IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15073

Opportunity and Inequality across Generations

Winfried Koeniger
Carlo Zanella

FEBRUARY 2022

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15073

Opportunity and Inequality across Generations

Winfried Koeniger
University of St.Gallen, CESifo, CFS and IZA

Carlo Zanella
University of Zurich

FEBRUARY 2022

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Opportunity and Inequality across Generations*

We analyze how intergenerational mobility and inequality would change relative to the status quo if dynasties had access to optimal insurance against low ability of future generations. Based on a dynamic, dynastic Mirrleesian model, we find that insurance against intergenerational ability risk increases in the social optimum relative to the status quo. This implies less intergenerational mobility in terms of welfare but no quantitatively significant change in earnings mobility. Earnings mobility is thus similar across economies with different incentives and welfare, illustrating that changes in earnings mobility cannot be interpreted readily in welfare terms without further analysis.

JEL Classification: E24, H21, I24, J24, J62

Keywords: asymmetric information, intergenerational mobility, inequality, human capital, schooling, bequests

Corresponding author:
Winfried Koeniger
Department of Economics
University of St.Gallen
SEW-HSG
Varnbüelstraße 14
9000 St.Gallen
Switzerland

E-mail: winfried.koeniger@unisg.ch

* We thank Jonathan Heathcote, anonymous referees and participants at various seminars for very helpful comments, and Alessandro Di Nola and Julien Prat for insightful discussions.

1 Introduction

Dynasties are exposed to the risk that their future generations may have low ability and thus low productivity in the labor market. Optimal insurance of ability risk has to take into account incentives which depend both on inequality and intergenerational mobility. Insurance of ability risk in turn shapes inequality and intergenerational mobility.

In the public debate, high intergenerational mobility is usually considered to be desirable, in the sense that high ability should allow individuals to rise to the top independent of their background. From the perspective of altruistic dynasties interested in sharing intergenerational risk, however, intergenerational mobility in terms of consumption decreases welfare.

A central question is thus to which extent optimal policy should provide insurance against the ability risk that each generation of a dynasty faces and whether optimal policy changes the patterns of inequality and mobility relative to those observed in the status quo. To shed light on this question, we contribute to the literature by providing a quantitative analysis in a dynamic, dynastic Mirrleesian model in which parents' nurture through bequests and schooling investments may partially insure a generation of a dynasty against the risk of receiving a bad ability draw by nature. Bequests and the schooling investment into the next generation are observable but the draw of ability is hidden, stochastic and persistent across generations. We characterize the social optimum as the solution of the dynamic Mirrleesian problem, in which asymmetric information constrains the insurance that a utilitarian planner can provide against the ability risk that dynasties face. This makes the quantitative comparison between the economy, which we calibrate to the U.S., and the social optimum non-trivial.

Inequality and less than full insurance are also a feature of the social optimum, in which the extent of intergenerational insurance and mobility is determined by optimal incentive provision as pointed out by Phelan (2006). Whether mobility or inequality increase or decrease in the social optimum relative to the status quo is thus a quantitative question, not only for the mobility of consumption and welfare but also for the mobility of earnings.

As part of our quantitative analysis, we investigate whether the intergenerational earnings mobility is too low in the status quo relative to the social optimum. We find that this is not the case as standard measures of earnings mobility have quantitatively similar values in the status quo and in the social optimum. The mechanisms, however, that drive earnings mobility are very different in the status quo and in the social optimum. On the one hand, the absence of borrowing constraints for families in the social optimum reduces the intergenerational correlation of schooling relative to the status quo. On the other hand, stronger incentives in the social optimum increase the correlation between labor effort and ability, within and across generations relative to the status quo. As labor earnings depend on schooling, ability and labor effort, the overall effect of implementing the social optimum on the intergenerational earnings mobility turns out to be quantitatively small.

In our analysis, we analyze a reform which implements the social optimum starting from the steady state of the economy calibrated to the U.S. We find that, after the reform, the consumption of each generation is less associated with its produced output implying more intergenerational insurance. This is achieved by increasing the role of nurture (in terms of bequests and human capital investment) relative to nature (in terms of ability). More intergenerational insurance against ability risk is thus associated with less social mobility, in the sense that the rank in the welfare distribution becomes more persistent across generations and depends less on the ability of the current generation. A contribution of our analysis is that we quantify the extent to which this is the case.

Our quantitative analysis further reveals that inequality of labor earnings increases after the reform but this does not make dynasties worse off ex ante because the reform also increases intergenerational insurance. Earnings mobility, as measured by the intergenerational earnings elasticity or rank-rank correlation of earnings, remains roughly unchanged instead. Thus, earnings mobility may be very similar across economies with very different incentives, insurance opportunities and welfare. These results illustrate that changes in earnings mobility and inequality cannot be interpreted readily as *good* or *bad* without further analysis.

How is more intergenerational insurance achieved in the social optimum? We find that, on average, labor income is taxed in the social optimum, at a rate of 36% in the second generation after the reform, whereas bequests and schooling are subsidized at rates of 25% and 46%, respectively. This relates to results in Farhi and Werning (2010) who show that, in an implementation with history-dependent tax schedules, bequests and human capital should be subsidized. In our model, subsidies for schooling investment and bequests play a different role in shaping inequality and opportunity also because we relax the assumptions in Farhi and Werning (2010) that children make no labor supply decision and that there is no uncertainty.[1] We show that bequest subsidies are phased out much more progressively across the earnings distribution than schooling subsidies. The intuition is that higher current ability is positively correlated with income, increases the expected ability of the offspring, and ability and schooling are complements for productivity. As a consequence, bequest subsidies are targeted relatively more towards the income poor than schooling subsidies.

[1] Farhi and Werning (2010) mention in their discussion of proposition 5 for history-dependent tax schedules that the symmetry in the taxation of bequests and human capital only holds under these assumptions.

In our quantitative analysis, we focus on insurance and refrain from adding motives for redistribution across dynasties by applying different weights in the planner's objective function. Specifically, we hold constant the present discounted value of the expected net costs for the allocation of each dynasty at the time of the reform and give equal weight to dynasties in the planner's objective function. Our reform experiment thus answers the question how much additional insurance the planner provides ex post by using an optimal policy without redistributing across dynasties at the time of the reform. This makes the reform also implementable from a politico-economic point of view. We discuss in subsection 5.4 that typical alternative welfare functions, which assign relatively more weight to families with lower ability, would imply even more intergenerational insurance in the social optimum relative to the calibrated economy.

Because the implementation of the social optimum requires a rather complex history-dependent tax and subsidy system (Farhi and Werning, 2010), we compare the results on insurance, mobility and welfare in the social optimum to those obtained in economies with simple tax and subsidy systems. To the best of our knowledge, such analysis of simpler tax systems has not been provided yet in the context of intergenerational models with persistence in unobservable ability.

Compared to the social optimum, the patterns for mobility and insurance differ significantly in the economies with linear taxes and subsidies on current labor income, bequests and schooling. We find that higher earnings mobility is not associated necessarily with higher welfare. Whereas earnings mobility is higher in the economies with the simple tax and subsidy systems than in the social optimum, only about half of the welfare gains of moving from the calibrated economy to the social optimum can be achieved in these economies. The smaller welfare gains illustrate the limited extent to which bequests and human capital (the endogenous state variables in the model) together with labor income, which depends on the ability draw of the current generation, can capture the history dependence of the optimal allocation in an environment with persistent ability shocks. Albanesi and Sleet (2006) have shown that the history can be fully summarized by conditioning optimal taxes on the endogenous state variable(s) if shocks to unobserved ability are i.i.d. and thus not persistent, and these results have been extended by Stantcheva (2017) in a life-cycle model with assets and human capital. In our model with persistent shocks to ability, history-independent tax schedules that condition on bequests, schooling and labor income only allow to implement the social optimum approximately.

Further relation to the literature

Our analysis builds on the classic literature on mobility and inequality across generations (Becker and Tomes, 1986). Inequality across generations is transmitted through parents' nurture and nature so that "one of the biggest risks in life is the family one is born into" (Farhi and Werning, 2010). Optimal insurance of this risk has to take into account that persistent inequality in "long-run income status" (Friedman, 1962) is inefficient if the lack of income mobility prevents more talented individuals in society to be more productive. Less income mobility may make the distribution of welfare, equal to the present value of period utilities, more unequal (Flinn, 2002) and feeds back into the demand for social insurance of those with currently low income (Bénabou and Ok, 2001).

Our analysis of the social optimum builds on Phelan (2006) and Farhi and Werning (2007) who have shown that the social optimum in a dynamic Mirrleesian economy with asymmetric information need not imply immiseration as in Atkeson and Lucas (1992) if the planner discounts the future but attaches more weight to future generations than implied by the altruism of a family dynasty. We proceed as in Farhi and Werning (2007) and assume that dynasties are weighed equally in the planner's problem. This implies a wedge between the discount rate that the planner and the family dynasty apply to the utility of each generation: the planner cares directly about the welfare of a future generation and also indirectly, given that family dynasties care about their offspring. The non-degenerate steady-state distribution of consumption in the social optimum, resulting from the wedge in discount rates, distinguishes our model setting from Koeniger and Prat (2018) and allows for a meaningful analysis of the transition from the steady state of the calibrated economy to the social optimum. This is essential to answer the central question on intergenerational insurance we pose in this paper. Because the wedge between the discount rates is an important determinant of the distribution in the social optimum, we explain in Section 2.1 how it naturally relates to the difference between the real interest rate and the discount rate of family dynasties in a small open economy with incomplete markets if we focus on the case with stationary (average) consumption in the long run. We discipline this difference in the calibration by matching median bequests observed in U.S. data.

The disciplining of the discount factor, which determines the altruistic motive of dynasties in the welfare function, matters for the trade-offs that arise depending on whether equality of opportunity is considered from a dynastic point of view, tilting the balance towards more insurance across generations increasing the role of nurture, or from the point of view of individual families within a generation, tilting the balance towards increasing the role of nature, more mobility and thus less intergenerational insurance.[2] Interpreta-

[2]See, for example, Lee and Seshadri (2018) and the recent discussion of different interpretations of equal opportunity in Arneson (2018).

tion of the empirical evidence on mobility patterns and inequality by Chetty et al. (2014), for example, thus depends on parameters of the social welfare function such as the discount factor.

Our analysis relates to the large strand of literature on optimal taxation of income, human capital or bequests. The special issue on human capital and inequality edited by Corbae et al. (2017) and the volume on inequality and redistribution of the Carnegie-Rochester-Conference (2016) provide a good overview over recent research. Gelber and Weinzierl (2016) analyze optimal income taxation when the ability type of the next generation depends on the resources of the current generation. In their model, the probability of children's types directly depends on parents' disposable income. Quantitatively, they find that a utilitarian planner then chooses more redistributive income taxation so that the socially-optimal marginal tax rates for most types are higher than observed in the U.S. Our model shares the feature that parents' resources may impact the earnings capacity of future generations. This is explicitly modeled because parents choose the amount of resources allocated to human capital accumulation and bequests. The utilitarian planner in our analysis maximizes welfare by choosing *jointly* how much to tax or to subsidize labor income, schooling and bequests. The additional policy instruments imply that income taxes can be less redistributive compared with Gelber and Weinzierl (2016) because intergenerational insurance is achieved also by incentivizing nurture through bequests and human capital accumulation.

Optimal taxation of human capital using a Mirrleesian approach has been analyzed by Findeisen and Sachs (2016), Kapička (2015), Kapička and Neira (2019), Stantcheva (2015, 2017), and Koeniger and Prat (2018). Heathcote et al. (2017), Krueger and Ludwig (2016), Lee and Seshadri (2019) and Peterman (2016) are examples for analyses based on a Ramsey approach to optimal taxation. Heathcote and Tsujiyama (2021) combine both approaches in their analysis of optimal income taxation. Farhi and Werning (2010), Pavoni and Yazici (2017) and Phelan and Rustichini (2018) analyze optimal taxation of intergenerational transfers such as bequests or inheritances. The contribution of our paper is to apply a dynastic Mirrleesian model with bequests and human capital accumulation to quantify the differences between intergenerational insurance, mobility and welfare in the social optimum and in an economy calibrated to the status quo in the U.S., using approximations of existing tax schedules. In this analysis, the social optimum does not imply immiseration in the long run as in Atkeson and Lucas (1992).

We use the calibrated model to complement the vast empirical literature on inequality and mobility. Recent examples are Chetty et al. (2014) or Gallipoli et al. (2020) for the U.S., Dodin et al. (2021) for Germany, Güell et al. (2018) for Italy, Markussen and Røed (2020) for Norway, Adermon et al. (2018) for Sweden, Chuard and Grassi (2020) for Switzerland, and their references provide a good overview of the literature. Quantifying the distance of the social optimum to the calibrated economy in terms of mobility, inequality and welfare, we compute measures which rely on the structural model because they require computing welfare, as well as measures that have counterparts in the reduced-form empirical literature such as earnings mobility. This allows us to provide a structural interpretation of what observable changes in insurance and mobility imply in terms of welfare.

The model further allows us to illustrate mechanisms through which nurture and nature affect mobility and the intergenerational transmission of inequality. This provides intuition for the channels through which intergenerational insurance can be provided in the model and in the transition to the social optimum. The illustrations have to be taken with a grain of salt because the structure of the model economy is tailored to keep the solution of the dynamic Mirrleesian problem tractable. The positive analyses in Cunha et al. (2006), Lee and Seshadri (2019) and Daruich and Kozlowski (2020) and Daruich (2020), for example, contain additional features such as early childhood investments or fertility from which we abstract in our analysis. We try instead to connect the optimal taxation literature to applied work on social insurance, similar in spirit to the analysis of Golosov et al. (2016) but with a focus on intergenerational insurance.

Our analysis proceeds in the following steps. In Section 2, we model the decision problem of family dynasties and provide a brief discussion of the social optimum and the implied optimal tax and transfer system. We then calibrate the model to U.S. data in Section 3. In Section 4, we inspect model mechanisms which provide intuition for our analysis of the constrained-efficient social optimum in Section 5. We investigate how a tax reform changes inequality and mobility on the transition from the calibrated economy to the socially optimal steady state. We compare these results to those obtained for economies with simple history-independent tax schedules and conclude in Section 6. The appendix contains a robustness analysis for alternative assumptions on the complementarity of schooling and ability in the production function, on the persistence of ability, on the Frisch elasticity of labor supply, and on the bequest target in the calibration.

2 The model

We use a dynasty model to capture key mechanisms through which nurture, in terms of schooling and bequests, and nature, in terms of ability, affect inequality and mobility across generations. We analyze decisions of family dynasties who are composed of parents and children in each generation, have an infinite planning horizon and a size normalized to one. Each generation of a dynasty chooses, conditional on the parents' ability draw, the labor supply of the parents, consumption, and the bequests and schooling for the children.

Our dynasty model builds on Koeniger and Prat (2018) and is deliberately simpler than the model by Lee and Seshadri (2019), for example, who analyze the sequence of decisions over the life cycle for each generation in more detail. The simplicity of our model keeps the dynamic Mirrleesian planner problem tractable when we solve for the social optimum. This allows us to make our main contribution, i.e., the analysis of how intergenerational insurance, mobility, inequality and welfare change on the transition from the calibrated steady state to the social optimum, and in economies with simpler tax schedules which implement the social optimum approximately.

Preferences are time separable across generations and 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:[3]

$$\mathbf{U}(c_t, l_t) = u(c_t) - \mathbf{v}(l_t),$$

where $u(c_t) \in \mathcal{C}^2(\mathbb{R}_+)$ is increasing in c_t and strictly concave, and $\mathbf{v}(l_t) \in \mathcal{C}^2(\mathbb{R}_+)$ is increasing in l_t and strictly convex.

Each generation of a family differs in its ability θ_t. Ability is not observable so that tax schedules cannot be conditioned on it. Bequests b_t and human capital h_t, think of years of (non-compulsory) schooling and high-school or college degrees, are public knowledge instead. Output y_t is produced with the technology $Y(h_t, l_t, \theta_t)$ which is increasing in its arguments and concave. Although output y_t is observable, actual labor supply l_t cannot be inferred from it because ability θ_t is stochastic and hidden.

The expenditure of schooling $g(h_{t+1}, h_t)$ depends on the amount of human capital investment h_{t+1} into the children and on the family background, which we summarize by the stock of human capital of parents h_t. This cost function follows from inverting a human-capital production function in the spirit of Ben-Porath (1967), where human capital of the next generation depends on the expenditure on schooling and parental background.[4]

At the beginning of each period, the dynasty learns the ability of the parents and then makes its choices about labor supply, consumption, bequests and human capital investment. Ability is drawn from the bounded interval $\Theta \equiv [\underline{\theta}, \overline{\theta}] \subset \mathbb{R}_+$, where we assume a

[3]We use labor effort and labor supply interchangeably when referring to l_t.

[4]We abstract from modeling a parental time input because such an input is plausibly unobservable which would render the analysis much less tractable. We also abstract from a direct influence of the childrens' ability on the cost of human capital investment for parsimony. This would add another channel through which output would depend on ability but would not add further insights as long as the observation of human capital investments does not provide information about ability.

continuously differentiable distribution $F : \Theta \to [0,1]$ with conditional density $f(\theta_t|\theta_{t-1})$. For the analysis of the planner's problem we make the further assumptions that $f(\theta_t|\theta_{t-1})$ has full support, that it is of class $\mathcal{C}^2$ with respect to its second argument θ_{t-1}, and that it has a bounded derivative $|\partial f(\theta_t|\theta_{t-1})/\partial\theta_{t-1}| \le B$ for some $B \in \mathbb{R}_+$. The dependence of the distribution $F(\theta_t|\theta_{t-1})$ on the type of the previous generation allows us to model intergenerational transmission of ability, which may occur because of genetic inheritance or nurture in early childhood. The quantitative effect of nature in this paper should thus be considered an upper bound, given that ability may also contain some nurture component. In the calibration discussed in Section 3, we discipline the transmission of ability by choosing the correlation between θ_t and θ_{t-1} such that the intergenerational correlation of earnings based on model simulations matches the empirical counterpart in U.S. data.

The stationary recursive problem of the family dynasty is

$$
\begin{aligned}
\widehat{W}(b,h,\theta) = & \max_{\{b',h',l,c\}} \left\{ \mathbf{U}(c,l) + \beta \int_{\Theta} \widehat{W}(b',h',\theta')\, dF(\theta'|\theta) \right\} \qquad (1)\\
\text{s.t. } b' &= (1+r)b - T^b(b) - c - g(h',h) - T^h(h',h) + y - T^y(y),\\
b' &\ge \max\{-\phi g(h',h), \underline{b'}\},\\
y &= Y(h,\theta,l),\\
\ln(\theta') &= \rho \ln(\theta) + \epsilon,
\end{aligned}
$$

where β is the discount factor of the family measuring the strength of the altruism towards future generations, $0 \le \rho \le 1$ captures the persistence of shocks to ability, and "$'$" denotes values of variables one period in the future. Families can pass on the fraction $\phi \in (0;1)$ of the schooling expenditures to the next generation but the borrowing constraint limits the overall student debt to $\underline{b'}$. The functions $T^i(\cdot)$, for $i = b,h,y$, denote the non-linear schedules for taxes and subsidies on bequests, education and labor income, respectively. Since $g(h',h)$ and $T^h(h',h)$ enter additively in the budget constraint, we set $T^h(h',h) = 0$ and interpret $g(h',h)$ as the *net* cost of human capital accumulation.

Concerning the intergenerational transmission of inequality in the (calibrated) model, there are important differences in how bequests, schooling and ability affect this transmission, as we illustrate in Section 4. Bequests or inheritances decrease the incentive to exert labor effort through a negative wealth effect and thus induce mean reversion in labor earnings.[5] Investment into schooling instead increases the persistence in labor earnings across

[5] Such a negative wealth effect on life-cycle labor supply, through early retirement or less labor force participation of some household members, is supported empirically by Holtz-Eakin et al. (1993) and Brown et al. (2010). See also the evidence for lottery winners by Imbens et al. (2001) for the U.S., by Cesarini et al. (2017) for Sweden, and the analysis of Kindermann et al. (2020) on the consequences of this wealth effect for the taxation of bequests and labor income. The wealth effect also implies less investment into schooling because schooling results in more resources only if combined with labor effort. Although the

generations because more human capital increases labor productivity and thus earnings, given that the income and substitution effect of the productivity increase on labor supply approximately balance each other. The transmission of inequality by nature (ability) will turn out to be more important than by nurture (bequests or human capital investment), in line with recent empirical findings of Bingley et al. (2018) and Gallipoli et al. (2020).[6]

2.1 Social optimum

We characterize key features of the allocation in the social optimum because our goal is to analyze changes in intergenerational insurance, mobility and inequality if we implement the social optimum starting from the steady state of the calibrated economy. In the social optimum, asymmetric information constrains the insurance provided by a utilitarian planner who discounts the future and weighs family dynasties equally. This implies non-degenerate inequality and mobility so that comparison between the calibrated economy and the social optimum is non-trivial. In the first best allocation instead, consumption is fully decoupled from production as derived in Appendix B. Whereas consumption is fully insulated from ability shocks in the first best, labor supply and human capital investment positively covary with ability. Indeed, as reported in Appendix B, the correlation of labor supply and schooling with ability is higher in the first best than in the second best.

