

Koeniger, Winfried; Prat, Julien

Conference Paper

Human Capital and Optimal Redistribution

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Labor Markets, No. B07-V1

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Koeniger, Winfried; Prat, Julien (2017) : Human Capital and Optimal Redistribution, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Labor Markets, No. B07-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/168073>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Human Capital and Optimal Redistribution

Winfried Koeniger* and Julien Prat**

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 compute the optimal allocation when ability is i.i.d. across generations, and show how it can be implemented with student loans or means-tested grants. The numerical results reveal that human capital investment should decline in parental income because parents with high income bequeath more and this lowers the labor supply of their children through a wealth effect. Taking existing income and inheritance taxes in the U.S. as given, we find that asset-poor families should contribute on average less to the education cost even though their contributions increase more progressively in income.

Keywords: human capital, optimal taxation, intergenerational equity.

JEL: E24, H21, I22, J24.

*University of St.Gallen (Swiss Institute for Empirical Economic Research), CESifo, CFS, IZA. **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.

1 Introduction

Of all the factors shaping inequality, one of the most debated is the transmission of financial 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 intergenerational inequality is 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 characterize efficient 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 future labor market productivity of children is uncertain and parents cannot insure against this risk, making it less attractive for families to invest in human capital if it provides a bad hedge against consumption risk. The planner then 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. In accordance with Mincer’s specification of the wage equation, we postulate a unit elasticity between ability and human capital. We find that the wedge for human capital required to implement the constrained-efficient allocation is much lower than the wedge for bequests, so that human capital should be subsidized for all but the very-low income families, while bequests should be taxed across all income levels.

Strikingly, our results suggest that the socially-optimal investment into children’s human capital should be decreasing in parental income. This result is explained by a wealth effect on labor supply. 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 wealth effect on labor supply, at the heart of this finding, is empirically plausible: Holtz-Eakin et al. (1993) document that the receipt of an inheritance reduces labor force participation, and Brown et al. (2010) provide evidence that it increases the probability of retirement.

We also illustrate how the social optimum can be implemented using means-tested grants or loans with contingent repayments. In particular, we describe how contributions to education costs should be designed to implement the social optimum, for given stylized representations of the U.S. income and inheritance tax schedules. This experiment shows that the social optimum can be implemented by optimal design of education finances alone. Hence it does not require a politically infeasible reform of the whole tax system since it relies on instruments that resemble actual repayments schemes. Given the redistribution implied by income and inheritance taxes in the U.S., we find that income-poor families should be recipients of net education subsidies whose size decreases more rapidly in income if they have less inherited wealth. In spite of the greater progressivity of their contributions, asset-poor families pay less of their education costs because they benefit from subsidies that are more generous on average.

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.

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 (2015) 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 her focus on a life-cycle problem in which human capital raises *today's* productivity, which generates a close relationship between the human capital wedge and the contemporaneous labor wedge. This is different in our intergenerational model in which the benefits of children's education accrue only *next* period, when these children become adults and participate in the labor market. This natural difference in timing in our dynastic family model generates a close relationship between the wedge for human capital and the intertemporal wedge for bequests.

The results in our model also 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 (2006, 2015) and Kapička and Neira (2015) extend this analysis by focussing on human capital investment or learning effort 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

2.1 Set-up

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 :

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

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 $h_t \in \mathbb{R}_+$ are public knowledge. Output y_t is produced according to the technology $Y(h_t, l_t, \theta_t)$ which is increasing in its arguments, concave and bounded below by zero. We will use the production function to substitute l_t in the utility function and write $U(c_t, y_t, h_t, \theta_t)$ instead of $\mathbf{U}(c_t, l_t)$ or, with assumption **[A1]**, $v(y_t, h_t, \theta_t) = \mathbf{v}(l_t)$. Note that the planner cannot use observable output y_t to infer actual labor supply l_t because ability θ_t is stochastic and hidden.

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

The timing of 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 innate 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 $\Theta \equiv [\underline{\theta}, \bar{\theta}]$ with $\underline{\theta} > 0$. In the absence of clear evidence on the genetic transmission of ability,² the i.i.d. case is a natural starting point in an intergenerational model. By contrast, the assumption of persistent ability is more plausible in life cycle problems, as for example in Farhi and Werning (2013) or Stantcheva (2014). 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 the online Appendix B.3.

2.2 The planner's problem

Incentive compatibility.—According to the revelation principle, we can solve for the social optima by focusing on a direct revelation mechanism such that families truthfully report

¹We 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} .

²See for example the empirical evidence in Black et al. (2015). Using information on adopted and biological parents in Swedish data, they find that biological parents are less important than adopted parents for the intergenerational transmission of wealth.

their types in each generation. Let $\theta^t \equiv \{\theta_0, \theta_1, \dots, \theta_t\}$ 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 schemes $x_t(\theta^t) \equiv \{c_t(\theta^t), h_{t+1}(\theta^t), y_t(\theta^t)\}$. Hence the feasible set \mathcal{X} contains all sequences $\mathbf{x} \equiv \{x_t(\theta^t)\}_{t=1}^T$ of measurable functions $x_t : \Theta^t \rightarrow \mathbb{R}_+^3$. The family's preferences over an allocation \mathbf{x} are given by

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

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 can choose any *reporting strategy* $\mathbf{r} \equiv \{r_t(\theta^t)\}_{t=1}^T$ from the set \mathcal{R} containing all sequences of measurable functions $r_t : \Theta^t \rightarrow \Theta$. Since types are private information, an allocation must be *incentive compatible*, i.e.,

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

where $(\mathbf{x} \circ \mathbf{r})(\theta^t) \equiv \{x_t(r^t(\theta^t))\}_{t=1}^T$ is the allocation \mathbf{x} resulting from the reporting strategy \mathbf{r} and history θ^t .

In order to write the planner's problem in a recursive form, we replace the ex-ante incentive constraint (1) with an ex-post requirement. For this purpose, we define the equilibrium continuation utility $\omega(\theta^t)$ for a given history θ^t as

$$\omega(\theta^t) \equiv U(c_t(\theta^t), y_t(\theta^t), h_t(\theta^{t-1}), \theta_t) + \beta \int_{\Theta} \omega(\theta^t, \theta_{t+1}) dF(\theta_{t+1}), \quad (2)$$

for all $t = 1, \dots, T$, and $\omega(\theta^{T+1}) = 0$ because families have horizon T . At the beginning of period t , families compare the continuation value $\omega(\theta^t)$ of truthful reporting to those derived from arbitrary reporting strategies

$$\omega^{\mathbf{r}}(\theta^t) \equiv U(c_t(r^t(\theta^t)), y_t(r^t(\theta^t)), h_t(r^{t-1}(\theta^{t-1})), \theta_t) + \beta \int_{\Theta} \omega^{\mathbf{r}}(\theta^t, \theta_{t+1}) dF(\theta_{t+1}), \quad (3)$$

with $\omega^{\mathbf{r}}(\theta^{T+1}) = 0$. Ex-post incentive compatibility is ensured when

$$\omega(\theta^t) \geq \omega^{\mathbf{r}}(\theta^t), \text{ for all } \theta^t \text{ and all } \mathbf{r} \in \mathcal{R}. \quad (4)$$

