest ability types. Moreover, the implicit tax on human capital can be locally progressive which is illustrated in Figure 5 by the positive slope of τ_h^* in ability at intermediate ability types.

6 Conclusion

We have shown that human capital investment by families is not constrained efficient if the ability of generations in a family dynasty is not observable. The wedge for human capital accumulation implied by the solution to the planner's problem depends on the labor wedges for the next generation, the wedge for bequests and an incentive term. We find that the wedge for human capital differs from the wedge of bequests at the social optimum because human capital carries more risk as parents cannot diversify the risk associated with their children's ability. Our numerical results illustrate that human capital investment thus should be encouraged more than bequests to achieve constrained efficiency.

A Appendix

A.1 Proofs

Proof. Proposition 1: Since the planner's Hamiltonian reads

$$\begin{aligned}\mathcal{H}= & \left[c\left(\omega\left(\theta\right)-\beta V'\left(\theta\right),y\left(\theta\right),\theta,h\right)+g(h'(\theta),h)-y\left(\theta\right)+q\Gamma\left(V'\left(\theta\right),h'(\theta),t+1\right)\right]f\left(\theta\right)\\ & +\lambda\left[V-\omega\left(\theta\right)f\left(\theta\right)\right]\\ & +\mu\left(\theta\right)\left[\partial U\left(c\left(\omega\left(\theta\right)-\beta V'\left(\theta\right),y\left(\theta\right),\theta,h\right),y\left(\theta\right),\theta,h\right)/\partial\theta\right],\end{aligned}$$

the first-order conditions are

$$\left[\frac{\partial c\left(\theta\right)}{\partial V'\left(\theta\right)}+q\frac{\partial\Gamma\left(V'\left(\theta\right),h'\left(\theta\right),t+1\right)}{\partial V'\left(\theta\right)}\right]f(\theta)=-\mu\left(\theta\right)\frac{\partial^{2}U\left(\cdot\right)}{\partial\theta\partial c\left(\theta\right)}\frac{\partial c\left(\theta\right)}{\partial V'\left(\theta\right)},\tag{21}$$

$$\left[\frac{\partial g(h'(\theta),h)}{\partial h'(\theta)}+q\frac{\partial\Gamma\left(V'\left(\theta\right),h'(\theta),t+1\right)}{\partial h'(\theta)}\right]f\left(\theta\right)=0,\tag{22}$$

$$\mu\left(\theta\right)\left[\frac{\partial^{2}U\left(\cdot\right)}{\partial\theta\partial c\left(\theta\right)}\frac{\partial c\left(\theta\right)}{\partial y\left(\theta\right)}+\frac{\partial^{2}U\left(\cdot\right)}{\partial\theta\partial l\left(\theta\right)}\frac{\partial l\left(\theta\right)}{\partial y\left(\theta\right)}\right]=-\left[\frac{\partial c\left(\theta\right)}{\partial y\left(\theta\right)}-1\right]f\left(\theta\right).\tag{23}$$

The costate variable satisfies

$$\frac{\partial\mu\left(\theta\right)}{\partial\theta}=-\left[\frac{\partial c\left(\theta\right)}{\partial\omega\left(\theta\right)}-\lambda+\frac{\mu\left(\theta\right)}{f(\theta)}\frac{\partial^{2}U\left(\cdot\right)}{\partial\theta\partial c\left(\theta\right)}\frac{\partial c\left(\theta\right)}{\partial\omega\left(\theta\right)}\right]f(\theta);\tag{24}$$

with the usual boundary conditions $\lim_{\theta\to\underline{\theta}}\mu\left(\theta\right)=0$ and $\lim_{\theta\to\overline{\theta}}\mu\left(\theta\right)=0$. We use assumption **[A1]** to invert the utility function

$$c\left(\omega\left(\theta\right)-\beta V'\left(\theta\right),y\left(\theta\right),\theta,h\right)=u^{-1}\left(\omega\left(\theta\right)-\beta V'\left(\theta\right)+v(y\left(\theta\right),\theta,h)\right).$$

It follows that

$$\begin{aligned}\frac{\partial c\left(\theta\right)}{\partial\omega\left(\theta\right)}&=\frac{1}{\partial u\left(c(\theta)\right)/\partial c(\theta)},\ \frac{\partial c\left(\theta\right)}{\partial V'\left(\theta\right)}=-\frac{\beta}{\partial u\left(c\left(\theta\right)\right)/\partial c(\theta)},\\ \frac{\partial c\left(\theta\right)}{\partial y\left(\theta\right)}&=\frac{\partial v(y\left(\theta\right),\theta,h)/\partial y\left(\theta\right)}{\partial u\left(c\left(\theta\right)\right)/\partial c(\theta)},\ \frac{\partial c\left(\theta\right)}{\partial h}=\frac{\partial v(y\left(\theta\right),\theta,h)/\partial h}{\partial u\left(c\left(\theta\right)\right)/\partial c(\theta)}.\end{aligned}$$

Condition for V': Since **[A1]** implies $\partial^{2}U\left(\cdot\right)/(\partial\theta\partial c)=0$, equation (21) simplifies to

$$\frac{1}{\partial u\left(c\left(\theta\right)\right)/\partial c(\theta)}=\frac{q}{\beta}\frac{\partial\Gamma\left(V'\left(\theta\right),h'(e(\theta),h),t\right)}{\partial V'\left(\theta\right)}=\frac{q}{\beta}\lambda'\left(\theta\right),$$

where we have used the envelope condition $\partial\Gamma\left(V,h,t\right)/\partial V=\lambda$.

Condition for y: Using $\partial^{2}U\left(\cdot\right)/(\partial\theta\partial l)=-\frac{\partial y}{\partial l}\frac{\partial^{2}v(y,\theta,h)}{\partial\theta\partial y}$ in (23) yields

$$1-\frac{\partial v(y\left(\theta\right),\theta,h)/\partial y\left(\theta\right)}{\partial u\left(c\left(\theta\right)\right)/\partial c(\theta)}=-\frac{\mu\left(\theta\right)}{f(\theta)}\frac{\partial^{2}v(y\left(\theta\right),\theta,h)}{\partial\theta\partial y\left(\theta\right)}.$$

Condition for h': The following envelope condition for human capital is obtained after substituting consumption using the promise-keeping constraint, noting that there is a continuum of incentive-compatibility constraints for all θ and that $\partial^2 U(\cdot) / (\partial c(\theta) \partial \theta) = 0$:

$$\begin{aligned}\frac{\partial \Gamma(V,h,t)}{\partial h} &= \int_{\Theta} \left(\frac{\partial c(\theta)}{\partial h} + \frac{\partial g(h'(\theta),h)}{\partial h} \right) dF(\theta) + \int_{\Theta} \mu(\theta) \frac{\partial^2 U(\cdot)}{\partial \theta \partial h} d\theta \\ &= \int_{\Theta} \left(\frac{\partial v(y(\theta),\theta,h)/\partial h}{\partial u(c(\theta))/\partial c(\theta)} + \frac{\partial g(h'(\theta),h)}{\partial h} \right) dF(\theta) - \int_{\Theta} \mu(\theta) \frac{\partial^2 v(y(\theta),\theta,h)}{\partial \theta \partial h} d\theta.\end{aligned}$$

Note the last term which captures the effect of human capital on the incentive compatibility constraint. Note further that for deriving the envelope condition we have inverted $h'(e,h)$ and substituted in $e = g(h',h)$ and we have used that for all θ

$$\left(\left(\frac{\partial c(\theta)}{\partial y} - 1 \right) f(\theta) + \mu(\theta) \left[\frac{\partial^2 U(\cdot)}{\partial \theta \partial c(\theta)} \frac{\partial c(\theta)}{\partial y(\theta)} + \frac{\partial^2 U(\cdot)}{\partial \theta \partial l(\theta)} \frac{\partial l(\theta)}{\partial y(\theta)} \right] \right) \frac{\partial y(\theta)}{\partial h} = 0,$$
$$\left(\frac{\partial g(h'(\theta),h)}{\partial h'(\theta)} + q \frac{\partial \Gamma(V'(\theta),h'(\theta)))}{\partial h'(\theta)} \right) \frac{\partial h'}{\partial h} f(\theta) = 0$$

by (22) and (23). The envelope condition for human capital can then be inserted into the optimality condition for human capital (22) to obtain

$$\begin{aligned}\frac{\partial g(h'(\theta),h)}{\partial h'(\theta)} &= -q \int_{\Theta} \left(\frac{\partial v(y'(\theta'),\theta',h')/\partial h'}{\partial u(c'(\theta'))/\partial c'(\theta')} + \frac{\partial g(h''(\theta'),h')}{\partial h'} \right) dF(\theta') \\ &\quad + q \int_{\Theta} \mu'(\theta') \frac{\partial^2 v(y'(\theta'),\theta',h')}{\partial \theta' \partial h'} d\theta'.\end{aligned}$$

For $\partial^2 U(\cdot) / (\partial c(\theta) \partial \theta) = 0$, equation (24) implies

$$\mu(\theta) = \int_{\underline{\theta}}^{\theta} \left[-\frac{1}{\partial u(c(x))/\partial c(x)} + \lambda \right] dF(x). \tag{25}$$

■

Remark 1 *Under assumptions [**A1'**] and [**A2**]:*

$$\frac{\partial v(y,\theta,h)}{\partial h} < 0, \; \frac{\partial v(y,\theta,h)}{\partial \theta} < 0, \; \frac{\partial v(y,\theta,h)}{\partial y} > 0,$$
$$\frac{\partial v(y,\theta,h)}{\partial \theta \partial h} \geq 0 \text{ iff } \chi \geq -\alpha, \; \frac{\partial v(y,\theta,h)}{\partial \theta \partial y} < 0.$$