We characterize the second best focusing on the planner problem with full commitment which provides an upper bound for the amount of insurance the planner can provide given the constraints. In such an environment, Farhi and Werning (2007) have analyzed allocations chosen by a utilitarian planner who discounts the future less than family dynasties and weighs dynasties equally. Denoting the planner's discount factor with ψ and the dynasties' discount factor with β, they considered the case in which $\psi > \beta$. They showed that this assumption breaks the immiseration result of Atkeson and Lucas (1992) and implies a non-degenerate stationary distribution of consumption and welfare in the planner problem.

We refer to Appendix A for details about the planner problem, its recursive formulation and solution. We emphasize two key equations, derived in the appendix, which show how the parameters ψ and β shape the solution of the planner problem. Let $\mathbb{E}$ denote the expectation operator, $u'(\cdot)$ the marginal utility of consumption, and let $c_t(\theta^t)$ denote consumption at time t as function of the sequence of abilities θ^t until time t, which is

negative wealth effect dominates on average, we find that more wealth relaxes borrowing constraints for some dynasties and thus increases their investment into schooling.

[6]The empirical importance of nature and nurture for the intergenerational transmission of inequality is still a matter of debate. Lee and Seshadri (2019) argue, based on a rich structural model, that the empirical estimates may overstate the importance of nature because they do not account for general equilibrium effects.

truthfully revealed by families to the planner in equilibrium. The evolution of consumption in the social optimum is characterized by

$$\mathbb{E}\left[\frac{1}{u'(c_t(\theta^t))}\right] = \frac{1/(1+r)}{\psi}\mathbb{E}\left[\frac{1}{u'(c_{t+1}(\theta^{t+1}))}\right]. \tag{2}$$

For stationarity it is thus necessary that $\psi = 1/(1+r)$, i.e., the planner's discount rate has to equal the real interest rate given that there is no aggregate risk and that the planner can diversify the idiosyncratic ability risk. Furthermore, we obtain an equation based on expectations conditional on the sequence θ^t:

$$\mathbb{E}\left[\frac{1}{u'(c_{t+1}(\theta^{t+1}))}\middle|\theta^t\right] = \frac{\beta}{\psi}\frac{1}{u'(c_t(\theta^t))} + \eta_t\left(1 - \frac{\beta}{\psi}\right), \tag{3}$$

where η is the multiplier attached to the constraint that captures the differences between the promises made by the planner and the promises which the planner would make if the dynasties' rate were applied to discount the welfare of future generations.

For $\beta/\psi < 1$ and $\eta > 0$, equation (3) shows that $1/u'(c(\theta^t))$ is mean-reverting. Because the planner cares more about providing equal opportunities for future generations, shocks to ability are not fully passed on to them. The mean reversion implies insurance against ability risk in the social optimum which contrasts the increase of inequality required for incentive provision. For $\beta = \psi$ instead, $1/u'(c(\theta^t))$ would follow a martingale, implying immiseration as in Atkeson and Lucas (1992).

As Farhi and Werning (2007), pp. 375-376, we focus on the case $\psi = 1/(1+r)$ in which the social optimum implies stationary (average) consumption. Our calibration of $\beta(1+r) < 1$ then disciplines the extent to which the planner cares more about providing opportunities for future generations than a dynasty itself. This implies a non-degenerate stationary *distribution* of consumption in the social optimum and thus makes it sensible, in our view, to analyze a tax reform which implements a transition towards the socially-optimal steady state, starting from the status quo characterized by the economy that we calibrate to the U.S.

2.2 Optimal taxes and subsidies to implement the social optimum

To interpret the allocation in the social optimum, it is useful to define wedges based on the first-order conditions of the laissez faire economy, as done in Appendix C. Evaluating the wedges at the social optimum reveals which choices the planner has to encourage or discourage to attain the social optimum. We show in Appendix C how the wedges relate to taxes or subsidies in an (approximate) implementation of the social optimum. For more detailed decompositions of the wedges, we refer to Stantcheva (2017) or Koeniger and

Prat (2018). In addition to the effects discussed in these papers, in our model the planner wants to encourage bequests and human capital relatively more because the planner is more patient than the dynasties. This implies relatively more Pigouvian subsidization of these choices.

The optimal labor tax in our model is qualitatively the same as in standard Mirrleesian models and the bequest subsidy is phased out progressively (Farhi and Werning, 2010). The schooling subsidy offsets the distortion of the schooling decision resulting from the taxation of labor income (Bovenberg and Jacobs, 2005), depends on the riskiness of human capital and corrects distortions of the accumulation motive and incentives. As emphasized by Jacobs and Bovenberg (2011), Stantcheva (2017) and Koeniger and Prat (2018), schooling has no effect on the information friction, i.e., on the incentive to truthfully reveal unobserved ability in the planner problem, if the elasticity between unobservable ability and observable human capital in the production function is unity. This is the case in our benchmark calibration. In the robustness analysis in which we recalibrate the economy, allowing for more complementarity between schooling and ability, schooling exacerbates the information friction by increasing information rents. The results reported in Appendix G.1 show that our main quantitative findings are unaffected by such recalibration.

3 Calibration

To quantify the changes in the intergenerational transmission of inequality and mobility, implied by a tax reform that implements the social optimum, we calibrate the model of family dynasties to the U.S. economy.

Preferences.—A key parameter in our calibration is the wedge between the rate at which the planner and the dynasty discount the welfare of future generations. Although this may seem a minor detail, it has major implications for the stationary distribution in the social optimum. As we have seen in Section 2.1, stationary (average) consumption in the planner problem requires that the discount rate of the planner equals the intertemporal marginal rate of transformation, i.e., the real interest rate. Whereas the planner can fully diversify the idiosyncratic ability risk, individual dynasties in the calibrated economy cannot because they only have access to a non-contingent bond. It is well known that, in such an economy with incomplete markets, stationarity requires that the discount rate of the families is higher than the real interest rate so that the discount factor $\beta < 1/(1+r)$. Thus, we discipline the wedge in the discount rate between the families and the planner by calibrating β so that, for a given real interest rate of 3%, the stationary distribution for bequests in the model implies a median for bequests that matches the median observed in the data, conditional on receiving a positive bequest.[7]

Evidence in Table 2 of Wolff and Gittleman (2014), based on the Survey of Consumer Finances (SCF) in the time period 1989 – 2007, shows that the median wealth transfer among households in the U.S. has been \$73,600, conditional on receiving a transfer.[8] We adjust this figure for household size dividing it by 1.4.[9]

The rest of our calibration strategy follows Koeniger and Prat (2018) by and large and we repeat it here for completeness. We specify the utility function as $\mathbf{U}(c,l) = \ln(c) - l^{\alpha}/\alpha$, which satisfies the assumptions for the utility function made in Section 2. The estimate for the Frisch elasticity of 0.5 documented in Chetty (2012) implies that $\alpha = \varepsilon^{-1} + 1 = 3$.

Technology.—The length of a period in the model is 30 years to approximate the time until labor-market entry of a new-born generation and the length of the labor-market career. We set the annualized real interest rate to 3% and assume the production technology

$$Y(h,l,\theta) = A(\theta,h)\, l$$

with labor productivity

$$A(\theta,h) = \left[\xi\theta^{\frac{\chi-1}{\chi}} + (1-\xi)h^{\frac{\chi-1}{\chi}}\right]^{\frac{\chi}{\chi-1}}$$

and $\chi \in [0,\infty)$, $\xi \in (0,1)$.

The linearity of the production technology in labor effort and the assumption of a given interest rate, which is not influenced by accumulation behavior within the U.S., imply that we can solve the problem of the dynasties separately from each other. As a benchmark, we assume that the elasticity of substitution $\chi = 1$ so that labor productivity is a Cobb-Douglas function of ability and human capital: $A(\theta,h) = \theta^{\xi}h^{1-\xi}$. We will check the robustness of our results for a different degree of complementarity between ability and schooling.

Cobb-Douglas productivity has the advantage that, for competitive labor markets,

[7] In the robustness analysis discussed in Section 5 we report results if we match the conditional *mean* of bequests instead. The interest rate of 3% approximates the inflation-indexed long-term average yield before the Great Recession, reported by the Federal Reserve in Table H.15.

[8] The value is expressed in terms of dollars in 2007. Table 1 in Wolff and Gittleman (2014) shows that 84% of the wealth transfers are classified as inheritances. Most other transfers are classified as gifts and most transfers are from family members. Given that the timing of intergenerational transfers before or after death is difficult to map into our model, we consider all wealth transfers.

[9] This number is reported in Table 8 of Hintermaier and Koeniger (2011) who compute household size for the waves of the SCF in the same time period, based on an equivalence scale that assigns a weight of 1 to the first person in the household, 0.34 to the second person and approximately 0.3 to each additional member of the household. See also Fernández-Villaverde and Krueger (2007), Table 1, last column.

wages $w(\theta,h)$ are log-linear in human capital and unobserved ability:

$$\ln w(\theta,h) = \ln A(\theta,h) = (1-\xi)\ln h + \xi \ln \theta, \tag{4}$$

so that it is straightforward to use the variance of residual wages as target to calibrate the variance of unobserved ability θ.[10] We assume that θ is drawn from a log-normal distribution, approximated by a truncation to obtain a compact support as in Farhi and Werning (2013). The mean and standard deviation specified in Table 1 imply a variance of residual wages of 0.2. This corresponds to the cross-sectional residual variance of earnings of the parent and children generation in the PSID for the U.S. reported in Gallipoli et al. (2020), Table 17, and the part of the variance of residual log-wages that is generated by persistent shocks reported in Heathcote et al. (2010) for the U.S. in 2005.[11] The larger variance of ability compared to Lee and Seshadri (2019) and Daruich and Kozlowski (2020) tries to capture that we do not model investments of parents into children during early childhood, and as shown in Table 10 of Lee and Seshadri (2019) such investments increase the heterogeneity substantially until labor market entry.[12] We also report results of a robustness check in which we target a lower intergenerational earnings elasticity and thus calibrate a lower persistence of the ability shock, which makes the shock easier to insure.

We refer to the vast empirical evidence on Mincerian wage regressions to calibrate the parameter ξ of the production function. In his survey, Card (1999) shows that the marginal return to schooling is quite robustly estimated across studies and close to 10%. Equation (4) thus implies $1-\xi = 0.1$, given that years of schooling S correspond to $\ln h$ in our model, where compulsory schooling is defined as $h = 1$ in which case the chosen years of schooling $S = \ln 1 = 0$. Because schooling and ability are positively correlated,

[10]Given (4), the variance of residual wages is the variance of wages which remains after regressing log-wages on years of schooling where, in our model, chosen years of schooling S beyond the compulsory pre-high school years correspond to $\ln h$.

[11]See panel C of Figure 3 in Heathcote et al. (2008). We focus on the variance resulting from persistent shocks because a generation's labor-market career takes place within a period in our model so that transitory shocks (at least partially) cancel out and θ, within a period, is fully persistent. Note further that the variance of labor earnings approximately equals the variance of wages in our model because the income and substitution effect of an increase in labor productivity on labor effort balance each other in our calibration.

[12]The importance of initial conditions for the welfare of generations has been emphasized by Keane and Wolpin (1997) and more recently by Huggett et al. (2011), De Nardi and Yang (2016) and Lee and Seshadri (2019). Structurally estimating career decisions in the U.S., Keane and Wolpin (1997) find that initial conditions at age 16 explain 90 percent of the total variance in expected lifetime utility. Based on a model with risky human capital, Huggett et al. (2011) find that differences in initial conditions at labor market entry account for more than 60% of the variation in lifetime utility. This percentage increases to more than 70% in Lee and Seshadri (2019) who model human capital formation early in life. The dominance of initial conditions for each generation's welfare also motivates our focus on the opportunities and the transmission of inequality across generations rather than on differences that arise due to shocks within the labor-market career of a generation emphasized in Storesletten et al. (2004).

schooling explains 23.4% of the variation in a regression of log wages on schooling based on the simulated model data. This is broadly in line with the empirical evidence by Card (1999) who reports that 20 – 35% of the variation in earnings data are explained by a linear schooling term and a low-order polynomial for experience.

Borrowing opportunities.—We set the parameters of the borrowing constraint in problem (1) to $\phi = 0.5$ and $\underline{b}' = -\$30,000$. This implies that families can finance up to 50% of their human capital investment into their children, with a maximal amount of debt of \$30,000. At the time the next generation makes its choices the accrued interest then implies a maximal total debt of \$72,818 so that the amount for outstanding student loans broadly matches the amounts reported in Lee et al. (2014).

Approximation of tax schedules.—We use the parametrization proposed by Heathcote et al. (2017) to approximate labor income taxes in the U.S.: $T^y(y) = y - \delta y^{1-t_y}$, with $t_y = 0.181$ and $\delta = 0.9276$.[13] We approximate taxes on bequests using the parametrization for estate taxes proposed by De Nardi and Yang (2016) because our model does not distinguish estates from bequests. Thus, families pay 20% tax if the bequest exceeds the exemption of \$756,000. The function $g(h', h)$ captures net education costs after subsidies and we discuss the calibration of its parameters in the next subsection. In the calibrated economy, households' tax payments exceed transfers. We assume that the surplus finances an exogenous stream of government expenditures other than the transfers to households. The net present value of these expenditures amounts to 27 percent of average labor earnings, and the flow of these government expenditures equals 16 percent of average labor earnings broadly in line with Heathcote et al. (2017). We hold the expenditures constant when analyzing the transition to the social optimum, as discussed further in Appendix F on the implementation of the reform.

Stochastic process for ability and education costs.—The parameters for the persistence of ability shocks and the education cost function are calibrated jointly together with the discount factor to match the following target statistics: median bequests, the average years of schooling beyond the compulsory eight years of pre-high school education, the average net cost of an additional year of secondary/tertiary education, the correlation between years of schooling across generations and the intergenerational earnings elasticity. Appendix D contains further details on the implementation of the calibration.

[13] $T^y(y)$ is negative if $y < \delta^{\frac{1}{t_y}} \approx 2/3$. A unit in our model corresponds to mean earnings of high-school dropouts, as explained further below. Thus, in our calibrated model, workers receive positive transfers if their annual income is below \$14,423. Otherwise they pay taxes.

Parameters	Target	Source
Preferences		
Discount factor (annualized): $\beta = 0.966$	Median bequest: \$52,571 (conditional on receiving one, equivalized)	Wolff and Gittleman (2014)
Disutility of labour $v(l) = l^{\alpha}/\alpha$: $\alpha = 3$	Frisch elasticity: 1/2	Chetty (2012)
Storage technology		
$r = 0.03$	Annualized real interest rate	Federal Reserve, Table H.15
Production technology: $y/l = \theta^{\xi} h^{1-\xi}$		
$\xi = 0.9$	Returns to education: 10%	Card (1999)
Borrowing opportunities		
$\phi = 0.5$, $\underline{b'} = -\$30,000$	Student loans in FRBNY Consumer Credit Panel	Lee et al. (2014)
Taxes		
$T^{y}(y) = y - \delta y^{1-t_y}$, $t_y = 0.181$, $\delta = 0.9276$	Parametric approximation of the U.S. labor income tax schedule	Heathcote et al. (2017)
$T^{b}(b) = \max\{t_b(b - x_b), 0\}$, $t_b = 0.2$, $x_b = \$756{,}000$	Parametric approximation of the U.S. estate tax schedule	De Nardi and Yang (2016)
AR(1)-process for ability: $\ln(\theta') = \rho \ln(\theta) + \epsilon$, $\ln\theta' \sim \mathcal{N}\left(-\frac{\sigma_{\epsilon}^2}{2(1-\rho^2)}, \frac{\sigma_{\epsilon}^2}{1-\rho^2}\right)$		
$\rho = 0.44$	Intergenerational earnings elasticity: $\iota = 0.45$	Chetty et al. (2014)
$\frac{\sigma_{\epsilon}^2}{1-\rho^2} = \frac{0.2}{\xi^2}$	Variance of residual wages: 0.2	See discussion in main text
Education cost: $g(h', h) = \kappa (h')^{\varsigma_1} h^{\varsigma_2}$		
$\kappa = 0.0014$	Average years of schooling: 4.86 (beyond 8 years of pre-high school education)	Barro and Lee (2013)
$\varsigma_1 = 0.7469$	Average net cost for an additional year of education: \$13,845	OECD (2011), Stantcheva (2017)
$\varsigma_2 = -0.0004$	Intergenerational correlation of years of schooling: 0.46	Hertz et al. (2008)

Table 1: Calibrated parameter values

We have chosen the target moments so that they are tightly related to the parameters we calibrate. Although jointly calibrated, each target is closely related to the calibration of one of the parameters. As mentioned above, calibration of the discount factor β helps to match the median bequest. The persistence in the stochastic process of ability allows us to match the intergenerational earnings elasticity ι (IGE), resulting from a linear regression of $\ln y'$ on $\ln y$. This is intuitive because ability θ affects labor earnings through changes in labor productivity $A(\theta, h)$. Given that the exponent of ability ξ in the Cobb-Douglas function for labor productivity is nine times higher than the exponent of human capital, the IGE is mostly determined by the persistence of ability that is fed into the model. The endogenous choices of human capital and labor supply quantitatively matter much less for labor earnings and thus for the IGE.

The parameters κ and ς_1 of the cost function $g(h', h)$ in Table 1 allow us to match average years of schooling and the corresponding average net cost for schooling. The population above age 15 in the US on average has 12.86 years of schooling, as reported in the database on educational attainment provided by Barro and Lee (2013).[14] This corresponds to 4.86 years of schooling (beyond eight years of compulsory pre-high school education) reported in Table 1, which we use as counterpart in the data for the years of education chosen by families in the model. We call it (non-compulsory) schooling in the rest of the paper.[15] The annual expenditure per student and year in the U.S. is $12,690 for upper-secondary education and $29,910 for tertiary education, as documented in tables B.1.2 and B.1.6 of OECD (2011). The average cost for an additional year of schooling is thus $21,300. We assume, as in Stantcheva (2017), that 35% of expenses for human capital investment related to higher education are subsidized so that we get a target of $13,845 for the cost net of the subsidy for a student at the time of high-school graduation.

We have to convert the monetary costs observed in the data into units of the model. We make the empirically plausible assumption that the average family *without* any non-compulsory schooling does not receive, or leave, any bequests and does not spend significant amounts on education. Such a family generation then approximately consumes all resources in a hand-to-mouth fashion so that income per model period corresponds to 0.9356 in model units.[16] Expressed in dollars, this amount equals the mean annual earn-

[14]We refer to the statistic for 2005 in version 2.2 of the database available at http://www.barrolee.com/. The statistic for the population above age 15 in 2010, or for the population above age 25 in 2005 or 2010 is similar with values for the average years of schooling of 13.18, 13.13 and 13.42, respectively.

[15]Legally, compulsory schooling includes at least some high school but more than 5% of 16 to 25 year olds drop out of high school, and only about 80% of students in a cohort complete high school according to the National Center for Education Statistics.

[16]A hand-to-mouth consumer without bequests consumes net income, $c = y - T^y(y) = \delta y^{1-t_y}$ where the last step follows using the tax schedule $T^y(y) = y - \delta y^{1-t_y}$. The optimal labor supply for a hand-to-mouth consumer without bequests is $l^*(\theta, h) \equiv \arg\max \left\{ \ln(\delta [A(\theta, h) l]^{1-t_y}) - v(l) \right\}$. For $v(l) = l^\alpha / \alpha$, we obtain

Variable	*Data* (1)	*Model* (2)
Median bequests, conditional on $b > 0$	\$52,571	\$51,595
Average years of (non-compulsory) schooling S	4.86	4.76
Correlation(S', S)	0.46	0.47
Intergenerational earnings elasticity	0.45	0.44
Average net cost of an additional year of schooling	\$13,845	\$13,858

Table 2: Target statistics in the data and model predictions

ings of high-school dropouts of \$20,241 in 2010 which have a present value for a 30-year period of \$436,762.[17]

The calibrated value of ς_1 implies sufficient convexity of the cost function so that the calibrated economy and the planner problem analyzed in Sections 2.1 and 5 are concave.[18] The calibrated parameter ς_2 of the cost function is close to zero. This implies that the model matches the intergenerational correlation in the years of schooling although parental background reduces the net cost of education only very mildly: if parents have five years of non-compulsory schooling, this reduces the cost of educating their children only by 2 per mille.

Numerical solution and predictions of the simulated model.—We solve the problem by applying the endogenous gridpoint method in our model with an occasionally binding constraint as proposed in Hintermaier and Koeniger (2010). The application of the algorithm is described further in the online Appendix C of Koeniger and Prat (2018). For the simulations, we draw 500,000 observations, simulate the respective paths based on the model solution for 100 generations to obtain a stationary distribution. We provide further implementation details on the numerical solution and calibration in Appendix D.

Table 2 shows that the calibrated model matches the data targets quite closely. For earnings across generations, we can provide further comparison with empirical results for the U.S. reported in Chetty et al. (2014). The transition matrix generated by our calibrated model, reported in Table G.1 of Appendix G, is remarkably similar to the estimated transition matrix in Table 2 of Chetty et al. (2014). The association between the parent's

$l^*(\theta, h) = \left(1 - t_y\right)^{1/\alpha}$. $A(1,1) = 1$ then implies that the period income of the average worker without any non-compulsory education is $y^*(1,1) = \left(1 - t_y\right)^{1/\alpha} = 0.9356$, once we insert the parameter values documented in Table 1.