We denote by \mathbf{x}^{IC} the set of all allocations \mathbf{x} satisfying (4).³

Planner's objective.—The planner discounts future utility with the factor q which equals the inverse interest factor.⁴ As Farhi and Werning (2013), we abstract from general equilibrium feedbacks so that the allocation problem can be analyzed separately for each family. The expected costs of an allocation is given by

$$\Pi(\mathbf{x}) \equiv \mathbb{E}_0 \left[\sum_{t=1}^T q^{t-1} (c_t(\theta^t) + e_t(\theta^t) - y_t(\theta^t)) \right],$$

and an allocation is efficient when it minimizes the costs of delivering a certain utility

$$\begin{aligned} \Gamma^*(V, h_1) &\equiv \min_{\mathbf{x} \in \mathcal{X}} \Pi(\mathbf{x}), \\ \text{s.t. (i)} \quad &\mathcal{U}(\mathbf{x}) \geq V; \quad \text{(ii)} \quad \mathbf{x} \in \mathbf{x}^{\text{IC}}. \end{aligned} \tag{5}$$

Condition (i) is the promise keeping constraint which ensures that the expected utility of truthful families is at least as high as the exogenously given level V , while condition (ii) excludes all the allocations that are not incentive compatible.

First-order approach.—Problem (5) requires to keep track of all the out-of-equilibrium payoffs in order to check the incentive constraint (4). We circumvent this issue by replacing the incentive constraint (4) with an envelope condition which only depends on the marginal utility of truth-tellers.

This necessary condition can be heuristically derived considering the following one shot deviations.⁵ For a given history θ^t , let us define the set of strategies $\mathbf{r}^{\sigma, t}$, indexed by $\sigma \in \Theta$ and $t \in \{1, \dots, T\}$, such that agents are truthful in all previous and future periods

³Note that allocations in \mathbf{x}^{IC} are incentive compatible 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. As shown by Theorem 2.1 of Fernandes and Phelan (2000), 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. See also Lemma 1 in Kapička (2013) for the case with persistent stochastic ability.

⁴We 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 labor supply of the next generation does not depend on human capital. We deliberately neutralize this motive in order to isolate the impact that human capital has on the incentives to exert labor effort.

⁵See Kapička (2013) for a formal derivation in the general case with persistent stochastic ability.

but report σ in period t whatever is their actual type:

$$r_s^{\sigma,t}(\theta^s) = \begin{cases} \theta_s, & \text{for } s \neq t. \\ \sigma, & \text{for } s = t. \end{cases}$$

Since types are i.i.d., next period continuation utilities only depend on the report and not on the current type of the family, hence $\omega^{\mathbf{r}^{\sigma,t}}(\theta^t, \theta_{t+1}) = \omega(\theta^{t-1}, \sigma, \theta_{t+1})$. Reinserting this equality into (3), we see that incentive compatibility requires that

$$\begin{aligned} \omega(\theta^t) &= \max_{\sigma} \omega^{\mathbf{r}^{\sigma,t}}(\theta^t) \\ &= \max_{\sigma} \left\{ U(c_t(\theta^{t-1}, \sigma), y_t(\theta^{t-1}, \sigma), h_t(\theta^{t-1}), \theta_t) + \beta \int_{\Theta} \omega(\theta^{t-1}, \sigma, \theta_{t+1}) dF(\theta_{t+1}) \right\}. \end{aligned} \quad (6)$$

Differentiating (2) and using (6), we obtain an envelope condition⁶ that is reminiscent of the one derived by Mirrlees in his static setting

$$\frac{\partial \omega(\theta^t)}{\partial \theta_t} = \frac{\partial U(c_t(\theta^t), y_t(\theta^t), h_t(\theta^{t-1}), \theta_t)}{\partial \theta_t}. \quad (7)$$

Relaxed problem.—Replacing the incentive constraint by (7) greatly simplifies the optimization problem because it now only depends on the continuation utility of truth-tellers instead of that of all possible types. If we denote by \mathbf{x}^{FOA} the set of allocations such that condition (7) holds for all θ^t , we have $\mathbf{x}^{\text{IC}} \subseteq \mathbf{x}^{\text{FOA}}$ and the following problem is a relaxed formulation of the original objective defined in (5)

$$\begin{aligned} \Gamma(V, h_1) &\equiv \min_{\mathbf{x} \in \mathcal{X}} \Pi(\mathbf{x}), \\ \text{s.t. } &(i) \mathcal{U}(\mathbf{x}) \geq V; \quad (ii) \mathbf{x} \in \mathbf{x}^{\text{FOA}}. \end{aligned} \quad (8)$$

We explain in Appendix A how to write (8) in recursive form so as to solve it through a sequence of standard optimal control problems. Although the envelope condition (7) is only necessary, the validity of the solution can be checked quite easily because ability types are i.i.d. and preferences satisfy the single-crossing condition. Then the first-order approach is valid when the allocation is monotone in ability.⁷

⁶The envelope condition (7) applies because problem (6) satisfies the conditions of Theorem 2 in Milgrom and Segal (2002).

⁷See 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).

2.3 Optimality conditions

In the first best allocation, families are fully insured against changes in ability. Consumption remains constant across agents and is therefore separated from production. With information asymmetries, perfect insurance is precluded because the planner faces an insurance-incentive trade-off. We relegate the formal analysis of the planner problem, including the derivation of the optimality conditions as well as the proofs for the subsequent results, to Appendix A. We now discuss the optimality conditions for consumption and human capital and refer to Appendix A for the standard optimality condition for output (30). For ease of notation, we shall henceforth use a prime ' to denote next period realizations.

Remark 1 *Under assumption [A1], the first-order conditions of the relaxed problem (8) imply that the reciprocal Euler equation holds as*

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

The inverse Euler equation implies that the marginal utility of consumption follows a martingale when the planner and family share the same discount factor, i.e. when $q = \beta$. Our analysis shows that this well known feature of optimal tax systems continues to hold when human capital accumulation is taken into account. Turning our attention to education, we find that its accumulation should be governed by the following optimality condition.

Proposition 1 *Let $g(h', h)$ denote the cost of acquiring education h' . If [A1] holds, the first-order condition with respect to human capital of the relaxed problem (8) reads*

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

where $\mu(\theta)$ is the costate variable of the planner's problem so that

$$\mu(\theta) = \int_{\underline{\theta}}^{\theta} \left[\lambda - \frac{1}{\partial u(c(x))/\partial c(x)} \right] dF(x), \quad \text{with } \lim_{\theta \rightarrow \underline{\theta}} \mu(\theta) = \lim_{\theta \rightarrow \bar{\theta}} \mu(\theta) = 0, \quad (10)$$

and $\lambda = \mathbb{E} [[\partial u(c(\theta)) / \partial c(\theta)]^{-1}]$ is the multiplier associated with the promise keeping constraint.