Proof. Inverting the production function $y = Y(h, l, \theta) = A(\theta, h)\, l$, we get $l = y/A(\theta, h)$ with $A(\theta, h) = [\xi\theta^\chi + (1-\xi)\, h^\chi]^{1/\chi}$ so that

$$\begin{aligned}
\frac{\partial v(y,\theta,h)}{\partial y} &= \frac{\partial \mathbf{v}\left(\frac{y}{A(\theta,h)}\right)}{\partial y} = \frac{\partial \mathbf{v}(l)}{\partial l}\frac{1}{A} > 0, \\
\frac{\partial v(y,\theta,h)}{\partial h} &= \frac{\partial \mathbf{v}\left(\frac{y}{A(\theta,h)}\right)}{\partial h} = -\frac{\partial \mathbf{v}(l)}{\partial l}\frac{y}{A^2}\frac{\partial A(\theta,h)}{\partial h} \\
&= -\frac{\partial \mathbf{v}(l)}{\partial l} l \frac{\frac{\partial A(\theta,h)}{\partial h}}{A} = -\frac{\partial \mathbf{v}(l)}{\partial l} l (1-\xi)\, h^{\chi-1} A^{-\chi} < 0, \\
\frac{\partial v(y,\theta,h)}{\partial \theta} &= \frac{\partial \mathbf{v}\left(\frac{y}{A(\theta,h)}\right)}{\partial \theta} = -\frac{\partial \mathbf{v}(l)}{\partial l}\frac{y}{A^2}\frac{\partial A(\theta,h)}{\partial \theta} \\
&= -\frac{\partial \mathbf{v}(l)}{\partial l} l \frac{\frac{\partial A(\theta,h)}{\partial \theta}}{A} = -\frac{\partial \mathbf{v}(l)}{\partial l} l \xi \theta^{\chi-1} A^{-\chi} < 0.
\end{aligned}$$

Differentiating these expressions a second time, we get

$$\begin{aligned}
\frac{\partial^2 v(y,\theta,h)}{\partial\theta\partial y} &= \frac{\partial^2 \mathbf{v}\left(\frac{y}{A(\theta,h)}\right)}{\partial\theta\partial y} = -\frac{\partial^2 \mathbf{v}(l)}{\partial l^2}\frac{y}{A^3}\frac{\partial A(\theta,h)}{\partial\theta} - \frac{\partial \mathbf{v}(l)}{\partial l}\frac{1}{A^2}\frac{\partial A(\theta,h)}{\partial\theta} \\
&= -\frac{\frac{\partial A(\theta,h)}{\partial\theta}}{A(\theta,h)^2}\frac{\partial \mathbf{v}(l)}{\partial l}\left(1 + \frac{l\partial^2 \mathbf{v}(l)/\partial l^2}{\partial \mathbf{v}(l)/\partial l}\right) = -\frac{\xi\theta^{\chi-1}}{A^{1+\chi}}\frac{\partial \mathbf{v}(l)}{\partial l}\alpha < 0,
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 v(y,\theta,h)}{\partial\theta\partial h} &= \frac{\partial^2 \mathbf{v}\left(\frac{y}{A(\theta,h)}\right)}{\partial\theta\partial h} \\
&= \frac{\partial^2 \mathbf{v}(l)}{\partial l^2}\left(\frac{y}{A^2}\right)^2 \frac{\partial A(\theta,h)}{\partial\theta}\frac{\partial A(\theta,h)}{\partial h} \\
&+ \frac{\partial \mathbf{v}(l)}{\partial l}\frac{2y}{A^3}\frac{\partial A(\theta,h)}{\partial\theta}\frac{\partial A(\theta,h)}{\partial h} - \frac{\partial \mathbf{v}(l)}{\partial l}\frac{y}{A^2}\frac{\partial^2 A(\theta,h)}{\partial\theta\partial h} \\
&= \frac{\partial \mathbf{v}(l)}{\partial l}\frac{y}{A^3}\frac{\partial A(\theta,h)}{\partial\theta}\frac{\partial A(\theta,h)}{\partial h}\left(\underbrace{1 + \frac{l\partial^2 \mathbf{v}(l)/\partial l^2}{\partial \mathbf{v}(l)/\partial l}}_{\text{Additional term}} + \underbrace{1 - \frac{\frac{\partial^2 A(\theta,h)}{\partial\theta\partial h}A(\theta,h)}{\frac{\partial A(\theta,h)}{\partial\theta}\frac{\partial A(\theta,h)}{\partial h}}}_{\text{Stantcheva (2014a)}}\right) \\
&= \frac{\partial \mathbf{v}(l)}{\partial l}\frac{y}{A}\frac{\xi\theta^{\chi-1}}{A^\chi}\frac{(1-\xi)\, h^{\chi-1}}{A^\chi}(\alpha+\chi)\ .
\end{aligned}$$

Thus, $\partial^2 v(y,\theta,h)/(\partial\theta\partial h) > 0$ iff $\chi \geq -\alpha$. ∎

Proof. Corollary 1: Follows immediately from Remark 1. ∎

Proof. Proposition 2: *Bequests.* The first-order condition for bequests reads

$$-\frac{\partial \mathbf{U}\left(c,l\right)}{\partial c}+\beta\int_{\Theta}\frac{\partial W\left(\theta',b',h'\right)}{\partial b'}dF\left(\theta'\right)=0,$$

which, reinserting the envelope condition

$$\frac{\partial W\left(\theta,b,h\right)}{\partial b}=(1+r)\frac{\partial \mathbf{U}\left(c,l\right)}{\partial c},$$

yields the Euler equation

$$\begin{aligned}\frac{\partial \mathbf{U}\left(c,l\right)}{\partial c}&=\beta(1+r)\frac{\partial \mathbf{U}\left(c',l'\right)}{\partial c'}dF\left(\theta'\right)\\&=\beta(1+r)\mathbb{E}\left[\frac{\partial \mathbf{U}\left(c',l'\right)}{\partial c'}\right].\end{aligned}$$

Labor supply. The first-order condition for labor supply reads

$$\frac{\partial \mathbf{U}\left(c,l\right)}{\partial l}+\beta\int_{\Theta}\left[\frac{\partial W\left(\theta',b',h'\right)}{\partial b'}\frac{\partial y}{\partial l}\right]dF\left(\theta'\right)=0\ .$$

The results above imply

$$\beta\int_{\Theta}\left[\frac{\partial W\left(\theta',b',h'\right)}{\partial b'}\frac{\partial y}{\partial l}\right]dF\left(\theta'\right)=\frac{\partial y}{\partial l}\frac{\partial \mathbf{U}\left(c,l\right)}{\partial c}$$

so that the first-order condition for labour supply simplifies to the standard intratemporal condition

$$\frac{\partial \mathbf{U}\left(c,l\right)}{\partial l}+\frac{\partial y}{\partial l}\frac{\partial \mathbf{U}\left(c,l\right)}{\partial c}=0.$$

Human capital. The first-order condition for human capital accumulation is

$$\beta\int_{\Theta}\left[-\frac{\partial g(h',h)}{\partial h'}\frac{\partial W\left(\theta',b',h'\right)}{\partial b'}+\frac{\partial W\left(\theta',b',h'\right)}{\partial h'}\right]dF\left(\theta'\right)=0.$$

The envelope condition is

$$\frac{\partial W\left(\theta',b',h'\right)}{\partial h'}=\frac{\partial y'}{\partial h'}\frac{\partial \mathbf{U}\left(c',l'\right)}{\partial c'}-\frac{\partial g(h'',h')}{\partial h'}\frac{\partial \mathbf{U}\left(c',l'\right)}{\partial c'}.$$

Noting that

$$\frac{\partial \mathbf{U}\left(c,l\right)}{\partial c}=\beta\int_{\Theta}\frac{\partial W\left(\theta',b',h'\right)}{\partial b'}dF\left(\theta'\right)$$

then implies that the first-order condition for human capital simplifies to

$$\frac{\partial g(h',h)}{\partial h'}\frac{\partial \mathbf{U}\left(c,l\right)}{\partial c}=\beta\int_{\Theta}\frac{\partial \mathbf{U}\left(c',l'\right)}{\partial c'}\left[\frac{\partial y'}{\partial h'}-\frac{\partial g(h'',h')}{\partial h'}\right]dF\left(\theta'\right).$$

■

Proof. Proposition 3: The wedge τ_l evaluated at the solution of the planner's problem follows immediately by using the definition for τ_l in the first-order condition (7) of the planner. To derive the analogous expression for τ_b, we recall that $\lambda'(\theta) = \mathbb{E}\left[\frac{1}{\partial u(c'(\theta'))/\partial c'(\theta')}\right]$ and rearrange the definition of τ_b to substitute $\partial u(c)/\partial c$ in condition (5). ■

Proof. Corollary 2: To compare the labor wedge in our model with the literature, we use definition (11) to derive

$$\frac{\tau_l}{1-\tau_l} = \frac{1-\frac{\partial v(y,\theta,h)/\partial y}{\partial u(c)/\partial c}}{\frac{\partial v(y,\theta,h)/\partial y}{\partial u(c)/\partial c}} = \frac{\partial u(c)/\partial c}{\partial v(y,\theta,h)/\partial y}\tau_l.$$

Thus, (14) implies that at the solution of the planner's problem,

$$\frac{\tau_l}{1-\tau_l} = -\frac{\partial u(c)/\partial c}{\partial v(y,\theta,h)/\partial y}\frac{\partial^2 v(y\theta,h)}{\partial\theta\partial y}\frac{\mu(\theta)}{f(\theta)}.$$

By Remark 1,

$$\frac{\tau_l}{1-\tau_l} = \frac{\partial u(c)/\partial c}{\frac{\partial \mathbf{v}(l)}{\partial l}\frac{1}{A}}\frac{\xi\theta^{\chi-1}}{A^{1+\chi}}\frac{\partial \mathbf{v}(l)}{\partial l}\alpha\frac{\mu(\theta)}{f(\theta)} = \alpha\frac{\xi\theta^{\chi}}{A^{\chi}}\frac{\partial u(c)/\partial c}{\theta f(\theta)}\int_{\underline{\theta}}^{\theta}\left[\lambda - \frac{1}{\frac{\partial u(c(x))}{\partial c(x)}}\right]dF(x),$$

where we have substituted in $\mu(\theta)$ using (25). ■

Proof. Proposition 4: The wedge for human capital implied by the solution to the planner's problem is obtained by adding τ_h on both sides of condition (6):

$$\begin{aligned}\tau_h = \tau_h &- \frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_{\Theta}\left(-\frac{\frac{\partial v(y',\theta',h')}{\partial h'}}{\frac{\partial u(c')}{\partial c'}} - \frac{\partial g(h'',h')}{\partial h'}\right)dF(\theta') + 1 \\ &- \frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_{\Theta}\mu'(\theta')\frac{\partial^2 v(y',\theta',h')}{\partial\theta'\partial h'}d\theta'.\end{aligned}$$

