Universität St.Gallen

Human Capital and Optimal Redistribution

Winfried Koeniger, Julien Prat

May 2015 Discussion Paper no. 2014-34

School of Economics and Political Science, Department of Economics

University of St.Gallen

Editor:	Martina Flockerzi University of St.Gallen School of Economics and Political Science Department of Economics Bodanstrasse 8 CH-9000 St. Gallen Phone +41 71 224 23 25 Fax +41 71 224 31 35 Email seps@unisg.ch
Publisher:	School of Economics and Political Science Department of Economics University of St.Gallen Bodanstrasse 8 CH-9000 St. Gallen Phone +41 71 224 23 25
Electronic Publication:	Fax +41 71 224 31 35 http://www.seps.unisg.ch

Human Capital and Optimal Redistribution[1]

Winfried Koeniger, Julien Prat

Author's address: Prof. Dr. Winfried Koeniger
HSG-SEW
Varnbüelstr. 14
CH-9000 St. Gallen
Phone +41 71 2242309
Fax +41 71 2242302
Email winfried.koeniger@unisg.ch
Website www.sew.unisg.ch/en

---

[1] Winfried Koeniger: University of St.Gallen (Swiss Institute for Empirical Economic Research), CEPR, CFS, IZA. Julien Prat: CNRS (CREST), CEPR, CESifo, IZA. 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 Kapicka, 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).

Abstract

We characterize optimal redistribution in a dynastic economy with observable human capital and hidden ability. The government can use education to improve the insurance-incentive trade-off because there is a wedge between human capital investment in the laissez faire and the social optimum. This wedge differs from the wedge for bequests because: (i) returns to human capital are risky; (ii) human capital may change informational rents. We illustrate numerically that, if ability is i.i.d. across generations, human capital investment declines in parents' income in the social optimum, and show how this optimum can be implemented with student loans or means-tested grants.

Keywords

Human capital, Optimal taxation, Intergenerational Equity.

JEL Classification

E24, H21, I22, J24.

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 financial 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 lays out a rigorous framework to analyze this trade-off. It shows that asymmetric information about labor market productivity prevents full insurance because productive agents would not find it optimal to reveal their 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 draw from the ability distribution, and hence productivity, is private information.

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.

The constrained efficient wedge for human capital turns out to be closely related, but not identical, to the wedge for bequests. The similarity is intuitive because parents can substitute financial with human capital when they transfer resources to their offspring. But the productivity of children is uncertain and parents cannot diversify this risk. This additional source of uncertainty makes it less attractive for families to invest in human capital as it provides a bad hedge against consumption risk. It then follows that the planner does not have to discourage human capital investment to the same extent as bequests.

The planner also takes into account the impact that education has on the trade-off between equality and incentives. Intuitively, if talented agents benefit more from human capital investments, increasing education raises their informational rents and worsens the incentive problem. This is why the wedge for human capital contains an additional term that is proportional to the degree of complementarity between ability and human capital.

We illustrate these results numerically, focussing on the case in which innate ability is uncorrelated across generations. The solution for the constrained-efficient allocation delivers a striking result. We find that the socially-optimal human capital investment into children should be decreasing in parents' ability and thus income. This result is explained by a wealth effect. In the constrained-efficient allocation without full insurance, children from a privileged background inherit larger bequests. This reduces their labor supply so that it becomes relatively less efficient for the planner to invest into their human capital.

The wedge for human capital required to implement this allocation is much lower than the wedge for bequests, in our numerical illustration with a unit elasticity between ability and human capital, so that human capital should be subsidized for all but the very-low income families while bequests should be taxed across all income levels. We illustrate numerically how the wedges can be implemented with means-tested grants or loans with contingent repayments and illustrate the extent to which both implementations provide insurance against ability risk.

Related literature.—Our paper relates to the two large literatures on human capital and optimal taxation. For brevity, we focus only on a number of recent contributions and refer to their literature reviews for further discussion of previous research.

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. It differs from the wedges for bequests because human capital carries more risk than bequests and, as explained above, may change the power of incentives. In the social optimum the planner equates the social return on bequests and on human capital as in Farhi and Werning (2010), who abstract from risk and do not endogenize the labor supply of children, or in the independent and complementary work by Stantcheva (2015). Implementation of the social optimum requires that the choice between bequests and human capital has to be distorted if, and only if, human capital investment changes the incentives to truthfully reveal ability.

Our results relate to recent 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) use the generalized envelope condition derived by Kapička (2013) and Pavan et al. (2014) to characterize the planner's necessary conditions, as we do in this paper.

Stantcheva (2014) extends the analysis of Findeisen and Sachs (2012) to a multi-period setting with training time, and possibly unobservable human capital. She proposes a decomposition of the human capital wedge that is similar, but not identical, to ours. The discrepancy arises because of differences in the timing of the models. Since Stantcheva (2014) analyzes a life-cycle problem, she assumes that human capital raises today's productivity which generates an interaction of human capital with the contemporaneous labor wedge. This mechanism is absent in our model because we focus on the education of children whose benefits accrue only next period, when these children become adults and participate to the labor market. The different timing allows us to simplify the human capital wedge further, thereby highlighting its close relationship with the intertemporal wedge for bequests in a dynastic family model.

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 when 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.

Boháček and Kapička (2008) characterize optimal education and tax policies when agents have heterogenous ability that remains constant over time. Kapička (2015) and Kapička and Neira (2015) extend this analysis by focussing on human capital investment or learning effort over the life cycle that are unobservable. By contrast, 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), Heathcote et al. (2014), Krueger and Ludwig (2013), Lee and Seshadri (2014) and Stantcheva (2015), Sections 2–5, 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 change the equality-efficiency trade-off over time by adjusting the amount of 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 shows how education may reduce the disutility of labor of future generations if human capital is not too complementary to innate ability. Then, consumption of leisure is less attractive and incentives to exert effort are stronger.

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 present the numerical solution for a calibrated version of the model in Section 4 and discuss implementation of the constrained-efficient allocation 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, has a planning horizon T and a size normalized to one. 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:

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

As in the seminal paper of Mirrlees (1971), agents differ in their ability θ_t which cannot be observed by the planner. Instead, both bequests b_t and human capital h_t are public knowledge. Output y_t is produced according to the technology $Y\left(h_t, l_t, \theta_t\right)$ which is increasing in its arguments and concave. We will use the production function to substitute l_t in the utility function and write $U\left(c_t, y_t, h_t, \theta_t\right)$ instead of $\mathbf{U}(c_t, l_t)$ or, with assumption **[A1]**, $v(y_t, h_t, \theta_t) = \mathbf{v}\left(l_t\right)$. 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. We briefly discuss the extension of our model to persistent ability shocks in Section 3.1 and delegate the presentation of the results for this case to appendix A.2.3.

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 arbitrary restrictions on the allocation. In particular, we do not rule out history dependent allocations summarized by $\mathbf{x} \equiv \left\{x_t\left(\theta^t\right)\right\}_{t=0}^{T}$ where $x_t\left(\theta^t\right) \equiv$

[1] Human capital investment affects productivity in the next period (for the next generation) and not in the current period as in Stantcheva (2014). 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 enables agents to use joint deviations. Furthermore h_{t+1} does not depend on the children's realized ability θ_{t+1}. This assumption could be relaxed but is imposed for parsimony: allowing h_{t+1} to depend on θ_{t+1} would add a channel through which output depends on ability but would not add substantial insights to our analytical and numerical results as long as observation of h_{t+1} does not allow the planner to infer θ_{t+1}.

$\left\{c_t\left(\theta^t\right), h_{t+1}\left(\theta^t\right), y_t\left(\theta^t\right)\right\}$. The family's preferences over an allocation $\mathbf{x}$ are given by

$$\mathcal{U}(\mathbf{x}) \equiv \mathbb{E}_0\left[\sum_{t=0}^{T} \beta^t U\left(c_t\left(\theta^t\right), y_t\left(\theta^t\right), h_t\left(\theta^{t-1}\right), \theta_t\right)\right],$$

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* $\mathbf{r} \equiv \left\{r_t\left(\theta^t\right)\right\}_{t=0}^{T}$ 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 $(\mathbf{x} \circ \mathbf{r})\left(\theta^t\right) \equiv \left\{x_t\left(r^t\left(\theta^t\right)\right)\right\}_{t=0}^{T}$ 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 allocations. 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

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

[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 the incentives to exert labor effort and the consequences of the different risk of bequests and human capital investment.

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

$$\omega^{\mathbf{r}}\left(\theta^{t}\right) \equiv U\left(c_{t}\left(r^{t}\left(\theta^{t}\right)\right), y_{t}\left(r^{t}\left(\theta^{t}\right)\right), h_{t}\left(r^{t-1}\left(\theta^{t-1}\right)\right), \theta_{t}\right)+\beta \int_{\Theta} \omega^{\mathbf{r}}\left(\theta^{t}, \theta_{t+1}\right) dF\left(\theta_{t+1}\right) .$$

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]

$$\begin{aligned}
\Gamma(V, h, t) &= \min_{\{c,y,h',V'\}}\left\{\int_{\Theta}\left[c(\theta)+g(h'(\theta), h)-y(\theta)+q\Gamma\left(V'(\theta), h'(\theta), t+1\right)\right] dF(\theta)\right\} \\
\text{s.t. } \omega(\theta) &= U(c(\theta), y(\theta), h, \theta)+\beta V'(\theta), && (2)\\
V &= \int_{\Theta} \omega(\theta)\, dF(\theta), && (3)\\
\frac{\partial \omega(\theta)}{\partial \theta} &= \frac{\partial U(c, y, h, \theta)}{\partial \theta}, && (4)
\end{aligned}$$

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 $\partial g(h', h)/\partial h' > 0$, whereas $\partial g(h', h)/\partial h < 0$ if costs are smaller for parents with more human capital.