[17]See Table 232 in the *Statistical Abstract of the United States 2012* available at https://www.census.gov.

[18]In particular, $\varsigma_1 > 1 - \xi$ where $1 - \xi$ is the exponent of human capital in the production function. Concavity makes our problem tractable because otherwise we could no longer rely on the first-order approach to characterize the social optimum. We verify the validity of this approach numerically as explained in Appendix E.

and children's income rank predicted by the model, as illustrated in Figure G.1 in Appendix G, captures key features of the empirical counterpart reported in the left panel of Figure 2 in Chetty et al. (2014). The empirical results in Chetty et al. (2014) imply somewhat less persistence than the model at the bottom and top of the income distribution. Indeed, the rank-rank correlation in our benchmark calibration is 0.31 and thus similar to 0.34 as reported in Chetty et al. (2014), Table 1, if we trim the top and bottom 10% of the parent income distribution. The left panel in Figure 2 of Chetty et al. (2014) suggests that such trimming would not change the rank-rank correlation much for their sample. In our calibrated model it takes on average slightly more than five generations until the offspring of a family in the bottom decile of the income distribution reaches the mean income, in line with results in OECD (2018) for the U.S.

In an alternative calibration, we target a smaller persistence parameter at the lower end of estimates for the intergenerational earnings elasticity reported in Table 1 of Chetty et al. (2014). This calibration improves the match of the transition matrix reported in Table 2 of Chetty et al. (2014), as shown in Table G.5 in Appendix G.1, with a model-implied rank-rank correlation of earnings equal to 0.28.

Appendix G.1 contains the results of this and other robustness checks. In these alternative calibrations we target the conditional mean instead of the conditional median of bequests, we target a lower intergenerational earnings elasticity, we calibrate a higher Frisch elasticity, and we allow for a higher complementarity between human capital and ability in the aggregator for productivity.

4 Model mechanisms

We highlight the endogenous transmission mechanisms in the model through which nurture provides insurance, reduces mobility in the welfare distribution and transmits inequality across generations. We then show for different types of families whether leaving a bequest or investing into schooling is more cost-effective to provide insurance. These findings provide intuition for how more insurance may be provided efficiently in the social optimum, as analyzed in the next section.

Mechanisms through which nurture provides insurance and transmits inequality.—In the model, parents can pass on resources to their children through nurture in terms of bequests and schooling. These resources allow children to smooth consumption but also transmit to the inequality of earnings. Bequests transmit to earnings inequality by reducing labor effort through a wealth effect. Through this channel, the model generates mean reversion in labor earnings because parents with higher labor earnings leave more bequests, thus inducing less labor effort and earnings of their children. We compute the average steady-state elasticity of labor earnings with respect to bequests to gauge how much the endogenous mean reversion reduces the persistence in labor earnings across generations. We find that the average elasticity of earnings with respect to bequests is indeed negative at -0.033. Interestingly the order of magnitude of this elasticity is in line with the evidence on lottery winners by Imbens et al. (2001) for the U.S. and by Cesarini et al. (2017) for Sweden. They estimate remarkably similar marginal propensities to earn out of changes in unearned income. Cesarini et al. (2017) report that (pre-tax) earnings decrease by 1.1 percent of the change in wealth, and this effect is very persistent. The average earnings response in our calibrated model is very similar at 1.4 percent.

Another channel, through which bequests affect inequality, is that bequests relax financial constraints for human capital investment. The relatively modest value of median bequests, resulting from a highly concentrated empirical distribution, implies that in the calibrated economy the spending of nearly half of the families is financially constrained.[19] To illustrate the consequences for human capital investment, Figure 1 plots schooling and bequests for children as a function of bequests that parents received, for a representative family with median ability and high-school education. The figure shows that schooling is a very non-monotonic function of bequests because there are various effects at work. At a low level of bequests, schooling increases in bequests because the borrowing constraint is binding, illustrated by the flat portion of the policy function for the bequests left to children. More financial resources thus allow more human capital investment. Once the financial constraint is slack, the negative wealth effect on labor effort of the next generation implies that it is less attractive for parents to invest into their children's human capital: more schooling for the children only increases children's welfare in our economy if they work so that this investment generates income. Figure 1 further shows that once bequests start to be taxed, it becomes more attractive again to invest additional resources into schooling rather than to leave further bequests. This changes the slope of the plotted functions because relatively more human capital accumulation then ensures that the endogenous (risk-adjusted) return to human capital equals the after-tax return on bequests.

Compared to bequests, human capital affects the transmission of earnings inequality very differently in our calibrated economy. It makes labor earnings more persistent and

[19] In our model, the incidence of the financial constraint does not imply that the intergenerational earnings elasticity deviates much from the intergenerational correlation of ability as in the classic Becker-Tomes model discussed in Lee and Seshadri (2019). As explained in Section 3, the empirical estimates on the returns to schooling imply that the effect of ability on labor productivity is much stronger than the effect of schooling on productivity in our model calibration. In Appendix G.1 we check robustness of our results if we target the mean of bequests, conditionally on receiving one. In this calibration 6.5% of families are financially constrained. As discussed at the end of Section 5, this increases the amount of insurance in the calibrated steady state relative to the benchmark calibration.

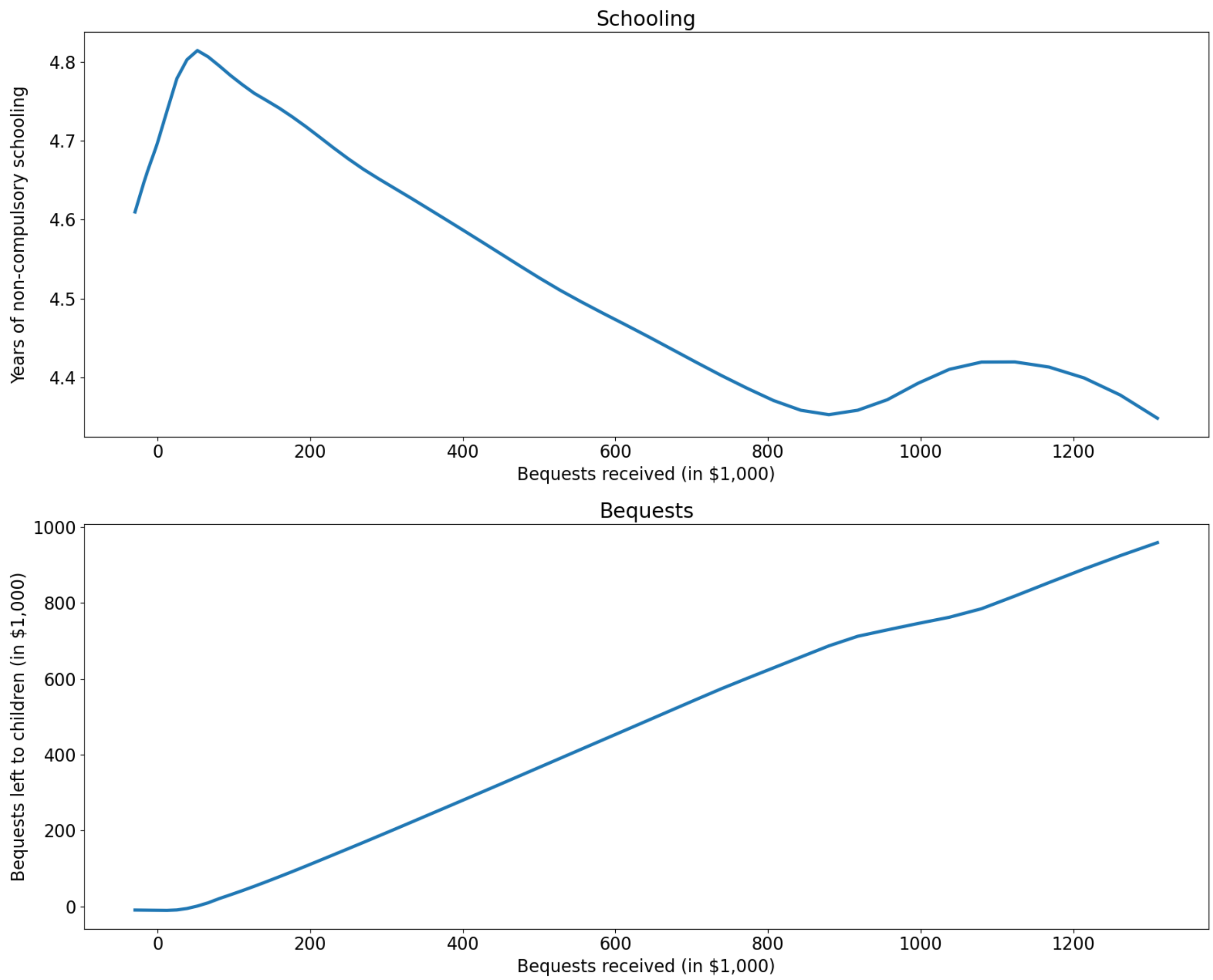

Figure 1: Schooling and bequests as a function of received bequests. *Notes:* We condition on parents with four years of (non-compulsory) schooling and median ability. Bequests are in units of $1,000.

more unequal across generations. Given that parents with higher labor earnings invest more into the human capital of their children, high earnings are transmitted to their children.[20] Indeed, we find that the average elasticity of earnings with respect to schooling investment is 0.51 in the calibrated economy, and we find that the elasticity has the highest value of 0.55 at the top of the earnings distribution where families have an ability above average.[21] The elasticity of labor earnings with respect to nature (ability) is even

[20]The income and substitution effect of an increase in labor productivity on labor effort approximately balance each other in our calibration, as suggested by the analytic results in Heathcote et al. (2014) for an economy with zero bond holdings in equilibrium.

[21]Behrman and Rosenzweig (2004) provide evidence that schooling investment indeed differs across siblings, as predicted already by classic theories of human capital investment (Becker and Tomes, 1986) if siblings have heterogeneous ability. Behrman and Rosenzweig (2004) find that bequests instead do not differ significantly across siblings and thus do not seem to compensate for differences in schooling investment as would be predicted by human capital theory in the first best. Although our model does not attempt

higher at 0.93 because of the stronger effect of ability on labor productivity implied by the empirical estimates on the returns to schooling in the calibration.

Cost-effective insurance provision through nurture.—Given these transmission channels of nurture and nature, one may ask what size of changes in nurture and nature generates the same welfare effect. This is of interest for efficient provision of intergenerational insurance and provides intuition for how more insurance is achieved in the transition to the social optimum analyzed in Section 5.

To answer this question, we report results of the following experiment in Table 3. Consider a family characterized by the initial conditions (b, h, θ). Then compute the welfare increase if that family receives additional $\$10,000$ as bequest. Table 3 displays the increase in years of schooling or the increase in ability, in units of its standard deviation, which would generate the same welfare increase. In the different rows of the table, we show the average results of this experiment for families in different quintiles of the earnings distribution. We have chosen the earnings distribution because earnings are observable but the results are very similar for quintiles of the welfare distribution, as shown in Table G.2 in Appendix G.[22]

Earnings quintile	Increase in years of schooling	Cost of additional schooling	Increase in ability (in units of standard deviation)
1	0.92	18,783	0.09
2	0.68	13,781	0.10
3	0.56	11,420	0.10
4	0.46	9,520	0.11
5	0.35	7,289	0.12

Table 3: Average increase in schooling and ability that is welfare equivalent to receiving an additional $10,000 as bequests, by earnings quintile

The results in Table 3 illustrate the cost effectiveness of nurture (b, h) in compensating the welfare consequences of changes of ability θ. The increase in schooling, reported in the first column of the table, implies direct costs for the current generation that are reported in the second column.[23] The direct costs of the additional years of schooling,

to explain the well-known equal division puzzle for bequests (Bernheim and Severinov, 2003), the policy functions for bequests and schooling investment in Figure 1 illustrate that borrowing constraints may contribute to explaining this empirical finding because such constraints reduce the negative correlation between human capital investment and bequests.

[22] The correlation between earnings and welfare is very high at 0.82 so that the values in the last column of Tables 3 and G.2 do not differ up until the second decimal place.

[23] For simplicity, we do not consider in these calculations that more human capital reduces the cost of investing into the human capital of the next generation. The reported costs can thus be considered an upper bound. Note that the cost of an additional year of schooling are approximately $\$20,000$ and thus

that are welfare-equivalent to obtaining additional $10,000 as bequest, are larger than $10,000 for families in the lower three quintiles of the earnings distribution. These families have relatively less ability, given the nearly perfect correlation between ability and earnings of 0.99, and schooling is less cost-effective to generate additional welfare for them than bequests. For families in the top of the earnings distribution instead, schooling is more cost-effective than bequests to generate additional welfare, given the complementarity of ability and human capital for labor productivity. For these families, the welfare-equivalent direct costs of the additional schooling are smaller than $10,000. Finally, the last column of Table 3 shows that the required changes of ability, which are equivalent in welfare terms to additional bequests of $10,000, are larger for higher quintiles of the earnings distribution. This indicates the decreasing returns in ability and shows the extent to which a given shock to ability has a stronger welfare impact at the bottom of the earnings distribution.

These findings complement the evidence in De Nardi and Yang (2016) who calibrate an intergenerational model with estate taxation to the U.S. and find that parental background matters most for life-time utility at the top of the distribution of parental earnings, given that the calibrated distribution is very unequal at the top. The difference of being born into a family in the lowest earnings state compared to the second-lowest earnings state is very small instead. Our results show that family background in terms of parent's schooling investment is most effective at the top of the ability distribution because of the complementarity of ability and schooling. At the bottom of the ability distribution, bequests are much more effective in providing insurance against ability risk instead.

5 Quantitative analysis of a tax reform implementing the social optimum

We compare insurance, mobility and inequality in the calibrated economy with the social optimum. We show that more intergenerational insurance and more effective incentive provision in the social optimum relative to the calibrated economy are associated with more cross-sectional inequality of labor earnings but quantitatively similar earnings mobility across generations. We elaborate on these results by decomposing the intergenerational earnings elasticity, by illustrating the role of nature and nurture in providing insurance, and by showing how bequests and schooling are incentivized in the social optimum across quintiles of the income and bequest distribution. We gauge the robustness of our

higher than the cost of an additional school year at high-school graduation targeted in the calibration. The reason is that many families invest into schooling beyond high-school graduation and that the cost of schooling is convex.

results, also by comparing the findings to those obtained for economies with simpler tax and subsidy systems than in the social optimum.

5.1 The transition to the social optimum

We analyze the transition to the social optimum, starting from the steady state of the calibrated economy. We implement this hypothetical reform by ensuring that the allocation for each dynasty has the same present discounted value of the expected net costs in the planner problem as in the calibrated economy. Although there is no redistribution across dynasties at the time of reform, the planner may redistribute towards future generations within a dynasty, given that the planner cares relatively more about welfare of future generations ($\psi > \beta$). The design of the reform ensures that the effects on insurance and mobility are not confounded by wealth effects which would shift the consumption distribution. Appendices D and F provide further details on how we implement the numerical solution and the reform.

We focus mostly on results on the transition path rather than in the steady state of the social optimum because our interest is on the changes of mobility and inequality in the first decades after the reform. We also report key results for the new steady state after the reform as further benchmark. The steady state is approximated by the period 100 generations after the reform. This is conservative because we have found that in our experiments convergence happens much faster.

Table 4 shows how the averages and standard deviations of key variables evolve after the reform. Average consumption and average labor earnings increase after the reform although average labor effort decreases slightly, illustrating the efficiency gains in the social optimum compared to the calibrated economy. Average bequests of dynasties decrease if we compare bequests in the calibrated economy to the net costs for providing a promised allocation after the reform.[24] Average schooling instead increases by approximately a year. This may seem surprising because nurture in terms of schooling and bequests are both subsidized on average in the social optimum, as we will show subsequently. It is the relaxation of the intergenerational borrowing constraint in the social optimum which explains the different evolution of bequests relative to schooling after reform.

The distributions in Figure G.2 of Appendix G show that the averages hide substantial heterogeneity. The figure plots distributions up to three generations after the reform and reveals that labor supply and human capital become much more dispersed after the reform. The planner decouples production of families from their consumption, the distribution of which changes much less. Figures G.3 and G.4 show the evolution of the

[24]We use the standard definition for the counterpart of assets in the planner problem, as explained in Appendix F.

	Averages				
	Bequests†	Schooling	Labor effort	Earnings	Consumption
Calibrated economy	24,924	4.72	1.0	32,169	27,723
Social optimum					
2nd generation after reform	-68,589	5.44	0.98	35,253	28,530
Steady state after reform	-76,961	5.44	0.99	35,816	28,537
	Standard deviations of logarithms				
	Bequests†	Schooling	Labor effort	Earnings	Consumption
Calibrated economy	2.6	0.07	0.06	0.44	0.38
Social optimum					
2nd generation after reform	1.2	0.06	0.10	0.51	0.36
Steady state of reform	2.7	0.07	0.18	0.53	0.60

Table 4: Summary statistics of key variables in the calibrated and reformed economy

Notes: † For bequests we report the coefficient of variation rather than the standard deviation of the logarithm because bequests can take negative values. Bequests after the reform correspond to the present value of net costs for an allocation, as explained further in Appendix F.
The units of bequests, earnings and consumption in the top panel are US dollar in 2010. Earnings and consumption are annualized. Labor effort is normalized by the average level in the calibrated steady state. The unconditional mean for bequests in the calibrated economy is smaller than the median conditional on receiving bequests targeted in the calibration.

distributions for families, conditioning on the top and bottom quartile of the ability distribution in each generation. The different changes of the distribution of labor supply across ability types in the first generation after the reform ($t = 0$), when the reform is implemented and assets and human capital are still given by pre-reform decisions, shows that social optimality requires an increase of labor supply of high ability types and more investment into the human capital of their children (visible in the distribution of non-compulsory school years plotted for the second generation after the reform ($t = 1$)). Labor supply of low ability types is reduced instead so that low-ability families obtain more of their welfare through enjoying more leisure. Moreover, some dynasties with currently high ability previously had low ability and vice versa, which affects the expected present value of net costs for providing a promised allocation (which we call bequests after the reform). Thus, labor supply and human capital expenditures become more dispersed within the top quartile of the ability distribution, and the dispersion of labor supply increases also within the bottom quartile.

The bottom panel of Table 4 shows that the transition after the reform is associated with an increase in the cross-sectional inequality of earnings, labor effort, and consumption, as measured by the respective standard deviation. We now provide further evidence on the larger dispersion within ability types on the transition path to the social optimum by analyzing the effects of the reform on mobility.

	(1) Pass-through coefficient	(2) Rank-rank correlation of ability and welfare	(3) Correlation of ability and labor effort	(4) Intergen. correlation of schooling	(5) Intergen. earnings elasticity (IGE)	(6) Intergen. rank-rank correlation of welfare
Calibrated economy	0.67	0.90	0.03	0.47	0.44	0.67
Social optimum						
2^{nd} generation after reform	0.58	0.36	0.75	0.30	0.41	0.80
Steady state after reform	0.63	0.22	0.37	0.41	0.43	0.92

Table 5: Insurance and Mobility

Notes: The pass-through coefficient captures the effect of unexpected changes in ability on consumption, obtained from a linear regression of log consumption on the ability shock ϵ.

5.2 Insurance and mobility

Table 5 shows how insurance and mobility evolve after the reform, starting from the steady state of the calibrated economy. The table contains statistics on insurance and mobility that have been prominent in the empirical literature: the pass-through of a productivity or ability shock to consumption in column (1), the intergenerational correlation of schooling in column (4) and the intergenerational earnings elasticity in column (5). We also add two statistics in columns (2) and (3) of Table 5 which show how ability correlates with the welfare of dynasties and with labor effort. The extent to which the rank in the welfare distribution depends on the ability draw indicates the degree of insurance in our model with consumption, labor supply and state variables which capture the role of nurture and nature. Finally, column (6) shows the rank-rank correlation of welfare across generations, which illustrates the persistence of dynasties' position in the welfare distribution.

Regarding insurance, column (1) in Table 5 shows that the pass-through coefficient, obtained from a linear regression of log consumption on the ability shock ϵ, decreases from 0.67 in the calibrated economy to 0.58 in the second generation after the reform and 0.63 in the steady state, illustrating the increase of consumption insurance.[25] That is, in the steady state after the reform 37 percent of ability shocks are insured compared to 33 percent in the calibrated economy.[26] Similarly, the rank-rank correlation between ability

[25] Given the process for ability specified in Table 1 and the assumption of perfect competition, equation (4) shows how the ability shock translates into unexpected wage changes. The differences between the statistics reported in Table 5 are statistically significant because we simulate the economy for a sample of $500,000$ families.

[26] Although the size of the pass-through coefficient in the calibrated intergenerational economy is not comparable directly to estimates obtained in a life-cycle context, a common finding is that U.S. households are partially insured against persistent income shocks. The size of the pass-through coefficient in our intergenerational model is of similar size as the estimated pass-through coefficient for permanent shocks in Blundell et al. (2008) and higher than the coefficient of 0.4 in Heathcote et al. (2014) who analyze partial insurance in a model with consumption, endogenous labor supply and preferences of the same class as in our model. The substantial pass-through of intergenerational ability risk to consumption in our calibrated

and welfare decreases in column (2) because of the stronger influence of nurture on the position in the welfare distribution in the social optimum than in the calibrated economy.