Substituting in the definition of the wedge $\tau_h(\theta)$ on the right-hand side, we get

$$\begin{aligned}\tau_h = &\frac{\beta}{\frac{\partial g(h',h)}{\partial h'}}\int_{\Theta}\left[\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}}\left(\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right)\right]dF(\theta') - 1 \\ &- \frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_{\Theta}\left(-\frac{\frac{\partial v(y',\theta',h')}{\partial h'}}{\frac{\partial u(c')}{\partial c'}} - \frac{\partial g(h'',h')}{\partial h'}\right)dF(\theta') + 1 \\ &- \frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_{\Theta}\mu'(\theta')\frac{\partial^2 v(y',\theta',h')}{\partial\theta'\partial h'}d\theta'.\end{aligned}$$

Since the derivatives of the multivariate function $v(y,\theta,h)$ in the proof of Remark 1 imply that $\frac{\partial v(y',\theta',h')}{\partial h'} = -\frac{\partial y'}{\partial h'}\frac{\partial v(y',\theta',h')}{\partial y'}$, this can be rearranged to

$$\begin{aligned}\tau_h =& \frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_\Theta \frac{\partial y'}{\partial h'}\left(1-\frac{\frac{\partial v(y',\theta',h')}{\partial y'}}{\frac{\partial u(c')}{\partial c'}}\right)dF(\theta') \\ &+\frac{1}{\frac{\partial g(h',h)}{\partial h'}}\int_\Theta\left(\beta\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}}-q\right)\left(\frac{\partial y'}{\partial h'}-\frac{\partial g(h'',h')}{\partial h'}\right)dF(\theta') \\ &-\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_\Theta \mu'(\theta')\frac{\partial^2 v(y',\theta',h')}{\partial\theta'\partial h'}d\theta'.\end{aligned}$$

The first term equals Δ_l using the definition of the labor wedge (11). The second term equals Δ_b using that $\mathbb{E}(xy) = \text{Cov}(x,y)+\mathbb{E}(x)\mathbb{E}(y)$ and $\mathbb{E}\left[\beta\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}}-q\right] = q\frac{\tau_b}{1-\tau_b}$ using the definition (10) of the wedge for bequests.

In the remaining part of the proof, we focus on the last term of τ_h to derive Δ_i. Integrating the integral of the last term by parts,

$$\int_\Theta \mu'(\theta')\frac{\partial^2 v(y',\theta',h')}{\partial\theta'\partial h'}d\theta' = \mu'(\theta')\frac{\partial v(y',\theta',h')}{\partial h'}\Big|_{\underline{\theta}'}^{\overline{\theta}'} - \int_\Theta \frac{\partial\mu'(\theta')}{\partial\theta'}\frac{\partial v(y',\theta',h')}{\partial h'}d\theta'.$$

The first term on the right-hand side is equal to zero because of the boundary conditions for $\mu'(\theta')$. Thus, using (24) and imposing assumption **[A1]**, the last term of the wedge τ_h becomes

$$\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_\Theta \frac{\partial\mu'(\theta')}{\partial\theta'}\frac{\partial v(y',\theta',h')}{\partial h'}d\theta' = -\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_\Theta\left[\frac{1}{\partial u(c')/\partial c'}-\lambda'(\theta)\right]\frac{\partial v(y',\theta',h')}{\partial h'}f(\theta')d\theta'.$$

Since by (5),

$$\lambda'(\theta) = \frac{\beta}{q\partial u(c(\theta))/\partial c(\theta)},$$

we get

$$\Delta_i = -\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_\Theta\left[\frac{1}{\partial u(c')/\partial c'}-\frac{\beta}{q\partial u(c)/\partial c}\right]\frac{\partial v(y',\theta',h')}{\partial h'}f(\theta')d\theta'$$

The integral simplifies since it is equivalent to

$$\begin{aligned}&\mathbb{E}\left[\frac{1}{\partial u(c')/\partial c'}-\frac{\beta}{q\partial u(c)/\partial c}\right]\mathbb{E}\left[\frac{\partial v(y',\theta',h')}{\partial h'}\right] \\ &+\text{Cov}\left(\frac{1}{\partial u(c')/\partial c'}-\frac{\beta}{q\partial u(c)/\partial c},\frac{\partial v(y',\theta',h')}{\partial h'}\right) \\ &=\text{Cov}\left(\frac{1}{\frac{\partial u(c')}{\partial c'}},\frac{\partial v(y',\theta',h')}{\partial h'}\right),\end{aligned}$$

where the second equality follows from the reciprocal Euler equation

$$\mathbb{E}\left[\frac{1}{\partial u\left(c'\right)/\partial c'} - \frac{\beta}{q\partial u\left(c\right)/\partial c}\right] = 0.$$

This concludes the proof. ■

Proof. Proposition 5: The social optimality condition for human capital (6) and equation (18) imply that

$$\Delta_i = -1 - \frac{q}{\partial g(h',h)/\partial h'}\int_\Theta \left(\frac{\partial v\left(y',\theta',h'\right)/\partial h'}{\partial u\left(c'\right)/\partial c'} + \frac{\partial g(h'',h')}{\partial h'}\right) dF\left(\theta'\right) \ .$$

Given the definition of Δ_l in Proposition 4, we have

$$\begin{aligned}
\Delta_i + \Delta_l &= -1 - \frac{q}{\partial g(h',h)/\partial h'}\int_\Theta \left(\frac{\partial v\left(y',\theta',h'\right)/\partial h'}{\partial u(c')/\partial c'(\theta')} + \frac{\partial g(h'',h')}{\partial h'} - \frac{\partial y'}{\partial h'}\tau_l'\right) dF\left(\theta'\right) \\
&= -1 - \frac{q}{\partial g(h',h))/\partial h'}\int_\Theta \left(\frac{\partial v\left(y',\theta',h'\right)/\partial h'}{\partial u(c')/\partial c'} + \frac{\partial g(h'',h')}{\partial h'}\right) dF\left(\theta'\right) \\
&\quad + \frac{q}{\partial g(h',h))/\partial h'}\int_\Theta \left(\frac{\partial y'}{\partial h'}\left(1 - \frac{\partial v(y',\theta',h')/\partial y'}{\partial u\left(c'\right)/\partial c'}\right)\right) dF\left(\theta'\right),
\end{aligned}$$

where the second equality uses the definition of the labor wedge in (11). As the derivatives of the multivariate function $v\left(y,\theta,h\right)$ in the proof of Remark 1 imply $\frac{\partial v(y',\theta',h')}{\partial h'} = -\frac{\partial v(y',\theta',h')}{\partial y'}\frac{\partial y'}{\partial h'}$, the equation simplifies to

$$\Delta_i + \Delta_l = -1 + \frac{q}{\partial g(h',h)/\partial h'}\int_\Theta \left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right] dF\left(\theta'\right) \ .$$

Thus, $\Delta_i = -\Delta_l$ iff

$$\frac{\partial g(h',h)}{\partial h'} = q\int_\Theta \left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right] dF\left(\theta'\right) \ . \tag{26}$$

We now show that an allocation cannot be socially optimal if (26) is not satisfied. We consider a perturbation of h' which leaves consumption and labor effort unchanged so that the allocation remains incentive compatible. For social optimality, it has to be case that the planner cannot increase his payoff with such a perturbation. Let us start from an incentive-compatible allocation $\mathbf{x} = \left\{c_t\left(\theta^t\right), y_t\left(\theta^t\right), h_{t+1}\left(\theta^t\right)\right\}_{t=1}^T$ and consider a perturbation $\mathbf{x}^{s,\delta}$ such that $c_t^{s,\delta}\left(\theta^t\right) = c_t\left(\theta^t\right)$ for all $\theta^t \in \Theta^t$, while

$$h_{t+1}^{s,\delta}\left(\theta^t\right) = \begin{cases} h_{s+1}\left(\theta^s\right) + \delta, \text{ with } \delta > 0 \\ h_{t+1}\left(\theta^t\right), \text{ whenever } t \neq s \end{cases},$$

and

$$y_{t+1}^{s,\delta}\left(\theta^{t+1}\right) = \begin{cases} y_{s+1}\left(\theta^{s+1}\right) A\left(\theta_{s+1}, h_{s+1}^{s,\delta}\left(\theta^s\right)\right) A\left(\theta_{s+1}, h_{s+1}^{s,\delta}\left(\theta^s\right)\right)^{-1} \\ y_{t+1}\left(\theta^{s+1}\right), \text{ whenever } t \neq s \end{cases},$$

for all $\theta^t \in \Theta^t$. The perturbation has been defined so as to ensure that

$$l_{s+1}^{s,\delta}\left(\theta^{s+1}\right) = \frac{y_{s+1}^{s,\delta}\left(\theta^{s+1}\right)}{A\left(\theta_{s+1}, h_{s+1}^{s,\delta}\left(\theta^s\right)\right)} = \frac{y_{s+1}\left(\theta^{s+1}\right)}{A\left(\theta_{s+1}, h_{s+1}\left(\theta^s\right)\right)} = l_{s+1}\left(\theta^{s+1}\right).$$

Given that neither output nor human capital are affected in any other period, we have $l_t^{s,\delta}\left(\theta^t\right) = l_t\left(\theta^t\right)$ for all $t = 1, ..., T$. But this implies that the perturbed allocation $\mathbf{x}^{s,\delta}$ leaves consumption and labor supply unchanged. Hence, the allocation remains incentive compatible after the perturbation. Evaluating the expected planner's payoffs for the two allocations, we find that

$$\begin{aligned}\frac{\mathbf{\Pi}\left(\mathbf{x}\right) - \mathbf{\Pi}\left(\mathbf{x}^{s,\delta}\right)}{q^s} &= \mathbb{E}_0\left[g(h_{s+1}\left(\theta^s\right), h_s\left(\theta^{s-1}\right)) - g\left(h_{s+1}^{s,\delta}\left(\theta^s\right), h_s\left(\theta^{s-1}\right)\right)\right] \\ &+ q\mathbb{E}_0\left[g(h_{s+2}\left(\theta^{s+1}\right), h_{s+1}\left(\theta^s\right)) - g(h_{s+2}\left(\theta^{s+1}\right), h_{s+1}^{s,\delta}\left(\theta^s\right))\right] \\ &- q\mathbb{E}_0\left[y_{s+1}\left(\theta^{s+1}\right)\left(\frac{A\left(\theta_{s+1}, h_{s+1}\left(\theta^s\right)\right) - A\left(\theta_{s+1}, h_{s+1}^{s,\delta}\left(\theta^s\right)\right)}{A\left(\theta_{s+1}, h_{s+1}\left(\theta^s\right)\right)}\right)\right].\end{aligned}$$