The first constraint defines the continuation value $\omega(\theta)$ as the sum of the current and next-period promised utilities $U(\cdot)$ and $V'(\theta)$, 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. The validity of condition (4) can be checked quite easily, however, if ability θ is

[4] Note that we impose 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

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

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.

i.i.d. and preferences satisfy the single-crossing condition. Then the first-order approach is valid when the allocation is monotone in ability.[8]

2.2 Optimality conditions

In the first best allocation, families 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*

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

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

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

with

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

Consumption and Output.—Equation (5) implies that the reciprocal Euler equation holds 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

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

Using $\partial c(\theta)/\partial\omega(\theta) = [\partial u(c(\theta))/\partial c(\theta)]^{-1}$ 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

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

[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).

is satisfied and the inverse of the marginal utility of consumption follows a martingale when $q = \beta$. Since the condition for optimal production (7) is analogous to the optimality condition in the standard Mirrlees problem, 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

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

The marginal cost of human capital investment on the left hand side is 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 investment reduces the cost of accumulating human capital for the next generation. Thirdly, human capital affects the incentive compatibility constraint, as captured by the second integral on the right hand side of (9). 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 types instead, information revelation is profitable solely if

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

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 the informational rents and mitigates the incentive problem.

These gains are 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 investment decision is made.[10]

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

[10] According to its definition in (8), the costate variable $\mu'(\theta')$ also captures the probability weight for each type.

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*

$$\textbf{[A1']:}\quad \mathbf{U}(c,l) = u(c) - \mathbf{v}(l), \ \textit{where } \mathbf{v}(l) = \zeta l^{\alpha}, \textit{ with } \zeta > 0 \textit{ and } \alpha > 1,$$
$$\textbf{[A2]:}\quad Y(h,l,\theta) = A(\theta,h)\,l,$$
$$\textit{with } A(\theta,h) = \left[\xi\theta^{\chi} + (1-\xi)\,h^{\chi}\right]^{1/\chi}, \ \chi \in (-\infty, 1] \textit{ and } \xi \in (0,1).$$

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

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,[11] is greater than the degree of complementarity χ between ability and human capital.

Thus the effect of h on the informational rents enjoyed by high-ability families depends on two, potentially opposite, channels. The first one is driven by the adjustment in labor supply following an increase in human capital. Raising h ensures that any level of output can be produced with less labor. Since the disutility of effort is convex, the returns to ability, and thus the informational rents, are reduced. This *labor supply effect* is unambiguously positive, in the sense that it relaxes the incentive constraint, and its size is proportional to the convexity of $v(\cdot)$, as measured by the elasticity parameter α.

The second effect depends on the technology of production $Y(\cdot)$. When human capital and ability are complementary factors, families with a high ability benefit more from any given increase in human capital. They find it more attractive to imitate less able agents, which raises their informational rents. This is why an increase in the degree of substitutability, as measured by an increase in the parameter χ, relaxes the incentive constraint, thereby reinforcing the positive influence of the labor supply channel.

3 The wedges

We now compare the optimality conditions in the laissez faire to those of 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.

[11] The Frisch elasticity of labor supply is equal to $1/(\alpha - 1)$.

In the laissez faire each family solves the maximization problem

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

where b is the bequest and the agent chooses functions $b', h', l : \Theta \to \mathbb{R}^{+}$.

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.[12] Then the 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*

$$
\begin{aligned}
\tau_b\left(\theta^t\right) &\equiv 1 - \frac{q}{\beta}\frac{\partial u(c)/\partial c}{\mathbb{E}\left[\partial u(c')/\partial c'\right]}, && (10) \\
\tau_l\left(\theta^t\right) &\equiv 1 - \frac{\partial v(y,h,\theta)/\partial y}{\partial u(c)/\partial c}, && (11) \\
\tau_h\left(\theta^t\right) &\equiv \frac{\beta}{\frac{\partial g(h',h)}{\partial h'}}\mathbb{E}\left[\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}}\left(\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right)\right] - 1. && (12)
\end{aligned}
$$

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

[12] 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. We have not been able, however, to derive simple conditions that establish concavity in our dynamic model with the additional bequest choice.

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 are such that $\tau_j > 0$ ($\tau_j < 0$), $j = b, h, l$.

Bequest and labor wedges.—Combining the conditions for the social optimum with the definition of the wedges allows us to derive the wedges at 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 labor τ_l^* are given by*

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

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

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, $\partial^2 v(y,h,\theta)/(\partial\theta\partial y) < 0$, and τ_l^* is positive whenever $\mu\left(\theta\right) > 0$. 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).[13]

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

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

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

[13] 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, $\lim_{\theta\to\underline{\theta}}\mu\left(\theta\right) = 0$ and $\lim_{\theta\to\overline{\theta}}\mu\left(\theta\right) = 0$ imply that

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

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

Proposition 4 *Under assumption* ***[A1']*** *and* ***[A2]****, the constrained efficient wedge* τ_h^* *for human capital can be decomposed as*

$$\tau_h^* = \Delta_b + \Delta_i,$$

where

$$\Delta_b \equiv \frac{q}{\frac{\partial g(h',h)}{\partial h'}} \mathbb{E}\left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right] \frac{\tau_b^*}{1-\tau_b^*} + \frac{\beta}{\frac{\partial g(h',h)}{\partial h'}\frac{\partial u(c)}{\partial c}} \mathrm{Cov}\left(\frac{\partial u(c')}{\partial c'}, \frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right), \tag{15}$$

and

$$\Delta_i \equiv -\frac{q}{\frac{\partial g(h',h)}{\partial h'}} \chi \mathbb{E}\left[l'(\theta') \frac{d\mathbf{v}(l'(\theta'))}{dl'} \frac{\frac{\partial A(\theta',h')}{\partial \theta'}\frac{\partial A(\theta',h')}{\partial h'}}{A(\theta',h')^2} \mu'(\theta')\right]. \tag{16}$$

The first component Δ_b relates the wedge for human capital to the wedge for bequests τ_b^*. It should not be surprising that the two wedges are closely related since both forms of capital transfer resources from one generation to the next. The first term in Δ_b is of the same sign as τ_b^*.[14] 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 $\beta\frac{\partial u(c')}{\partial c'}/\frac{\partial u(c)}{\partial c}$ and the discount factor of the planner q. The two discount factors differ because the reciprocal (not the standard) Euler equation holds at the social optimum. As $\tau_b^* \in (0,1)$, the difference is expected to be positive. In order to correct that distortion, 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.[15]

However, bequests and human capital are not perfect substitutes because the return to human capital depends on future ability and is thus risky. The risk adjustment is captured by the second term in (15) 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 the former provides a bad hedge against consumption risk, rendering its accumulation less attractive to families.

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

[15] In other words, the component of the wedge Δ_b is positive if the risk-adjusted return to human capital investment is higher for families than for the planner. See also the expression for Δ_b in equation (26) in the proof of Proposition 4.

The second component Δ_i corresponds to the incentive term in (9) net of the labor supply effect discussed at the end of Section 2.2. We show explicitly in the proof of Proposition 4 that the effect on the incentive constraint through changes in labor supply is exactly offset at the social optimum by the distortion of the human capital decision introduced by the labor wedge. Intuitively, the planner neutralizes the intratemporal distortions for the human capital decision so that the socially optimal investment in human capital differs from the laissez faire only because of the intertemporal distortions. Hence, the decomposition in Proposition 4 does not contain a component relating the human capital wedge to the labor wedge. As in Stantcheva (2014), Δ_i can be interpreted as the *net wedge* because it captures the implicit tax on human capital once the distortions introduced by the wedges for bequests and labor have been compensated for.

When the parameter $\chi = 0$, the production technology is Cobb-Douglas and the net wedge $\Delta_i = 0$. To understand why, it is instructive to rewrite the planner's optimality condition as[16]

$$1 = \frac{q}{\frac{\partial g(h',h)}{\partial h'}} \mathbb{E}\left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right] - \Delta_i \ . \tag{17}$$

If we divide by q, equation (17) shows that the planner equates the social return on bequests $1/q$ to the social return on human capital. If $\Delta_i = 0$, the necessary condition for human capital (17) corresponds to the one prevailing in first-best environments: the planner simply equates the marginal costs and returns of human capital investment.

Figure 1 illustrates why incentive provision does not distort the planner's optimality conditions when the technology is Cobb-Douglas. The formal foundations for Figure 1 are derived in Appendix A.2.1. The concave curves in the figure illustrate a typical allocation in the $\{y, u(c) + \beta V'\}$ plane. The convex curves are the indifference curves of a randomly chosen family. Local incentive compatibility holds when the indifference curve is tangent to the allocation. Figure 1 also reports the effect of a perturbation (derived formally in Appendix A.2.1) which increases human capital and output holding labor supply, consumption and promised utility constant. The allocation resulting from such a perturbation remains incentive compatible because both curves shift and tilt such that incentive compatibility is maintained. The planner can therefore set human capital so as to maximize revenues without affecting the incentive compatibility of the allocation.

This result does not hold when χ differs from zero. Then the perturbation described above does not preserve incentive compatibility. For example, if ability and human capital are so complementary that $\chi < 0$, high types are able to use human capital relatively more efficiently and the indifference curve intersects the perturbed allocation from above. It is not optimal anymore for the family to keep its labor supply unchanged. Instead, it

[16] Equation (17) follows replacing τ_h by $\Delta_b + \Delta_i$ on the left hand side of the definition (12), and noticing that

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

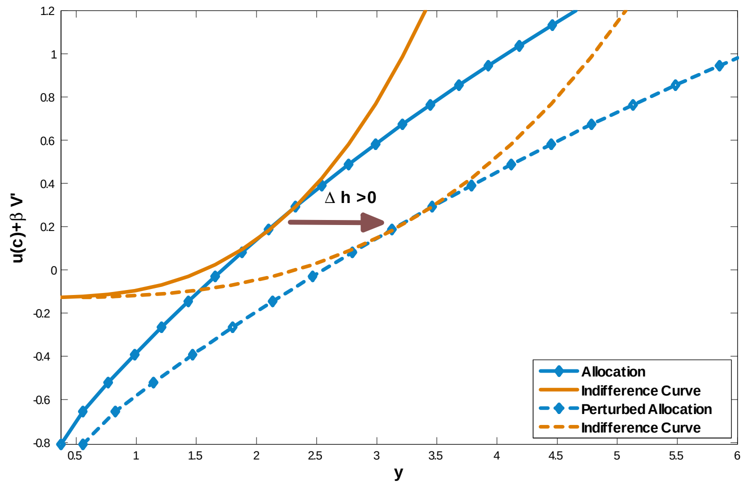

Figure 1: Effect of perturbation $\Delta h = 4$ when $\chi = 0$. Notes: Plot for a family with zero assets and 13 years of human capital; parameter values specified as in Table 1, Section 4.

will reduce its effort and imitate lower types.[17] In other words, increasing human capital makes it more difficult to elicit truthful reporting from families when $\chi < 0$.