The marginal cost of human capital investment on the left hand side of (9) 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.⁸ 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 and so $\partial \omega(\theta) / \partial \theta = 0$. With hidden types instead, information revelation is profitable solely if

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

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.⁹

The sign of the cross derivative $\partial^2 v(\cdot) / (\partial \theta \partial h)$ is determined by (i) the elasticity ε_{vl} of the marginal disutility of labor with respect to changes in labor supply (which is inversely related to the Frisch elasticity of labor supply) and (ii) the Hicksian coefficient

⁸As shown in Appendix A, this benefit for the planner is captured by $-\frac{\partial v(y'(\theta'), \theta', h'(\theta))}{\partial h'(\theta)} / \frac{\partial u(c'(\theta'))}{\partial c'(\theta')} > 0$.

⁹According to its definition in (10), the costate variable $\mu'(\theta')$ also captures the probability weight for each type.

$\rho_{h\theta}$ of complementarity between human capital and ability. Both are captured by a single parameter if we assume functional forms for the disutility of labor and the production function for output as specified in the following corollary.

Corollary 1 *The effect of human capital on information rents is determined by*

$$\text{sgn} \left(\frac{\partial^2 v(y, h, \theta)}{\partial \theta \partial h} \right) = \text{sgn} (1 + \varepsilon_{vl} + 1 - \rho_{h\theta}).$$

Assuming that

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

$$[\mathbf{A2}]: \quad Y(h, l, \theta) = A(\theta, h)l,$$

$$\text{with } A(\theta, h) = [\xi \theta^\chi + (1 - \xi) h^\chi]^{1/\chi}, \chi \in (-\infty, 1] \text{ and } \xi \in (0, 1),$$

$\varepsilon_{vl} = \alpha - 1$ and $\rho_{h\theta} = 1 - \chi$ so that $\partial^2 v(y, h, \theta) / (\partial \theta \partial h) \geq 0$ 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,¹⁰ 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.

¹⁰The Frisch elasticity of labor supply is equal to $1/(\alpha - 1)$.

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.

In the laissez faire each family solves the maximization problem

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

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

Remark 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 Remark 2 are necessary and sufficient.¹¹ Then the conditions in Remark 2 can be combined with the social-optimality conditions (29)-(31) in Appendix A to derive interpretable expressions for the wedges between the choices in the laissez faire and the constrained-efficient allocation of the planner. We start with the following definition.

¹¹Note 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.

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

$$\tau_b(\theta^t) \equiv 1 - \frac{q}{\beta} \frac{\partial u(c)/\partial c}{\mathbb{E}[\partial u(c')/\partial c']}, \quad (11)$$

$$\tau_l(\theta^t) \equiv 1 - \frac{\partial v(y, h, \theta)/\partial y}{\partial u(c)/\partial c}, \quad (12)$$

$$\tau_h(\theta^t) \equiv \frac{\beta}{\frac{\partial g(h', h)}{\partial h'}} \mathbb{E} \left[\frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} \left(\frac{\partial y'}{\partial h'} - \frac{\partial g(h'', h')}{\partial h'} \right) \right] - 1. \quad (13)$$

Wedges are defined as the deviations from the laissez faire. In general, the wedges depend on the whole history of shocks since the allocation $\{c, h', y\}$ is a function of θ^t which we suppressed in the notation for convenience. In the following we denote the wedges as $\tau_j \equiv \tau_j(\theta^t)$, and the corresponding leads and lags of the wedges as $\tau'_j \equiv \tau'_j(\theta^{t+1})$ and $\tau_{j-} \equiv \tau_{j-}(\theta^{t-1})$, $j = b, l, h$. 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. We first characterize the standard wedges for bequests and labor supply which we also obtain in our model with human capital.

Remark 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*

$$\tau_b^* = 1 - \frac{1}{\mathbb{E} \left[\frac{1}{\frac{\partial u(c')}{\partial c'}} \right] \mathbb{E} \left[\frac{\partial u(c')}{\partial c'} \right]}, \quad (14)$$

$$\tau_l^* = - \frac{\partial^2 v(y, h, \theta)}{\partial \theta \partial y} \frac{\mu(\theta)}{f(\theta)}. \quad (15)$$

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 $\mu(\theta) > 0$. 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).¹²

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

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

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

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

Proposition 2 *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

$$\Delta_b \equiv \frac{q}{\frac{\partial g(h', h)}{\partial h'}} \mathbb{E} \left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'', h')}{\partial h'} \right] \frac{\tau_b^*}{1 - \tau_b^*} + \frac{\beta}{\frac{\partial g(h', h)}{\partial h'} \frac{\partial u(c)}{\partial c}} \text{Cov} \left(\frac{\partial u(c')}{\partial c'}, \frac{\partial y'}{\partial h'} - \frac{\partial g(h'', h')}{\partial h'} \right), \quad (16)$$

and

$$\Delta_i \equiv -\frac{q}{\frac{\partial g(h', h)}{\partial h'}} \chi \mathbb{E} \left[l'(\theta') \frac{d\mathbf{v}(l'(\theta'))}{dl'} \frac{\frac{\partial A(\theta', h')}{\partial \theta'} \frac{\partial A(\theta', h')}{\partial h'}}{A(\theta', h')^2} \mu'(\theta') \right]. \quad (17)$$

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

¹²Compared 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, $\lim_{\theta \rightarrow \underline{\theta}} \mu(\theta) = 0$ and $\lim_{\theta \rightarrow \bar{\theta}} \mu(\theta) = 0$ imply that

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

of capital transfer resources from one generation to the next. The first term in Δ_b is of the same sign as τ_b^* .¹³ 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.¹⁴

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 (16) which depends on the covariance between the return to human capital and the marginal utility of consumption. Since the return to human capital and consumption of the next generation are both 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.

The second component Δ_i corresponds to the incentive term in (9) net of the labor supply effect discussed at the end of Section 2.3. We show explicitly in the proof of Proposition 2 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. This is why the decomposition in Proposition 2 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

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

¹⁴In 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 (34) in the proof of Proposition 2.

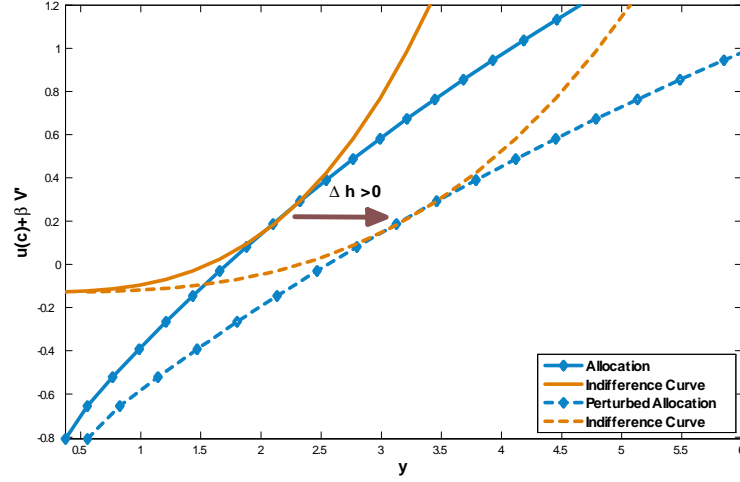


Figure 1: Effect of perturbation $\Delta h = 4$ when $\chi = 0$. Notes: Plot for a family with zero assets and 13 years of human capital; parameter values specified as in Table 1, Section 4.