As

$$\frac{\partial y_t}{\partial h_t} = y_t \frac{\partial A\left(\theta_t, h_t\right)/\partial h_t}{A\left(\theta_t, h_t\right)},$$

the monotone convergence theorem implies that

$$\lim_{\delta \to 0} \mathbb{E}_0\left[\frac{y_{s+1}\left(\theta^{s+1}\right)}{\delta}\left(\frac{A\left(\theta_{s+1}, h_{s+1}\left(\theta^s\right)\right) - A\left(\theta_{s+1}, h_{s+1}^{s,\delta}\left(\theta^s\right)\right)}{A\left(\theta_{s+1}, h_{s+1}\left(\theta^s\right)\right)}\right)\right] = \mathbb{E}_0\left[\frac{\partial y_{s+1}\left(\theta^{s+1}\right)}{\partial h_{s+1}\left(\theta^s\right)}\right].$$

Letting δ go to zero, we therefore obtain

$$\begin{aligned}&\lim_{\delta \to 0} \frac{\mathbf{\Pi}\left(\mathbf{x}\right) - \mathbf{\Pi}\left(\mathbf{x}^{s,\delta}\right)}{\delta} \\ &= q^s \mathbb{E}_0\left[\frac{\partial g(h_{s+1}\left(\theta^s\right), h_s\left(\theta^{s-1}\right))}{\partial h_{s+1}\left(\theta^s\right)} - q\left(\frac{\partial y_{s+1}\left(\theta^{s+1}\right)}{\partial h_{s+1}\left(\theta^s\right)} - \frac{\partial g(h_{s+2}\left(\theta^{s+1}\right), h_{s+1}\left(\theta^s\right))}{\partial h_{s+1}\left(\theta^s\right)}\right)\right].\end{aligned}$$

Thus the initial allocation $\mathbf{x}$ cannot be optimal if the expression above differs from 0. Furthermore, since the time of the perturbation s has been chosen arbitrarily, condition (26) must hold for all $t = 1, ..., T$. ■

A.2 Liquidity constraints

In this subsection we show how our results modify if we impose the constraint $b' \geq 0$. In the laissez faire each family then solves the maximization problem

$$\begin{aligned} W(\theta, b, h, t) = & \max_{\{b', h', l\}} \left\{ \mathbf{U}(c, l) + \beta \int_{\Theta} W(\theta', b', h', t+1)\, dF(\theta') \right\} \\ \text{s.t. } b' &= (1+r)b - c - e + y, \\ b' &\geq 0, \\ y &= Y(h, \theta, l), \\ h' &= h'(e, h) \text{ so that } e = g(h', h), \end{aligned}$$

where the multiplier $\eta > 0$ if the liquidity constraint is binding.

Proposition 6 *If bequests are required to be non-negative, the laissez faire is characterized by the following first-order conditions for bequests, human capital and labor supply:*

$$\begin{aligned} \frac{\partial \mathbf{U}(c, l)}{\partial c} &= \beta(1+r)\mathbb{E}\left[\frac{\partial \mathbf{U}(c', l')}{\partial c'}\right] + \eta \\ \frac{\partial g(h', h)}{\partial h'}\left(\frac{\partial \mathbf{U}(c, l)}{\partial c} - \eta\right) &= \beta \int_{\Theta} \left[\frac{\partial y'}{\partial h'} \frac{\partial \mathbf{U}(c', l')}{\partial c'}\right] dF(\theta') \\ &\quad - \beta \int_{\Theta} \left[\frac{\partial g(h'', h')}{\partial h'} \frac{\partial \mathbf{U}(c', l')}{\partial c'}\right] dF(\theta') \\ -\frac{\partial \mathbf{U}(c, l)}{\partial l} &= \frac{\partial y}{\partial l}\left(\frac{\partial \mathbf{U}(c, l)}{\partial c} - \eta\right) \end{aligned}$$

Proof. *Bequests.* The first-order condition for bequests reads

$$-\frac{\partial \mathbf{U}(c, l)}{\partial c} + \beta \int_{\Theta} \frac{\partial W(\theta', b', h')}{\partial b'} dF(\theta') + \eta = 0,$$

which, reinserting the envelope condition

$$\frac{\partial W(\theta, b, h)}{\partial b} = (1+r)\frac{\partial \mathbf{U}(c, l)}{\partial c},$$

yields the Euler equation

$$\begin{aligned} \frac{\partial \mathbf{U}(c, l)}{\partial c} &= \beta(1+r)\frac{\partial \mathbf{U}(c', l')}{\partial c'} dF(\theta') + \eta \\ &= \beta(1+r)\mathbb{E}\left[\frac{\partial \mathbf{U}(c', l')}{\partial c'}\right] + \eta\ . \end{aligned}$$

Labor supply. The first-order condition for labor supply reads

$$\frac{\partial \mathbf{U}(c, l)}{\partial l} + \beta \int_{\Theta} \left[\frac{\partial W(\theta', b', h')}{\partial b'} \frac{\partial y}{\partial l}\right] dF(\theta') = 0\ .$$

The results above imply

$$\beta \int_{\Theta} \left[\frac{\partial W(\theta', b', h')}{\partial b'} \frac{\partial y}{\partial l} \right] dF(\theta') = \frac{\partial y}{\partial l} \left(\frac{\partial \mathbf{U}(c,l)}{\partial c} - \eta \right)$$

so that the first-order condition for labour supply simplifies to the standard intratemporal condition

$$\frac{\partial \mathbf{U}(c,l)}{\partial l} + \frac{\partial y}{\partial l} \left(\frac{\partial \mathbf{U}(c,l)}{\partial c} - \eta \right) = 0.$$

Human capital. The first-order condition for human capital accumulation is

$$\beta \int_{\Theta} \left[-\frac{\partial g(h', h)}{\partial h'} \frac{\partial W(\theta', b', h')}{\partial b'} + \frac{\partial W(\theta', b', h')}{\partial h'} \right] dF(\theta') = 0.$$

The envelope condition is

$$\frac{\partial W(\theta', b', h')}{\partial h'} = \frac{\partial y'}{\partial h'} \frac{\partial \mathbf{U}(c', l')}{\partial c'} - \frac{\partial g(h'', h')}{\partial h'} \frac{\partial \mathbf{U}(c', l')}{\partial c'}.$$

Noting that

$$\frac{\partial \mathbf{U}(c,l)}{\partial c} - \eta = \beta \int_{\Theta} \frac{\partial W(\theta', b', h')}{\partial b'} dF(\theta')$$

then implies that the first-order condition for human capital simplifies to

$$\begin{aligned} &\frac{\partial g(h', h)}{\partial h'} \left(\frac{\partial \mathbf{U}(c,l)}{\partial c} - \eta \right) \\ &= \beta \int_{\Theta} \frac{\partial \mathbf{U}(c', l')}{\partial c'} \left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'', h')}{\partial h'} \right] dF(\theta'). \end{aligned}$$

■

The modified definitions of the wedges are as follows:

Definition 2 *If bequests are required to be non-negative, the wedges for bequests τ_b^c, labor supply τ_l^c and human capital τ_h^c are*

$$\tau_b^c \equiv 1 - \frac{q}{\beta} \frac{\partial u(c)/\partial c - \eta}{\mathbb{E}\left[\partial u(c')/\partial c'\right]}, \tag{27}$$

$$\tau_l^c \equiv 1 - \frac{\partial v(y, \theta, h)/\partial y}{\partial u(c)/\partial c - \eta}, \tag{28}$$

$$\tau_h^c \equiv \frac{\beta}{\frac{\partial g(h', h)}{\partial h'}} \int_{\Theta} \left[\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c} - \eta} \left(\frac{\partial y'}{\partial h'} - \frac{\partial g(h'', h')}{\partial h'} \right) \right] dF(\theta') - 1. \tag{29}$$

Combining the results of Propositions 1 and 6, we then find:

Proposition 7 *If bequests are required to be non-negative, the first-order conditions of the planner's problem imply under assumption* ***[A1]*** *that*

$$\tau_b^c = 1 - \frac{1}{\mathbb{E}\left[\frac{1}{\frac{\partial u(c')}{\partial c'}}\right]\mathbb{E}\left[\frac{\partial u(c')}{\partial c'}\right]} + \frac{\eta}{\frac{\beta}{q}\mathbb{E}\left[\frac{\partial u(c')}{\partial c'}\right]}, \tag{30}$$

$$\tau_l^c = -\frac{\mu(\theta)}{f(\theta)}\frac{\partial^2 v(y,\theta,h)}{\partial\theta\partial y} - \frac{\eta}{\frac{\partial u(c)}{\partial c} - \eta}\frac{\frac{\partial v(y,\theta,h)}{\partial y}}{\frac{\partial u(c)}{\partial c}}, \tag{31}$$

$$\tau_h^c = \Delta_l^c + \Delta_b^c + \Delta_i^c + \Delta_c, \tag{32}$$

with

$$\begin{aligned}
\Delta_l^c &\equiv \frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_\Theta \frac{\partial y'}{\partial h'}\tau_l' dF(\theta'),\\
\Delta_b^c &\equiv \frac{q}{\frac{\partial g(h',h)}{\partial h'}}\mathbb{E}\left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right]\frac{\tau_b}{1-\tau_b}\\
&\quad + \frac{\beta}{\frac{\partial g(h',h)}{\partial h'}\frac{\partial u(c)}{\partial c}}\mathrm{Cov}\left(\frac{\partial u(c')}{\partial c'}, \frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right),\\
\Delta_i^c &\equiv -\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\mathrm{Cov}\left(\frac{1}{\frac{\partial u(c')}{\partial c'}}, \frac{\partial v(y',\theta',h')}{\partial h'}\right),\\
\Delta_c &\equiv \frac{\eta}{\frac{\partial u(c)}{\partial c} - \eta}\frac{\beta}{\frac{\partial g(h',h)}{\partial h'}}\int_\Theta\left[\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}}\left(\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right)\right]dF(\theta').
\end{aligned}$$