To summarize, Δ_i measures the wedge between the first and second best investment rules due to the impact that human capital has on the implementability of the allocation. This effect is not internalized by families because they take the allocation as given. They ignore the impact of their investments on the incentive compatibility of the allocation which drives a wedge between their optimal choice and that of the planner. Whether this net wedge is positive or negative depends on the degree of complementarity between human capital and ability.

3.1 Extensions

Liquidity constraints.—We have not ruled out negative bequests. Given our focus on a family unit of parents and children, this allows us to capture children of low-income families who enter their working life with debt because they take on loans to finance education and do not receive bequests. One may argue, however, that a negative b is not plausible since parents cannot require children to make transfers to them and that children within a family may not be able to take on debt obligations. In our model this corresponds to the constraint $b' \geq 0$ ensuring that bequests cannot be negative. We characterize in Appendix A.2.2 how the possibility of a binding liquidity constraint affects the wedges. It 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

[17]The opposite adjustment occurs when $\chi > 0$ as the indifference curve intersects the allocation from below, making it more easy to motivate families.

become larger instead to offset that a binding constraint increases resources of the future generation.

Persistent types.—Our main results hold in a model where, instead of being independent, types are correlated across generations. Adding this feature captures the genetic transmission of characteristics from parents to their offspring. We show in Appendix A.2.3 that, if we let the density $f(\cdot)$ from which children's abilities are drawn vary with the type of the parents, then the incentive effect of human capital also depends on whether the probability distribution satisfies the monotone likelihood-ratio property.[18] This restriction is commonly assumed to ensure that the value of private information is positive, since families who underreport their types have higher expectations about the ability of their children than the planner. Furthermore, we show that an expression analogous to that in Proposition 4 holds for the constrained-efficient human capital wedge, so that it continues to be closely related to the wedge for bequests even when we allow for persistent ability shocks.

3.2 Comparison with the literature

We find that socially optimal distortions of human capital investment are tightly related to the optimal distortions of bequests. Compared to bequests, human capital carries risk and may change incentives. If $\chi < 0$, human capital worsens incentives ($\Delta_i > 0$) so that the social return on human capital in equation (17) is reduced (vice versa if $\chi > 0$). It follows that the planner distorts the family's decision between bequests and human capital investment to equate their social return. This result is different from Farhi and Werning (2010), who abstract from risk and do not consider the effect of human capital on incentives, and is analogue to the independent complementary research in Section 6 of Stantcheva (2015).

Findeisen and Sachs (2012) find a negative incentive effect of human capital assuming that more human capital and higher innate ability both favorably shift the distribution function of labor market productivity. 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. If innate ability and human capital are less complementary than in the Cobb-Douglas case, 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

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

stronger for agents with low innate ability. In this case more human capital alleviates the incentive problem, opposite to the result in Findeisen and Sachs (2012).

As Stantcheva (2014), we find that human capital leaves the incentive constraint unchanged solely if the technology of production is Cobb-Douglas. We reach similar conclusions although, in our specification of the allocation problem, the planner chooses output while the planner in Stantcheva (2014) chooses unobservable labor effort. In accordance with the revelation principle, the wedges do not depend on the choice of control variable but their decomposition differs. More precisely, our derivations identify an additional term which captures how human capital investment affects incentives to produce a given level of output by lowering the required labor input. We show in the proof of Proposition 4 that, at the social optimum, this effect exactly offsets the distortion of human capital investment induced by the labor wedge.

Besides this similarity, human capital investment has an immediate effect on productivity in Stantcheva (2014), while it only becomes productive next period in our model. The different timing in Stantcheva (2014) implies that human capital has a direct intratemporal impact on labor supply, leading to a tight relationship between the labor and human capital wedges. Such a link is absent in our economy because the effect of human capital is purely intergenerational, tying it instead to the wedge for bequests. We illustrate this relationship in the next section where we calibrate and simulate the model.

4 Numerical analysis

We uncover further interesting features of the allocation and wedges by solving the model numerically when ability is i.i.d. across generations. This not only facilitates interpretation but also contains the computational burden. 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.

4.1 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 assumption of an annual discount rate of 3%, this implies that $\beta = 0.412$. 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:

$$\ln w(\theta, h) = \ln A(\theta, h) = (1 - \xi) \ln h + \xi \ln \theta. \tag{18}$$

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).[19] 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.[20] 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 (18) implies that $1 - \xi = 0.1$ and $\xi = 0.9$.

TABLE 1: Calibration

Parameters	Model	Target	Source
Utility function			
$\beta = q = 0.412$	Discount rate	Annualized 3%	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)
$\kappa = .087$	$g(h', h) = \kappa [h'^{\varsigma} - 1]$	secondary education	

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

[19] The log-normal distribution is approximated by a truncated density so that we draw the next-period ability from a compact interval $[\underline{\theta}, \overline{\theta}]$.

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

schooling starting from the first year of upper-secondary education, i.e., grade 9 in the U.S.

We 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 $\varsigma = 0.214$.[21]

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 numéraire.[22] According to census data, the mean annual earnings of high-school dropouts have been equal to \$20,241 in 2010.[23] 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 $\kappa = 0.0208/\left[\exp(1)^{0.214} - 1\right] = 0.087$.

We initialize the level of human capital so that h_1 corresponds to high-school graduation ($S = 4$). The specification of the initial promised value ensures that the planner breaks even when we account for the cost of compulsory education, i.e., $\Gamma(V_1, h_1, 1) = 0$. We solve the model for a dynasty of four generations (120 years) and focus on the decisions of the second generation.[24] The algorithm, discussed further in Appendix A.3, follows Farhi and Werning (2013) closely.

[21] 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 $(\exp(8)^{\varsigma} - 1) / (\exp(4)^{\varsigma} - 1)$.

[22] 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 $y^*(1,1) = 1$.

[23] 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)

[24] We show in the Appendix A.3 that, due to the high discount factor, increasing the horizon of the dynasty has a negligible effect when T is larger or equal to 3.

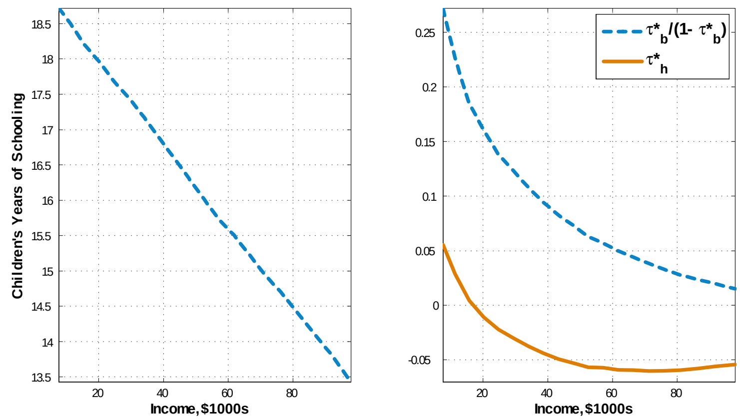

Figure 2: Investment in children's human capital and wedges for human capital and bequests, as a function of labor income for parents with zero assets and 13 years of schooling.

4.2 Results

Human capital investment and parental income.—The left panel of Figure 2 shows that optimal investment into education decreases in the income, and thus 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 is constant in parents' ability because types are uncorrelated across generations. Under 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 this by giving children of high ability parents more consumption and by letting them produce less output, thus reducing their disutility of labor. In the decentralized allocation, children inherit higher bequests when the productivity of their parents is high, which entails a negative wealth effect on their labor supply.

The smaller labor effort of children of high-ability parents makes it, in turn, less attractive for the planner to invest into their human capital. This is why, in Figure 2, children of very able parents only receive 13.5 years of education, roughly corresponding to high-school graduation, whereas hard-working children of low-ability parents complete 5 more years of education to obtain a master's degree. Optimal education hence exerts a mean-reverting influence across generations by ensuring that labor market productivity of parents and expected labor market productivity of their offspring are negatively correlated.

This prediction for the social optimum is strikingly at odds with the observed positive correlation between the college attendance rate of children and the percentile rank of parental income documented in Chetty et al. (2014), Table 1. Taking the model at face value, this suggests that the observed allocation in the U.S. is not constrained efficient: low-income families undereducate their children while high-income families overeducate them. Alternatively, one may question the assumption in the model that ability is i.i.d. across generations. We expect the difference between the empirical observations and the model predictions to be smaller when ability is persistent across generations. Although there is no consensus on 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 investment 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. An interesting question for future research is how persistent ability has to be in order to make human capital investment increase in parent's ability, and whether such persistence across generations is at all plausible.

The human capital wedge.—The right panel of Figure 2 plots τ_h^* against labor income and shows that the wedges for bequests and human capital are tightly related but not identical. As is to be expected from results on wedges for savings in Kocherlakota (2010), chapters 3 and 4, the wedge for bequests is regressive: it is decreasing in income because the planner wants to discourage families with bequests to shirk and report a low type. It follows from Proposition 4 that the wedge for human capital in the calibrated Cobb-Douglas case reads

$$\tau_h^* = \frac{\tau_b^*}{1-\tau_b^*} + \frac{\beta}{\frac{\partial g(h',h)}{\partial h'}\frac{\partial u(c)}{\partial c}} \mathrm{Cov}\left(\frac{\partial u\left(c'\right)}{\partial c'}, \frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right). \tag{19}$$

It differs from the wedge for bequests because human capital carries risk. Since $\tau_h^* < \tau_b^*/(1-\tau_b^*)$ in Figure 2, the covariance term in equation (19) is negative, showing that human capital is a bad hedge for consumption risk.

The planner does not have to discourage human capital accumulation as much as bequests because parents cannot diversify the risk associated with their children's ability. This explains why 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 very low-ability types who have the lowest income. Moreover, the gap between the two wedges narrows as labor income increases because the offspring will enjoy higher consumption on average. This lowers the absolute value of the covariance term in (19) if $u'''(c) > 0$, as is the case for log utility in our numerical solution, so that marginal utility is decreasing at a decreasing rate. Then variation in consumption levels will result in smaller fluctuations of marginal utility at higher consumption levels. For middle-income families this effect dominates the regressivity induced by the wedge for bequests τ_b^*, implying an implicit tax on human capital that is locally flat or even progressive, as illustrated in Figure 2 by the positive slope of τ_h^* at high income levels.