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 as¹⁵

$$1 = \frac{q}{\frac{\partial g(h', h)}{\partial h'}} \mathbb{E} \left[\frac{\partial y'}{\partial h'} - \frac{\partial g(h'', h')}{\partial h'} \right] - \Delta_i . \quad (18)$$

If we divide by q , equation (18) 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 (18) 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 the online Appendix B.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 representative family. Local incentive compatibility holds when the indifference curve

¹⁵Equation (18) follows replacing τ_h by $\Delta_b + \Delta_i$ on the left hand side of the definition (13), and noticing that

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

is tangent to the allocation. Figure 1 also reports the effect of a perturbation 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 in such a way 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 will reduce its effort and imitate lower types.¹⁶ 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 the online Appendix B.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

¹⁶The opposite adjustment occurs when $\chi > 0$ as the indifference curve intersects the allocation from below, making it easier to motivate families.

and human capital 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 the online Appendix B.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.¹⁷ 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 2 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 Discussion

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 (18) 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*.

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

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 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 2 that, at the social optimum, this effect exactly offsets the distortion of human capital investment induced by the labor wedge.

We have already discussed in the Introduction how the economic application to a life-cycle or intergenerational economy implies a natural difference in the assumption about the timing of the productivity increases resulting from human capital investment. The contemporaneous productivity effect in the life-cycle model of Stantcheva (2014) implies a close relationship between the wedges for human capital and labor. In our intergenerational economy, investment into the human capital of children increases their productivity in the next period, when these children enter the labor market, so that there is a close relationship between the wedges for human capital and bequests.

4 Numerical analysis

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

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:

$$\ln w(\theta, h) = \ln A(\theta, h) = (1 - \xi) \ln h + \xi \ln \theta. \quad (19)$$

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).¹⁸ They show that the variance of residual log-wages among U.S. workers due to persistent shocks has been equal to 0.2 in 2005.¹⁹ We use the variance resulting from persistent shocks because θ is fully persistent in our model during a generation’s labor-market career and transitory shocks (at least partially) wash out.

In order to calibrate the parameter ξ of the production function, we use the large body of empirical evidence on Mincerian wage regressions. As surveyed by Card (1999), the literature shows that the marginal returns of an additional year of schooling are remarkably consistent across studies and close to 0.1. Since years of schooling S correspond to $\ln h$ in our model, equation (19) implies that $1 - \xi = 0.1$.

¹⁸The log-normal distribution is approximated by a truncated density so that we draw the next-period ability from a compact interval $[\underline{\theta}, \bar{\theta}]$.

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

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 <i>et al.</i> (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 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 $\varsigma = 0.214$.²⁰

The parameter κ is calibrated to match the actual cost of the first year of upper-secondary education. We thus have to relate the monetary costs observed in the data to units of the model. We make the empirically plausible assumption that the median worker

²⁰Annual 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 $(\exp(8)^\varsigma - 1) / (\exp(4)^\varsigma - 1)$.

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.²¹ According to census data, the mean annual earnings of high-school dropouts have been equal to \$20,241 in 2010.²² By comparison, the annual expenditure per year for upper-secondary students was \$12,690. Computing the cost-income ratio, we find that the cost of an additional year of upper-secondary education amounts to 62.6% of annual income or, given our 30-year period, to 2.08% of lifetime income of the median worker without non-compulsory education. It follows that $\kappa = 0.0208 / [\exp(1)^{0.214} - 1] = 0.087$.

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., $\Gamma(V, h_1) = 0$. We solve the model for a dynasty of four generations (120 years) and focus on the decisions of the second generation.²³ The algorithm, discussed further in the online Appendix C, follows Farhi and Werning (2013) closely.

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.

²¹For 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 $y^*(1, 1) = 1$.

²²See 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>

²³We show in the online Appendix C 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.

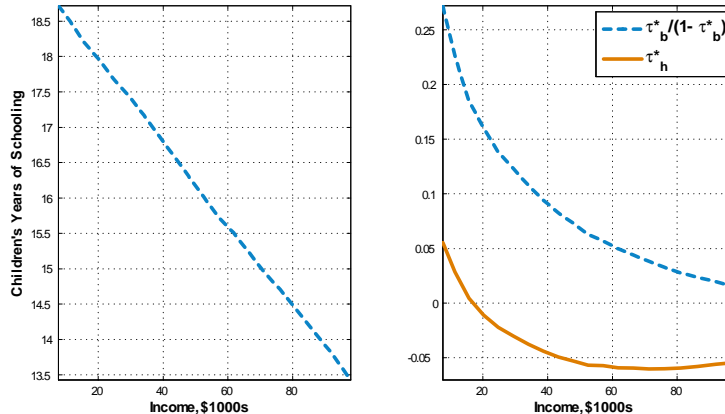


Figure 2: Investment in children’s human capital and wedges for human capital and bequests, as a function of labor income for parents with zero assets and 13 years of schooling.

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. Empirical evidence by Holtz-Eakin et al. (1993) and Brown et al. (2010) shows that such a negative wealth effect of inheritance on labor supply is plausible.

This wealth effect is often used in the public debate to advocate higher taxes on inherited wealth so that the currently young generation maintains a high labor effort. The insight from our dynamic intergenerational model is that such proposals should be treated with caution: the incentive scheme in the constrained-efficient allocation exploits that bequests increase children’s consumption and lower their labor effort. This allows the planner to redistribute resources in the current generation while ensuring incentive compatibility.

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 feature of 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 2 that the wedge for human capital in the calibrated Cobb-Douglas case reads

$$\tau_h^* = \frac{\tau_b^*}{1 - \tau_b^*} + \frac{\beta}{\frac{\partial g(h', h)}{\partial h'} \frac{\partial u(c)}{\partial c}} \text{Cov} \left(\frac{\partial u(c')}{\partial c'}, \frac{\partial y'}{\partial h'} - \frac{\partial g(h'', h')}{\partial h'} \right). \quad (20)$$

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 (20) is negative, showing that human capital provides a bad hedge against 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

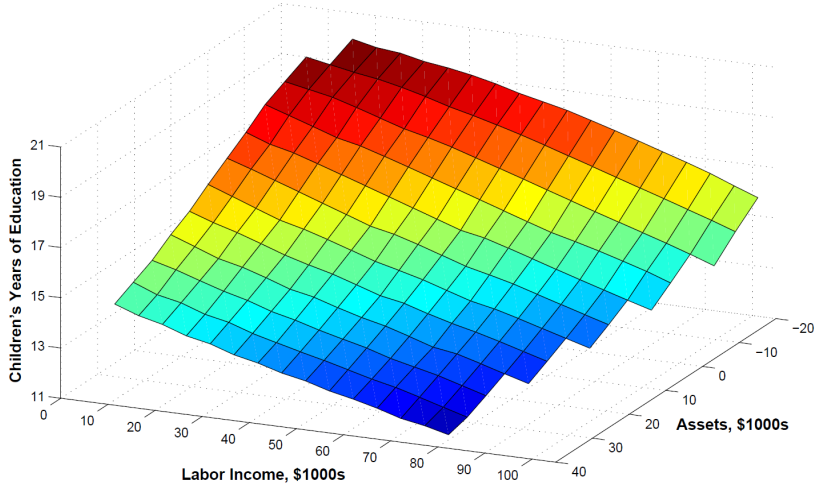


Figure 3: Human capital accumulation as a function of families' inherited wealth and labor income for families with parents that have 13 years of education.