Proof. We derive the wedge τ_l^c evaluated at the solution of the planner's problem using the definition for τ_l^c in the first-order condition (7) of the planner. Condition (7) implies

$$1 - \frac{\frac{\partial v(y,\theta,h)}{\partial y}}{\frac{\partial u(c)}{\partial c}} + \frac{\frac{\partial v(y,\theta,h)}{\partial y}}{\frac{\partial u(c)}{\partial c} - \eta} - \frac{\frac{\partial v(y,\theta,h)}{\partial y}}{\frac{\partial u(c)}{\partial c} - \eta} = -\frac{\mu(\theta)}{f(\theta)}\frac{\partial^2 v(y,\theta,h)}{\partial\theta\partial y}$$

which, using the definition of the wedge τ_l^c, becomes

$$\tau_l^c = -\frac{\mu(\theta)}{f(\theta)}\frac{\partial^2 v(y,\theta,h)}{\partial\theta\partial y} + \frac{\frac{\partial v(y,\theta,h)}{\partial y}}{\frac{\partial u(c)}{\partial c}} - \frac{\frac{\partial v(y,\theta,h)}{\partial y}}{\frac{\partial u(c)}{\partial c} - \eta}.$$

Simplifying, we get

$$\tau_l^c = -\frac{\mu(\theta)}{f(\theta)}\frac{\partial^2 v(y,\theta,h)}{\partial\theta\partial y} - \frac{\eta}{\frac{\partial u(c)}{\partial c} - \eta}\frac{\frac{\partial v(y,\theta,h)}{\partial y}}{\frac{\partial u(c)}{\partial c}},$$

where $\frac{\partial u(c)}{\partial c} - \eta > 0$ since $\int_\Theta \frac{\partial W(\theta',b',h')}{\partial b'}dF(\theta') > 0$. To derive the analogous expression for τ_b, we recall that $\lambda'(\theta) = \mathbb{E}\left[\frac{1}{\frac{\partial u(c')}{\partial c'}}\right]$ and rearrange the definition of τ_b^c to substitute

$\partial u\left(c\right)/\partial c$ in condition (5). Condition (5) implies

$$\frac{\partial u\left(c\right)}{\partial c}=\frac{\frac{\beta}{q}}{\mathbb{E}\left[\frac{1}{\frac{\partial u\left(c'\right)}{\partial c'}}\right]}.$$

The definition of the wedge τ_{b}^{c} can be rearranged to

$$\partial u\left(c\right)/\partial c=\left(1-\tau_{b}^{c}\right)\frac{\beta}{q}\mathbb{E}\left[\partial u\left(c'\right)\right)/\partial c'\right]+\eta.$$

so that substituting out $\partial u\left(c(\theta)\right)/\partial c(\theta)$ yields

$$\frac{\frac{\beta}{q}}{\mathbb{E}\left[\frac{1}{\frac{\partial u\left(c'\right))}{\partial c'}}\right]}=\left(1-\tau_{b}^{c}\right)\frac{\beta}{q}\mathbb{E}\left[\partial u\left(c'\right)/\partial c'\right)\right]+\eta.$$

Solving this expression for τ_{b}^{c} results in

$$\tau_{b}^{c}=1-\frac{1}{\mathbb{E}\left[\frac{1}{\frac{\partial u\left(c'\right))}{\partial c'}}\right]\mathbb{E}\left[\partial u\left(c'\right)\right)/\partial c'\right]}+\frac{\eta}{\frac{\beta}{q}\mathbb{E}\left[\partial u\left(c'\right)/\partial c'\right]}.$$

The wedge for human capital implied by the solution to the planner's problem is obtained by adding τ_{h}^{c} on both sides of condition (6):

$$\begin{aligned}\tau_{h}^{c}=\tau_{h}^{c}&-\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_{\Theta}\left(-\frac{\frac{\partial v(y',\theta',h')}{\partial h'}}{\frac{\partial u(c')}{\partial c'}}-\frac{\partial g(h'',h')}{\partial h'}\right)dF\left(\theta'\right)+1\\&-\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_{\Theta}\mu'\left(\theta'\right)\frac{\partial^{2}v(y',\theta',h')}{\partial\theta'\partial h'}d\theta'.\end{aligned}$$

Substituting in the definition of the wedge τ_{h}^{c} on the right-hand side, we get

$$\begin{aligned}\tau_{h}^{c}=&\frac{\beta}{\frac{\partial g(h',h)}{\partial h'}}\int_{\Theta}\left[\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}-\eta}\left(\frac{\partial y'}{\partial h'}-\frac{\partial g(h'',h')}{\partial h'}\right)\right]dF\left(\theta'\right)-1\\&-\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_{\Theta}\left(-\frac{\frac{\partial v(y',\theta',h')}{\partial h'}}{\frac{\partial u(c')}{\partial c'}}-\frac{\partial g(h'',h')}{\partial h'}\right)dF\left(\theta'\right)+1\\&-\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_{\Theta}\mu'\left(\theta'\right)\frac{\partial^{2}v(y',\theta',h')}{\partial\theta'\partial h'}d\theta'\end{aligned}$$

which, using

$$\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}-\eta}=\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}}+\frac{\eta}{\frac{\partial u(c)}{\partial c}-\eta}\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}},$$

can be rearranged to

$$\begin{aligned}
\tau_h^c = & \frac{q}{\frac{\partial g(h',h)}{\partial h'}} \int_\Theta \frac{\partial y'}{\partial h'} \left(1 - \frac{\frac{\partial v(y'),\theta',h')}{\partial y'}}{\frac{\partial u(c')}{\partial c'}} \right) dF(\theta') \\
& + \frac{1}{\frac{\partial g(h',h)}{\partial h'}} \int_\Theta \left(\beta \frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} - q \right) \left(\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} \right) dF(\theta') \\
& - \frac{q}{\frac{\partial g(h',h)}{\partial h'}} \int_\Theta \mu'(\theta') \frac{\partial^2 v(y',\theta',h')}{\partial \theta' \partial h'} d\theta' \\
& + \frac{\eta}{\frac{\partial u(c)}{\partial c} - \eta} \frac{\beta}{\frac{\partial g(h',h)}{\partial h'}} \int_\Theta \left[\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} \left(\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} \right) \right] dF(\theta').
\end{aligned}$$

The first term equals Δ_l^c using the definition of the labor wedge (11) in the unconstrained case. The second term equals Δ_b^c using that $\mathbb{E}(xy) = \text{Cov}(x,y) + \mathbb{E}(x)\mathbb{E}(y)$ and that $\mathbb{E}\left[\beta \frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} - q\right] = q\frac{\tau_b}{1-\tau_b}$, where we use the definition (10) of the wedge for bequests in the unconstrained case. The third term of τ_h^c can be shown to yield Δ_i^c, as derived in the proof of Proposition 3. The fourth term equals Δ_c. ■

Thus, if the liquidity constraint for a family is binding ($\eta > 0$), the wedge on labor decreases ceteris paribus as the planner encourages more labor earnings to alleviate the constraint. The wedge for bequests and human capital increase ceteris paribus since a binding liquidity constraint implies that the future generation has more resources than would be socially optimal.

A.3 Persistent types

We now turn our attention to the general case where types are correlated from one generation to the next. For simplicity we abstract from liquidity constraints. The analysis with persistent types draws on results by Kapička (2013), applied to dynamic optimal taxation problems by Farhi and Werning (2013), Golosov et al. (2013) and, in work independent from ours, by Stantcheva (2014a). Following Pavan, Segal and Toikka (2014), the envelope condition in the problem with persistent shocks is:

$$\frac{\partial \omega(\theta)}{\partial \theta} = \frac{\partial U(c,y,\theta,h)}{\partial \theta} + \beta \int_\Theta \omega(\theta') \frac{\partial f(\theta'|\theta)}{\partial \theta} d\theta'. \tag{33}$$

This condition serves as local incentive compatibility constraint in the relaxed problem based on the first-order approach. The recursive formulation with persistent types requires that Δ and V are treated as state variables where

$$\Delta(\theta) \equiv \int_\Theta \omega(\theta) \frac{\partial f(\theta|\theta_-)}{\partial \theta_-} d\theta,$$

so that

$$\frac{\partial \omega(\theta)}{\partial \theta} = \frac{\partial U(c, y, \theta, h)}{\partial \theta} + \beta \Delta'.$$

As before we consider the relaxed planner's problem, with local constraints evaluated at the truthful equilibrium reports, and apply optimal control techniques. The recursive problem is

$$\begin{aligned} &\Gamma(V, \Delta, \theta_-, h, t) \qquad (34)\\ &= \min_{\{c,y,h',\Delta',V'\}} \left\{ \int_\Theta \left[c + g(h', h) - y(\theta) + q\Gamma(V', \Delta', \theta, h', t+1)\right] dF(\theta|\,\theta_-) \right\} \\ s.t.\ \omega(\theta) &= U(c, y, \theta, h) + \beta V', \\ V &= \int_\Theta \omega(\theta)\, dF(\theta|\,\theta_-), \\ \Delta &= \int_\Theta \omega(\theta) \frac{\partial f(\theta|\,\theta_-)}{\partial \theta_-} d\theta, \\ \frac{\partial \omega(\theta)}{\partial \theta} &= \frac{\partial U(c, y, \theta, h)}{\partial \theta} + \beta \Delta'. \end{aligned}$$

As before, we substitute consumption with the promise-keeping constraint, defining consumption $c(\omega(\theta) - \beta V', y, \theta, h)$ as an implicit function of other control and state variables. This enables us to write the Hamiltonian associated with the planner's problem as

$$\begin{aligned} \mathcal{H} = &\left[c(\omega(\theta) - \beta V', y, \theta, h) + g(h', h) - y + q\Gamma(V', \Delta', \theta, h', t+1)\right] f(\theta|\,\theta_-) \\ &+ \lambda(\theta_-)\left[V - \omega(\theta) f(\theta|\,\theta_-)\right] + \gamma(\theta_-)\left[\Delta - \omega(\theta) \frac{\partial f(\theta|\,\theta_-)}{\partial \theta_-}\right] \\ &+ \mu(\theta)\left[\frac{\partial U(c(\omega(\theta) - \beta V, y, \theta, h), y, \theta, h)}{\partial \theta} + \beta \Delta'\right]. \end{aligned}$$