Education and family background.—Besides differences in labor income, the model also

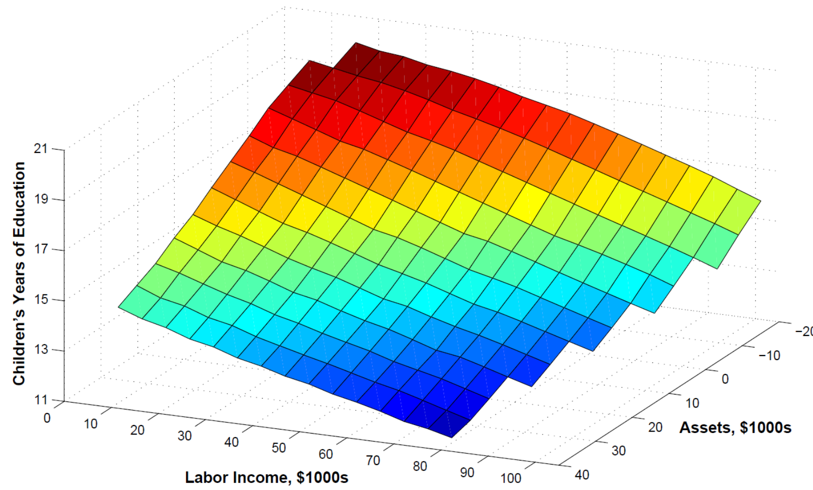

Figure 3: Human capital accumulation as a function of families' inherited wealth and labor income for families with parents that have 13 years of education.

allows us to evaluate the influence of inherited wealth on educational choices.[25] Figure 3 shows that parents' wealth and labor income have qualitatively similar effects on the optimal level of education. This is intuitive since both increase the resources available for the next generation. As explained above, the resulting wealth effect lowers the labor effort of the offspring, making it less attractive for the planner to provide them with education.

Complementarity and incentive wedge.—The numerical results are based on the canonical Mincerian specification of the wage equation for which wages are log-linear in schooling, with a constant slope across ability groups. Our model allows us to check the importance of the Cobb-Douglas assumption if we change the degree of complementarity between θ and h by changing the parameter χ of the production function. As shown in Proposition 4, $\chi \neq 0$ introduces an additional motive for the planner to tax or subsidize human capital investment. The results presented in Appendix A.3 show that the shape of the human capital wedge remains very similar and its level is shifted by the incentive effect of human capital investment Δ_i. Since labor effort is highest for the children of low-ability parents, the incentive effect is monotonically decreasing in parents' ability.

[25]We vary the promised value V_1 which corresponds to varying assets of families in the decentralized allocation. To simplify the interpretation of the figures, we thus replace the promised values by the corresponding levels of bequest, $b(V_1, h_1, 1) = \Gamma(V_1, h_1, 1)$, in the decentralized allocation. The human capital of parents does not matter much quantitatively in our calibration so that we hold it constant at 13 years of education.

5 Implementation

An important question is how the solution of the planner's problem can be implemented in a decentralized economy. One possibility is to rely on education loans that are contingent on the whole history of loans and earnings. The socially optimal allocation can then be implemented by combining these history-dependent repayments with taxes on labor income and bequests that condition only on contemporaneous variables.

The history dependence of the tax system becomes much simpler, however, when ability types are i.i.d. Then the history can be summarized by two state variables, bequests and human capital. This makes it possible to implement the constrained efficient allocation 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 that only depend on current income.

Existing tax and subsidy systems for student loans in continental Europe and Anglo-Saxon countries contain elements which resemble these implementation schemes. 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).[26]

Given the tax schedule $T(b, h, y, h')$, agents solve the maximization problem

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

along with the terminal conditions $b_{T+1} = 0$ and $W_{T+1} = 0$. The proof that the optimal allocation can be implemented in this way follows directly from the argument in Albanesi and Sleet (2006), and its extension by Stantcheva (2014) to a setting with human capital. The history dependence in our model is summarized by V, h and t. But one can define a mapping between the state vector $\{V, h, t\}$ and inherited wealth b such that, when its image is inserted in the decentralized program above, the problem $W(b, h, t)$ corresponds to the dual of the planner's program $\Gamma(V, h, t)$. For the duality principle to be satisfied, the resources of the family, as measured by the amount of assets at the beginning of the period, must equal the value of the planner's cost minimization problem, i.e., $b(V, h, t) = \Gamma(V, h, t)$.

Marginal taxes and wedges.—Before presenting the simulation of the decentralized allocation, we discuss how marginal taxes relate to the respective wedges. The first-

[26] 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.

order condition for labor supply and the definition of the labor wedge (11) imply that the marginal income tax equals the labor wedge: $\partial T(\cdot)/\partial y = \tau_l$. Concerning τ_b, the first-order condition with respect to b' implies that

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

The wedge for bequests τ_b generally has to be implemented by taxes that ensure that it also holds *ex post*. Then, the Euler equation of families is satisfied for each consumption level at the reported 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.

To see how the marginal tax on human capital investment is related to the wedge τ_h, we combine the first-order condition for human capital and the definition of the wedge for human capital (12):

$$\frac{\partial T(\cdot)}{\partial h'} = \frac{\partial g(h', h)}{\partial h'}\tau_h - \beta \mathbb{E}\left[\left(\frac{\partial y'}{\partial h'}\frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'}\right)\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}}\right]. \tag{20}$$

As pointed out by Stantcheva (2014), 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 the latter 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 (20) allows us to relate our results to Bovenberg and Jacobs (2005) who show that education expenses should be fully tax deductible to avoid distortions of human capital investment. We recover the analogon of this result in our model: if, as in Bovenberg and Jacobs (2005), there are no bequests and productivity is Cobb-Douglas ($\chi = 0$), $\tau_h = 0$ and human capital accumulation is socially optimal in the laissez faire without tax distortions. Then equation (20) reads

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

The current marginal tax on human capital accumulation $\partial T(\cdot)/\partial h'$ is negatively related to the expected change in the risk-adjusted return to human capital for the next generation caused by the change in taxes. This expected change consists of two terms. The first term is the expected change in taxes $\mathbb{E}\left[\frac{\partial y'}{\partial h'}\frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'}\right]$ 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 in taxes due to the higher human capital stock of the next generation. The second term captures that the returns to human capital are uncertain in our model so that it matters whether the tax changes reduce consumption risk. Hence, education should be subsidized if human capital investment increases the tax burden and the future tax changes caused by human capital accumulation do not reduce consumption risk too much.

This is not the whole story in our model, however, since $\tau_h^* = \Delta_b + \Delta_i \neq 0$ at the social optimum by Proposition 4. Hence, optimal taxes or subsidies do not only try to offset how human capital alter future tax payments but also account for (i) the distortions at the intertemporal margin relative to bequests (captured by Δ_b), and (ii) the distortions due to changes in the power of incentives (captured by Δ_i).

Illustration of the optimal tax schedule.—We define means-tested grants as

$$G(y, h'|b, h) = -T(b, h, y, h').$$

Figure 4 shows how means-tested grants depend on parents' assets and labor income. Families that are poor in both dimensions receive the largest grants because they invest more in the education of their children (see Figure 3). The grant decreases in assets and labor income for poor families as one would expect. When families have sufficient resources in terms of assets and income, the grant becomes negative so that families have to pay for the education of their children.

The grant is always decreasing in labor income but its progressivity is decreasing in the level of assets. Given the mapping between assets and promised utility, the planner's possibility to redistribute across families with higher assets is more constrained by incentives. This is why the grant is progressive for most but not all combinations of assets and income: the exception are families with high income who have to pay less if they hold more assets.

Loans with contingent repayments.— We have degrees of freedom in the implementation of the social optimum. This flexibility allows us to decompose the consolidated tax schedule $T(b, h, y, h')$ into a set of fiscal instruments resembling actual ones.

We introduce a tax/transfer schedule that depends solely on current labor income, and follow the parametrization proposed by Heathcote et al. (2014) for the U.S.: $T^y(y) = y - y^{1-\tau}$, with $\tau = 0.15$.[27] Assuming that loans fully fund education expenditures, i.e., $L(h', h) = g(h', h)$, the optimal debt-repayment schedule solves

$$\underbrace{T^y(y)}_{\text{Income tax}} - \underbrace{L(h', h)}_{\text{Education loan}} + \underbrace{D(b, h, y, h')}_{\text{Net payment}} = \underbrace{T(b, h, y, h')}_{\text{Gross payment}}.$$

The consolidated repayment schedule $D(\cdot)$ includes not only the repayment of the education loan of the parents but also any residual payments not covered by the income tax $T^y(\cdot)$. We spell out the dynamic program for this implementation and sketch properties of $D(\cdot)$ in appendix A.2.4.

[27] Observe that the schedule $T^y(y)$ is negative whenever $y < 1$. Given that a unit in our model corresponds to mean annual earnings of high-school dropouts, as explained in Section 4, workers whose yearly income is below \$20,241 receive positive transfers while others pay income taxes.

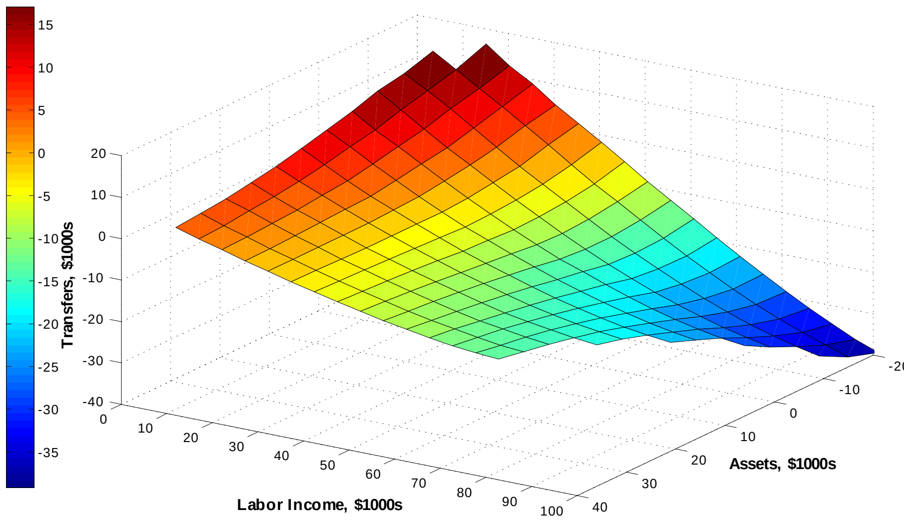

Figure 4: Implementation of the social optimum with means-tested grants for families with parents that have 13 years of education.