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 (20) 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 allows us to evaluate the influence of inherited wealth on educational choices.²⁴ 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 canon-

²⁴We vary the promised value V which corresponds to varying assets of families in the decentralized allocation. To simplify the interpretation of the figures, we replace the promised values by the corresponding levels of bequest, $b(V, h_1) = \Gamma(V, h_1)$, 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.

ical 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 2, $\chi \neq 0$ introduces an additional motive for the planner to tax or subsidize human capital investment. The results presented in the online Appendix C show that the shape of the human capital wedge remains very similar but 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.

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).²⁵

²⁵An 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. Implementation of fully history-contingent optimal policy, as would be required if ability were persistent, is particularly implausible in intergenerational settings in which this would require to condition on the output produced by great-grandparents, great-

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

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

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 $\Gamma(V, h, t)$ defined in Appendix A. 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-order condition for labor supply and the definition of the labor wedge (12) 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

great-grandparents and so on.

for human capital (13):

$$\frac{\partial T(\cdot)}{\partial h'} = \frac{\partial g(h', h)}{\partial h'} \tau_h - \beta \mathbb{E} \left[\left(\frac{\partial y'}{\partial h'} \frac{\partial T'(\cdot)}{\partial y'} + \frac{\partial T'(\cdot)}{\partial h'} \right) \frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c)}{\partial c}} \right]. \quad (21)$$

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 (21) allows us to relate our results to Bovenberg and Jacobs' (2005) finding 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 (21) reads

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

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 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 2. 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 ability 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 being families with high income who have to pay less if they hold more assets. As we are going to see in the implementation with education loans that decomposes $T(\cdot)$, its shape crucially depends on the differences in human capital investment across families with different income and assets, and thus on the wealth effect of bequests on their children's labor supply.

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 $\tau = 0.15$.²⁶ 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

²⁶Observe 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.

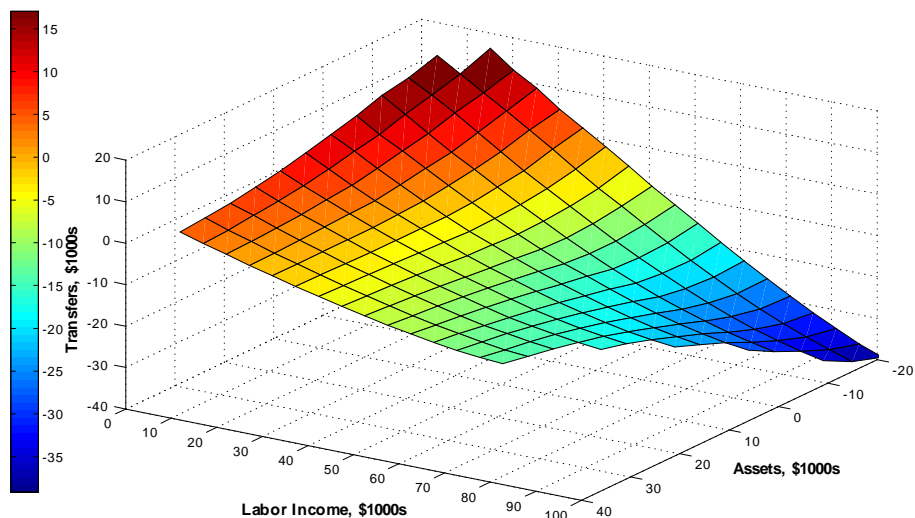


Figure 4: Implementation of the social optimum with means-tested grants for families with parents that have 13 years of education.

of $D(\cdot)$ in the online Appendix B.4.

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 and even benefit from small net transfers, while receiving the largest loans for the education of their children.

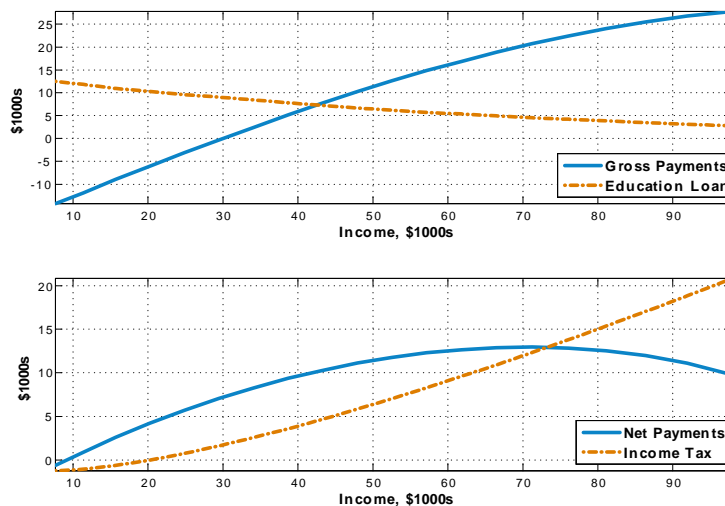


Figure 5: Implementation of the social optimum with loans and contingent payments for families with zero assets and parents with 13 years of education.

This is illustrated further in Figure 6 which plots the consolidated contribution, $D(b, h, y, h') - L(h', h)$, of parents to education costs. The consolidated contribution is particularly interesting because it shows how the social optimum can be implemented by optimal design of education finances alone, taking the existing income and inheritance tax schedules in the U.S. as given. This may be of interest if reforming the whole tax system is politically infeasible but a separate reform of education finances is possible. We approximate the income tax schedule $T^y(y)$ for the U.S. as before using the parametrization by Heathcote et al. (2014) and assume an estate tax of zero.²⁷

Figure 6 displays the consolidated contribution to education costs, $D(\cdot) - L(\cdot)$, as a function of income y and inherited wealth b .²⁸ As already seen in Figure 4, the progressivity of gross transfers $-T(\cdot)$ with respect to labor income is decreasing in the level of bequests. Since there is no estate tax and income taxes $T^y(y)$ only condition on current labor income y , this pattern is also borne out in the consolidated contribution to the

²⁷This is our stylized approximation of the tax schedule for bequests in the U.S. De Nardi and Yang (2015) report that the high threshold of \$675,000 in 2000 above which estates started to be taxed in the U.S. can be considered a conservative lower bound since households have many ways to reduce effective estate taxation.