The costate variable satisfies

$$\frac{\partial \mu(\theta)}{\partial \theta} = -\left[\frac{1}{\partial u(c)/\partial c} - \lambda(\theta_-) - \gamma(\theta_-) \frac{\frac{\partial f(\theta|\theta_-)}{\partial \theta_-}}{f(\theta|\,\theta_-)} + \frac{\mu(\theta)}{f(\theta|\,\theta_-)} \frac{\partial^2 U(\cdot)}{\partial \theta \partial c} \frac{\partial c}{\partial \omega(\theta)}\right] f(\theta|\,\theta_-), \qquad (35)$$

with $\lim_{\theta \to \underline{\theta}} \mu(\theta) = 0$ and $\lim_{\theta \to \bar{\theta}} \mu(\theta) = 0$. The first-order conditions read

$$\begin{aligned} \frac{\partial \mathcal{H}(\cdot)}{\partial V'} &= \left[\frac{\partial c}{\partial V'} + q \frac{\partial \Gamma(V', \Delta', \theta, h', t+1)}{\partial V'}\right] f(\theta|\,\theta_-) + \mu(\theta) \frac{\partial^2 U(\cdot)}{\partial \theta \partial c} \frac{\partial c}{\partial V'} = 0, \\ \frac{\partial \mathcal{H}(\cdot)}{\partial \Delta'} &= \left[q \frac{\partial \Gamma(V', \Delta', \theta, h', t+1)}{\partial \Delta'}\right] f(\theta|\,\theta_-) + \beta \mu(\theta) = 0, \\ \frac{\partial \mathcal{H}(\cdot)}{\partial y} &= \left[\frac{\partial c}{\partial y} - 1\right] f(\theta|\,\theta_-) + \mu(\theta)\left[\frac{\partial^2 U(\cdot)}{\partial \theta \partial c} \frac{\partial c}{\partial y} + \frac{\partial^2 U(\cdot)}{\partial \theta \partial l} \frac{\partial l}{\partial y}\right] = 0, \\ \frac{\partial \mathcal{H}(\cdot)}{\partial h'} &= \frac{\partial g(h', h)}{\partial h'} + q \frac{\partial \Gamma(V', \Delta', \theta, h', t+1)}{\partial h'} = 0. \end{aligned}$$

For the optimality condition for human capital, we use the envelope condition

$$\begin{aligned}\frac{\partial\Gamma\left(V,\Delta,\theta_{-},h,t\right)}{\partial h} &= \int_{\Theta}\left(\frac{\partial c}{\partial h}+\frac{\partial g(h',h)}{\partial h}\right)dF(\theta|\,\theta_{-})+\int_{\Theta}\mu(\theta)\frac{\partial^{2}U\left(\cdot\right)}{\partial\theta\partial h}d\theta\\ &= \int_{\Theta}\left(\frac{\partial v(y,\theta,h)/\partial h}{\partial u\left(c\right)/\partial c}+\frac{\partial g(h',h)}{\partial h}\right)dF(\theta|\,\theta_{-})-\int_{\Theta}\mu(\theta)\frac{\partial^{2}v(y,\theta,h)}{\partial\theta\partial h}d\theta\\ &+\int_{\Theta}\mu(\theta)\frac{\partial^{2}u\left(\cdot\right)}{\partial\theta\partial c}\frac{\partial c}{\partial h}d\theta\ .\end{aligned}$$

Imposing [**A1**] and using the envelope conditions $\partial\Gamma\left(\cdot\right)/\partial V=\lambda(\theta_{-})$ and $\partial\Gamma\left(\cdot\right)/\partial\Delta=\gamma(\theta_{-})$ allows us to derive the system of first-order conditions analogous to Proposition 1 but with persistent types:

$$\frac{\partial\mathcal{H}\left(\cdot\right)}{\partial V'}=\left[-\frac{\beta}{\partial u\left(c\left(\theta\right)\right)/\partial c(\theta)}+q\lambda'(\theta)\right]f(\theta|\,\theta_{-})=0\ , \tag{36}$$

$$\frac{\partial\mathcal{H}\left(\cdot\right)}{\partial\Delta'}=q\gamma'\left(\theta\right)f(\theta|\,\theta_{-})+\beta\mu\left(\theta\right)=0\ , \tag{37}$$

$$\frac{\partial\mathcal{H}\left(\cdot\right)}{\partial y}=\left[\frac{\partial v(y,\theta,h)/\partial y}{\partial u\left(c\right)/\partial c}-1\right]f(\theta|\,\theta_{-})-\frac{\partial^{2}v(y,\theta,h)}{\partial\theta\partial y}\mu\left(\theta\right)=0\ , \tag{38}$$

$$\begin{aligned}\frac{\partial\mathcal{H}\left(\cdot\right)}{\partial h'} &= \frac{\partial g(h',h)}{\partial h'}+q\int_{\Theta}\left(\frac{\frac{\partial v(y',\theta',h')}{\partial h'}}{\frac{\partial u(c')}{\partial c'}}+\frac{\partial g(h'',h')}{\partial h'}\right)dF(\theta'|\,\theta)\\ &-q\int_{\Theta}\mu'\left(\theta'\right)\frac{\partial^{2}v(y',\theta',h')}{\partial\theta'\partial h'}d\theta'=0.\end{aligned} \tag{39}$$

The system of equations is similar to the system derived for i.i.d. types but note that persistence of types alters the multiplier of the incentive compatibility constraint $\mu\left(\theta\right)$. Using (35) to substitute out $\mu(\theta)$ in equation (38), we get

$$\left[\frac{\frac{\partial v(y,\theta,h)}{\partial y}}{\frac{\partial u(c)}{\partial c}}-1\right]f(\theta|\,\theta_{-})=\frac{\partial^{2}v(y,\theta,h)}{\partial\theta\partial y}\int_{\underline{\theta}}^{\theta}\left[-\frac{1}{\frac{\partial u(c(x))}{\partial c(x)}}+\lambda(\theta_{-})+\gamma(\theta_{-})\frac{\frac{\partial f(\,\theta|\theta_{-})}{\partial\theta_{-}}}{f(x|\,\theta_{-})}\right]f(x|\,\theta_{-})dx.$$

Proposition 8 *If types θ are persistent, and assumptions [**A1**] and [**A2**] hold, the human capital wedge can be decomposed as*

$$\tau_{h}^{p}=\Delta_{l}^{p}+\Delta_{b}^{p}+\Delta_{i}^{p}$$

with

$$
\begin{aligned}
\Delta_l^p &\equiv \frac{q}{\frac{\partial g(h',h)}{\partial h'}} \int_\Theta \frac{\partial y'}{\partial h'} \tau_l^{p\prime} f(\theta' | \theta) d\theta', \\
\Delta_b^p &\equiv \frac{q}{\frac{\partial g(h',h)}{\partial h'}} \mathbb{E}_\theta \left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} \right] \frac{\tau_b^p}{1-\tau_b^p} \\
&+ \frac{\beta}{\frac{\partial g(h',h)}{\partial h'} \frac{\partial u(c)}{\partial c}} \mathrm{Cov}_\theta \left(\frac{\partial u(c')}{\partial c'}, \frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} \right), \\
\Delta_i^p &\equiv - \frac{q}{\frac{\partial u(c)}{\partial c} \frac{\partial g(h',h)}{\partial h'}} \mathrm{Cov}_\theta \left(\frac{1}{\frac{\partial u(c')}{\partial c'}}, \frac{\partial v(y',\theta',h')}{\partial h'} \right) \\
&+ \frac{\beta}{\frac{\partial u(c)}{\partial c} \frac{\partial g(h',h)}{\partial h'}} \frac{A^\chi}{\alpha \xi} \frac{\tau_l^p}{1-\tau_l^p} \theta^{1-\chi} \mathrm{Cov}_\theta \left(\frac{\frac{\partial f(\theta'|\theta)}{\partial \theta}}{f(\theta'|\theta)}, \frac{\partial v(y,\theta',h')}{\partial h'} \right),
\end{aligned}
$$

where the dependence of the expectations and covariance on the current realization of θ is denoted by the subscript.

Proof. Adding the wedge for human capital, analogous to the definition in (12), on both sides of (39) and rearranging, we find that

$$
\tau_h^p = \Delta_l^p + \Delta_b^p + \Delta_i^p
$$

with

$$
\Delta_l^p \equiv \frac{q}{\frac{\partial g(h',h)}{\partial h'}} \int_\Theta \frac{\partial y'}{\partial h'} \tau_l^{p\prime} f(\theta' | \theta) d\theta',
$$

$$
\begin{aligned}
\Delta_b^p &\equiv \frac{1}{\frac{\partial g(h',h)}{\partial h'}} \int_\Theta \left(\beta \frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} - q \right) \left(\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} \right) f(\theta'|\theta) d\theta' \\
&= \frac{q}{\frac{\partial g(h',h)}{\partial h'}} \mathbb{E}_\theta \left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} \right] \frac{\tau_b^p}{1-\tau_b^p} \\
&+ \frac{\beta}{\frac{\partial g(h',h)}{\partial h'} \frac{\partial u(c)}{\partial c}} \mathrm{Cov}_\theta \left(\frac{\partial u(c')}{\partial c'}, \frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} \right),
\end{aligned}
$$

$$
\Delta_i^p \equiv - \frac{q}{\frac{\partial g(h',h)}{\partial h'}} \int_\Theta \mu'(\theta') \frac{\partial^2 v(y',\theta',h')}{\partial \theta' \partial h'} d\theta'. \tag{40}
$$

While Δ_l^p and Δ_b^p are straightforward counterparts to the respective terms that apply if types are not persistent (see Δ_l and Δ_b in Proposition 3, developing Δ_i^p yields further insights. We elaborate on the term Δ_i^p integrating by parts:

$$
\int_\Theta \mu'(\theta') \frac{\partial^2 v(y',\theta',h')}{\partial \theta' \partial h'} d\theta' = \left[\mu'(\theta') \frac{\partial v(y',\theta',h')}{\partial h'} \right] |_{\underline{\theta}'}^{\overline{\theta}'} - \int_\Theta \frac{\partial \mu'(\theta')}{\partial \theta'} \frac{\partial v(y',\theta',h')}{\partial h'} d\theta'.
$$

The first term on the right-hand side is equal to zero because of the boundary conditions for $\mu'(\theta')$. Thus,

$$\begin{aligned}
&\Delta_i^p \\
&= \frac{q}{\frac{\partial g(h',h)}{\partial h'}} \int_\Theta \frac{\partial \mu'(\theta')}{\partial \theta'} \frac{\partial^2 v(y',\theta',h')}{\partial h'} d\theta' \\
&= -\frac{q}{\frac{\partial g(h',h)}{\partial h'}} \int_\Theta \left[\frac{1}{\partial u(c)/\partial c} - \lambda'(\theta) - \gamma'(\theta) \frac{\frac{\partial f(\theta'|\theta)}{\partial \theta}}{f(\theta'|\theta)} \right] \frac{\partial v(y',\theta',h')}{\partial h'} f(\theta'|\theta) d\theta'.
\end{aligned}$$