Regarding mobility, column (5) shows that the mobility of labor earnings across generations changes very little although the intergenerational correlation of schooling in column (4) decreases.[27] Instead, the intergenerational rank-rank correlation of welfare in column (5) increases, illustrating that the planner achieves more insurance while maintaining incentives for able families to exert effort as shown in column (3). The correlation between ability and labor effort increases after the reform and the incentives to work for able families are stronger two generations after the reform than in the steady state.

Decomposing the intergenerational earnings elasticity

To better understand why the intergenerational earnings elasticity does not change much after the reform, we provide a decomposition. The intergenerational earnings elasticity (IGE), as obtained from a linear regression of $\ln y'$ on $\ln y$, is given by

$$\iota \equiv \frac{\mathrm{Cov}(\ln y', \ln y)}{\mathrm{Var}(\ln y)} \,. \tag{5}$$

Earnings $y(h,\theta,l) = h^{1-\xi}\theta^{\xi} l$ in the benchmark calibration and thus are a function of schooling, ability, and labor effort. We decompose the numerator and denominator of the IGE into its subcomponents, using that

$$\ln y = (1-\xi)S + \xi \ln \theta + \ln l \,, \tag{6}$$

and the definition of schooling $S \equiv \ln h$. Thus,

$$\mathrm{Cov}(\ln y', \ln y) = \underbrace{(1-\xi)^2 \mathrm{Cov}(S', S)}_{\equiv C_1} + \underbrace{\xi^2 \mathrm{Cov}(\ln \theta', \ln \theta)}_{\equiv C_2} + \underbrace{\mathrm{Cov}(\ln l', \ln l)}_{\equiv C_3} + \mathcal{C} \,, \tag{7}$$

where we collect covariances across the production inputs across generations in $\mathcal{C}$ with

model results from the modest amount of median bequests and the returns to schooling observed in the data, given the small response of labor supply to productivity shocks. In the robustness check discussed further below, in which we target mean bequests in the calibration at the cost of predicting too high bequests for large parts of the population, the pass-through coefficient falls to 0.4 in our intergenerational model.

[27] We also find quantitatively small changes of earnings mobility between 0.42 and 0.38 if we use the intergenerational rank-rank correlation of earnings as an alternative measure. Note that the change of earnings mobility on the transition after the reform is non-monotonic. As emphasized by Nybom and Stuhler (2013), such non-monotonicity of earnings mobility is quite common after structural changes such as the tax reform in our analysis. Figure G.5 in Appendix G shows the transition path of intergenerational income mobility for different mobility measures and provides further discussion of the non-monotonicity.

$$
\begin{aligned}
\mathcal{C} \equiv\ & \underbrace{(1-\xi)\xi\mathrm{Cov}(S',\ln\theta)}_{\equiv\mathcal{C}_1}+\underbrace{(1-\xi)\mathrm{Cov}(S',\ln l)}_{\equiv\mathcal{C}_2}+\underbrace{\xi\mathrm{Cov}(\ln\theta',\ln l)}_{\equiv\mathcal{C}_3}\\
& \underbrace{(1-\xi)\xi\mathrm{Cov}(S,\ln\theta')}_{\equiv\mathcal{C}_4}+\underbrace{(1-\xi)\mathrm{Cov}(S,\ln l')}_{\equiv\mathcal{C}_5}+\underbrace{\xi\mathrm{Cov}(\ln\theta,\ln l')}_{\equiv\mathcal{C}_6}\ .
\end{aligned}
$$

Analogously,

$$
\mathrm{Var}(\ln y)=\underbrace{(1-\xi)^2\mathrm{Var}(S)}_{\equiv V_1}+\underbrace{\xi^2\mathrm{Var}(\ln\theta)}_{\equiv V_2}+\underbrace{\mathrm{Var}(\ln l)}_{\equiv V_3}+\mathcal{V}\ , \tag{8}
$$

where we collect covariances across the production inputs within generations in $\mathcal{V}$ with

$$
\mathcal{V}\equiv\underbrace{2(1-\xi)\xi\mathrm{Cov}(S,\ln\theta)}_{\equiv\mathcal{V}_1}+\underbrace{2(1-\xi)\mathrm{Cov}(S,\ln l)}_{\equiv\mathcal{V}_2}+\underbrace{2\xi\mathrm{Cov}(\ln\theta,\ln l)}_{\equiv\mathcal{V}_3}\ .
$$

Table 6 displays the results of the decomposition. The decomposition reveals that both the variance and the covariance of log earnings increase in (the transition to) the social optimum, implying only a mild change of the IGE. Interestingly, Table 6 shows that the composition of the variance and the covariance changes substantially. In particular, the decomposition provides insights to which extent the changes in the correlation between ability and labor effort and the intergenerational correlation of schooling, documented in columns (3) and (4) of Table 5, affect the IGE quantitatively.

Table 6 shows that the share of the total variance attributed to the variance of ability of the current generation V_2 and the share of the covariance attributed to the intergenerational covariance of ability C_2 both decrease in the social optimum compared with the calibrated economy. Instead, the shares attributed to the covariance of labor effort and ability within and across generations, i.e. $\mathcal{V}_3$, $\mathcal{C}_3$ and $\mathcal{C}_6$, increase. The share attributed to the variance of labor effort V_3 and the intergenerational covariance of labor effort C_3 also increase, particularly in the steady state of the social optimum. The decomposition further shows that the other changes in the (co-)variance of schooling and the covariances of schooling with labor effort and ability, within and across generations, contribute little quantitatively to explaining the IGE in the calibrated economy as well as in the social optimum.

The decomposition thus shows that the stronger contribution of the (co-)variance of labor effort and the covariance between ability and labor effort in the social optimum

	Calibrated economy	Social optimum (second generation)	Social optimum (steady state)
$\text{Cov}(\ln y', \ln y)$	0.08	0.11	0.12
shares (%):			
C_1	0.6	0.3	0.4
C_2	92.7	67.5	62.5
C_3	2.7	1.7	17.5
$\mathcal{C}$	3.9	30.5	19.6
$\mathcal{C}_1$	15.8	11.6	10.3
$\mathcal{C}_2$	-0.1	2.9	4.0
$\mathcal{C}_3$	0.0	15.1	10.9
$\mathcal{C}_4$	3.0	2.2	2.0
$\mathcal{C}_5$	-1.0	-0.4	0.4
$\mathcal{C}_6$	-13.9	-0.8	-7.9
$\text{Var}(\ln y)$	0.19	0.28	0.28
shares (%):			
V_1	0.5	0.4	0.4
V_2	92.2	62.7	62.2
V_3	1.9	3.9	10.8
$\mathcal{V}$	5.4	33.1	26.5
$\mathcal{V}_1$	6.0	4.1	3.9
$\mathcal{V}_2$	-0.7	0.9	1.3
$\mathcal{V}_3$	0.1	28.1	21.4

Table 6: Decomposition of the Intergenerational Earnings Elasticity

Notes: This table decomposes the IGE into $(C_1+C_2+C_3+\mathcal{C})/(V_1+V_2+V_3+\mathcal{V})$, where $\mathcal{C} = \mathcal{C}_1+\cdots+\mathcal{C}_6$ and $\mathcal{V} = \mathcal{V}_1+\mathcal{V}_2+\mathcal{V}_3$ as defined in the main text. All shares are percentages of the total covariance or variance, respectively. Rounding error may imply that the sum of the shares does not equal 100.

relative to the calibrated economy crowds out some of the contribution of the variance and intergenerational covariance of ability in the social optimum.[28] This illustrates that work incentives in the social optimum are quite different relative to the calibrated economy although the change in the IGE after the reform is quantitatively small.

These results are interesting for at least two reasons. First, they show that changes in the IGE, or earnings mobility more generally, have to be interpreted with care. In particular, incentives, insurance, efficiency and ultimately welfare may differ substantially in economies with a quantitatively similar IGE. A decomposition, along the lines presented above, helps to uncover the economic mechanisms that shape earnings mobility. Empirical evidence on the mechanisms in Bolt et al. (2021), based on U.K. data, reveals an important role of cognitive ability at age 16 for explaining life-time earnings, consistent with the quantitatively important role of ability for the IGE in the calibrated economy, reported in column (1) of Table 6.

Second, the results illustrate the importance of endogenous labor supply in our analysis. In much of the literature on the IGE instead, as for example in the classic contributions by Becker and Tomes (1986) and Mulligan (1999), generations inelastically supply labor. Given that labor supply is more correlated with ability in the social optimum and that utility of dynasties depends on consumption and labor supply, current consumption or promises for future consumption have to compensate high-ability families relatively more to provide insurance against ability shocks. Indeed, we find that consumption and labor effort are less negatively correlated in the social optimum than in the calibrated economy illustrating that the planner incentivizes labor effort differently.

Finally, let us put the dominant contribution of the (co-)variance of ability for the IGE in the calibrated economy into a broader context. Consistent with this result, we find that the effect of nurture through schooling and bequests on the intergenerational transmission of earnings inequality is modest in the calibrated model. The variation in bequests and schooling given to a generation explains $0.9-1.3$ percent of the cross-sectional variance of that generation's earnings, depending on whether the covariance is split proportionately or equally across the determinants. The bequests and schooling *received by parents* explain at most 5.3 percent of the part of the variation of children's earnings that can be attributed to parents' nature and nurture. Given that ability at labor market entry in our model may also contain some nurture component, the model attributes a large role to ability and the quantitative results have to be interpreted as an upper bound. By

[28] $\mathcal{C}_3 \equiv \xi \text{Cov}(\ln\theta', \ln l)$ increases in the social optimum (also) because (i) the contemporaneous correlation between θ and l increases and (ii) the autocorrelation between θ' and θ is positive. The difference of the sign of the covariances $\mathcal{C}_6 \equiv \xi \text{Cov}(\ln\theta, \ln l')$ and $\mathcal{C}_3 \equiv \xi \text{Cov}(\ln\theta', \ln l)$ in the social optimum obtains because $\ln l'$ also depends on the endogenous state variables b' and h', which in turn depend on θ, whereas the probability distribution of the draw of the exogenous state variable θ' only depends on θ.

and large, they seem compatible with the important role of nature in the transmission of earnings emphasized in recent empirical research by Bingley et al. (2018) based on a credible "Children of Twins" design, and with the finding of Gallipoli et al. (2020), Table 4, that inequality among parents explains only 8% of the inequality of earnings among their children.

Although the effect of nurture on (the inequality of) earnings is quantitatively small in the calibrated economy, nurture may insure generations against ability risk by decoupling production from consumption. We now elaborate on this point.

5.3 The role of nurture for insurance provision

We investigate in more detail the role of nurture for providing insurance in the calibrated economy and on the transition path to the social optimum. We first quantify how important nature, in terms of ability, is compared to nurture, in terms of bequests and schooling, for the rank in the welfare distribution. We then inspect in more detail how nurture through bequests and schooling insures dynasties against ability shocks.

The role of nature and nurture for the position in the welfare distribution.—We regress the rank in the welfare distribution on the ranks in the distributions for bequests, human capital, and ability. Table 7 displays the results for the calibrated economy in column (1) and for the economy two generations after the reform in column (2). The linear specification explains most of the variation in welfare ranks, e.g., 92% in the calibrated economy according to the R^2 statistic. The regression coefficients in Table 7 show how moving up one decile in the distribution of bequests, human capital, or ability, respectively, changes the rank in the welfare distribution. For example, the coefficient of 0.77 for θ in column (1) implies that if ability were one decile higher in the ability distribution, then the family would move up 0.77 deciles in the welfare distribution.

Table 7 indicates that, in the calibrated economy, nature θ plays a more important role for a family's place in the welfare distribution than nurture b or S. The results in column (1) show that the rank in the ability distribution is approximately 2.5 to 5 times as important as bequests or schooling, respectively. In the second generation after the reform ($t = 1$), more intergenerational insurance against ability shocks increases the importance of nurture relative to nature dramatically: as shown in column (2), nurture is two orders of magnitude more important than nature for the position in the welfare distribution. The results in column (3) show that the same is true in the new steady state after the reform where, among the two variables capturing the effect of nurture, bequests (the expected present value of net costs for the promised allocation in the terminology of the planner problem) become more important than schooling for the position in the welfare distribu-

Rank in distribution of...	Welfare rank		
	(1) Calibrated economy	(2) 2nd generation after reform	(3) steady state after reform
b	0.31	0.62	1.02
s	0.15	0.59	0.32
θ	0.77	0.02	0.01
N	500000	500000	500000
R^2	0.92	0.90	0.98

Table 7: Welfare rank regressions

Notes: The estimation results are obtained with an OLS-regression of welfare ranks on an intercept and the ranks in the distributions of bequests (b), schooling (S), and ability (θ). Column (1) uses simulated data of the calibrated economy. Columns (2) and (3) use simulated data of the socially optimal economy in the second generation after the reform and in the steady state, respectively.

tion. As the history of ability draws becomes longer after the reform, the expected present value of net costs for the promised allocation becomes the dominant determinant of the rank in the welfare distribution. The role of ability for welfare beyond these promised allocation costs becomes small.

Note that the promised costs imply less assets on average after the reform, as previously shown in Table 4 and Figures G.2 to G.4. As the borrowing constraint is absent in the social optimum compared with the constrained economy, labor effort is more strongly correlated with ability in the social optimum as shown in column (3) of Table 5.

The effect of nature and nurture on welfare.—We provide further details on how nurture shields dynasties against shocks to nature in terms of ability in the calibrated economy and in the social optimum. Table 8 displays probability matrices, for which each cell shows the probability that a family in a quintile of the ability distribution is in a specific quintile of the welfare distribution. The top panel reports results for the calibrated economy and the bottom panel for the social optimum. The welfare measure is dynastic and includes the discounted welfare of future generations.

If ability θ fully determined the position in the welfare distribution, the matrix would be an identity matrix. In this case, nurture in terms of received bequests b and obtained human-capital investment h would be irrelevant for the position in the welfare distribution. The more weight is on the off-diagonal elements, the more nurture dampens the effect of nature on the position in the welfare distribution and thus insures generations from an intergenerational perspective.[29]

For example, the cell in the first row and fifth column of the matrix in the top panel shows that a family in the first quintile of the ability distribution has a one percent probability of being in the top quintile of the welfare distribution in the calibrated economy. These families have received $\$268,130$ as bequests and have obtained 4.8 years of non-compulsory schooling on average. This shows that nurture can compensate for bad draws of nature. Bequests are more effective in compensating for low ability than human capital investments because ability and schooling are complements in making labor effort more productive. Thus, the differences in average bequests across columns are relatively larger than the differences in average additional school years, in particular for low-ability quintiles.

The matrix shows that there is much less than full insurance against ability risk in the calibrated economy. For example, the cell in the first row and first column of the matrix shows that 85% of families currently in the lowest ability quintile are also in the lowest quintile of the welfare distribution.

The bottom panel of Table 8 shows that the probability matrix after the reform has less weight on the diagonal, implying more insurance and a stronger dependence of welfare on nurture rather than nature compared with the calibrated economy. The correlation between the rank in the ability distribution and the rank in the welfare distribution is indeed much smaller at 0.36, compared with 0.90 in the calibrated economy, as shown in column (2) of Table 5. More insurance across generations and less mobility in the welfare distribution after an ability shock are two sides of the same coin.

More intergenerational insurance on the transition towards the social optimum is achieved with a larger dispersion of wealth and human capital across families with different abilities,[30] and the correlation between wealth (or promises) and schooling falls after the reform: in the second generation after the reform ($t = 1$), $cor(b, S) = 0.12$ compared with 0.43 in the calibrated economy. The stronger dependence of welfare on nurture increases the persistence of the rank in the welfare distribution across generations. The rank-rank correlation of welfare increases from 0.67 in the calibrated economy to 0.80 in the second generation after the reform and to 0.92 in the steady state, as reported in column (6) of Table 5.

[29]Because labor supply is approximately uncorrelated with ability in the calibrated model, changes in welfare then result mostly from changes in consumption.

[30]In the social optimum, families do not face borrowing constraints as in the calibrated economy so that wealth (or promises) can take more negative values.

Ability / Welfare	*Calibrated economy* Quintiles 1	2	3	4	5
Quintiles					
1	0.85	0.09	0.04	0.02	0.01
	(-6.41,4.5)	(38.47,4.9)	(92.59,4.9)	(156.41,4.8)	(268.13,4.8)
2	0.15	0.63	0.12	0.06	0.03
	(-8.58,4.4)	(-5.04,4.6)	(47.07,4.9)	(115.18,4.9)	(232.3,4.9)
3	0.0	0.28	0.5	0.15	0.07
	(.)	(-8.77,4.5)	(0.43,4.7)	(71.64,5.0)	(195.53,5.0)
4	0.0	0.0	0.34	0.48	0.18
	(.)	(.)	(-8.25,4.6)	(13.52,4.8)	(147.76,5.0)
5	0.0	0.0	0.0	0.29	0.71
	(.)	(.)	(.)	(-5.4,4.7)	(72.54,5.0)
Ability / Welfare	*Social optimum (2nd generation after reform)* Quintiles 1	2	3	4	5
Quintiles					
1	0.37	0.24	0.17	0.13	0.08
	(-277.1,5.0)	(-259.76,5.3)	(-211.23,5.4)	(-93.8,5.5)	(182.45,5.5)
2	0.25	0.23	0.2	0.18	0.14
	(-283.93,5.1)	(-270.42,5.3)	(-232.67,5.5)	(-131.05,5.5)	(131.67,5.6)
3	0.19	0.21	0.21	0.21	0.18
	(-287.04,5.1)	(-275.28,5.3)	(-240.37,5.5)	(-148.04,5.6)	(113.4,5.7)
4	0.13	0.19	0.21	0.23	0.24
	(-290.13,5.1)	(-280.31,5.3)	(-247.34,5.5)	(-162.17,5.6)	(97.47,5.7)
5	0.06	0.13	0.19	0.25	0.36
	(-295.06,5.1)	(-286.0,5.3)	(-256.77,5.5)	(-178.45,5.7)	(94.42,5.8)

Table 8: Effect of nature and nurture on welfare

Notes: Each cell contains the probability of a family in an ability quintile to be in a specific quintile of the families' welfare distribution. In brackets for each cell, we report the average values of the state variables other than ability. Bequests are in units of $1,000 and school years are non-compulsory. Families do not face borrowing constraints in the social optimum compared to the calibrated economy so that bequests (or promises) can take more negative values. The probabilities across columns in each row may not add up to 1 because of rounding.

Progressivity of bequest and schooling subsidies

How is intergenerational insurance achieved in the social optimum? To answer this question, we inspect the wedges between the laissez faire and the socially optimal allocation based on comparison of the respective first-order conditions. Non-zero wedges imply that choices in the laissez faire need to be modified by taxes or subsidies to implement the social optimum. Equations (C.5)-(C.10) in Appendix C show explicitly how the *average* wedges can be mapped into linear taxes or subsidies that have a straightforward interpretation, starting from the auxiliary problem (C.4).[31]

We find that insurance is provided by taxing labor income and by subsidizing bequests and schooling on average, consistent with the discussion in Section 2. In the second generation after the reform, the average implied labor tax rate is 0.36 whereas the average subsidy rate is 0.46 for schooling and 0.25 for bequests. As mentioned in subsection 2.2, the optimal taxes and subsidies are jointly determined. For example, the schooling subsidies offset distortions of the schooling decision resulting from the taxation of labor income and from distortions of the accumulation motive and incentives.[32] The strong increase of the schooling subsidies after the reform, at a labor tax rate that is on average only seven percentage points higher than the average marginal tax rate in the calibrated economy, suggests that schooling decisions are distorted in the calibrated economy and that this distortion is alleviated by the schooling subsidies in the social optimum.[33] The average rates for taxes and subsidies hide interesting heterogeneity across the earnings and bequest distribution, which we illustrate in Table 9.

Table 9 shows that bequest subsidies are phased out much more progressively across the earnings distribution than schooling subsidies. This is visible in the top panel of Table 9 which shows the average tax rates across earnings quintiles. The difference in the progressivity of the subsidies across quintiles in the earnings distribution is intuitive. The position in the earnings distribution is highly correlated with ability. This implies a higher expected ability of the offspring in the higher quintiles of the earnings distribution because of the persistence of ability draws, which makes schooling subsidies more cost effective, as we have illustrated in Table 3 in Section 4. Bequests instead are relatively more cost effective for providing insurance to families with currently low ability.

As shown in the bottom panel of Table 9, bequest and schooling subsidies are quite

[31]Of course, these taxes based on the average wedges would not implement the social optimum exactly, as we discuss further in the subsection 5.5. We refer to them here to summarize the rich information contained in the wedges.

[32]See equation (C.9) in Appendix C for an explicit derivation of the schooling subsidy and further discussion.

[33]Note that schooling subsidies in the calibrated economy are included in the net cost for education. Thus, the schooling subsidies after the reform correspond to the difference between subsidies in the social optimum and the calibrated economy.