²⁸The consolidated contribution also conditions on h and h' . We report representative results for parents with h corresponding to 13 years of education. In equilibrium, human capital investment into the children h' does not differ across families with the same inherited wealth b , parental education background h , and income y .

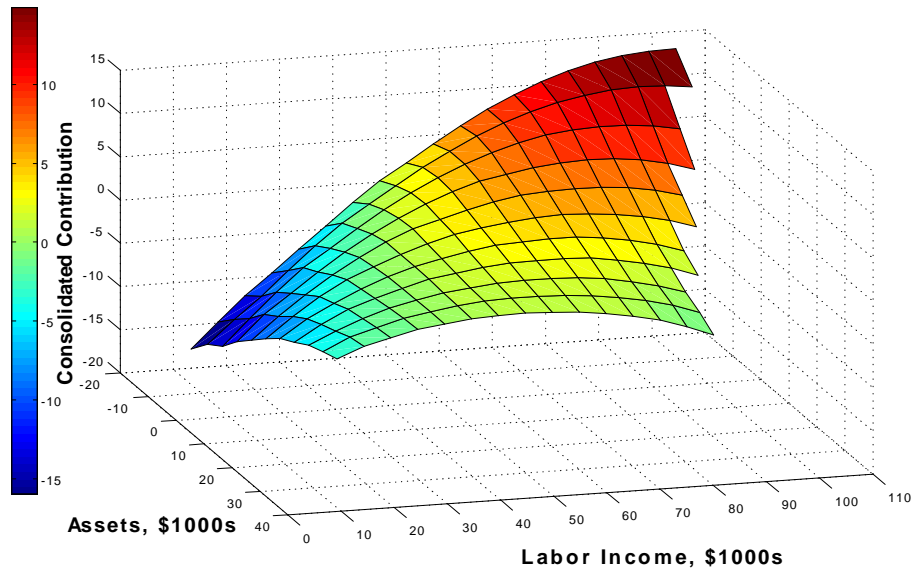


Figure 6: Consolidated contribution to education costs as a function of inherited assets and labor income for parents with 13 years of education.

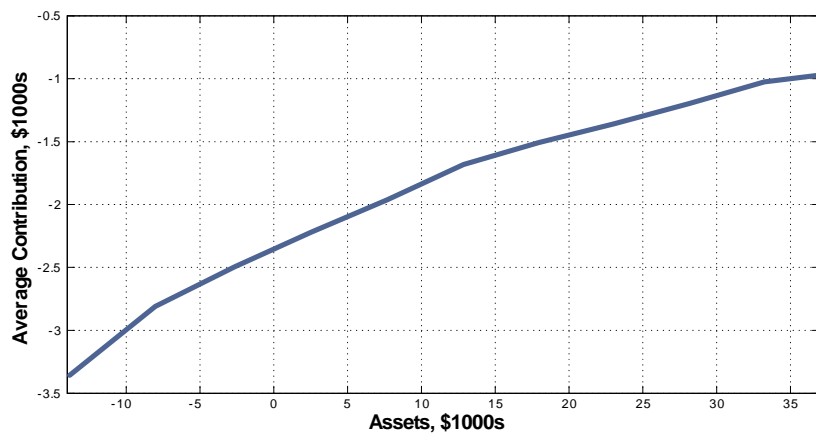


Figure 7: Average contribution as a function of inherited assets for parents with 13 years of education.

education cost. Incentives require that the contribution increases more progressively in income for asset-poor families. Yet there is substantial redistribution since income-poor families do not have to bear education cost: D is close to 0 and L is large for families with very low income whose children receive a lot of human capital because they do not inherit much, and thus, work more. From an *ex ante* perspective under the veil of ignorance, Figure 7 shows that the expected consolidated contribution, averaged across the realized labor income of different types, is increasing in assets.²⁹ This implies that, on average, families with less inherited assets contribute less to education costs.

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. Taking as given the existing income and estate taxes in the U.S., we find that income-poor families should not contribute to education costs to attain the social optimum. Incentive provision implies that the contributions are more progressive in income for asset-poor families.

A striking finding in the numerical analysis is that low-income families invest more into the human capital of their children than high-income families in the social optimum: 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

²⁹The average consolidated contribution is negative across all asset levels in Figure 7, partially offsetting the income tax payments $T^y(y)$.

for intergenerational analysis in many countries, but vital from a policy perspective.

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] Black, Sandra E., Paul J. Devereux, Petter Lundborg and Kaveh Majlesi (2015): “Poor Little Rich Kids? The Determinants of the Intergenerational Transmission of Wealth,” IZA Discussion Paper No. 9227.
- [7] Bohaček, Radim and Marek Kapička (2008): “Optimal Human Capital Policies,” *Journal of Monetary Economics*, vol. 55, 1–16.
- [8] Bovenberg, A. Lans and Bas Jacobs (2005): “Redistribution and Education Subsidies are Siamese Twins,” *Journal of Public Economics*, vol. 89, 2005–2035.
- [9] Brown, Jeffrey R., Courtney C. Coile and Scott J. Weisbenner (2010): “The Effect of Inheritance Receipt on Retirement,” *Review of Economics and Statistics*, vol. 92, 425–434.
- [10] 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.

- [11] Chetty, Raj (2012): “Bounds on Elasticities with Optimization Frictions: a Synthesis of Micro and Macro Evidence on Labor Supply,” *Econometrica*, vol. 80, 969–1018.
- [12] 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.
- [13] 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.
- [14] De Nardi, Mariacristina, and Fang Yang (2015): “Wealth Inequality, Family Background, and Estate Taxation,” Manuscript prepared for the Carnegie-Rochester Conference Series on Public Policy, University College London.
- [15] Erosa, Andrés and Tatyana Koreshkova (2007): “Progressive Taxation in a Dynastic Model of Human Capital,” *Journal of Monetary Economics*, vol. 54, 667–685.
- [16] Farhi, Emmanuel and Iván Werning (2007): “Inequality and Social Discounting,” *Journal of Political Economy*, vol. 115, 365–402.
- [17] Farhi, Emmanuel and Iván Werning (2010): “Progressive Estate Taxation,” *Quarterly Journal of Economics*, vol. 125, 635–673.
- [18] Farhi, Emmanuel and Iván Werning (2013): “Insurance and Taxation over the Life Cycle,” *Review of Economic Studies*, vol. 80, 596–635.
- [19] Fernandes, Ana and Christopher Phelan (2000): “Recursive Formulation for Repeated Agency with History Dependence,” *Journal of Economic Theory*, vol. 91, 223–247.
- [20] Findeisen, Sebastian and Dominik Sachs (2012): “Education and Optimal Dynamic Taxation: the Role of Income-Contingent Loans,” Working Paper, University of Konstanz.
- [21] Gary-Bobo, Robert and Alain Trannoy (2015): “Optimal Student Loans and Graduate Tax under Moral Hazard and Adverse Selection,” *RAND Journal of Economics*, vol. 46, 546–576.
- [22] 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.