Since by (36) and (37),

$$\gamma'(\theta) = -\frac{\beta \mu(\theta)}{q f(\theta|\theta_-)}$$

and

$$\lambda'(\theta) = \frac{\beta}{q \partial u(c(\theta))/\partial c(\theta)},$$

we get

$$\begin{aligned}
&\Delta_i^p \\
&= -\frac{q}{\frac{\partial g(h',h)}{\partial h'}} \int_\Theta \left[\frac{1}{\frac{\partial u(c')}{\partial c'}} - \frac{\beta}{q \partial \frac{\partial u(c)}{\partial c}} + \frac{\beta \mu(\theta)}{q f(\theta|\theta_-)} \frac{\frac{\partial f(\theta'|\theta)}{\partial \theta}}{f(\theta'|\theta)} \right] \frac{\partial v(y',\theta',h')}{\partial h'} f(\theta'|\theta) d\theta' \\
&= -\frac{q}{\frac{\partial g(h',h)}{\partial h'} \frac{\partial u(c)}{\partial c}} \int_\Theta \left[\frac{\frac{\partial u(c)}{\partial c}}{\frac{\partial u(c')}{\partial c'}} - \frac{\beta}{q} + \frac{A^\chi \theta^{1-\chi}}{\alpha \xi} \frac{\beta}{q} \frac{\tau_l^p}{1-\tau_l^p} \frac{\frac{\partial f(\theta'|\theta)}{\partial \theta}}{f(\theta'|\theta)} \right] \frac{\partial v(y',\theta',h')}{\partial h'} f(\theta'|\theta) d\theta',
\end{aligned}$$

where the second equality follows from

$$\left[\frac{\frac{\partial u(c)}{\partial c}}{\frac{\partial v(y,\theta,h)}{\partial y}} - 1 \right] \frac{f(\theta|\theta_-)}{\frac{\partial u(c(\theta))}{\partial c(\theta)}} = \left[\frac{\tau_l^p}{1-\tau_l^p} \right] \frac{f(\theta|\theta_-)}{\frac{\partial u(c(\theta))}{\partial c(\theta)}} = \mu(\theta) \alpha \frac{\xi \theta^{\chi-1}}{A^\chi}.$$

In order to further simplify, note that, as in the case without persistent types, we have

$$\begin{aligned}
&\int_\Theta \left[\frac{1}{\partial u(c')/\partial c'} - \frac{\beta}{q \partial u(c)/\partial c} \right] \frac{\partial v(y',\theta',h')}{\partial h'} f(\theta'|\theta) d\theta' \\
&= \underbrace{\mathbb{E}_\theta \left[\frac{1}{\partial u(c')/\partial c'} - \frac{\beta}{q \partial u(c)/\partial c} \right]}_{=0} \mathbb{E}_\theta \left[\frac{\partial v(y',\theta',h')}{\partial h'} \right] \\
&+ \mathrm{Cov}_\theta \left(\frac{1}{\partial u(c')/\partial c'} - \frac{\beta}{q \partial u(c)/\partial c}, \frac{\partial v(y',\theta',h')}{\partial h'} \right) \\
&= \mathrm{Cov}_\theta \left(\frac{1}{\frac{\partial u(c')}{\partial c'}}, \frac{\partial v(y',\theta',h')}{\partial h'} \right).
\end{aligned}$$

Moreover, the changes $\partial f(\theta'|\theta)/\partial\theta$ in the density have to sum to zero across all θ' so that

$$\mathbb{E}_\theta\left[\frac{\frac{\partial f(\theta'|\theta)}{\partial\theta}}{f(\theta'|\theta)}\right] = \int_\Theta \frac{\frac{\partial f(\theta'|\theta)}{\partial\theta}}{f(\theta'|\theta)} f(\theta'|\theta)d\theta' = \int_\Theta \frac{\partial f(\theta'|\theta)}{\partial\theta}d\theta' = 0.$$

It follows that

$$\begin{aligned}
&\int_\Theta \frac{\frac{\partial f(\theta'|\theta)}{\partial\theta}}{f(\theta'|\theta)} \frac{\partial v(y',\theta',h')}{\partial h'} f(\theta'|\theta)d\theta' \\
&= \underbrace{\mathbb{E}_\theta\left[\frac{\frac{\partial f(\theta'|\theta)}{\partial\theta}}{f(\theta'|\theta)}\right]}_{=0} \mathbb{E}_\theta\left[\frac{\frac{\partial v(y',\theta',h')}{\partial h'}}{f(\theta'|\theta)}\right] + \mathrm{Cov}_\theta\left(\frac{\frac{\partial f(\theta'|\theta)}{\partial\theta}}{f(\theta'|\theta)}, \frac{\partial v(y',\theta',h')}{\partial h'}\right) \\
&= \mathrm{Cov}_\theta\left(\frac{\frac{\partial f(\theta'|\theta)}{\partial\theta}}{f(\theta'|\theta)}, \frac{\partial v(y',\theta',h')}{\partial h'}\right).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\Delta_i^p \\
&= -\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\mathrm{Cov}_\theta\left(\frac{1}{\frac{\partial u(c')}{\partial c'}}, \frac{\partial v(y',\theta',h')}{\partial h'}\right) \\
&\quad - \frac{\beta}{\frac{\partial u(c)}{\partial c}\frac{\partial g(h',h)}{\partial h'}} \frac{A^\chi}{\alpha\xi} \frac{\tau_l^p}{1-\tau_l^p}\theta^{1-\chi}\mathrm{Cov}_\theta\left(\frac{\frac{\partial f(\theta'|\theta)}{\partial\theta}}{f(\theta'|\theta)}, \frac{\partial v(y',\theta',h')}{\partial h'}\right).
\end{aligned}$$

■

Analogous to Proposition 3, Proposition 8 shows that the wedge for human capital is affected by the expected labor wedges in the next period in term Δ_l^p, the wedge for bequests in term Δ_b^p and the effect of human capital on the incentive-compatibility constraint in term Δ_i^p. Compared with the results for i.i.d types, the effect of human capital on the incentive-compatibility constraint in Δ_i^p also depends on the current labor wedge τ_l^p if ability types are persistent and $\gamma(\theta) > 0$. The sign of this additional effect depends on how the likelihood ratio $\frac{\partial f(\theta'|\theta)}{\partial\theta}/f(\theta'|\theta)$ covaries with the effect of human capital on the disutility of labor $\partial v(y',\theta',h')/\partial h'$ as θ' changes. We find:

Corollary 4 *Under assumptions [**A1**], [**A1'**] and [**A2**], $\Delta_i^p < 0$ if $\chi \geq -\alpha$ and $\frac{\partial f(\theta'|\theta)}{\partial\theta}/f(\theta'|\theta)$ monotonically increases in θ'. The planner then has a motive to increase human capital accumulation in order to relax the incentive compatibility constraint.*

Proof. The proof follows directly from Remark 1. ■

It seems natural that $\frac{\partial f(\theta'|\theta)}{\partial\theta}/f(\theta'|\theta)$ increases in θ' since this implies that the planner is more likely to observe higher future output of dynasties that have high current ability. See, for example, the interpretation of the monotone likelihood ratio assumption in

Rogerson (1985). With persistence of ability types, the planner thus has an additional incentive to subsidize education for reducing information rents of the future generation, and this incentive is stronger the larger is the current labor wedge τ_l^p.

Concerning the **wedge for bequests**, we impose assumption **[A1]**, use equation (35) and follow the steps of the derivations of the reciprocal Euler equation noting that $\mathbb{E}_\theta\left[\frac{\partial f(\theta'|\theta)}{\partial\theta}/f(\theta'|\theta)\right]=0$. This establishes that the wedge for bequests $\tau_b^p>0$ also in the case with persistent types. See also Stantcheva (2014a).

Finally, we derive the analogon of Proposition 5 for the case with persistent shocks.

Proposition 9 *At the optimum, $\Delta_l^p=-\Delta_i^p$ and so the constrained efficient human capital wedge $\tau_h^{p*}=\Delta_b^p$, i.e.,*

$$\tau_h^{p*}=\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\mathbb{E}_\theta\left[\frac{\partial y'}{\partial h'}-\frac{\partial g(h'',h')}{\partial h'}\right]\frac{\tau_b^{p*}}{1-\tau_b^{p*}}$$
$$+\frac{\beta}{\frac{\partial g(h',h)}{\partial h'}\frac{\partial u(c)}{\partial c}}\mathrm{Cov}_\theta\left(\frac{\partial u\left(c'\right)}{\partial c'},\frac{\partial y'}{\partial h'}-\frac{\partial g(h'',h')}{\partial h'}\right).$$

where the dependence of the expectations and covariance on the current realization of θ is denoted by the subscript.