	Earnings quintile				
	1	2	3	4	5
Bequest subsidy	0.30	0.27	0.25	0.23	0.20
Schooling subsidy	0.48	0.46	0.45	0.46	0.46
Labor income tax	0.39	0.38	0.37	0.35	0.30
	Bequest quintile				
	1	2	3	4	5
Bequest subsidy	0.24	0.28	0.27	0.23	0.24
Schooling subsidy	0.47	0.47	0.47	0.46	0.46
Labor income tax	0.37	0.38	0.37	0.36	0.32

Table 9: Average rates for subsidies or taxes for bequests, schooling, and labor per quintile

Note: Tax and subsidy rates are reported for the social optimum in the second generation after the reform.

stable across the bequest distribution. The intuition is that the current wealth level due to bequests is much less correlated with current ability than current earnings. The correlation coefficient $cor(y,\theta) = 0.99$ is much higher than $cor(b,\theta) = 0.16$. Thus, the targeting of bequest subsidies towards families with currently low ability is less visible across quintiles of the bequest distribution.

Farhi and Werning (2010) noted that the progressivity of bequest subsidies in their analysis in general also depends on the redistribution achieved by other taxes and subsidies. Our results illustrate that in a model with a tax and subsidy system that jointly considers subsidies to bequests and schooling as well as labor taxes, the progressivity of subsidies is quantitatively most pronounced across the earnings distribution, and for bequests rather than schooling. Table 9 also shows the regressivity of labor taxes, confirming results in the literature on one-dimensional optimal income taxation problems, with a log-normal distribution for earnings as in our paper (Mankiw et al., 2009).[34]

5.4 Robustness

Table G.3 in Appendix G.1 shows that the results on insurance and mobility reported in Table 5 are quantitatively robust if we target a lower intergenerational earnings elasticity of 0.3 instead of 0.45, and if we recalibrate the economy with a higher complementarity between ability and schooling than in the benchmark. When we calibrate the model for

[34] Heathcote and Tsujiyama (2021) find that the progressivity of labor taxes depends on the fiscal pressure, in terms of the size of the revenues which have to be financed by taxes. Table G.6 in Appendix G.1 shows that the level of the average rates of taxes or subsidies in our model increases if we recalibrate the economy allowing for stronger complementarity between ability and schooling. The patterns of progressivity and regressivity described for the benchmark are robust for such recalibration.

a higher Frisch elasticity of 0.86 instead of 0.5, the pass-through of the ability shock to consumption first falls after the reform and then increases to a higher value in the steady state of the social optimum. This illustrates that, in the long run, a higher Frisch elasticity of labor supply may make it more costly to provide insurance while maintaining incentives.

In Table G.3 in Appendix G.1 we do not report the results on income mobility because it does not change much after the reform for all cases but for the case in which we target the mean instead of the median of bequests in the calibration. In this case, the larger amount of bequests in the calibrated economy implies that there is an extended time period after the reform in which these bequests are run down at the same time as human capital increases. This has a positive effect on labor supply and is associated with an increase of the persistence of labor earnings from 0.44 to 0.64. As shown in column (2) of Table G.3 in Appendix G.1, a higher target level of bequests dampens the effect of ability on the position in the welfare distribution in the calibrated economy because families are better insured against ability shocks relative to the benchmark case. Qualitatively as in the benchmark case, ability becomes less important for the position in the welfare distribution after the reform. As in the benchmark, the pass-through coefficient increases in the transition to the steady state after an initial decrease right after the reform.

Concerning the properties of the welfare function, intergenerational insurance would be valued even more in the social optimum compared to the benchmark if the wedge were larger between the rate at which the planner and the dynasty discount the welfare of future generations. Similarly, a larger weight in the social welfare function on families with low ability would increase subsidies to insure these families against ability risk. As shown in sections IV.A and IV.B of Farhi and Werning (2010), such social welfare functions imply that bequest subsidies increase for low productivity parents and that these subsidies become more progressive.

Given that we have constructed the reform without redistribution of resources across dynasties from an ex-ante perspective, our results on insurance and mobility show to which extent, at the time of the reform, dynasties are willing to increase the influence of nurture relative to nature to obtain more intergenerational insurance. As is common in settings with risk sharing and insurance, generations with high ability may be better off ex post with less insurance but are bound by the commitment in our model environment. Without such commitment less insurance could be achieved if generations with high ability would have to be made indifferent to an outside option, which they would obtain if they reneged on the risk sharing arrangement.[35]

[35] Lack of the commitment of the government to stick to the implemented tax schedules also would impose a further constraint to achieve ex-ante credibility of the implemented tax schedule, in the sense that deviations have to be made sufficiently unattractive ex post. Farhi et al. (2012) show that such a constraint

It is worth noting that relative to the transition to a social optimum with immiseration, there is more insurance and a smaller increase in inequality to maintain incentives. As discussed, for example in Kocherlakota (2010), pp. 158-159, the inequality-increasing incentive effect is kept in check by the motive to provide opportunities for later generations if, as in our model, the planner discounts the future less than families and not at the same rate as in models that imply a social optimum with immiseration.

In the quantitative analysis of the tax reform, we assume that the status quo in the U.S. is best described as a stationary environment (the steady state of the calibrated economy) rather than by a transition path. This may matter for the results of our quantitative analysis. For example, depending on whether reforms in the past have increased or decreased earnings mobility in the U.S., the value of the IGE which we have used as target in the calibration may be above or below the steady-state value as pointed out by Nybom and Stuhler (2013). Indeed, the IGE on the transition to the socially optimal steady state is non-monotonic also in our analysis of the tax reform (see column (5) of Table 5). Although we cannot say much about how modeling the calibrated economy as being on a transition path would change our quantitative findings, the robustness analysis reported in Appendix G.1 for alternative values of the data targets and parameter values suggests that the result of more intergenerational insurance in the social optimum is quite robust to alternative assumptions.

5.5 Simpler tax and subsidy systems

Because the implementation of the social optimum requires a complex tax and subsidy system that conditions on the history of ability shocks, we provide results on insurance, mobility and welfare in economies with simpler tax and subsidy systems. The simpler tax systems achieve only part of the welfare gain attained in the social optimum but have the advantage that they are easier to implement because they only require information on the respective current generation, i.e., not the history of previous generations beyond the information contained in the state variables of the current generation.

We compare economies with simple tax and subsidy systems to the social optimum using two complementary approaches. We solve for the optimal linear taxes and subsidies that do not vary across generations after the reform, based on the problem (C.4) specified in Appendix C and modified to include linear, constant rates for taxes and subsidies. As an alternative, we approximate linear schedules that may vary across generations analo-

induces progressivity of the marginal tax on capital and can make the level of the capital tax positive. As shown by Findeisen and Sachs (2018) this result may not extend to human capital. Pavoni and Yazici (2017) show that children who are less patient than their parents provide a further rationale for positive taxes on intergenerational transfers. Moser and Olea de Souza e Silva (2019) analyze the implication of present bias for optimal saving policies in a life-cycle context.

gous to the approximation of Farhi and Werning (2013) or Stantcheva (2017) in a life-cycle context. We thus set the tax or subsidy rates for income, bequests and schooling in each generation to their cross-sectional weighted averages in the second best, as detailed in Appendix C.[36] Analogous to the reform that implemented the social optimum, each dynasty in state $s = (b, h, \theta)$ has to contribute the same amount of net-taxes as in the calibrated economy, at the time at which the tax system is introduced. This ensures that there is no confounding wealth effect and all tax systems we present thus include a lump-sum component.

	Bequest	Schooling	Labor Income
Optimal linear taxes / subsidies, constant across generations	-0.28	-0.31	0.18
Approximated linear taxes / subsidies, varying across generations:			
2nd generation after reform	-0.25	-0.46	0.36
Steady state after reform	-0.32	-0.49	0.40

Table 10: Simple linear taxes and subsidies

Notes: The taxes and subsidies varying across generations are cross-sectional averages derived in Appendix C. A positive value implies a tax, a negative value implies a subsidy.

Table 10 displays the resulting linear taxes or subsidies. Bequests and schooling are subsidized while labor income is taxed. The qualitative features of this tax and subsidy system are intuitive, as discussed in Section 2. Table 10 shows that bequests and schooling are subsidized more compared with the calibrated economy, in which bequests are taxed only above the large exemption of $756,000 and schooling subsidies are included in the net cost for education. Thus, the level of the subsidy rates reported in Table 10 corresponds to the quantitative difference of the rates to the calibrated economy. A robust finding in Table 10 is that the optimal subsidy rate for schooling is higher than for bequests.

Concerning labor income taxation, the optimal linear tax rate reported in the first row of Table 10 is eleven percentage points smaller than the average marginal income tax in

[36]Note that solving for the optimal history-independent taxes and subsidies is numerically feasible if we restrict the tax schedules to be linear and constant in the post-reform period. We then perform a global search for the three optimal tax or subsidy rates and locally apply the Nelder-Meade optimization algorithm. A higher dimensionality of the parameter space characterizing the tax and subsidy system quickly makes this procedure prohibitively costly in terms of computing time, e.g. if one attempts to solve for an optimal non-linear tax and subsidy system or if one allows for different taxes and subsidies across generations on the transition to the social optimum.

	(1) Pass-through coefficient	(2) Rank-rank correlation of ability and welfare	(3) Correlation of ability and labor effort	(4) Intergen. correlation of schooling	(5) Intergen. earnings elasticity (IGE)	(6) Intergen. rank-rank correlation of welfare
Calibrated economy	0.67	0.90	0.03	0.47	0.44	0.67
2nd generation after reform:						
Optimal linear taxes / subsidies, constant across generations	0.70	0.82	0.18	0.13	0.33	0.86
Approx. linear taxes / subsidies, varying across generations	0.57	0.79	0.38	0.12	0.33	0.89
Social optimum	0.58	0.36	0.75	0.30	0.41	0.80

Table 11: Insurance and mobility in economies with simple tax systems relative to the calibrated economy and the social optimum

Notes: The pass-through coefficient captures the effect of unexpected changes in ability on consumption, obtained from a linear regression of log consumption on the ability shock ϵ.

the calibrated economy. The bottom part of Table 10 shows that, for the approximated tax and subsidy rates which vary across generations but are not optimized, the absolute size of the taxes or subsidies increases slightly on the transition to the new steady state to provide insurance through nurture in terms of bequests and human capital at the same time as the inequality in labor earnings increases. Comparing the results in the bottom part with those in the top part of Table 10 reveals that the level of the approximated rates differs substantially from the optimal but constant rates. We now investigate the implications of these differences for insurance, mobility and welfare, compared with the social optimum.

Table 11 shows the results for insurance and mobility in the second generation after the reform in the economies with the simpler tax systems. The results for the calibrated economy in the top row and for the social optimum in the bottom row, for the second generation after the reform, are repeated from Table 5 for convenience. Concerning the mobility of earnings and schooling, columns (4) and (5) show that earnings and schooling are less persistent across generations in the economies with the simpler tax systems relative to the social optimum *and* the calibrated economy.[37] At the same time, column (2) shows that the welfare of families in the social optimum is shielded more from the ability shocks than in the economies with the simpler tax systems. The rank in the welfare distribution depends less on the rank in the ability distribution in the social optimum

[37] On the one hand, the mobility of schooling depends on financial constraints for schooling choices in the calibrated economy. In the other economies there is no exogenous borrowing limit instead. On the other hand, the incentives for schooling investment differ across the economies because of the different taxes and subsidies.

whereas the rank in the welfare distribution is slightly less persistent across generations as shown in column (5). This illustrates once again that more intergenerational insurance, and as we are going to see below also higher welfare, are not necessarily associated with more mobility in earnings or schooling, as measured by the IGE or the intergenerational correlation of schooling.

In the economy with approximated linear, time-varying taxes consumption insurance is similar as in the social optimum, as indicated by the pass-through coefficient in column (1) of Table 11, whereas consumption insurance is lower, and hence the pass-through coefficient higher, in the economy with optimal linear taxes. We find in column (3) of Table 11 that this relates to the stronger incentive to exert labor effort for families with higher ability, and thus with higher productivity, in the social optimum relative to the economies with the simpler tax systems. The correlation of labor effort with ability is 0.75 in the social optimum compared with 0.38 in the economy with approximated linear, time-varying taxes, 0.18 in the economy with optimal linear taxes, and 0.03 in the calibrated economy. In the socially optimal allocation, consumption and labor effort are less negatively correlated than in the economies with the simpler tax systems. Hence, changes in ability affect welfare less in the social optimum as observed in column (2).

Given that the economies with the simpler tax systems imply different insurance and mobility patterns compared to the social optimum, one may wonder to which extent these differences matter for welfare. To answer this question, we compare the welfare gains achieved by the linear taxes and subsidies and the welfare gains of the second-best allocation relative to the calibrated economy.[38] We compute the welfare gains at the time of the reform. By the design of the reform, we keep the present value of the expected net cost of the allocation per dynasty unchanged at the time of reform to focus on insurance and mobility by abstracting from redistribution across dynasties.

[38] In the working-paper version we provide comparisons with the laissez faire without distortionary taxes, starting from the steady state of the calibrated economy. The welfare gain for the social optimum is a bit smaller then, at 4.8%, because of the borrowing constraint that is present in the calibrated economy but not in the laissez-faire economy. The welfare gain of 4.8% is larger than the gain between 1% and 3% reported in the life-cycle models of Farhi and Werning (2013) and Stantcheva (2017). By replicating their gains, we have verified that the difference comes from the larger initial cross-sectional heterogeneity in our setting, given that we start from the calibrated steady-state distribution, and from the intergenerational rather than life-cycle model implying different parameter values for the variance and the persistence of the shocks. The wedge between the discount factor of the planner and the dynasties instead does not explain the difference, as we discuss further below.

	Welfare gain in percent of consumption equivalents
Second best	6.0
Optimal linear taxes / subsidies, constant across generations	2.8
Approximated linear taxes / subsidies, varying across generations	2.2

Table 12: Welfare gains

Notes: Welfare gains at the time of the reform relative to the calibrated economy, holding constant the present value of expected net costs for each dynasty at the time of the reform. The consumption equivalents are computed holding labor supply constant and applying the discount factor ψ.

The first row of Table 12 shows the welfare gain of 6.0% for the social optimum, in which the planner is more patient than the dynasties ($\psi > \beta$). We report the welfare gains in percentage changes of consumption equivalents from the perspective of the planner, applying the discount factor ψ.[39]

The second and third row of Table 12 show the welfare gains achieved in the economies with the simple linear tax and subsidy systems. The economy with the approximated, varying tax and subsidy rates, in the third row of Table 12, achieves 37% of the welfare gains obtained by moving from the steady state of the calibrated economy to the second best after the reform. The economy with optimized but constant linear taxes and subsidies, in the second row of Table 12, achieves 47% of the welfare gains instead. This is a sizable part but less than in the calibrated life-cycle models of Farhi and Werning (2013) and Stantcheva (2017) who find that simple linear taxes deliver more than 90% of the welfare gains.[40]

In order to gauge the importance of $\psi > \beta$ for these results, i.e., the difference of the discount factor of the dynasties β and the discount factor of the planner ψ, we also compute the welfare gains for the case $\psi = \beta$, keeping the interest rate r unchanged.[41] As

[39]If we evaluate the welfare gains from the perspective of the family, applying the discount factor β in the objective function, the welfare gains are larger by a factor 1.05 as gains accruing in the present then receive more weight.

[40]We have found that an approximated quadratic tax schedule, which captures the progressive phasing out of bequest subsidies emphasized by Farhi and Werning (2010), does not achieve higher welfare gains. This may be a consequence of approximation error and, unfortunately, computing the optimal quadratic tax schedule is prohibitively costly. Appendix C provides further details on the non-linear approximation.

[41]For a stationary distribution in the calibrated economy, we need to maintain the assumption that $\beta <$

mentioned before, $\psi = \beta$ implies immiseration in the social optimum. In this case, the welfare gains of the socially optimal allocation after the reform compared with the calibrated economy are 9.6%. Approximated linear taxes and subsidies, which vary across generations, imply welfare gains of 4.9%, thus achieving 51% of the welfare gains of the reform. The share of welfare gains achieved with simpler tax and subsidies is thus quantitatively similar to our benchmark case, illustrating that the assumption $\psi > \beta$ is not critical for this result.

Although a substantial part of the welfare gains can be generated with the simple taxes and subsidies, one may expect that the history dependence of optimal taxes and subsidies is not fully captured by the endogenous state variables bequests b and human capital h in our model because the unobserved shocks to ability θ are persistent in our model and not i.i.d. as in Albanesi and Sleet (2006). Allowing for further history dependence of the taxes and subsidies, while maintaining tractability, seems viable within a life cycle of a generation, as shown in Kapička (2017), but less so across generations where this would require information on past generations for determining taxes and subsidies of the current generation.

6 Conclusion

We have analyzed intergenerational insurance, mobility and inequality in a dynastic model, comparing the economy calibrated to the U.S. with the social optimum, in which insurance by the utilitarian planner is constrained by incentive compatibility. We have disciplined the well-known wedge between the weights for future generations applied by the planner and the dynasties, targeting observable data on bequests in the calibration.

We have found that, relative to the calibrated U.S. economy, social optimality implies more intergenerational insurance against ability risk which is achieved with a stronger influence of nurture, in terms of schooling and bequests, on the family's position in the welfare distribution. At the same time, labor effort is more strongly correlated with ability in the social optimum relative to the calibrated economy, indicating that both more insurance and stronger incentives are provided in the social optimum.

Our quantitative analysis has revealed that the intergenerational mobility of schooling increases in the social optimum whereas income mobility across generations remains quite stable. A decomposition of the intergenerational earnings elasticity (IGE) reveals that (i) the lower intergenerational correlation of schooling contributes little to the IGE and (ii) the much higher correlation between labor effort and ability in the social optimum contributes both to a higher variance and intergenerational covariance of earnings, thus

$1/(1+r)$.

implying almost no change of the IGE. These results illustrate that economies with very different incentives, insurance and welfare may have a similar IGE. Changes in earnings mobility thus cannot readily be judged as *good* or *bad* without further analysis.

We have illustrated this point further by analyzing economies with simpler linear schedules for taxes and subsidies than in the social optimum. In these economies, families are less insured than in the social optimum but more insured than in the calibrated economy. We have found that the simpler tax and subsidies achieve about half of the welfare gains of the socially optimal tax system relative to the calibrated economy. The IGE in these economies, however, is lower than in the social optimum *and* the calibrated economy.

Our analysis of the transition to the social optimum shows that a weaker influence of nature on the position in the welfare distribution and more inequality of labor earnings ex post do not imply necessarily lower welfare ex ante because they may be the flip side of more intergenerational insurance. This illustrates further that interpretation of descriptive evidence on the evolution of inequality and mobility require the usual assumptions about preferences and technology, and about the social welfare function. For a plausible set of these assumptions, we have shown how intergenerational insurance and mobility may be shaped by the tax and subsidy system.

On average, bequests and schooling are subsidized in the social optimum to insure future generations against ability risk. We have shown that bequest subsidies are phased out more progressively across the earnings distribution than schooling subsidies because ability and schooling are complements for productivity, higher current ability is positively correlated with income, and higher current ability increases the expected ability of the offspring. Thus, bequest subsidies are targeted relatively more to the income poor than schooling subsidies.

It would be fruitful in further research, albeit computationally demanding, to extend the dynamic, dynastic Mirrlessian model by modeling the effects of socially optimal policy on decisions within a generation's life-cycle. Such analysis would allow to distinguish the effects of intergenerational transfers at different stages of the life cycle, e.g., at retirement or early adulthood. This would permit to quantify the relevance of the timing of such transfers in the presence of uninsurable risk or capital market imperfections.

References

Adermon, A., Lindahl, M., and Waldenström, D. (2018). Intergenerational Wealth Mobility and the Role of Inheritance: Evidence from Multiple Generations. *Economic Journal*, 128:F482–F513.

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

Arneson, R. (2018). Four Conceptions of Equal Opportunity. *Economic Journal*, 128:F152–F173.

Atkeson, A. and Lucas, R. E. (1992). On Efficient Distribution With Private Information. *Review of Economic Studies*, 59(3):427–453.

Barro, R. J. and Lee, J. W. (2013). A New Data Set of Educational Attainment in the World, 1950–2010. *Journal of Development Economics*, 104(C):184–198.

Becker, G. S. and Tomes, N. (1986). An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Labor Economics*, 87(6):1153–1189.

Behrman, J. R. and Rosenzweig, M. R. (2004). Parental Allocations to Children: New Evidence on Bequest Differences among Siblings. *Review of Economics and Statistics*, 86(2):637–640.

Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy*, 75(4):352–365.

Bénabou, R. and Ok, E. A. (2001). Social Mobility and the Demand for Redistribution: the POUM Hypothesis. *Quarterly Journal of Economics*, 116(2):447–487.

Bernheim, B. D. and Severinov, S. (2003). Bequests as Signals: an Explanation of the Equal Division Puzzle. *Journal of Political Economy*, 111(4):733–764.

Bingley, P., Cappellari, L., and Tatsiramos, K. (2018). On the Origins of Socio-Economic Inequalities: Evidence from a "Children of Twins" Design. Manuscript.

Blundell, R., Pistaferri, L., and Preston, I. (2008). Consumption Inequality and Partial Insurance. *American Economic Review*, 98(5):1887–1921.

Bolt, U., French, E., Hentall MacCuish, J., and O'Dea, C. (2021). The Intergenerational Elasticity of Earnings: Exploring the Mechanisms. Mimeograph, University of Cambridge.

Bovenberg, A. L. and Jacobs, B. (2005). Redistribution and Education Subsidies are Siamese Twins. *Journal of Public Economics*, 89(11-12):2005–2035.