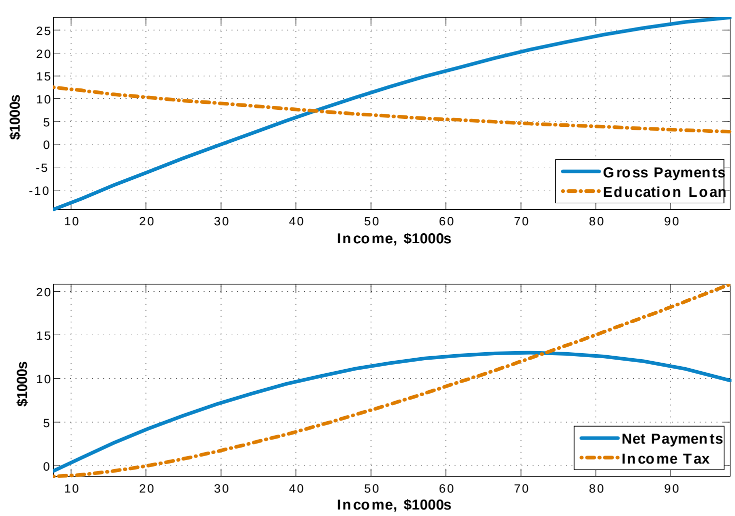

Figure 5: Implementation of the social optimum with loans and contingent payments for families with zero assets and parents with 13 years of education.

Figure 5 illustrates how the implementation scheme depends on labor income for our representative family with no assets and parents with 13 years of education. Its upper-panel shows that education loans decrease in income because years of schooling are also decreasing in income in the social optimum (see the left panel of Figure 2). Consistent with the results on means-tested grants reported in Figure 4, the gross payment in Figure 5 is increasing in labor income with a declining slope. Given the progressive U.S. income tax schedule, the net payment has to balance insurance and incentives to implement the gross payment required in the social optimum. Since the education loan decreases in income and the income tax is convex in income, the lower panel of Figure 5 shows that the schedule for the net payment becomes regressive for families with more than $\$70,000$ in labor income. This prediction is in line with the finding in Heathcote et al. (2014) that the value of progressivity of the income tax that maximizes social welfare is lower than the one observed in the data. Although the net payment compensates some of the progressivity of observed income taxes, it provides substantial insurance: low-income families do not have to repay the education loans of the parents but even benefit from small net transfers, while receiving the largest loans for the education of their children.

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 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, and because human capital may change incentives. Our numerical results illustrate how the constrained efficient allocation can be implemented by means-tested grants or loans with contingent repayments. Constrained efficiency requires that low-income families invest more into the human capital of their children than high-income families: since children from a privileged background receive larger bequests, a wealth effect lowers their labor supply so that it is less efficient for the planner to invest into their human capital.

Why, instead, is human capital investment positively related to family income in the real world? It is important to investigate this question in future research. If intergenerational ability is persistent and not i.i.d. as in our simulations, then the empirically observed positive correlation between family income and human capital may be (close to) socially optimal. If instead income-rich families invest more into their children because credit markets are imperfect, or the timing or taxation of bequests and other intergenerational transfers is suboptimal, social welfare may be improved by redistributing resources. Distinguishing between these two possibilities is challenging, given the scarcity of data for intergenerational analysis in many countries, but vital from a policy perspective.

A Appendix

A.1 Proofs

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

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

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

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

It follows that

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

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

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

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 and that $\partial^2 U(\cdot)/(\partial c(\theta)\,\partial\theta) = 0$:

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

The last term 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

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

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

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

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]**:*

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

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

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

Differentiating these expressions a second time, we get

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

and

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

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

Proof. Corollary 1: Follows immediately from Remark 1 above. ■

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

$$-\frac{\partial \mathbf{U}(c,l)}{\partial c} + \beta\frac{\partial W(b',h',t+1)}{\partial b'} = 0,$$

which, reinserting the envelope condition

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

yields the Euler equation

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

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

$$\frac{\partial \mathbf{U}\left(c,l\right)}{\partial l}+\beta\frac{\partial W\left(b',h',t+1\right)}{\partial b'}\frac{\partial y}{\partial l}=0.$$

The results above imply

$$\beta\frac{\partial W\left(b',h',t+1\right)}{\partial b'}\frac{\partial y}{\partial l}=\frac{\partial y}{\partial l},$$

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

$$-\frac{\partial g(h',h)}{\partial h'}\frac{\partial \mathbf{U}\left(c,l\right)}{\partial c}+\beta\frac{\partial W\left(b',h',t+1\right)}{\partial h'}=0.$$

Using the envelope condition

$$\frac{\partial W\left(b',h',t+1\right)}{\partial h'}=\int_{\Theta}\frac{\partial \mathbf{U}\left(c',l'\right)}{\partial c'}\left[\frac{\partial y'}{\partial h'}-\frac{\partial g(h'',h')}{\partial h'}\right]dF\left(\theta'\right),$$

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

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

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

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

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

By Remark 1,

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

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 adding τ_h on both sides of condition (6), and substituting its definition on the right-hand side to get

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

Since the derivatives of the multivariate function $v\left(y,h,\theta\right)$ 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 as

$$\begin{aligned}
\tau_h &= \underbrace{\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_{\Theta}\frac{\partial y'}{\partial h'}\left(1 - \frac{\frac{\partial v(y',\theta',h')}{\partial y'}}{\frac{\partial u(c')}{\partial c'}}\right)dF\left(\theta'\right)}_{\equiv\Delta_l} \\
&\quad \underbrace{+\frac{1}{\frac{\partial g(h',h)}{\partial h'}}\int_{\Theta}\left(\beta\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} - q\right)\left(\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right)dF\left(\theta'\right)}_{\equiv\Delta_b} \\
&\quad -\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_{\Theta}\frac{\partial^2 v(y',\theta',h')}{\partial\theta'\partial h'}\mu'\left(\theta'\right)d\theta'.
\end{aligned} \qquad (26)$$

The terms Δ_l and Δ_b capture the cross-distortion induced by the wedges for labor and bequests, respectively. Since $\mathbb{E}\left[\beta\frac{\partial u(c')/\partial c'}{\partial u(c)/\partial c} - q\right] = q\frac{\tau_b}{1-\tau_b}$, the definition (10) of the wedge for bequests implies that Δ_b is given by (15).

The third term on the right hand side of (26) is the overall (gross) incentive effect of h. We now show that part of it offsets Δ_l and that the remaining term corresponds to the net wedge Δ_i. To see this, we use results from Remark 1 to rewrite

$$\begin{aligned}\frac{\partial^2 v(y,\theta,h)}{\partial\theta\partial h} &= -\frac{\partial^2 v(y,\theta,h)}{\partial\theta\partial y}\frac{\partial y}{\partial h} \\ &+\frac{\partial \mathbf{v}(l)}{\partial l}\frac{l}{A^2}\frac{\partial A(\theta,h)}{\partial\theta}\frac{\partial A(\theta,h)}{\partial h}\left(1-\frac{\frac{\partial^2 A(\theta,h)}{\partial\theta\partial h}A(\theta,h)}{\frac{\partial A(\theta,h)}{\partial\theta}\frac{\partial A(\theta,h)}{\partial h}}\right) .\end{aligned} \tag{27}$$

The first-order condition (7) for output implies that

$$\int_\Theta \frac{\partial^2 v(y',\theta',h')}{\partial\theta'\partial y'}\frac{\partial y'}{\partial h'}\mu'(\theta')d\theta' = \int_\Theta\left[\frac{\frac{\partial v(y',\theta',h')}{\partial y'}}{\frac{\partial u(c')}{\partial c'}}-1\right]\frac{\partial y'}{\partial h'}f(\theta')\,d\theta'.$$

Hence, using the decomposition (27) to compute the integral below, we find that, under assumptions **[A1']** and **[A2]**,

$$-\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_\Theta \frac{\partial^2 v(y',\theta',h')}{\partial\theta'\partial h'}\mu'(\theta')\,d\theta' = -\Delta_l+\Delta_i\ , \tag{28}$$

where we have used that $\chi = 1 - A_{\theta h}(\theta,h)\,A(\theta,h)\,/\,[A_\theta(\theta,h)\,A_h(\theta,h)]$. This decomposes the overall incentive effect of human capital investment into two components. The first component, when evaluated at the optimal level of output, exactly offsets the distortion for human capital investment induced by the labor wedge. This is why, when substituting (28) into (26), we obtain (16). ■

A.2 Additional results (Web-Appendix)

A.2.1 Foundations for Figure 1

We show that, after the perturbation described in the main text, the allocation remains incentive compatible only if $A(\theta,h)$ is Cobb Douglas, i.e., if $\chi=0$. Let

$$w(r) = u(c(r)) + \beta\Omega'(r)\ ,$$

where r is the report. The reporting problem of the dynasty is

$$\begin{aligned}\Omega(\theta) &= \max_r w(r) - v(y(r),\theta,h)\\ \text{s.t. } v(y(r),\theta,h) &= \mathbf{v}\left(\frac{y(r)}{A(\theta,h)}\right) .\end{aligned}$$

The first-order condition evaluated at the truthful θ reads

$$\left.\frac{\partial w(r)/\partial r}{\partial y(r)/\partial r}\right|_{r=\theta} = \left.\frac{\partial v(y(r),\theta,h)}{\partial y}\right|_{r=\theta} . \tag{29}$$

Consider the following perturbation of increasing human capital, increasing output such that labor supply remains constant, and holding consumption and the continuation value w constant:

$$h^{\delta} = h + \delta,\ y^{\delta}(\theta) = y(\theta)\frac{A(\theta, h^{\delta})}{A(\theta, h)},\ l^{\delta}(\theta) = l(\theta),\ w^{\delta}(\theta) = w(\theta).$$