Proof. The social optimality condition for human capital (39) and equation (40) imply that

$$\Delta_i=-1-\frac{q}{\partial g(h',h)/\partial h'}\int_\Theta\left(\frac{\partial v\left(y',\theta',h'\right)/\partial h'}{\partial u\left(c'\right)/\partial c'}+\frac{\partial g(h'',h')}{\partial h'}\right)dF(\theta'|\,\theta)\ .$$

Given the definition of Δ_l^p in Proposition 8, we have

$$\begin{aligned}\Delta_i^p+\Delta_l^p&=-1-\frac{q}{\partial g(h',h)/\partial h'}\int_\Theta\left(\frac{\partial v\left(y',\theta',h'\right)/\partial h'}{\partial u(c')/\partial c'(\theta')}+\frac{\partial g(h'',h')}{\partial h'}-\frac{\partial y'}{\partial h'}\tau_l^{p\prime}\right)dF\left(\theta'|\,\theta\right)\\&=-1-\frac{q}{\partial g(h',h))/\partial h'}\int_\Theta\left(\frac{\partial v\left(y',\theta',h'\right)/\partial h'}{\partial u(c')/\partial c'}+\frac{\partial g(h'',h')}{\partial h'}\right)dF\left(\theta'|\,\theta\right)\\&+\frac{q}{\partial g(h',h))/\partial h'}\int_\Theta\left(\frac{\partial y'}{\partial h'}\left(1-\frac{\partial v(y',\theta',h')/\partial y'}{\partial u\left(c'\right)/\partial c'}\right)\right)dF\left(\theta'|\,\theta\right),\end{aligned}$$

where the second equality uses the definition of the labor wedge in (11). As the derivatives of the multivariate function $v\left(y,\theta,h\right)$ in the proof of Remark 1 are such that $\frac{\partial v(y',\theta',h')}{\partial h'}=-\frac{\partial v(y',\theta',h')}{\partial y'}\frac{\partial y'}{\partial h'}$, the equation simplifies to

$$\Delta_i^p+\Delta_l^p=-1+\frac{q}{\partial g(h',h)/\partial h'}\int_\Theta\left[\frac{\partial y'}{\partial h'}-\frac{\partial g(h'',h')}{\partial h'}\right]dF\left(\theta'|\,\theta\right)\ .$$

Thus, $\Delta_i^p = -\Delta_l^p$ iff

$$\frac{\partial g(h',h)}{\partial h'} = q \int_{\Theta} \left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} \right] dF\left(\theta' | \theta\right) \ . \tag{41}$$

We then can follow the same steps as in the proof of Proposition 5 to show that an allocation cannot be socially optimal if (41) is not satisfied. ■

A.4 Implementation

Substituting $\tau_h = \Delta_b$ into equation (19), we get

$$\begin{aligned}
\frac{\partial T(\cdot)}{\partial h'} &= \frac{\partial g(h',h)}{\partial h'}\Delta_b - \beta \mathbb{E}\left[\left(\frac{\partial y'}{\partial h'} \frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'} \right) \frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} \right] \\
&= \mathbb{E}\left[\beta \frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} - q \right] \mathbb{E}\left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} \right] \\
&+ \frac{\beta}{\frac{\partial u(c)}{\partial c}} \mathrm{Cov}\left(\frac{\partial u\left(c'\right)}{\partial c'}, \frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} \right) \\
&- \beta \mathbb{E}\left[\frac{\partial y'}{\partial h'} \frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'} \right] \mathbb{E}\left[\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} \right] \\
&- \frac{\beta}{\frac{\partial u(c)}{\partial c}} \mathrm{Cov}\left(\frac{\partial y'}{\partial h'} \frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'}, \frac{\partial u\left(c'\right)}{\partial c'} \right),
\end{aligned}$$

where in the second equality we substitute Δ_b as defined in Proposition 3. Using the definition of the wedge for bequests (10) implies that

$$\mathbb{E}\left[\beta \frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} - q \right] = q \frac{\tau_b}{1-\tau_b}$$

and

$$\mathbb{E}\left[\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} \right] = \frac{q}{\beta} \frac{1}{1-\tau_b}$$

so that

$$\begin{aligned}
\frac{\partial T(\cdot)}{\partial h'} &= q\frac{\tau_b}{1-\tau_b}\mathbb{E}\left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right] + \frac{\beta}{\frac{\partial u(c)}{\partial c}}\text{Cov}\left(\frac{\partial u\left(c'\right)}{\partial c'}, \frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right) \\
&- q\frac{1}{1-\tau_b}\mathbb{E}\left[\frac{\partial y'}{\partial h'}\frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'}\right] - \frac{\beta}{\frac{\partial u(c)}{\partial c}}\text{Cov}\left(\frac{\partial y'}{\partial h'}\frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'}, \frac{\partial u\left(c'\right)}{\partial c'}\right) \\
&= q\mathbb{E}\left[\left(1-\frac{\partial T'(\cdot)}{\partial y'}\right)\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} - \frac{\partial T'(\cdot)}{\partial h'}\right]\frac{\tau_b}{1-\tau_b} \\
&- q\mathbb{E}\left[\frac{\partial y'}{\partial h'}\frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'}\right] \\
&+ \frac{\beta}{\frac{\partial u(c)}{\partial c}}\text{Cov}\left(\frac{\partial u\left(c'\right)}{\partial c'}, \left(1-\frac{\partial T'(\cdot)}{\partial y'}\right)\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} - \frac{\partial T'(\cdot)}{\partial h'}\right).
\end{aligned}$$

Applying the logic of the proof of Proposition 5 implies that, at the social optimum, the discounted expected after-tax returns to human capital have to equal the marginal cost of human capital investment:

$$\frac{\partial g(h',h)}{\partial h'} = q\mathbb{E}\left[\left(1-\frac{\partial T'(\cdot)}{\partial y'}\right)\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} - \frac{\partial T'(\cdot)}{\partial h'}\right].$$

Thus,

$$\begin{aligned}
\frac{\partial T(\cdot)}{\partial h'} &= \frac{\partial g(h',h)}{\partial h'}\frac{\tau_b^*}{1-\tau_b^*} \\
&- q\mathbb{E}\left[\frac{\partial y'}{\partial h'}\frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'}\right] \\
&+ \frac{\beta}{\frac{\partial u(c)}{\partial c}}\text{Cov}\left(\frac{\partial u\left(c'\right)}{\partial c'}, \left(1-\frac{\partial T'(\cdot)}{\partial y'}\right)\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} - \frac{\partial T'(\cdot)}{\partial h'}\right).
\end{aligned}$$

References

[1] Abreu Dilip, David Pearce and Ennio Stacchetti (1990): "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," Econometrica, vol. 58(5), 1041-1063

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

[3] Atkeson, Andrew and Robert E. Lucas, Jr. (1992): "On Efficient Distribution with Private Information," Review of Economic Studies, vol. 59, 427–453.

[4] Battaglini, Marco and Rohit Lamba (2014): "Optimal Dynamic Contracting: the First-Order Approach and Beyond," Manuscript, Princeton University.

[5] Ben-Porath, Yoram (1967): "The Production of Human Capital and the Life Cycle of Earnings," Journal of Political Economy, vol. 75, 352–365.

[6] Bohaček, Radim and Marek Kapička (2008): "Optimal Human Capital Policies," Journal of Monetary Economics, vol. 55, 1–16.

[7] Bovenberg, A. Lans and Bas Jacobs (2005): "Redistribution and Education Subsidies are Siamese Twins," Journal of Public Economics, vol. 89, 2005–2035.

[8] Card, David (1999): "The Causal Effect of Education on Earnings," in: *Handbook of Labor Economics*, Ashenfelter, Orley and David Card (eds.), vol. 3A, 1801–1863.

[9] Chetty, Raj (2012): "Bounds on Elasticities with Optimization Frictions: a Synthesis of Micro and Macro Evidence on Labor Supply," Econometrica, vol. 80, 969–1018.

[10] Cunha, Flavio, James J. Heckman, Lance Lochner and Dimitriy V. Masterov (2006): "Interpreting the Evidence on Life Cycle Skill Formation," ch. 12, in: *Handbook of the Economics of Education*, Hanushek, Eric A. and Finis Welch (eds.), vol. 1, 697–812.

[11] Erosa, Andrés and Tatyana Koreshkova (2007): "Progressive Taxation in a Dynastic Model of Human Capital," Journal of Monetary Economics, vol. 54, 667–685.

[12] Farhi, Emmanuel and Iván Werning (2007): "Inequality and Social Discounting," Journal of Political Economy, vol. 115, 365–402.

[13] Farhi, Emmanuel and Iván Werning (2010): "Progressive Estate Taxation," Quarterly Journal of Economics, vol. 125, 635–673.

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

[15] Findeisen, Sebastian and Dominik Sachs (2012): "Education and Optimal Dynamic Taxation: the Role of Income-Contingent Loans," Working Paper, University of Konstanz.

[16] Gary-Bobo, Robert and Alain Trannoy (2014): "Optimal Student Loans and Graduate Tax under Moral Hazard and Adverse Selection," Working Paper, CREST, Paris.

[17] Gelber, Alexander and Matthew Weinzierl (2014): "Equalizing Outcomes and Equalizing Opportunities: Optimal Taxation when Children's Abilities Depend on Parents' Resources," Working Paper No. 13-014, Harvard Business School.

[18] Golosov, Mikhail, Maxim Troshkin, Aleh Tsyvinski (2011): "Optimal Dynamic Taxes," NBER Working Paper No. 17642.

[19] Heathcote, Jonathan, Kjetil Storesletten and Giovanni L. Violante (2008): "The Macroeconomic Implications of Rising Wage Inequality in the United States," NBER Working Paper No. 14052.

[20] Heathcote, Jonathan, Kjetil Storesletten and Giovanni L. Violante (2010): "The Macroeconomic Implications of Rising Wage Inequality in the United States," Journal of Political Economy, vol. 118, 681–722.

[21] Kapička, Marek (2006): "Optimal Income Taxation with Human Capital Accumulation and Limited Record Keeping," Review of Economic Dynamics, vol. 9, 612–639.

[22] Kapička, Marek (2013): "Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: a First-Order Approach," Review of Economic Studies, vol. 80, 1027–1054.

[23] Kocherlakota, Narayana R. (2010): *The New Dynamic Public Finance*, Princeton University Press, Princeton, New Jersey.

[24] Koehne, Sebastian and Moritz Kuhn (2014): "Optimal Taxation in a Habit Formation Economy," CESifo Working Paper No. 4581.

[25] Krueger, Dirk and Alexander Ludwig (2013): "Optimal Capital and Progressive Labor Income Taxation with Endogenous Schooling Decisions and Intergenerational Transfers," Manuscript, University of Pennsylvania.

[26] Lee, Sang Yoon (Tim) and Ananth Seshadri (2014): "On the Intergenerational Transmission of Economic Status," Manuscript, University of Mannheim.

[27] Mirrlees, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," Review of Economic Studies, vol. 38, 175–208.

[28] OECD (2011): *Education at a Glance*, OECD, Paris.

[29] Pavan, Alessandro, Ilya Segal and Juuso Toikka (2014): "Dynamic Mechanism Design: A Myersonian Approach," Econometrica, vol. 82, 601–653.

[30] Rogerson, William P. (1985): "The First-Order Approach to Principal-Agent Problems," Econometrica, vol. 53, 1357–1367.

[31] Saez, Emmanuel (2001): "Using Elasticities to Derive Optimal Income Tax Rates," Review of Economic Studies, vol. 68, 205–229.

[32] Stantcheva, Stefanie (2014a): "Optimal Taxation and Human Capital Policies over the Life Cycle," Manuscript, MIT.

[33] Stantcheva, Stefanie (2014b): "Optimal Income, Education and Bequest Taxes in an Intergenerational Model," Manuscript, MIT