Brown, J. R., Coile, C. C., and Weisbenner, S. J. (2010). The Effect of Inheritance Receipt on Retirement. *Review of Economics and Statistics*, 92:425–434.

Card, D. (1999). *The Causal Effect of Education on Earnings*, volume 3A of *Handbook of Labor Economics*, chapter 30, pages 1801–1863. Elsevier.

Carnegie-Rochester-Conference (2016). Inequality, Institutions and Redistribution. *Journal of Monetary Economics*, 77(C):1–148.

Cesarini, D., Lindqvist, E., Notowidigdo, M., and Östling, R. (2017). The Effect of Wealth on Individual and Household Labor Supply: Evidence from Swedish Lotteries. *American Economic Review*, 107(12):3917–3946.

Chetty, R. (2012). Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply. *Econometrica*, 80(3):969–1018.

Chetty, R., Guren, A., Manoli, D., and Weber, A. (2013). Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities. *NBER Macroeconomics Annual*, 27(1):1–56.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *Quarterly Journal of Economics*, 129(4):1553–1623.

Chuard, P. and Grassi, V. (2020). Switzer-Land of Opportunity: Intergenerational Income Mobility in the Land of Vocational Education. SEPS Discussion Paper Series 2020-11, University of St.Gallen.

Corbae, D., Nardi, M. D., and Lochner, L. (2017). RED Special Issue on Human Capital and Inequality, an Introduction. *Review of Economic Dynamics*, 25:1–3.

Cunha, F., Heckman, J. J., and Lochner, L. (2006). *Interpreting the Evidence on Life Cycle Skill Formation*, volume 1 of *Handbook of the Economics of Education*, chapter 12, pages 697–812. Elsevier.

Daruich, D. (2020). The Macroeconomic Consequences of Early Childhood Development Policies. Manuscript, University of Southern California.

Daruich, D. and Kozlowski, J. (2020). Explaining Intergenerational Mobility: the Role of Fertility and Family Transfers. *Review of Economic Dynamics*, 36:220–245.

De Nardi, M. and Yang, F. (2016). Wealth Inequality, Family Background, and Estate Taxation. *Journal of Monetary Economics*, 77(C):130–145.

Dodin, M., Findeisen, S., Henkel, L., Sachs, D., and Schüle, P. (2021). Social Mobility in Germany. Mimeograph, University of St.Gallen.

Farhi, E., Sleet, C., Werning, I., and Yeltekin, S. (2012). Non-linear Capital Taxation Without Commitment. *The Review of Economic Studies*, 79(4):1469–1493.

Farhi, E. and Werning, I. (2007). Inequality and Social Discounting. *Journal of Political Economy*, 115:365–402.

Farhi, E. and Werning, I. (2010). Progressive Estate Taxation. *The Quarterly Journal of Economics*, 125(2):635–673.

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

Fernández-Villaverde, J. and Krueger, D. (2007). Consumption over the Life Cycle: Facts from Consumer Expenditure Survey Data. *The Review of Economics and Statistics*, 89(3):552–565.

Findeisen, S. and Sachs, D. (2016). Education and Optimal Dynamic Taxation: the Role of Income-Contingent Student Loans. *Journal of Public Economics*, 138(C):1–21.

Findeisen, S. and Sachs, D. (2018). Education Policies and Taxation without Commitment. *The Scandinavian Journal of Economics*, 120(4):1075–1099.

Flinn, C. J. (2002). Labour Market Structure and Inequality: a Comparison of Italy and the U.S. *Review of Economic Studies*, 69(3):611–645.

Friedman, M. (1962). *The Distribution of Income*, chapter 10, pages 161–176. Capitalism and Freedom. University of Chicago Press.

Gallipoli, G., Low, H., and Mitra, A. (2020). Consumption and Income Inequality across Generations. CEPR Discussion Papers 15166, C.E.P.R. Discussion Papers.

Gelber, A. and Weinzierl, M. (2016). Optimal Taxation When Children's Abilities Depend on Parents' Resources. *National Tax Journal*, 69(1):11–40.

Golosov, M., Troshkin, M., and Tsyvinski, A. (2016). Redistribution and Social Insurance. *American Economic Review*, 106(2):359–386.

Güell, M., Pellizzari, M., Pica, G., and Rodríguez Mora, J. V. (2018). Correlating Social Mobility and Economic Outcomes. *Economic Journal*, 128:F353–F403.

Heathcote, J., Storesletten, K., and Violante, G. L. (2008). The Macroeconomic Implications of Rising Wage Inequality in the United States. NBER Working Papers 14052, National Bureau of Economic Research, Inc.

Heathcote, J., Storesletten, K., and Violante, G. L. (2010). The Macroeconomic Implications of Rising Wage Inequality in the United States. *Journal of Political Economy*, 118(4):681–722.

Heathcote, J., Storesletten, K., and Violante, G. L. (2014). Consumption and Labor Supply with Partial Insurance: An Analytical Framework. *American Economic Review*, 104(7):2075–2126.

Heathcote, J., Storesletten, K., and Violante, G. L. (2017). Optimal Tax Progressivity: An Analytical Framework. *The Quarterly Journal of Economics*, 132(4):1693–1754.

Heathcote, J. and Tsujiyama, H. (2021). Optimal Income Taxation: Mirrlees meets Ramsey. *Journal of Political Economy*, 129(11).

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N., and Verashchagina, A. (2008). The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends. *The B.E. Journal of Economic Analysis & Policy*, 7(2):1–48.

Hintermaier, T. and Koeniger, W. (2010). The Method of Endogenous Gridpoints with Occasionally Binding Constraints among Endogenous Variables. *Journal of Economic Dynamics and Control*, 34(10):2074–2088.

Hintermaier, T. and Koeniger, W. (2011). On the Evolution of the US Consumer Wealth Distribution. *Review of Economic Dynamics*, 14(2):317–338.

Holtz-Eakin, D., Joulfaian, D., and Rosen, H. S. (1993). The Carnegie Conjecture: Some Empirical Evidence. *The Quarterly Journal of Economics*, 108(2):413–435.

Huggett, M., Ventura, G., and Yaron, A. (2011). Sources Of Lifetime Inequality. *American Economic Review*, 101(7):2923–2954.

Imbens, G. W., Rubin, D. B., and Sacerdote, B. I. (2001). Estimating the Effect of Unearned Income on Labor Earnings, Savings, and Consumption: Evidence from a Survey of Lottery Players. *American Economic Review*, 91(4):778–794.

Jacobs, B. and Bovenberg, A. L. (2011). Optimal Taxation of Human Capital and the Earnings Function. *Journal of Public Economic Theory*, 13:957–971.

Kapička, M. (2013). Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach. *Review of Economic Studies*, 80(3):1027–1054.

Kapička, M. (2015). Optimal Mirrleesean Taxation in a Ben-Porath Economy. *American Economic Journal: Macroeconomics*, 7(2):219–248.

Kapička, M. (2017). Quantifying the Welfare Gains from History Dependent Income Taxation. Manuscript.

Kapička, M. and Neira, J. (2019). Optimal Taxation with Risky Human Capital. *American Economic Journal: Macroeconomics*, 11(4):271–309.

Keane, M. P. and Wolpin, K. I. (1997). The Career Decisions of Young Men. *Journal of Political Economy*, 105(3):473–522.

Kindermann, F., Mayr, L., and Sachs, D. (2020). Inheritance Taxation and Wealth Effects on the Labor Supply of Heirs. *Journal of Public Economics*, 191:104127.

Kocherlakota, N. R. (2010). *The New Dynamic Public Finance*. Princeton University Press.

Koeniger, W. and Prat, J. (2018). Human Capital and Optimal Redistribution. *Review of Economic Dynamics*, 27:1–26.

Krueger, D. and Ludwig, A. (2016). On the Optimal Provision of Social Insurance: Progressive Taxation versus Education Subsidies in General Equilibrium. *Journal of Monetary Economics*, 77(C):72–98.

Lee, D., Van der Klaauw, W., Haughwout, A. F., Brown, M., and Scally, J. (2014). Measuring Student Debt and its Performance. Staff Reports 668, Federal Reserve Bank of New York.

Lee, S. Y. T. and Seshadri, A. (2018). Economic Policy and Equality of Opportunity. *Economic Journal*, 128:F114–F151.

Lee, S. Y. T. and Seshadri, A. (2019). On the Intergenerational Transmission of Economic Status. *Journal of Political Economy*, 127(2):855–921.

Mankiw, G. N., Weinzierl, M. C., and Yagan, F. (2009). Optimal Taxation in Theory and Practice. *Journal of Economic Perspectives*, 23(4):147–174.

Markussen, S. and Røed, K. (2020). Economic Mobility Under Pressure. *Journal of the European Economic Association*, 18(4).

Mazumder, B. (2016). *Estimating the Intergenerational Elasticity and Rank Association in the United States: Overcoming the Current Limitations of Tax Data*, volume 43 of *Research in Labor Economics, Inequality: Causes and Consequences*, pages 83–129. Emerald Group Publishing Ltd.

Moser, C. and Olea de Souza e Silva, P. (2019). Optimal Paternalistic Saving Policies. Mimeograph, Columbia University.

Mulligan, C. (1999). Galton versus the Human Capital Approach to Inheritance. *Journal of Political Economy*, 107(6):S184–S224.

Nybom, M. and Stuhler, J. (2013). Interpreting Trends in Intergenerational Income Mobility. IZA Discussion Paper 7514, IZA.

OECD (2011). *Education at a Glance*. OECD.

OECD (2018). *A Broken Social Elevator? How to Promote Social Mobility*. OECD.

Pavan, A., Segal, I., and Toikka, J. (2014). Dynamic Mechanism Design: A Myersonian Approach. *Econometrica*, 82(2):601–653.

Pavoni, N. and Yazici, H. (2017). Intergenerational Disagreement and Optimal Taxation of Parental Transfers. *Review of Economic Studies*, 84(3):1264–1305.

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

Phelan, C. (2006). Opportunity and Social Mobility. *Review of Economic Studies*, 73:487–504.

Phelan, C. and Rustichini, A. (2018). Pareto Efficiency and Identity. *Theoretical Economics*, 13:979–1007.

Stantcheva, S. (2015). Optimal Income, Education, and Bequest Taxes in an Intergenerational Model. NBER Working Papers 21177, National Bureau of Economic Research, Inc.

Stantcheva, S. (2017). Optimal Taxation and Human Capital Policies over the Life Cycle. *Journal of Political Economy*, 125(6):1931–1990.

Storesletten, K., Telmer, C. I., and Yaron, A. (2004). Consumption and Risk Sharing over the Life Cycle. *Journal of Monetary Economics*, 51(3):609–633.

Wolff, E. N. and Gittleman, M. (2014). Inheritances and the Distribution of Wealth or Whatever Happened to the Great Inheritance Boom? *Journal of Economic Inequality*, 12:439–468.

A Planner problem: recursive formulation and derivations

In this appendix we present the problem of a utilitarian planner who maximizes the welfare of generations under incentive compatibility constraints. The socially-optimal solution of this problem requires that families truthfully reveal their hidden ability. We first present both the primal and the dual problem of the planner. We then provide the recursive formulation of the relaxed dual problem, based on the first-order approach.[42] We use the first-order conditions of the relaxed problem to derive the key equations (2) and (3) discussed in Section 2.1 of the main text. Stating the problem of the planner requires that we discuss incentive compatibility.

Incentive compatibility.—We focus on a direct revelation mechanism which ensures that families truthfully report their type in each generation. We denote the history of types within a given family as $\theta^t \equiv \{\theta_0, \theta_1, ..., \theta_t\}$ and history dependent allocations as $x_t(\theta^t) \equiv \{c_t(\theta^t), h_{t+1}(\theta^t), y_t(\theta^t)\}$. The feasible set $\mathcal{X}$ contains all sequences $\mathbf{x} \equiv \{x_t(\theta^t)\}_{t=1}^T$ of measurable functions $x_t : \Theta^t \to \mathbb{R}_+^3$. Using the production function to substitute l_t in the utility function and writing $U(c_t, y_t, h_t, \theta_t)$ instead of $\mathbf{U}(c_t, l_t)$, preferences of a family dynasty for an allocation $\mathbf{x}$ are

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

where $\mathbb{E}_0$ is the expectation operator conditional on information available at time 0 and β is the discount factor of the family.

Family dynasties 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 $r_t : \Theta^t \to \Theta$. The types are private information so that an allocation must be incentive compatible to ensure truthful reporting, i.e.,

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

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

Primal problem.—We assume a utilitarian planner who weighs the welfare of each family dynasty equally and discounts the future less than a dynasty, i.e. $\psi > \beta$. See Kocherlakota (2010), chapter 5, p. 146, for a textbook treatment. The planner can fully diversify the idiosyncratic ability risk that family dynasties face. Since there are no general

[42]This approach replaces the incentive-compatibility constraints with an envelope condition that needs to be satisfied on the equilibrium path on which families truthfully reveal their type. Given that shocks to ability are unobservable and persistent, our recursive formulation of the planner problem relies on results of Kapička (2013) and Pavan et al. (2014).

equilibrium feedbacks that link the problems of the dynasties, the planner can maximize aggregate welfare by maximizing welfare of each dynasty. For a utilitarian planner, the problem of insuring family dynasties under the veil of ignorance (from the perspective of period 0) is equivalent to the problem of optimal redistribution across family dynasties with different initial conditions. The primal problem of the planner is

$$
\begin{aligned}
W = \max_{\{c_t, y_t, h_{t+1}\}} \mathbb{E}_0 \sum_{t=1}^{\infty} \psi^{t-1} U\left(c_t, y_t, h_t, \theta_t\right) && \text{(A.2)}\\
\text{s.t. } \mathbb{E}_0 \sum_{t=1}^{\infty} \beta^{t-1} U(\cdot) \geq V, && \text{(A.3)}\\
(IC), &&\\
\mathbb{E}_0 \sum_{t=1}^{\infty} q^{t-1} z_t \leq \Gamma_0, &&
\end{aligned}
$$

where z_t is the per-period net cost (i.e., $z_t \equiv c_t + g(h_{t+1}, h_t) - y_t$), Γ_0 is a given level of average discounted costs, V is a (promised) utility level and $q \equiv 1/(1+r)$. Without loss of generality, we can assume an initial (distribution of) promised utility.

We consider the constant discount factor ψ in the planner's objective. In the derivation of the planner's objective in Kocherlakota (2010), p. 147, the discount factor is time varying and converges to ψ if $\psi > \beta$ and $t \to \infty$. See also Kocherlakota (2010), p. 157. We abstract from possible time variation in the planner's discount factor assuming that the planning objective has converged. This has the advantage that our transition analysis after the tax reform in Section 5 is not confounded by changes in the discount factor over time on the transition path.

Dual problem.—The dual cost minimization problem of the planner is

$$
\begin{aligned}
\Gamma_0 = \min_{\{c_t, y_t, h_{t+1}\}} \mathbb{E}_0 \sum_{t=1}^{\infty} q^{t-1} z_t && \text{(A.4)}\\
\text{s.t. } \mathbb{E}_0 \sum_{t=1}^{\infty} \beta^{t-1} U(\cdot) \geq V &&\\
(IC), &&\\
\mathbb{E}_0 \sum_{t=1}^{\infty} \psi^{t-1} U(\cdot) \geq W. &&
\end{aligned}
$$

Incentive constraints for the recursive formulation and the first-order approach.—We replace the ex-ante incentive constraint (A.1) with an ex-post requirement to write the planner's dual problem in recursive form.[43] For this purpose, we define the equilibrium continuation utility $\omega(\theta^t)$ for a given history θ^t as

$$\omega\left(\theta^t\right) \equiv U\left(c_t\left(\theta^t\right), y_t\left(\theta^t\right), h_t\left(\theta^{t-1}\right), \theta_t\right) + \beta \int_\Theta \omega\left(\theta^t, \theta_{t+1}\right) dF\left(\theta_{t+1}|\theta_t\right), \tag{A.5}$$

for all $t = 1, \ldots, \infty$. Families compare the continuation value $\omega(\theta^t)$ of truthful reporting to the values derived from arbitrary reporting strategies

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

Incentives are compatible ex post if

$$\omega\left(\theta^t\right) \geq \omega^{\mathbf{r}}\left(\theta^t\right), \text{ for all } \theta^t \text{ and all } \mathbf{r} \in \mathcal{R}. \tag{A.6}$$

We use $\mathbf{x}^{\mathbf{IC}}$ to denote the set of all allocations $\mathbf{x}$ satisfying (A.6).[44]

Problem (A.4) requires to keep track of all the out-of-equilibrium payoffs to check the incentive constraint (A.6). Applying the first-order approach, we reduce the complexity of the problem by replacing the incentive constraint with an envelope condition which only depends on the marginal utility of truth-tellers. The envelope condition for the considered problem is

$$\frac{\partial \omega(\theta^t)}{\partial \theta_t} = \frac{\partial U\left(c_t(\theta^t), y_t(\theta^t), h_t\left(\theta^{t-1}\right), \theta_t\right)}{\partial \theta_t} + \beta \int_\Theta \omega\left(\theta^{t+1}\right) \frac{\partial f\left(\theta_{t+1}|\theta_t\right)}{\partial \theta_t} d\theta_{t+1}. \tag{A.7}$$

Intuitively, if one considers a one-shot perturbation of the type θ_t in equation (A.5), the sum of all the derivatives of terms with respect to the report of the type is zero, once the derivatives are evaluated on the equilibrium path where truthful reporting is optimal. Equation (A.7) reduces to the condition prevailing in Mirrlees' static setting if types are i.i.d. In this case, the second term on the right-hand side of (A.7) vanishes. The second term on the right-hand side is relevant instead if types are persistent because unobserved ability then generates additional private information. For example, parents who underreport their type become more optimistic than the planner about the ability of their children if types are positively correlated.

Replacing the incentive constraint by (A.7) greatly simplifies the optimization prob-

[43]In this part we draw heavily on material in Koeniger and Prat (2018) which we present here for completeness.

[44]Note that allocations in $\mathbf{x}^{\mathbf{IC}}$ are incentive compatible for all $\theta^t \in \Theta^t$. This requires truth telling to be optimal after *any history* of shocks, whereas the incentive constraint (A.1) only requires truth telling to be ex-ante optimal. 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.

lem because it only depends on the continuation utility of truth-tellers and not on the continuation utility of all possible types. Defining $\mathbf{x}^{\mathbf{FOA}}$ as the set of allocations so that condition (A.7) holds for all θ^t, we note that $\mathbf{x}^{\mathbf{IC}} \subseteq \mathbf{x}^{\mathbf{FOA}}$. Replacing the incentive constraint in problem (A.4) by $\mathbf{x} \in \mathbf{x}^{\mathbf{FOA}}$ thus relaxes this problem.

Recursive relaxed problem.—We write the relaxed problem in recursive form so that we can solve it as sequence of standard optimal control problems. Denoting with "′" values of variables one period in the future and with "−" values of variables with a one-period lag, the stationary recursive problem is:

$$\Gamma(V, W, \Phi, h, \theta_-) = \min_{\{c,y,h',V',W',\Phi'\}} \{$$
$$\int_\Theta [c(\theta) + g(h'(\theta), h) - y(\theta) + q\Gamma(V'(\theta), W'(\theta), \Phi'(\theta), h'(\theta), \theta)]\, dF(\theta|\theta_-)\},$$

$$\text{s.t.} \qquad \omega(\theta) = U(c(\theta), y(\theta), h, \theta) + \beta V'(\theta), \tag{A.8}$$
$$\tilde{\omega}(\theta) = U(c(\theta), y(\theta), h, \theta) + \psi W'(\theta) \tag{A.9}$$
$$V = \int_\Theta \omega(\theta) dF(\theta|\theta_-), \tag{A.10}$$
$$W = \int_\Theta \tilde{\omega}(\theta) dF(\theta|\theta_-), \tag{A.11}$$
$$\Phi = \int_\Theta \omega(\theta) \frac{\partial f(\theta|\theta_-)}{\partial \theta_-} d\theta, \tag{A.12}$$
$$\frac{\partial \omega(\theta)}{\partial \theta} = \frac{\partial U(c, y, h, \theta)}{\partial \theta} + \beta \Phi'(\theta). \tag{A.13}$$

Equations (A.8) and (A.9) define the continuation values using the discount factor of the family and planner, respectively. Equations (A.10) and (A.11) are the respective promise keeping constraints. Because of the persistence of ability, the planner keeps track how reports of ability in the last period change promised utility so that the problem also has a threat-keeping constraint (A.12). The envelope condition (A.13) is the incentive compatibility constraint as explained above.

The recursive problem is standard but for the additional constraints (A.9) and (A.11) which enter the problem because the planner discounts the future at a different rate than the family dynasties. Substituting (A.9) into (A.11), and substituting $U(\cdot)$ using (A.8) and (A.10), we obtain

$$W = V + \int_\Theta \Big(\psi W'(\theta) - \beta V'(\theta)\Big) dF(\theta|\theta_-). \tag{A.14}$$

The problem has therefore one more state variable W and equation (A.14) as additional constraint, which replaces (A.9) and (A.11). The additional state variable and constraint (A.14) would be redundant if the planner and the dynasties discounted the future at the same rate because $W = V$ and $\int_\Theta \psi W'(\theta) - \beta V'(\theta) dF(\theta|\theta_-) = 0$ in this case.