Then

$$\frac{\partial v\left(y^{\delta}(\theta), \theta, h^{\delta}\right)}{\partial y} = \frac{\partial \mathbf{v}\left(\frac{y^{\delta}(\theta)}{A(\theta, h^{\delta})}\right)}{\partial y} = \frac{\partial \mathbf{v}(l(\theta))/\partial\theta}{A(\theta, h^{\delta})} = \frac{\partial v(y(\theta), \theta, h)}{\partial y}\frac{A(\theta, h)}{A(\theta, h^{\delta})}. \tag{30}$$

Since

$$\begin{cases} \partial y(\theta)/\partial\theta = \frac{\partial l(\theta)}{\partial\theta}A(\theta, h) + l(\theta)\frac{\partial A(\theta, h)}{\partial\theta} \\ \partial y^{\delta}(\theta)/\partial\theta = \frac{\partial l(\theta)}{\partial\theta}A(\theta, h^{\delta}) + l(\theta)\frac{\partial A(\theta, h^{\delta})}{\partial\theta} \end{cases},$$

(29) implies

$$\left.\frac{\partial w(r)/\partial r}{\partial y(r)/\partial r}\right|_{r=\theta} = \left[\frac{\frac{\partial l(\theta)}{\partial\theta}A(\theta, h) + l(\theta)\frac{\partial A(\theta, h)}{\partial\theta}}{\frac{\partial l(\theta)}{\partial\theta}A(\theta, h^{\delta}) + l(\theta)\frac{\partial A(\theta, h^{\delta})}{\partial\theta}}\right]\left.\frac{\partial w^{\delta}(r)/\partial r}{\partial y^{\delta}(r)/\partial r}\right|_{r=\theta}.$$

Reinserting this condition along with (30) into the first-order condition (29), we get

$$\left[\frac{\frac{\partial l(\theta)}{\partial\theta}A\left(\theta, h^{\delta}\right) + l(\theta)\frac{\partial A(\theta, h^{\delta})}{\partial\theta}}{\frac{\partial l(\theta)}{\partial\theta}A(\theta, h) + l(\theta)\frac{\partial A(\theta, h)}{\partial\theta}}\right]\left.\frac{\partial w^{\delta}(r)/\partial r}{\partial y^{\delta}(r)/\partial r}\right|_{r=\theta} = \frac{\partial v\left(y^{\delta}(\theta), \theta, h^{\delta}\right)}{\partial y}\frac{A\left(\theta, h^{\delta}\right)}{A(\theta, h)}. \tag{31}$$

In general, this condition is not compatible with the first-order condition for truthful reporting after the perturbation:

$$\left.\frac{\partial w^{\delta}(r)/\partial r}{\partial y^{\delta}(r)/\partial r}\right|_{r=\theta} = \frac{\partial v\left(y^{\delta}(\theta), \theta, h^{\delta}\right)}{\partial y}. \tag{32}$$

Hence, the allocation after the perturbation is not incentive compatible.

Let us now illustrate this using the CES production technology specified in [**A2**]. Then $A_{\theta}(\theta, h) = \xi\theta^{\chi-1}A(\theta, h)^{1-\chi}$ which can be reinserted into (31) to obtain

$$\left[\frac{\frac{\partial l(\theta)}{\partial\theta}A\left(\theta, h^{\delta}\right) + l(\theta)\xi\theta^{\chi-1}A\left(\theta, h^{\delta}\right)^{1-\chi}}{\frac{\partial l(\theta)}{\partial\theta}A(\theta, h) + l(\theta)\xi\theta^{\chi-1}A(\theta, h)^{1-\chi}}\right]\left.\frac{\partial w^{\delta}(r)/\partial r}{\partial y^{\delta}(r)/\partial r}\right|_{r=\theta} = \frac{\partial v\left(y^{\delta}(\theta), \theta, h^{\delta}\right)}{\partial y}\frac{A\left(\theta, h^{\delta}\right)}{A(\theta, h)},$$

which is equivalent to

$$\left[\frac{\frac{\partial l(\theta)}{\partial\theta} + l(\theta)\xi\theta^{\chi-1}A\left(\theta, h^{\delta}\right)^{-\chi}}{\frac{\partial l(\theta)}{\partial\theta} + l(\theta)\xi\theta^{\chi-1}A(\theta, h)^{-\chi}}\right]\left.\frac{\partial w^{\delta}(r)/\partial r}{\partial y^{\delta}(r)/\partial r}\right|_{r=\theta} = \frac{\partial v\left(y^{\delta}(\theta), \theta, h^{\delta}\right)}{\partial y}.$$

This condition differs from (32) but for the special case where $A(\cdot)$ is Cobb-Douglas since then $\chi = 0$, and the ratio in square brackets collapses to unity as one recovers (32). The perturbed allocation thus remains incentive compatible only for the Cobb-Douglas specification. Conversely when $\chi \neq 0$ we have

$$\frac{\frac{\partial l(\theta)}{\partial \theta} + l(\theta)\,\xi\theta^{\chi-1}A\left(\theta, h^{\delta}\right)^{-\chi}}{\frac{\partial l(\theta)}{\partial \theta} + l(\theta)\,\xi\theta^{\chi-1}A(\theta, h)^{-\chi}} \gtrless 1 \text{ when } \chi \lessgtr 0$$

and

$$\left.\frac{\partial w^{\delta}(r)}{\partial r}\right|_{r=\theta} \lessgtr \frac{\partial v\left(y^{\delta}(\theta), \theta, h^{\delta}\right)}{\partial y} \left.\frac{\partial y^{\delta}(r)}{\partial r}\right|_{r=\theta} \text{ when } \chi \lessgtr 0.$$

If $\chi < 0$, this proves that the marginal cost of reporting one's type on the right-hand side exceeds the marginal gain on the left-hand side. Hence, families find it optimal to underreport their type.

For the graphical representation, we illustrate the problem in the $\{y, w\}$ plane. Then the maximization problem reads

$$\begin{aligned} \Omega(\theta) &= \max_{y} w(y) - v(y, \theta, h) \\ \text{s.t. } v(y, \theta, h) &= \mathbf{v}\left(\frac{y}{A(\theta, h)}\right), \end{aligned}$$

and the first-order condition reads

$$\frac{\partial w(y)}{\partial y} = \frac{\partial v(y, \theta, h)}{\partial y}. \tag{33}$$

In the figures, the term on the left-hand side, $\partial w(y)/\partial y$, determines the slope of the allocation curve. The term on the right-hand side, $\partial v(y, \theta, h)/\partial y$, determines the slope of the indifference curve.

A.2.2 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

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

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

Proposition 5 *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:*

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

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

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

which, reinserting the envelope condition

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

yields the Euler equation

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

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

$$\frac{\partial \mathbf{U}(c,l)}{\partial l} + \beta\frac{\partial W(b',h')}{\partial b'}\frac{\partial y}{\partial l} = 0\ .$$

The results above imply

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

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

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

The envelope condition is

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

Noting that

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

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

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

■

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*

$$\tau_b^c \equiv 1 - \frac{q}{\beta}\frac{\partial u(c)/\partial c - \eta}{\mathbb{E}\left[\partial u(c')/\partial c'\right]}, \tag{34}$$

$$\tau_l^c \equiv 1 - \frac{\partial v(y,h,\theta)/\partial y}{\partial u(c)/\partial c - \eta}, \tag{35}$$

$$\tau_h^c \equiv \frac{\beta}{\frac{\partial g(h',h)}{\partial h'}} \int_{\Theta}\left[\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c} - \eta}\left(\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right)\right] dF(\theta') - 1. \tag{36}$$

Combining the results of Propositions 1 and 5, we then find:

Proposition 6 *If bequests are required to be non-negative, the first-order conditions of the planner's problem imply under assumptions **[A1]** and **[A2]** that*

$$\tau_b^{c,*} = 1 - \frac{1}{\mathbb{E}\left[\frac{1}{\frac{\partial u(c')}{\partial c'}}\right]\mathbb{E}\left[\frac{\partial u(c')}{\partial c'}\right]} + \frac{\eta}{\frac{\beta}{q}\mathbb{E}\left[\frac{\partial u(c')}{\partial c'}\right]}, \tag{37}$$

$$\tau_l^{c,*} = -\frac{\partial^2 v(y,h,\theta)}{\partial\theta\partial y}\frac{\mu(\theta)}{f(\theta)} - \frac{\eta}{\frac{\partial u(c)}{\partial c} - \eta}\frac{\frac{\partial v(y,h,\theta)}{\partial y}}{\frac{\partial u(c)}{\partial c}}, \tag{38}$$

$$\tau_h^{c,*} = \Delta_b^c + \Delta_i^c + \Delta_c, \tag{39}$$

with

$$\Delta_b^c \equiv \frac{q}{\frac{\partial g(h',h)}{\partial h'}}\mathbb{E}\left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right]\frac{\tau_b^*}{1-\tau_b^*} + \frac{\beta}{\frac{\partial g(h',h)}{\partial h'}\frac{\partial u(c)}{\partial c}}\mathrm{Cov}\left(\frac{\partial u(c')}{\partial c'}, \frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'}\right),$$

$$\Delta_i^c \equiv -\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\chi\int_{\Theta} l'(\theta')\frac{d\mathbf{v}(l'(\theta'))}{dl'}\frac{\frac{\partial A(\theta',h')}{\partial\theta'}\frac{\partial A(\theta',h')}{\partial h'}}{A(\theta',h')^2}\mu'(\theta')\,d\theta',$$

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

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

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

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

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

Simplifying, we get

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

where $\frac{\partial u(c)}{\partial c}-\eta>0$ since $\frac{\partial W(b',h')}{\partial b'}>0$. To derive the analogous expression for τ_b^c, 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 $\partial u(c)/\partial c$ in condition (5). Condition (5) implies

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

The definition of the wedge τ_b^c can be rearranged to

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

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

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

Solving this expression for τ_b^c results in

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

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):