- [23] Golosov, Mikhail, Maxim Troshkin, Aleh Tsyvinski (2011): “Optimal Dynamic Taxes,” NBER Working Paper No. 17642.
- [24] 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.
- [25] 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.
- [26] Heathcote, Jonathan, Kjetil Storesletten and Giovanni L. Violante (2014): “Optimal Tax Progressivity: An Analytical Framework,” Manuscript, Federal Reserve Bank of Minneapolis.
- [27] Holtz-Eakin, Douglas, David Joulfaian and Harvey S. Rosen (1993): “The Carnegie Conjecture: Some Empirical Evidence,” *Quarterly Journal of Economics*, vol. 108, 413–435.
- [28] Kapička, Marek (2006): “Optimal Income Taxation with Human Capital Accumulation and Limited Record Keeping,” *Review of Economic Dynamics*, vol. 9, 612–639.
- [29] 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.
- [30] Kapička, Marek (2015): “Optimal Mirrleesian Taxation in a Ben-Porath Economy,” *American Economic Journal: Macroeconomics*, vol. 7(2), 219–248.
- [31] Kapička, Marek and Julian Neira (2015): “Optimal Taxation in a Life-Cycle Economy with Endogenous Human Capital Formation,” Manuscript, U.C. Santa Barbara.
- [32] Kocherlakota, Narayana R. (2010): *The New Dynamic Public Finance*, Princeton University Press, Princeton, New Jersey.
- [33] Koehne, Sebastian and Moritz Kuhn (2014): “Optimal Taxation in a Habit Formation Economy,” CESifo Working Paper No. 4581.
- [34] Krueger, Dirk and Alexander Ludwig (2013): “Optimal Capital and Progressive Labor Income Taxation with Endogenous Schooling Decisions and Intergenerational Transfers,” Manuscript, University of Pennsylvania.

- [35] Lee, Sang Yoon (Tim) and Ananth Seshadri (2014): “On the Intergenerational Transmission of Economic Status,” Manuscript, University of Mannheim.
- [36] Milgrom, Paul and Ilya Segal (2002): “Envelope Theorems for Arbitrary Choice Sets,” *Econometrica*, vol. 70, 583–601.
- [37] Mirrlees, J. A. (1971): “An Exploration in the Theory of Optimum Income Taxation,” *Review of Economic Studies*, vol. 38, 175–208.
- [38] OECD (2011): *Education at a Glance*, OECD, Paris.
- [39] Pavan, Alessandro, Ilya Segal and Juuso Toikka (2014): “Dynamic Mechanism Design: A Myersonian Approach,” *Econometrica*, vol. 82, 601–653.
- [40] Rogerson, William P. (1985): “The First-Order Approach to Principal-Agent Problems,” *Econometrica*, vol. 53, 1357–1367.
- [41] Saez, Emmanuel (2001): “Using Elasticities to Derive Optimal Income Tax Rates,” *Review of Economic Studies*, vol. 68, 205–229.
- [42] Stantcheva, Stefanie (2014): “Optimal Taxation and Human Capital Policies over the Life Cycle,” Manuscript, MIT.
- [43] Stantcheva, Stefanie (2015): “Optimal Income, Education and Bequest Taxes in an Intergenerational Model,” Manuscript, MIT

APPENDIX

A Analysis of the planner problem and proofs

A recursive representation of the relaxed problem (8) can be derived as follows. First define the promised value at the end of period t-1 as $V(\theta^{t-1}) \equiv \int_{\Theta} \omega(\theta^t) dF(\theta)$. The continuation values are by definition equal to

$$\omega(\theta^t) = U(c_t(\theta^t), y_t(\theta^t), h_t(\theta^{t-1}), \theta_t) + \beta V(\theta^t).$$

Taking the promised value $V(\theta^{t-1})$ as given and using the simpler notation V , we can rewrite (8) in recursive form:

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

$$\text{s.t. } \omega(\theta) = U(c(\theta), y(\theta), h, \theta) + \beta V'(\theta), \quad (22)$$

$$V = \int_{\Theta} \omega(\theta) dF(\theta), \quad (23)$$

$$\frac{\partial \omega(\theta)}{\partial \theta} = \frac{\partial U(c, y, h, \theta)}{\partial \theta}, \quad (24)$$

where we only keep a time index for the value function, and otherwise use a prime ' to denote next period realizations. We have inverted the human capital accumulation function $h'(e, h)$ to substitute $e(\theta)$ with $g(h'(\theta), h)$. Note that we do not need to condition on the history of θ because it follows an i.i.d. process.

The first constraint repeats the definition of the continuation value $\omega(\theta)$ as the sum of the current and next-period promised utilities $U(\cdot)$ and $V'(\theta)$, respectively. Equation (23) 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, or envelope condition.

Following Mirrlees' (1971) approach, we solve for Γ using standard optimal control techniques. First we use (22) to replace consumption with $c(\omega(\theta) - \beta V'(\theta), y(\theta), h, \theta)$. Then we treat $\omega(\theta)$ as the state variable obeying the law of motion given by the envelope condition (24) with costate variable $\mu(\theta)$. Associating the multiplier λ with the promise

keeping constraint (23), the resulting Hamiltonian reads

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

The first-order conditions are

$$\left[\frac{\partial c(\theta)}{\partial V'(\theta)} + q \frac{\partial \Gamma(V'(\theta), h'(\theta), t + 1)}{\partial V'(\theta)} \right] f(\theta) = -\mu(\theta) \frac{\partial^2 U(\cdot)}{\partial \theta \partial c(\theta)} \frac{\partial c(\theta)}{\partial V'(\theta)}, \quad (25)$$

$$\left[\frac{\partial g(h'(\theta), h)}{\partial h'(\theta)} + q \frac{\partial \Gamma(V'(\theta), h'(\theta), t + 1)}{\partial h'(\theta)} \right] f(\theta) = 0, \quad (26)$$

$$\mu(\theta) \left[\frac{\partial^2 U(\cdot)}{\partial \theta \partial c(\theta)} \frac{\partial c(\theta)}{\partial y(\theta)} + \frac{\partial^2 U(\cdot)}{\partial \theta \partial l(\theta)} \frac{\partial l(\theta)}{\partial y(\theta)} \right] = - \left[\frac{\partial c(\theta)}{\partial y(\theta)} - 1 \right] f(\theta). \quad (27)$$

The costate variable satisfies

$$\frac{\partial \mu(\theta)}{\partial \theta} = - \left[\frac{\partial c(\theta)}{\partial \omega(\theta)} - \lambda + \frac{\mu(\theta)}{f(\theta)} \frac{\partial^2 U(\cdot)}{\partial \theta \partial c(\theta)} \frac{\partial c(\theta)}{\partial \omega(\theta)} \right] f(\theta); \quad (28)$$

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

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

It follows that

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

Condition for V' : Since **[A1]** implies $\partial^2 U(\cdot) / (\partial \theta \partial c) = 0$, equation (25) simplifies to

$$\frac{1}{\partial u(c(\theta)) / \partial c(\theta)} = \frac{q}{\beta} \frac{\partial \Gamma(V'(\theta), h'(e(\theta), h), t)}{\partial V'(\theta)} = \frac{q}{\beta} \lambda'(\theta), \quad (29)$$