Optimality conditions.—We use the separability of utility in consumption and labor effort to solve constraint (A.8) for consumption. We then substitute the resulting consumption $c(\omega(\theta) - \beta V'(\theta), y(\theta), h, \theta)$ into the objective function. The Hamiltonian reads

$$\begin{aligned}\mathcal{H} =& [c(\omega(\theta) - \beta V'(\theta), y(\theta), h, \theta) + g(h'(\theta), h) - y(\theta) \\ &+ q\Gamma(V'(\theta), W'(\theta), \Phi'(\theta), h'(\theta), \theta)] f(\theta|\theta_-) \\ &+ \lambda[V - \omega(\theta) f(\theta, \theta_-)] + \gamma \left[\Phi - \omega(\theta) \frac{\partial f(\theta|\theta_-)}{\partial \theta_-}\right] \\ &+ \eta \left[W - V - \left(\psi W'(\theta) - \beta V'(\theta)\right) f(\theta|\theta_-)\right] \\ &+ \mu(\theta) \left[\frac{\partial U(c(\omega(\theta) - \beta V'(\theta), y(\theta), h, \theta), y(\theta), h, \theta)}{\partial \theta} + \beta \Phi'(\theta)\right].\end{aligned} \tag{A.15}$$

The first-order conditions for h', y and Φ' remain qualitatively unchanged compared with those reported in Appendix A of Koeniger and Prat (2018). We thus focus on the first-order necessary conditions which generate new insights. The first-order condition for V' is

$$\frac{\beta}{\partial u(c(\theta))/\partial c(\theta)} - \beta\eta = q \frac{\partial \Gamma(V'(\theta), W'(\theta), \Phi'(\theta), h'(\theta), \theta)}{\partial V'(\theta)} \tag{A.16}$$

and the first-order condition for W' is

$$\eta\psi = q \frac{\partial \Gamma(V'(\theta), W'(\theta), \Phi'(\theta), h'(\theta), \theta)}{\partial W'(\theta)}. \tag{A.17}$$

We now use these two equations to derive a modified reciprocal Euler equation and the key equations (2) and (3) discussed in Section 2.1 of the main text. As a first step, we substitute the envelope condition $\partial\Gamma/\partial W = \eta$ into equation (A.17) which implies

$$\eta\psi = q\eta'(\theta). \tag{A.18}$$

The shadow price of the new constraint (A.14) thus evolves deterministically which simplifies the numerical solution.

Next, we substitute the envelope condition $\partial\Gamma/\partial V = \lambda - \eta$ into first-order condition

(A.16) to obtain

$$\frac{1}{\partial u(c(\theta))/\partial c(\theta)} = \frac{q}{\beta}\lambda'(\theta) - \frac{q}{\beta}\eta'(\theta) + \eta. \tag{A.19}$$

Using (A.18) to substitute $\eta'(\theta)$,

$$\frac{1}{\partial u(c(\theta))/\partial c(\theta)} = \frac{q}{\beta}\lambda'(\theta) + \eta\left(1 - \frac{\psi}{\beta}\right). \tag{A.20}$$

This is the modified reciprocal Euler equation because $\lambda'(\theta) = \mathbb{E}[(\partial u(c'(\theta'))/\partial c'(\theta'))^{-1}|\hat{\theta}]$, as can be shown following the same steps as in the proof of Remark 1 in Koeniger and Prat (2018), for example. Note that the expectation conditions on the history of abilities until the current period which we denote by $\hat{\theta}$.

Next we derive equations (2) and (3) presented in Section 2.1 of the main text. We take unconditional expectations on both sides of (A.20) and, using again that $\lambda'(\theta) = \mathbb{E}[(\partial u(c'(\theta'))/\partial c'(\theta'))^{-1}|\hat{\theta}]$ at all leads and lags, we replace the unconditional expected value of the inverse marginal utility in the current period by the corresponding expected value of the multiplier λ_t. Hence,

$$\mathbb{E}[\lambda_{t+1}] = \frac{\beta}{q}\mathbb{E}[\lambda_t] + \eta_t\left(\frac{\psi}{q} - \frac{\beta}{q}\right). \tag{A.21}$$

For the implementation of the reform, discussed in further detail in Appendix F, $\lambda_0 = \eta_0$ at the time of the reform. Therefore, $\mathbb{E}[\lambda_1] = \frac{\psi}{q}\lambda_0$, which can be iterated forward so that $\mathbb{E}[\lambda_{t+1}] = \frac{\psi}{q}\mathbb{E}[\lambda_t]$ for all t.

Replacing the unconditional expectations of λ_t and λ_{t+1} by the corresponding expectations of the inverse marginal utilities and denoting the marginal utility of consumption as $u'(\cdot)$, we obtain the key equation (2) in Section 2.1 of the main text:

$$\mathbb{E}\left[\frac{1}{u'(c_t(\theta^t))}\right] = \frac{q}{\psi}\mathbb{E}\left[\frac{1}{u'(c_{t+1}(\theta^{t+1}))}\right].$$

Clearly, stationarity of consumption requires $\psi = q$. Substituting this into the modified reciprocal Euler equation (A.20), we obtain key equation (3) in Section 2.1 of the main text:

$$\mathbb{E}\left[\frac{1}{u'(c_{t+1}(\theta^{t+1}))}\middle|\theta^t\right] = \frac{\beta}{\psi}\frac{1}{u'(c_t(\theta^t))} + \eta_t\left(1 - \frac{\beta}{\psi}\right).$$

For $\beta/\psi < 1$, $1/u'(c(\theta^t))$ follows a mean-reverting process. This recovers, for our model setting, results in Farhi and Werning (2007), p. 380, or Kocherlakota (2010), p. 158.

Note that we obtain the standard reciprocal Euler equation if $\eta = 0$, in which case the constraint for providing a certain amount of welfare W is slack, or if $\psi = \beta$ so that the immiseration result applies. As we have just seen, stationarity of consumption in the planner problem requires $\psi = q$; and a stationary distribution in the decentralized calibrated economy with incomplete markets requires $q > \beta$. Thus, $\psi > \beta$ seems a rather natural assumption.

The steady state consumption level is determined by the resources of the planner, as mentioned in Farhi and Werning (2007), p. 385, and $\eta_0 = \partial\Gamma_0/\partial W_0$ measures the marginal cost for the planner of providing social welfare W_0 in period 0. We can index the planner problem by η_0 since the entire sequence of multipliers η_t is deterministic. This property not only simplifies the numerical solution but also the implementation of the reform, as we explain further in Appendix F.

B The first-best allocation

We characterize the first-best allocation for the utilitarian planner who observes ability and maximizes welfare under the veil of ignorance:

$$\max_{\{a_{t+1},c_t,y_t,h_{t+1}\}} \mathbb{E}_0 \sum_{t=1}^{\infty} \psi^{t-1}[u(c_t(\theta^t)) - v(l_t(\theta^t))]$$

s.t.

$$a_{t+1} = Ra_t + \mathbb{E}_t[y_t(\theta^t) - c_t(\theta^t) - g(h_{t+1}(\theta^t), h_t(\theta^{t-1}))], \text{ for all } t, \tag{B.1}$$

where a_t denotes the planner's assets and $y_t(\theta^t) = l_t(\theta^t)A(\theta_t, h_t(\theta^{t-1}))$ denotes output. *Average* output, consumption and education investment enter in the resource constraint (B.1) because the planner can diversify the idiosyncratic ability risk.

B.1 First-order conditions and their implications for the first-best allocation

Planner's assets a_{t+1}.— Letting ϱ_t denote the multiplier for constraint (B.1) at time t, the first-order condition with respect to a_{t+1} yields

$$\varrho_t = R\varrho_{t+1}, \tag{B.2}$$

and hence $\varrho_t = R^{-t}\varrho_0$.

Consumption c_t.— The first-order condition for consumption $c_t(\theta^t)$ is

$$\psi^t u'(c_t(\theta^t)) = \rho_t \tag{B.3}$$

for all histories θ^t. Because ρ_t is constant across all histories θ^t, (B.3) implies that the planner fully insures consumption. Combining (B.2) and (B.3), we obtain

$$u'(c_t(\theta^t)) = \frac{\rho_0}{(R\psi)^t}.$$

Output y_t and labor effort l_t.— The first-order condition for labor effort l_t required to produce output y_t is

$$\psi^t v'(l_t(\theta^t)) = \rho_t A(\theta_t, h_t(\theta^{t-1})) \tag{B.4}$$

or, using (B.2),

$$v'(l_t(\theta^t)) = \frac{\rho_0}{(R\psi)^t} A(\theta_t, h_t(\theta^{t-1})).$$

Hence, labor effort in the first best varies with productivity $A(\theta_t, h_t(\theta^{t-1}))$.

Human capital investment.— The first-order condition for human capital investment is

$$\begin{aligned} &-\rho_t \frac{\partial g(h_{t+1}(\theta^t), h_t(\theta^{t-1}))}{\partial h_{t+1}} \\ &+\rho_{t+1} \mathbb{E}\left[l_{t+1} \frac{\partial A(\theta_{t+1}, h_{t+1}(\theta^t))}{\partial h_{t+1}} - \frac{\partial g(h_{t+2}(\theta^{t+1}), h_{t+1}(\theta^t))}{\partial h_{t+1}} | \theta^t \right] = 0 \end{aligned} \tag{B.5}$$

or, using (B.2),

$$R = \frac{\mathbb{E}\left[l_{t+1} \frac{\partial A(\theta_{t+1}, h_{t+1}(\theta^t))}{\partial h_{t+1}} - \frac{\partial g(h_{t+2}(\theta^{t+1}), h_{t+1}(\theta^t))}{\partial h_{t+1}} | \theta^t \right]}{\frac{\partial g(h_{t+1}(\theta^t), h_t(\theta^{t-1}))}{\partial h_{t+1}}} \tag{B.6}$$

Human capital investment in the first best equates the expected return of human capital, per invested amount at the margin, to the return R on the planner's assets. As noted already by Friedman (1962) and derived, for example, in Becker and Tomes (1986), the first best implies that (expected) returns to human capital and physical capital are equal. Given that ability shocks are persistent and more ability increases the return to human capital, equation (B.6) implies that the planner invests more into human capital of currently more able dynasties.

Remark.— In the calibration, $\psi R = 1$. Equations (B.3) and (B.4) then imply that consumption in the first best is constant over time and labor effort is only a function of productivity $A(\theta_t, h_t(\theta^{t-1}))$. Compared with the constrained-efficient social optimum, consumption is fully decoupled from production in the first best. Production requires labor effort and human capital as inputs which optimally change according to the ability of dynasties. Quantitatively, the correlation between labor effort and ability is 0.98 in the steady state of the first best compared with 0.37 in the second best. The correlation between schooling and ability is 0.52 in the first best compared with 0.36 in the second best.

C The wedges and approximate implementation

In the laissez faire each dynasty solves the maximization problem[45]

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

where the family chooses functions $b', h', l, c : \Theta \to \mathbb{R}$. Note that, as in the planner problem but differently to the calibrated economy presented in Section 2, we make the common assumption that the dynasty faces no borrowing constraint in the laissez faire. Thus, below we obtain the standard definitions of the wedges based on the first-order conditions of the laissez faire problem.

The first-order conditions for bequests, human capital and labor supply are:

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

C.1 The wedges

Based on these first-order conditions and given the separability of the utility function, the history-dependent wedges at time t for bequests $\tau_{b,t}$, human capital $\tau_{h,t}$ and labor supply

[45] The first-order conditions of this problem are equivalent for the problem $W^L(b(\theta),h(\theta),\theta) = \max_{\{b'(\theta),h'(\theta),l(\theta),c(\theta)\}} \left\{ \left[\mathbf{U}(c(\theta),l(\theta)) + \beta \int_\Theta W^L(b'(\theta),h'(\theta),\theta') \right] dF(\theta'|\theta) \right\}$, s.t. the constraints.

$\tau_{l,t}$ are then defined as

$$\tau_{b,t}\left(\theta^t\right) \equiv 1 - \frac{q}{\beta}\frac{\partial u\left(c_t\left(\theta^t\right)\right)/\partial c_t}{\mathbb{E}\left[\partial u\left(c_{t+1}\left(\theta^{t+1}\right)\right)/\partial c_{t+1}\middle|\theta^t\right]}, \tag{C.1}$$

$$\tau_{h,t}\left(\theta^t\right) \equiv \frac{\beta}{\frac{\partial g\left(h_{t+1}\left(\theta^t\right),h_t\left(\theta^{t-1}\right)\right)}{\partial h_{t+1}}}\mathbb{E}\left[\frac{\frac{\partial u\left(c_{t+1}\left(\theta^{t+1}\right)\right)}{\partial c_{t+1}}}{\frac{\partial u\left(c_t\left(\theta^t\right)\right)}{\partial c_t}}\left(\frac{\partial y_{t+1}\left(\theta^{t+1}\right)}{\partial h_{t+1}} - \frac{\partial g\left(h_{t+2}\left(\theta^{t+1}\right),h_{t+1}\left(\theta^t\right)\right)}{\partial h_{t+1}}\right)\middle|\theta^t\right] - 1, \tag{C.2}$$

$$\tau_{l,t}\left(\theta^t\right) \equiv 1 - \frac{\partial v\left(y_t\left(\theta^t\right),h_t\left(\theta^{t-1}\right),\theta_t\right)/\partial y_t}{\partial u\left(c_t\left(\theta^t\right)\right)/\partial c_t}, \tag{C.3}$$

where the function $v(\cdot)$ denotes the disutility of labor once we have substituted labor using the production function.

C.2 Approximate implementation of the social optimum

The social optimum can be decentralized with a general, history-dependent tax schedule as shown, for example, in Stantcheva (2017). We approximate the implementation of the social optimum with a history-independent and linear tax schedule. With this goal in mind, we now specify an auxiliary, decentralized problem for each dynasty that helps us to explain how we approximate the linear history-independent tax schedule. The auxiliary problem is

$$\max_{\{b_{t+1},h_{t+1},l_t,c_t\}} \mathbb{E}_t\left[\sum_{s=t}^{\infty}\beta^{s-t}\mathbf{U}\left(c_s,l_s\right)\right] \tag{C.4}$$

$$\begin{aligned}
\text{s.t. } b_{t+1} &= (1+r)\left(1-t_{b,t}\left(\theta^{t-1}\right)\right)b_t + \left(1-t_{y,t}\left(\theta^t\right)\right)y_t - c_t - \left(1+t_{h,t}\left(\theta^t\right)\right)g(h_{t+1},h_t) - T\left(\theta^t\right),\\
y_t &= Y(h_t,\theta_t,l_t),\\
\ln(\theta) &= \rho\ln(\theta_-) + \epsilon,
\end{aligned}$$

where the tax shifter $T\left(\theta^t\right)$ becomes a transfer if negative, and the conditioning of that shifter and the taxes $t_{j,t}(\cdot)$, $j = b,h,y$, on the history imply general, possibly non-linear tax schedules across dynasties with different histories. These general tax schedules allow to implement the social optimum and this can be achieved also by conditioning taxes on the history of observable variables such as output and education expenditures. As discussed in Stantcheva (2017), for example, this requires that the history of these observable variables allows to identify θ^t.

Note that $t_{b,t}\left(\theta^{t-1}\right)$ is the tax rate applied at the time parents choose bequest b_t and $t_{h,t}(\theta^t)$ increases the cost of a year of schooling, $g(h_{t+1}, h_t)$, so that $t_{h,t}(\theta^t) < 0$ has the interpretation of a subsidy for human capital investment h_{t+1}. For realism, we implement the conditioning of taxes or transfers on human capital by linking the tax or subsidy for human capital to education expenditures.

We proceed by linking the taxes $t_{j,t}(\cdot)$, $j = b, h, y$ to the respective wedges and then use these relationships to approximate linear, history-independent taxes.

Labor income tax.—The first-order condition for labor supply or income y and the definition of the labor wedge (C.3) imply that the marginal income tax of a dynasty with a certain history θ^t equals the labor wedge, i.e. $t_{y,t}(\theta^t) = \tau_{l,t}(\theta^t)$. In subsection 5.5 we are interested in how well the social optimum can be approximated with simpler linear taxes that do not depend on history. Given the non-tractability of solving for the optimal linear taxes that are allowed to vary across generations, as mentioned in the main text, for the time-varying taxes we proceed as Farhi and Werning (2013) or Stantcheva (2017) and approximate the linear income taxes with the cross-sectional average of the labor wedge:

$$\bar{t}_{y,t} = \mathbb{E}\left[\tau_{l,t}\left(\theta^t\right)\right] \tag{C.5}$$

so that every dynasty faces the same labor income tax.

Bequest tax.— The first-order condition with respect to b_{t+1} implies that

$$\frac{\partial u(c_t)}{\partial c_t} = \beta(1+r)\left(1 - t_{b,t+1}\left(\theta^t\right)\right)\mathbb{E}_t\left[\frac{\partial u(c_{t+1})}{\partial c_{t+1}}\right]. \tag{C.6}$$

Thus, comparison with equation (C.1) implies that bequests of a dynasty with a certain history θ^t should be taxed at rate $\tau_{b,t}(\theta^t)$.[46] We approximate the bequest tax with the cross-sectional average of the bequest wedge, i.e.,

$$\bar{t}_{b,t+1} = \mathbb{E}\left[\tau_{b,t}\left(\theta^t\right)\right]. \tag{C.7}$$

Human capital tax.— We combine the first-order condition for human capital

[46] As discussed in Kocherlakota (2010), taxation of assets generally has to be implemented ex post, after realization of θ_{t+1}, thus ensuring that the Euler equation of families is satisfied for each consumption level at the reported ability. We approximate this ex-post heterogeneity in the tax rate when we consider nonlinear taxes below.

$$\frac{\partial g(h_{t+1},h_t)}{\partial h_{t+1}}\left(1+t_{h,t}\left(\theta^t\right)\right)\frac{\partial u\left(c_t\right)}{\partial c_t}=\beta\mathbb{E}_t\left[\left(\frac{\partial y_{t+1}}{\partial h_{t+1}}\left(1-t_{y,t+1}\left(\theta^{t+1}\right)\right)\right.\right.$$
$$\left.\left.-\frac{\partial g(h_{t+2},h_{t+1})}{\partial h_{t+1}}\left(1+t_{h,t+1}\left(\theta^{t+1}\right)\right)\right)\frac{\partial u\left(c_{t+1}\right)}{\partial c_{t+1}}\right] \tag{C.8}$$

with the definition of the wedge for human capital (C.2) and solve for $t_{h,t}\left(\theta^t\right)$:

$$t_{h,t}\left(\theta^t\right)=\tau_{h,t}\left(\theta^t\right)-\frac{\beta}{\frac{\partial g(h_{t+1},h_t)}{\partial h_{t+1}}}\mathbb{E}_t\left[\left(\frac{\partial y_{t+1}}{\partial h_{t+1}}t_{y,t+1}\left(\theta^{t+1}\right)+\frac{\partial g(h_{t+2},h_{t+1})}{\partial h_{t+1}}t_{h,t+1}\left(\theta^{t+1}\right)\right)\frac{\frac{\partial u(c_{t+1})}{\partial c_{t+1}}}{\frac{\partial u(c_t)}{\partial c_t}}\right]. \tag{C.9}$$

The equation shows that a positive wedge for human capital does not necessarily imply a positive current marginal tax on human capital accumulation. The second term inside the expectation operator on the right-hand side shows that the sign and size of the tax also depends on how human capital changes labor income and thus labor-income taxes in the next period, how human capital changes the cost for education in the next period, and how these changes are correlated with the marginal utility of consumption. In particular, the planner has to undo the distortion on human-capital accumulation implied by labor-income taxation, as shown in Bovenberg and Jacobs (2005), and the distortion implied by the tax/subsidy on human capital next period. Note that through the stochastic discount factor, any distortion of the bequest decision also influences the tax/subsidy on human capital. Furthermore, effects of human capital accumulation on incentives are captured as well through the wedge $\tau_{h,t}$. Such effects occur if productivity is not Cobb-Douglas as emphasized, for example, by Stantcheva (2017).

As for the other taxes, we approximate the linear tax or subsidy for human capital investment by taking the cross-sectional average, i.e.,

$$\bar{t}_{h,t}=\mathbb{E}\left[t_{h,t}\left(\theta^t\right)\right]. \tag{C.10}$$

Given the recursive nature of equation (C.9), we use the approximated taxes in $t+1$ when approximating taxes in t. This ensures consistent use of the linear tax approximations in the dynasties' problem to compare the decentralized problem with the approximated linear tax schedule to the social optimum. We thus solve problem (C.4) replacing the general tax schedules with the approximated linear taxes $\bar{t}_{j,t}$, $j=b,h,y$, and compare the welfare gains of this economy with simple linear taxes to the welfare gains of the social optimum with implicit non-linear and history-dependent taxes.