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

Substituting in the definition of the wedge τ_h^c on the right-hand side, we get

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

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

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

Using the definition of the labor wedge (11) in the unconstrained case, the first term captures the cross-distortion induced by the labor wedge as in the case without constraints (compare equation (26) in the proof of Proposition 4). The second term captures the cross-distortion introduced by the wedge for bequests. This term equals Δ_b^c using that $\mathbb{E}(xy) = \mathrm{Cov}(x,y) + \mathbb{E}(x)\mathbb{E}(y)$ and $\mathbb{E}\left[\beta \frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} - q\right] = q\frac{\tau_b}{1-\tau_b}$, where we use the definition (10) of the wedge for bequests in the unconstrained case. The third term of τ_h^c can be combined with the first term, following the same steps as in the proof of Proposition 4, to derive Δ_i^c. 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.2.3 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 (2014). Following Pavan et al. (2014), the envelope condition in the problem with persistent shocks is:

$$\frac{\partial \omega(\theta)}{\partial \theta} = \frac{\partial U(c, y, h, \theta)}{\partial \theta} + \beta \int_{\Theta} \omega(\theta') \frac{\partial f(\theta'|\,\theta)}{\partial \theta} d\theta'. \tag{40}$$

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

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

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

$$\begin{aligned} & \Gamma(V, \Delta, \theta_-, h, t) \\ = & \min_{\{c,y,h',\Delta',V'\}} \left\{ \int_{\Theta} [c + g(h', h) - y(\theta) + q\Gamma(V', \Delta', \theta, h', t+1)]\, dF(\theta|\,\theta_-) \right\} \\ s.t.\ \omega(\theta) = & \ U(c, y, h, \theta) + \beta V', \\ V = & \int_{\Theta} \omega(\theta)\, dF(\theta|\,\theta_-), \\ \Delta = & \int_{\Theta} \omega(\theta) \frac{\partial f(\theta|\,\theta_-)}{\partial \theta_-} d\theta, \\ \frac{\partial \omega(\theta)}{\partial \theta} = & \ \frac{\partial U(c, y, h, \theta)}{\partial \theta} + \beta \Delta'. \end{aligned} \tag{41}$$

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

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

The costate variable satisfies

$$\frac{\partial \mu(\theta)}{\partial \theta} = -\left[\frac{1}{\partial u(c)/\partial c} - \lambda(\theta_-) - \gamma(\theta_-)\frac{\frac{\partial f(\theta|\theta_-)}{\partial \theta_-}}{f(\theta|\theta_-)} + \frac{\mu(\theta)}{f(\theta|\theta_-)}\frac{\partial^2 U(\cdot)}{\partial \theta \partial c}\frac{\partial c}{\partial \omega(\theta)}\right] f(\theta|\theta_-), \tag{42}$$

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

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

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

$$\frac{\partial \mathcal{H}(\cdot)}{\partial V'} = \left[-\frac{\beta}{\partial u(c(\theta))/\partial c(\theta)} + q\lambda'(\theta)\right] f(\theta|\theta_-) = 0 , \tag{43}$$

$$\frac{\partial \mathcal{H}(\cdot)}{\partial \Delta'} = q\gamma'(\theta) f(\theta|\theta_-) + \beta\mu(\theta) = 0 , \tag{44}$$

$$\frac{\partial \mathcal{H}(\cdot)}{\partial y} = \left[\frac{\partial v(y, h, \theta)/\partial y}{\partial u(c)/\partial c} - 1\right] f(\theta|\theta_-) - \frac{\partial^2 v(y, h, \theta)}{\partial \theta \partial y}\mu(\theta) = 0 , \tag{45}$$

$$\begin{aligned}
\frac{\partial \mathcal{H}(\cdot)}{\partial h'} &= \frac{\partial g(h', h)}{\partial h'} + q\int_\Theta \left(\frac{\frac{\partial v(y', \theta', h')}{\partial h'}}{\frac{\partial u(c')}{\partial c'}} + \frac{\partial g(h'', h')}{\partial h'}\right) dF(\theta'|\theta) \\
&\quad - q\int_\Theta \mu'(\theta')\frac{\partial^2 v(y', \theta', h')}{\partial \theta' \partial h'} d\theta' = 0.
\end{aligned} \tag{46}$$

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 $\mu(\theta)$.

Using (42) to substitute out $\mu(\theta)$ in equation (45), we get

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

The labor wedge is analogous to the one derived by Farhi and Werning (2013), Proposition 2, so that we omit its derivation for brevity and focus on the wedge for human capital.

Proposition 7 *If types θ are persistent, and assumptions **[A1]** and **[A2]** hold, the human capital wedge can be decomposed as*

$$\tau_h^{p,*}=\Delta_b^p+\Delta_i^p$$

with

$$\begin{aligned}
\Delta_b^p &\equiv \frac{q}{\frac{\partial g(h',h)}{\partial h'}}\mathbb{E}_\theta\left[\frac{\partial y'}{\partial h'}-\frac{\partial g(h'',h')}{\partial h'}\right]\frac{\tau_b^p}{1-\tau_b^p}\\
&\quad+\frac{\beta}{\frac{\partial g(h',h)}{\partial h'}\frac{\partial u(c)}{\partial c}}\mathrm{Cov}_\theta\left(\frac{\partial u(c')}{\partial c'},\frac{\partial y'}{\partial h'}-\frac{\partial g(h'',h')}{\partial h'}\right),\\
\Delta_i^p &\equiv -\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\chi\int_\Theta l'(\theta')\frac{d\mathbf{v}(l'(\theta'))}{dl'}\frac{\frac{\partial A(\theta',h')}{\partial\theta'}\frac{\partial A(\theta',h')}{\partial h'}}{A(\theta',h')^2}\mu'(\theta')\,d\theta',
\end{aligned}$$

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 (46) and rearranging, we find that

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

Using the definition of the labor wedge, the first term captures the cross-distortion induced by the labor wedge analogous as in the case without persistence (compare to equation (26) in the proof of Proposition 4). The second term captures the cross-distortion introduced by the wedge for bequests. This term equals Δ_b^p using that $\mathbb{E}(xy)=\mathrm{Cov}(x,y)+\mathbb{E}(x)\,\mathbb{E}(y)$

and $\mathbb{E}\left[\beta\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}}-q\right]=q\frac{\tau_b}{1-\tau_b}$, where we use the definition (10) of the wedge for bequests in the unconstrained case. The third term of τ_h^p is the gross incentive wedge. Combining this wedge with the first term, following the same steps as in the proof of Proposition 4, we derive Δ_i^p. The first term equals Δ_l^p using the definition of the labor wedge analogous to (11) in the unconstrained case. The second term equals Δ_b^p 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^p can be combined with the first term, following the same steps as in the proof of Proposition 4, to derive Δ_i^p. ■

Besides the net wedge Δ_i^p emphasized in the previous proposition, also the gross incentive wedge contains interesting insights. Developing the gross wedge further shows how persistence alters the incentive problem.

Corollary 3 *The overall incentive effect of human capital accumulation, or the gross incentive wedge, reads*

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

Proof. As shown in the proof of Proposition 7, the gross incentive wedge is

$$\Delta_{i,\text{gross}}^p \equiv -\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_\Theta \mu'(\theta')\frac{\partial^2 v(y',\theta',h')}{\partial\theta'\partial h'}d\theta'. \tag{47}$$

Integrating by parts:

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

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

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

Since by (43) and (44),

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

and

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

we get

$$\begin{aligned}
&\Delta^p_{i,\text{gross}} \\
&= -\frac{q}{\frac{\partial g(h',h)}{\partial h'}}\int_\Theta\left[\frac{1}{\frac{\partial u(c')}{\partial c'}} - \frac{\beta}{q\partial\frac{\partial u(c)}{\partial c}} + \frac{\beta\mu(\theta)}{qf(\theta|\,\theta_-)}\frac{\frac{\partial f(\theta'|\theta)}{\partial\theta}}{f(\theta'|\,\theta)}\right]\frac{\partial v(y',\theta',h')}{\partial h'}f(\theta'|\,\theta)d\theta' \\
&= -\frac{q}{\frac{\partial g(h',h)}{\partial h'}\frac{\partial u(c)}{\partial c}}\int_\Theta\left[\frac{\frac{\partial u(c)}{\partial c}}{\frac{\partial u(c')}{\partial c'}} - \frac{\beta}{q} + \frac{A^\chi\theta^{1-\chi}}{\alpha\xi}\frac{\beta}{q}\frac{\tau^p_l}{1-\tau^p_l}\frac{\frac{\partial f(\theta'|\theta)}{\partial\theta}}{f(\theta'|\,\theta)}\right]\frac{\partial v(y',\theta',h')}{\partial h'}f(\theta'|\,\theta)d\theta',
\end{aligned}$$

where the second equality follows from

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

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

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

With persistence there is an additional term. Since 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

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

Hence, the expression for $\Delta^p_{i,\text{gross}}$ in the corollary follows. ■

Compared with the results for i.i.d types, the effect of human capital on the incentive-compatibility constraint in $\Delta^p_{i,\text{gross}}$ also depends on the current labor wedge τ^p_l if ability types are persistent and $\gamma(\theta) > 0$. For the sign of this additional effect, it matters 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. The following corollary specifies sufficient conditions under which human capital investment mitigates the incentive problem if ability is persistent across generations.

Corollary 4 *Under assumptions **[A1]**, **[A1']** and **[A2]**, $\Delta^p_{i,gross} \leq 0$ if: (i) $\chi \geq -\alpha$, (ii) $\frac{\partial f(\theta'|\theta)}{\partial \theta}/f(\theta'|\theta)$ monotonically increases in θ' and (iii) the labor wedge $\tau^p_l \geq 0$. The planner then has a motive to increase human capital accumulation in order to relax the incentive compatibility constraint.*

Proof. The sign of the gross incentive wedge $\Delta^p_{i,\text{gross}}$ depends on the sign of the two covariances. Since

$$sgn\left(\mathrm{Cov}_{\theta}\left(\frac{1}{\frac{\partial u(c')}{\partial c'}}, \frac{\partial v(y',\theta',h')}{\partial h'}\right)\right) = sgn\left(\frac{\partial^2 v(y',h',\theta')}{\partial \theta' \partial h'}\right),$$

the first term in $\Delta^p_{i,\text{gross}}$ is negative iff $\chi \geq -\alpha$, by Remark 1 in Appendix A.1. The second term is negative as well for a positive labor wedge τ^p_l if the ratio $\frac{\partial f(\theta'|\theta)}{\partial \theta}/f(\theta'|\theta)$ is increasing in θ' so that $\mathrm{Cov}_{\theta}\left(\frac{\frac{\partial f(\theta'|\theta)}{\partial \theta}}{f(\theta'|\theta)}, \frac{\partial v(y',\theta',h')}{\partial h'}\right) > 0$. ■

It seems natural that $f_\theta(\theta'|\theta)/f(\theta'|\theta)$ 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^p_l.