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

Condition for y : Using $\partial^2 U(\cdot) / (\partial \theta \partial l) = -\frac{\partial y}{\partial l} \frac{\partial^2 v(y, h, \theta)}{\partial \theta \partial y}$ in (27) yields

$$1 - \frac{\partial v(y(\theta), h, \theta) / \partial y(\theta)}{\partial u(c(\theta)) / \partial c(\theta)} = -\frac{\mu(\theta)}{f(\theta)} \frac{\partial^2 v(y(\theta), h, \theta)}{\partial \theta \partial y(\theta)}. \quad (30)$$

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 [A1] implies $\partial^2 U(\cdot) / (\partial c(\theta) \partial \theta) = 0$:

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

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 (26) and (27). The envelope condition for human capital can then be inserted into the optimality condition for human capital (26) to obtain

$$\begin{aligned} \frac{\partial g(h'(\theta), h)}{\partial h'(\theta)} &= -q \int_{\Theta} \left(\frac{\partial v(y'(\theta'), \theta', h') / \partial h'}{\partial u(c'(\theta')) / \partial c'(\theta')} + \frac{\partial g(h''(\theta'), h')}{\partial h'} \right) dF(\theta') \\ &\quad + q \int_{\Theta} \mu'(\theta') \frac{\partial^2 v(y'(\theta'), \theta', h')}{\partial \theta' \partial h'} d\theta'. \end{aligned} \quad (31)$$

For $\partial^2 U(\cdot) / (\partial c(\theta) \partial \theta) = 0$, equation (28) implies

$$\mu(\theta) = \int_{\underline{\theta}}^{\theta} \left[-\frac{1}{\partial u(c(x)) / \partial c(x)} + \lambda \right] dF(x). \quad (32)$$

Proof. Remark 1: Evaluating the law of motion (32) of the costate variable at the

upper bound of the ability distribution yields

$$\lambda - \int_{\underline{\theta}}^{\bar{\theta}} \frac{\partial c(\theta)}{\partial \omega(\theta)} dF(\theta) = \mu(\bar{\theta}) = 0. \quad (33)$$

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 $\lambda'(\theta) = \mathbb{E} \left[[\partial u(c'(\theta')) / \partial c'(\theta')]^{-1} \right]$. Combining this equality with (29), we finally obtain the reciprocal Euler equation. ■

Proof. Proposition 1: The proposition summarizes equations (31), (32), and (33) derived above. ■

Remark 4 Under assumptions [A1'] and [A2]:

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

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

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

Differentiating these expressions a second time, we get

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

and

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

Under assumptions [A1'] and [A2], $\varepsilon_{vl} = \alpha - 1$ and $\rho_{h\theta} = 1 - \chi$ so that

$$\frac{\partial^2 v(y, h, \theta)}{\partial \theta \partial h} = \frac{\partial \mathbf{v}(l)}{\partial l} \frac{y}{A} \frac{\xi \theta^{\alpha-1}}{A^\alpha} \frac{(1-\xi) h^{\alpha-1}}{A^\alpha} (\alpha + \chi).$$

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

Proof. Corollary 1: Follows immediately from Remark 4 above. ■

Proof. Remark 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

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

yields the Euler equation

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

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

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

The results above imply

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

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

condition

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

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

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

Using the envelope condition

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

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

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

■

Proof. Remark 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 (30) of the planner. To derive the analogous expression for τ_b , we recall that $\lambda'(\theta) = \mathbb{E} \left[\frac{1}{\frac{\partial u(c'(\theta'))}{\partial c'(\theta')}} \right]$ and rearrange the definition of τ_b to substitute $\partial u(c)/\partial c$ in condition (29). ■

Proof. Corollary 2: To compare the labor wedge in our model with the literature, we use definition (12) to derive

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

Thus, (15) implies that at the solution of the planner's problem,

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

By Remark 4,

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

where we have substituted in $\mu(\theta)$ using (32). ■

Proof. Proposition 2: The wedge for human capital implied by the solution to the planner's problem is obtained adding τ_h on both sides of condition (31), and substituting its definition on the right-hand side to get

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

Since the derivatives of the multivariate function $v(y, h, \theta)$ in the proof of Remark 4 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

$$\begin{aligned}\tau_h &= \underbrace{\frac{q}{\frac{\partial g(h',h)}{\partial h'}} \int_{\Theta} \frac{\partial y'}{\partial h'} \left(1 - \frac{\frac{\partial v(y',\theta',h')}{\partial y'}}{\frac{\partial u(c')}{\partial c'}} \right) dF(\theta')}_{\equiv \Delta_l} \\ &\quad + \underbrace{\frac{1}{\frac{\partial g(h',h)}{\partial h'}} \int_{\Theta} \left(\beta \frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c')}{\partial c'}} - q \right) \left(\frac{\partial y'}{\partial h'} - \frac{\partial g(h'',h')}{\partial h'} \right) dF(\theta')}_{\equiv \Delta_b} \\ &\quad - \frac{q}{\frac{\partial g(h',h)}{\partial h'}} \int_{\Theta} \frac{\partial^2 v(y',\theta',h')}{\partial \theta' \partial h'} \mu'(\theta') d\theta'.\end{aligned}\tag{34}$$

The terms Δ_l and Δ_b capture the cross-distortion induced by the wedges for labor and bequests, respectively. Since $\mathbb{E} \left[\beta \frac{\frac{\partial u(c')}{\partial c'}}{\frac{\partial u(c')}{\partial c'}} - q \right] = q \frac{\tau_b}{1-\tau_b}$, the definition (11) of the wedge for bequests implies that Δ_b is given by (16).

The third term on the right hand side of (34) 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 4 to rewrite

$$\begin{aligned}\frac{\partial^2 v(y, \theta, h)}{\partial \theta \partial h} &= -\frac{\partial^2 v(y, \theta, h)}{\partial \theta \partial y} \frac{\partial y}{\partial h} \\ &\quad + \frac{\partial \mathbf{v}(l)}{\partial l} \frac{l}{A^2} \frac{\partial A(\theta, h)}{\partial \theta} \frac{\partial A(\theta, h)}{\partial h} \left(1 - \frac{\frac{\partial^2 A(\theta, h)}{\partial \theta \partial h} A(\theta, h)}{\frac{\partial A(\theta, h)}{\partial \theta} \frac{\partial A(\theta, h)}{\partial h}} \right).\end{aligned}\tag{35}$$

The first-order condition (30) 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 (35) to compute the integral below, we find that, under assumptions **[A1']** and **[A2]**,

$$-\frac{q}{\frac{\partial g(h', h)}{\partial h'}} \int_{\Theta} \frac{\partial^2 v(y', \theta', h')}{\partial \theta' \partial h'} \mu'(\theta') d\theta' = -\Delta_l + \Delta_i, \quad (36)$$

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 (36) into (34), we obtain (17). ■