Concerning the wedge for bequests, we impose assumption **[A1]**, use equation (42) 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 (2014).

A.2.4 Implementation

We derive how the debt-repayment schedule depends on income and education expenditures if the social optimum is implemented with loans and contingent payments as in Section 5. The maximization problem of the family is

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

The first-order condition with respect to l is

$$\frac{\partial u(c)}{\partial l}\left(\frac{\partial y}{\partial l} - \left(\frac{\partial T^y(y)}{\partial y} + \frac{\partial D(b,h,y,h')}{\partial y}\right)\frac{\partial y}{\partial l}\right) - \frac{\partial \mathbf{v}(l)}{\partial l} = 0.$$

Using the definition of the labor wedge, we find that the debt-repayment schedule depends on income in the following way:

$$\frac{\partial D(b,h,y,h')}{\partial y} = \tau_l - \frac{\partial T^y(y)}{\partial y}. \tag{48}$$

The first-order condition with respect to h' is

$$\begin{aligned} &\left(\frac{\partial g(h',h)}{\partial h'} - \frac{\partial L(h',h)}{\partial h'} + \frac{\partial D(b,h,y,h')}{\partial h'}\right)\frac{\partial u(c)}{\partial c} \\ = \;& \beta\int_{\Theta}\left(1 - \frac{\partial T^{y\prime}(y')}{\partial y'} - \frac{\partial D'(b',h',y',h'')}{\partial y'}\right)\frac{\partial y'}{\partial h'}\frac{\partial u(c')}{\partial c'}dF(\theta') \\ & -\beta\int_{\Theta}\left(\frac{\partial g(h'',h')}{\partial h'} - \frac{\partial L'(h'',h')}{\partial h'} + \frac{\partial D'(b',h',y',h'')}{\partial h'}\right)\frac{\partial u(c')}{\partial c'}dF(\theta'). \end{aligned}$$

Using (48) and the assumption that loans fully fund education expenditures, so that $L(h',h) = g(h',h)$, the debt-repayment schedule depends on human capital accumulation as follows:

$$\frac{\partial D(b,h,y,h')}{\partial h'}\frac{\partial u(c)}{\partial c} = \beta\int_{\Theta}\left[(1-\tau_l')\frac{\partial y'}{\partial h'}\frac{\partial u(c')}{\partial c'} - \frac{\partial D'(b',h',y',h'')}{\partial h'}\right]\frac{\partial u(c')}{\partial c'}dF(\theta').$$

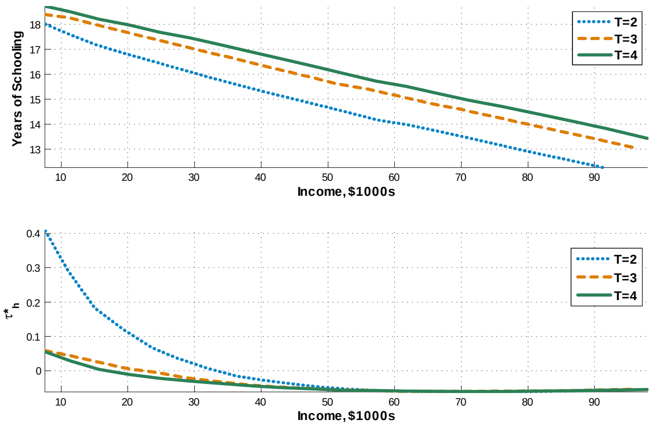

Figure 6: Investment in children's human capital and wedges for human capital, as a function of labor income for parents with zero assets and 13 years of schooling.

A.3 Numerical solution and further numerical results (Web-Appendix)

Since the algorithm closely follows Farhi and Werning (2013), we succinctly describe its structure and refer readers to their paper for a thorough description of the numerical approach. The simulation proceeds by backward induction starting from the last period/generation T, with the normalization $V_{T+1} = 0$. It proves convenient for computational purposes to replace the promised value V with the multiplier λ associated to the promise keeping constraint (3). In each period t, arbitrary values are set for the minimum and maximum values of the two states variables λ_t and h_t. Only the lower bound for human capital at zero has a natural interpretation. The resulting vectors are discretized and combined to construct a grid of points that cover a region of the state space.

Then the boundary value problem defined in equation (8) is solved for each point of the grid by applying the following procedure. To compute the law of motion of the costate μ_t, we first use the first-order conditions (6) and (7) to determine the optimal values of y_t and h_{t+1}. These solutions allow us to infer optimal consumption c_t and thus to determine the value of the multiplier λ_t using the first-order condition (5). This, in turn, gives us the law of motion of the costate μ_t. The solver iterates until the starting guess for $\omega_t(\underline{\theta})$ yields a solution that is consistent with the terminal condition in (8). The final solutions are stored and integrated over the interval $[\underline{\theta}, \bar{\theta}]$ so as to obtain the continuation values for the solution of the problem in the previous period (next iteration). These steps are repeated until the initial date $t = 1$ is reached.

The dynasty horizon T.—We solve the problem for a horizon of $T = 4$. This is equivalent to assuming that agents are altruistic towards their great-grandchildren but not towards their great-great-grandchildren. This arbitrary limit is chosen for computational

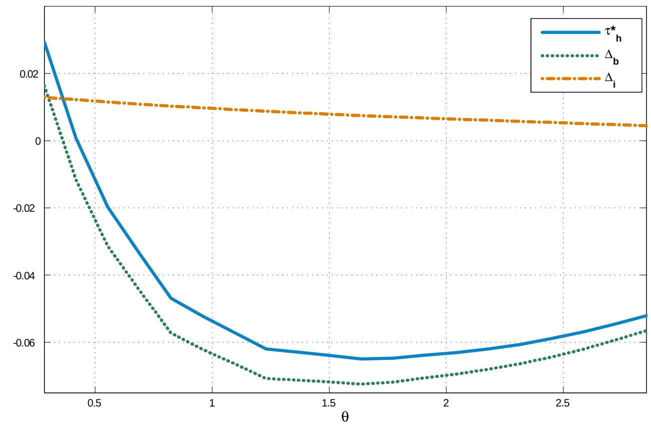

Figure 7: Decomposition of the wedge for human capital as a function of ability θ for parents with zero assets and 13 years of schooling when $\chi = -0.1$.

reasons, since the state space has to be enlarged as the horizon increases, so as to avoid corner solutions. Fortunately, the impact of T on the numerical solution decreases rapidly because generations are separated by a period that is 30 years long. Hence, the discount factor β^T converges quickly to zero as T increases so that in $t = 1$ the utility of the fourth and fifth generation only obtains a weight of 7% and 2.9%, respectively. Extending the horizon T to more than four generations should thus have only a small effect on the numerical solution.

This intuition is confirmed by Figure 6 which plots human capital investment and the related wedge τ^*_h for the different horizons $T = 2, 3, 4$. The upper-panel shows that the investment in education is increasing in T. This is intuitive since a longer horizon extends the benefits from human capital investment for the planner. As expected, the change in human capital investment is much larger for a change in the horizon from $T = 2$ to $T = 3$, than from $T = 3$ to $T = 4$. The small changes between $T = 3$ and $T = 4$ suggest that the solution for the horizon $T = 4$ is already quite a good approximation for longer horizons.

This is confirmed by the results for the human capital wedge, reported in the lower panel of Figure 6. We see that an increase in the horizon from $T = 2$ to $T = 3$ has a large effect on τ^*_h for low types with low income levels. But again, the impact of a longer horizon T is very small when the horizon increases from $T = 3$ to $T = 4$. The results reported in Figure 6 hence suggest that the chosen horizon $T = 4$ strikes a reasonable balance between computing time and numerical accuracy.

Complementarity between ability and human capital χ.—Since estimates on the degree of complementarity between ability and education are scarce and hard to map into our model, we have based our calibration on the Cobb-Douglas specification, setting $\chi = 0$ in [**A2**]. We now check robustness of our results if human capital and ability are more complementary than in the Cobb-Douglas case, as suggested by evidence in Cunha et al.

(2006). Then $\chi < 0$ and the constrained efficient wedge for human capital $\tau_h^* = \Delta_b + \Delta_i$, where the incentive term $\Delta_i > 0$.

Figure 7 reports τ_h^* along with its two components Δ_b and Δ_i, when $\chi = -0.1$, $T = 2$ and all the other parameters are as in Table 1. As expected, the incentive term Δ_i now drives a positive wedge between τ_h^* and Δ_b. This wedge shrinks as θ increases because children of talented parents provide less effort on average, and thus derive less benefits from the complementarity between θ and h. The shape of τ_h^* remains very similar compared with the benchmark case in Figure 6. By contrast, the level of τ_h^* is smaller because more complementarity makes human capital more risky: if children turn out to have low ability, more complementarity lowers the return to human capital investment.

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] Chetty, Raj, Nathaniel Hendren, Patrick Kline and Emmanuel Saez (2014): "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States", Quarterly Journal of Economics, vol. 29, 1553–1623.

[11] 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.

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

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

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

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

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

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

[18] 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.

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

[20] 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.

[21] 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.

[22] Heathcote, Jonathan, Kjetil Storesletten and Giovanni L. Violante (2014): "Optimal Tax Progressivity: An Analytical Framework," Manuscript, Federal Reserve Bank of Minneapolis.

[23] 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.

[24] Kapička, Marek (2015): "Optimal Mirrleesian Taxation in a Ben-Porath Economy," American Economic Journal: Macroeconomics, vol. 7(2), 219–248.

[25] Kapička, Marek and Julian Neira (2015): "Optimal Taxation in a Life-Cycle Economy with Endogenous Human Capital Formation," Manuscript, U.C. Santa Barbara.

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

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

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

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

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

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

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

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

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

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

[36] Stantcheva, Stefanie (2015): "Optimal Income, Education and Bequest Taxes in an Intergenerational Model," Manuscript, MIT