Non-linear approximation of the tax schedules.—For the welfare comparisons of the decentralized economy with a simple non-linear approximation of the tax schedules, mentioned in footnote 40, we assume $t_{y,t}(y_t, \theta^t)$, $t_{b,t+1}(b_{t+1}, \theta^t)$ and $t_{h,t}(h_{t+1}, \theta^t)$ in the auxiliary problem (C.4). I.e., we capture explicitly some of the non-linearities of the tax schedules by letting the income tax depend on the current income, by letting the bequest tax depend on the bequest level and by letting the subsidies for human capital expenditures depend on the size of these expenditures. This allows, for example, to capture explicitly the progressivity emphasized by Farhi and Werning (2010) in the context of bequest taxation. The tax schedules still condition on the history given that the ability shocks are persistent and not i.i.d. so that the history is not fully encoded in the endogenous state variables. The first-order conditions of the auxiliary problem (C.4), with the modified tax schedules, and the definition of the wedges then imply that

$$t_{y,t}\left(y_t, \theta^t\right) + \frac{\partial t_{y,t}(y_t, \theta^t)}{\partial y_t} y_t = \tau_{l,t}\left(\theta^t\right), \tag{C.11}$$

$$t_{b,t}\left(b_t, \theta^t\right) + \frac{\partial t_{b,t}(b_t, \theta^t)}{\partial b_t} b_t = 1 - \frac{q}{\beta} \frac{\partial u\left(c_{t-1}\left(\theta^{t-1}\right)\right)/\partial c_{t-1}}{\partial u\left(c_t(\theta^t)\right)/\partial c_t}, \tag{C.12}$$

and

$$\begin{aligned} t_{h,t}\left(h_{t+1}, \theta^t\right) + \frac{\partial t_{h,t}(h_{t+1}, \theta^t)}{\partial h_{t+1}} \frac{1}{\epsilon_{g(\cdot),h_{t+1}}} h_{t+1} = \tau_{h,t}\left(\theta^t\right) \\ - \frac{\beta}{\frac{\partial g(h_{t+1},h_t)}{\partial h_{t+1}}} \mathbb{E}_t \left[\left(\frac{\partial y_{t+1}}{\partial h_{t+1}} \left(t_{y,t+1}\left(y_{t+1}, \theta^{t+1}\right) + \frac{\partial t_{y,t+1}\left(y_{t+1}, \theta^{t+1}\right)}{\partial y_{t+1}} y_{t+1} \right) \right.\right. \\ \left.\left. + \frac{\partial g(h_{t+2}, h_{t+1})}{\partial h_{t+1}} t_{h,t+1}\left(h_{t+2}, \theta^{t+1}\right) \right) \frac{\frac{\partial u(c_{t+1})}{\partial c_{t+1}}}{\frac{\partial u(c_t)}{\partial c_t}} \right], \end{aligned} \tag{C.13}$$

where $\epsilon_{g(\cdot),h_{t+1}} \equiv \frac{\partial g(h_{t+1},h_t)}{\partial h_{t+1}} \frac{h_{t+1}}{g(h_{t+1},h_t)}$ is the elasticity of education expenditures with respect to human capital. Note the right-hand side of (C.12) differs from the right-hand side of (C.1) because we approximate the non-linear tax schedule for bequests based on the implementation discussed in Kocherlakota (2010), which conditions on the current realization of the shock. See also Farhi and Werning (2010), p. 664, for this type of implementation in an intergenerational model.

Approximating the non-linearity of each tax with a quadratic function, the respective tax *rate* is linear, i.e., $t(x) = \widetilde{\alpha} + \widetilde{\beta} x$. The left-hand side of equations (C.11), (C.12) and (C.13) then becomes

$$\widetilde{\alpha}_{y,t} + 2\widetilde{\beta}_{y,t} y_t, \tag{C.14}$$

$$\widetilde{\alpha}_{b,t} + 2\widetilde{\beta}_{b,t} b_{t+1}, \tag{C.15}$$

and

$$\widetilde{\alpha}_{S,t} + \widetilde{\beta}_{S,t}\frac{1}{\varsigma_1} + \widetilde{\beta}_{S,t} S_{t+1}, \tag{C.16}$$

where (C.16) follows from the parametric assumption for the cost function $g(h', h) = \kappa(h')^{\varsigma_1} h^{\varsigma_2}$ and from the change of variable $S = \ln(h)$ to express the tax rate as a function of years of schooling S, implying $\frac{\partial t(h)}{\partial h} h = \frac{\partial t(S)}{\partial S}$.[47]

Analogously to the approximation of the linear taxes (and constant tax rates), we can then approximate the simple non-linear tax schedules by regressing the right-hand side of equations (C.11), (C.12), (C.13) for each generation t on a constant and $x = y_t, b_{t+1}, S_{t+1}$, respectively.[48] Given the recursive nature of equation (C.13), we use the approximated taxes in $t+1$ when approximating taxes in t, as before.

The estimated regression coefficients $\widehat{\alpha}_{x,t}$ and $\widehat{\beta}_{x,t}$, for each generation t, then allow us to identify the parameters of interest $\widetilde{\alpha}_{x,t}$ and $\widetilde{\beta}_{x,t}$, $x = y, b, S$ with the following system of equations:

$$\widehat{\alpha}_{y,t} = \widetilde{\alpha}_{y,t}, \tag{C.17}$$

$$\widehat{\alpha}_{b,t} = \widetilde{\alpha}_{b,t}, \tag{C.18}$$

$$\widehat{\alpha}_{S,t} = \widetilde{\alpha}_{S,t} + \widetilde{\beta}_{S,t}\frac{1}{\varsigma_1}, \tag{C.19}$$

$$\widehat{\beta}_{y,t} = 2\widetilde{\beta}_{y,t}, \tag{C.20}$$

$$\widehat{\beta}_{b,t} = 2\widetilde{\beta}_{b,t}, \tag{C.21}$$

$$\widehat{\beta}_{S,t} = \widetilde{\beta}_{S,t}. \tag{C.22}$$

[47] See Section 3 for the explanation why $S = \ln(h)$ in our model.

[48] The approximation of the linear tax with the cross-sectional average obtains if we only regress the right-hand sides on a constant.

D Numerical solution

We solve the problem of dynasties applying the endogenous gridpoint method, as explained in the online Appendix C of Koeniger and Prat (2018). The planner problem is solved building on the programs of Farhi and Werning (2013).

For the dynastic problem we use a grid of 75 points for bequests, 100 points for human capital, and 12 points for ability θ. Consistent with our interpretations of $\log(h)$ as non-compulsory schooling years in the Mincer wage regression, the lowest grid point of human capital is $\exp(0) = 1$. Note that the cost function thus implies a very small minimum expenditure for education over a 30-year period of \$600 for the calibrated parameter values.

We calibrate the problem of dynasties by minimizing $D = \sum_{i=1}^{k} \left(\frac{x_i^m - x_i^d}{x_i^d} \right)^2$ where x_i^m is the i-th moment generated by the model and x_i^d is the corresponding target moment in the data. We compute a moment x_i^m in the stationary distribution by simulating the model for 500,000 dynasties. To calibrate the model, we start by performing a global search on a parameter grid to minimize this expression. Based on the best parameters thus obtained, we then use the Nelder-Meade optimization algorithm to further improve on the fit of the model. Proceeding this way, we are able to reduce D close to zero to match the data targets well.

For the numerical solution of the planner problem, we follow Farhi and Werning (2013), pp. 614-615, and replace the state variables (V, Φ) with the multipliers (λ, γ). This has computational advantages because the domain of the multipliers is bounded below. Furthermore, conditioning the problem on λ allows us to solve the first-order conditions to determine the allocation and then obtain V by integrating once over the utility of that allocation. Similarly, conditional on γ, we can obtain Φ by integrating once. This speeds up the numerical solution because otherwise we would have to ensure, for example, that the chosen allocation integrates to V and computationally expensive integration potentially would have to be performed many times.

We choose a grid of 17 points for λ, 12 points for γ, 18 points for h, 25 points for θ_-, and 26 points for ϵ. We solve the planner problem for $T = 8$ iterations corresponding to 240 years, after which the solution for the value function approximates the steady state accurately. Intuitively, we need few iterations to achieve a good approximation of the steady state because the discount factor for a period length of 30 years is much lower than for a period length of one year. All programs are implemented in *julia*. On a standard processor of the current vintage, solving the problem for a given parametrization of the calibrated economy takes two minutes and solving the planner problem takes 30 hours.

E Validity of the first-order approach

We show that the first-order approach used to solve the planner problem is valid in our setting. We explain how we validate the first-order approach and then present the results of the validation.

Let us denote expenditure for schooling as e. If a history-dependent allocation $\mathcal{A}_t = (c_t, y_t, e_t)$ is dynamically incentive compatible, it is a function $\mathcal{A}_t(\lambda_t, \gamma_t, h_t, \theta_{t-1}, \theta_t)$ where λ_t and γ_t are multipliers used as in Appendix D as auxiliary state variables in the recursive planning problem, h_t is human capital, and θ_t and θ_{t-1} are current and last period's ability. This follows from results in Fernandes and Phelan (2000) and Kapička (2013).

Incentive compatibility requires that reports are truthful. We now show how we verify incentive compatibility of the allocation that has been obtained by applying the first-order approach in the planning problem. The planner's problem results in the following mappings and depends on the initial state, where a "^" denotes a reported value so that $\hat{\theta}$ denotes reported ability and θ denotes true ability:

- allocation $\mathcal{A}_t(\lambda_t, \gamma_t, h_t, \hat{\theta}_{t-1}, \hat{\theta}_t)$ for all $t = 1 \dots T$,
- auxiliary transition function $(\lambda_t, \gamma_t, h_t, \hat{\theta}_{t-1}, \hat{\theta}_t) \mapsto (\lambda_{t+1}, \gamma_{t+1}, h_{t+1})$,
- initial state $(\lambda_1, \gamma_1, h_1, \theta_0)$ which is common knowledge (hence, $\hat{\theta}_0 = \theta_0$).

We set up the reporting problem to verify whether the allocation described by these mappings is indeed incentive compatible. In the reporting problem, the dynasty reports a sequence of abilities $\hat{\theta}^T$ so that utility from allocation $\mathcal{A}$ given by

$$\sum_{t=1}^{T} \mathbb{E}_0 \beta^t [\mathbf{U}(c_t, y_t / A(h_t, \theta_t))]$$

is maximized. Given the mappings obtained from the planner's problem and the initial state, all (endogenous) state variables are determined by the reports $\hat{\theta}^T$.

For given state variables λ_t, γ_t, h_t, the dynasty's utility only depends on current ability θ_t and the previous period's reported ability $\hat{\theta}_{t-1}$ and not on the entire history of abilities.

Therefore, the recursive formulation of the reporting problem is:

$$
\begin{aligned}
v_t(\lambda_t, \gamma_t, h_t, \hat{\theta}_{t-1}, \theta_t) = & \max_{\hat{\theta}_t \in \Theta} \mathbf{U}(c_t, y_t/A(h_t, \theta_t)) + \beta \mathbb{E}[v_{t+1}(\lambda_{t+1}, \gamma_{t+1}, h_{t+1}, \hat{\theta}_t, \theta_{t+1}) | \theta_t] \qquad \text{(E.1)} \\
& \text{s.t.} \\
& \quad (c_t, y_t, e_t) = \mathcal{A}_t(\lambda_t, \gamma_t, h_t, \hat{\theta}_{t-1}, \hat{\theta}_t), \\
& \quad (\lambda_t, \gamma_t, h_t, \hat{\theta}_{t-1}, \hat{\theta}_t) \mapsto (\lambda_{t+1}, \gamma_{t+1}, h_{t+1}), \\
& \text{for all } t = 1, \ldots, T \text{ with } v_{T+1}(\cdot) = 0.
\end{aligned}
$$

If the solution $\hat{\theta}^T$ to (E.1) is equal to θ^T, then the allocation obtained in the planning problem by applying the first-order approach is incentive compatible.

In the numerical implementation, the following numerical issues may cause $\hat{\theta}^T$ to differ from θ^T. First, the policy functions of the planner problem are computed on a grid and are therefore not exact. Second, we use linear interpolation both to solve and simulate the economy which adds numerical error. Third, the planner problem is solved using ordinary differential equations based on the density functions of the stochastic process for ability whereas the dynastic (reporting) problem uses a discretized transition matrix to approximate the stochastic process. These issues may lead to numerical differences between the solution of the planner problem and the reporting problem so that we cannot expect to verify truthful reporting exactly.[49]

Table E.1 presents the results of the numerical comparison of the allocation of the planner's problem with the allocation implied by the reporting problem. We compare the dynasty's welfare attained by optimal reporting to the welfare attained by truthful reporting in the planner problem. Table E.1 displays both the maximal and the mean deviation across three alternative measures in a simulation of 500,000 dynasties: the difference in utils $V - \hat{V}$, the relative difference in utils $(V - \hat{V})/|V|$, and the relative increase in consumption $\bar{C}$ across all time periods and states which is required for welfare based on truthful reporting to equal welfare based on optimal reporting.

The statistics in Table E.1 show that the allocation based on truthful reporting is very close to the allocation implied by optimal reporting. Our preferred measure in terms of consumption equivalents $\bar{C}$ is at most 0.05% and smaller than 0.01% on average for our simulated dynasties. We conclude that the first-order approach is valid in our environment. The small differences between the allocations have the order of magnitude which

[49] A further challenge is that Tauchen's discretization of the stochastic process with compact support may imply that the transition matrix is not consistent with the set of attainable points implied by the support of the density function. Such inconsistencies matter for the reporting problem because reports not consistent with the support of the density function in the planner problem could have positive probability in the transition matrix of the reporting problem. We account for this by forcing the transition matrix to map only into points that are contained in the support of the density function.

Statistic	max	mean
$\hat{V} - V$	0.0008	< 0.0001
$(\hat{V} - V)/\lvert V \rvert$	0.08%	$< 0.01\%$
$\bar{C}$	0.05%	$< 0.001\%$

Table E.1: Numerical validation of the first-order approach

Notes: $\hat{V}$ is the value based on optimal reporting, V is the value based on truthful reporting, and

$$\bar{C} \equiv \exp\left((\hat{V} - V)\frac{1-\beta}{1-\beta^{T+1}}\right) - 1,$$

where we use that utility derived from consumption is logarithmic and account for the finite number of iterations T in the numerical solution that we use to approximate the steady state. In accordance with the solution of the planner problem, discussed in Appendix D, we set $T = 8$ corresponding to 240 years also for the reporting problem. Statistics are calculated for 500,000 simulated dynasties at the time of reform.

one would expect because of numerical approximation error.

F Implementation of the reform to attain the social optimum

In this section we describe how we construct the reform to attain the social optimum starting from the steady state of the calibrated economy. We first simulate the calibrated economy for $M = 500,000$ households. We approximate the stationary distribution by simulating the economy for 100 generations. We label the steady-state of the calibrated economy as $t = -1$. The planner proposes the reform at $t = 0$. Note that the reform is proposed after consumption and human capital investment decisions by the parents have been made but before their childrens' types are realized and therefore before the children make their decisions.

To focus on the implications of the reform for insurance, we abstract from redistribution between dynasties by holding constant the present discounted value of net costs of each dynasty's allocation. The reported welfare gains are thus not confounded by wealth effects. We now describe in more detail how we implement the reform.

Resource constraint.— In the calibrated economy, households pay a positive amount of net taxes to the government. These taxes can be thought of being used to finance an exogenous stream of government expenditures. In the calibrated economy, the net present value of government expenditures amounts to 27 percent of average labor earnings and the flow of government expenditures equals 16 percent of average labor earnings (Heathcote et al., 2017). Since we do not model the expenditure side of the government, we assume that the planner has to continue to raise the amount of resources required for these government expenditures. In other words, the net government surplus in the reformed economy equals the surplus in the calibrated economy so that the planner does not have more resources available due to some arbitrary assumption about a change in the size of that surplus in the reform. Each dynasty in state $s = (b, h, \theta)$ at the time of the reform has to contribute the same amount of net-taxes $t(s)$ as in the calibrated economy. Then the planner's allocation satisfies for each state

$$\widehat{\Gamma}_\eta(\lambda(s), 0, h(s), \theta) = (1+r)b(s) + t(s), \tag{F.1}$$

where $\widehat{\Gamma}_\eta(\cdot)$ is the expected cost for the planner of providing the allocations for the dynasties conditional on the state variables of the relaxed planner problem derived in Appendix A. Note that $\widehat{\Gamma}_\eta(\cdot)$ is the cost function for any value of η once the state variables (V, Φ) have been replaced by their multipliers (λ, γ), analogous to Farhi and Werning (2013). The multiplier $\gamma(s) = 0$ for all dynasties in state s because, apart from the utility promise and the parents' ability, there is no further restriction from history so that the threat-keeping constraint is not binding in the reform period.

The remaining degree of freedom in equation (F.1) is η_0 at the time of the reform which depends on the resources available to the planner. We set $\eta_0(s) = \lambda_0(s)$ which, by the envelope condition $\partial\Gamma/\partial V = \lambda - \eta$ obtained in Appendix A, implies that a marginal variation of the promised utility V leaves the planner's cost for the allocation of a dynasty in state s unchanged. Equation (A.21) and the subsequent discussion in Appendix A imply that the unconditional expectation $\mathbb{E}[\lambda_t]$ then is constant across t after the reform, given that $\psi = q$ in our calibration.

Assets.— Given that the socially optimal allocation does not determine dynasties' assets or bequests, we briefly mention how we compute them after the reform. If one interprets assets as the difference between the net present value of expenditures and the net present value of earnings, as usually done in the literature, the counterpart of assets in the planner's problem is the expected present value of net costs $\tilde{\Gamma}(s)$ for providing an allocation.

G Further results and robustness analysis

y_t / y_{t+1}	Quintiles 1	2	3	4	5
1	0.41	0.25	0.18	0.12	0.05
2	0.24	0.24	0.22	0.18	0.11
3	0.16	0.21	0.22	0.22	0.19
4	0.11	0.17	0.21	0.24	0.27
5	0.08	0.13	0.17	0.24	0.38

Table G.1: Earnings quintile transition matrix in the steady state of the calibrated economy

Notes: The probabilities across columns in each row may not add up to 1 because of rounding.

Welfare quintile	Increase in years of schooling	Cost of additional schooling	Increase in ability (in standard deviations)
1	0.89	16,960	0.09
2	0.66	12,962	0.10
3	0.55	11,372	0.10
4	0.48	10,247	0.11
5	0.40	9,252	0.12

Table G.2: Average increase in schooling and ability that is welfare equivalent to receiving an additional $10,000 as bequests, by welfare quintile

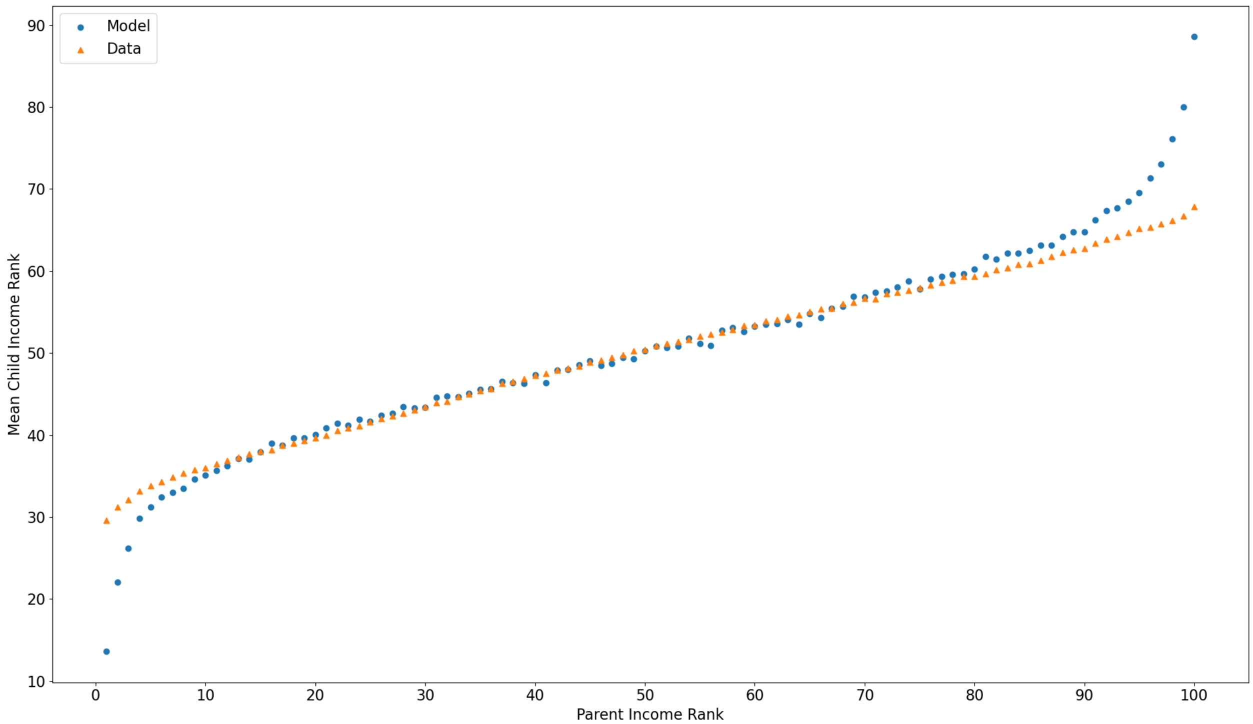

Figure G.1: Association between the rank of children and parents in the income distribution. *Notes:* The data points are taken from the left panel of Figure 2 in Chetty et al. (2014). The model predictions are based on the earnings distribution in the steady state of the calibrated economy. The figure shows the mean child percentile rank within each parent percentile rank bin.

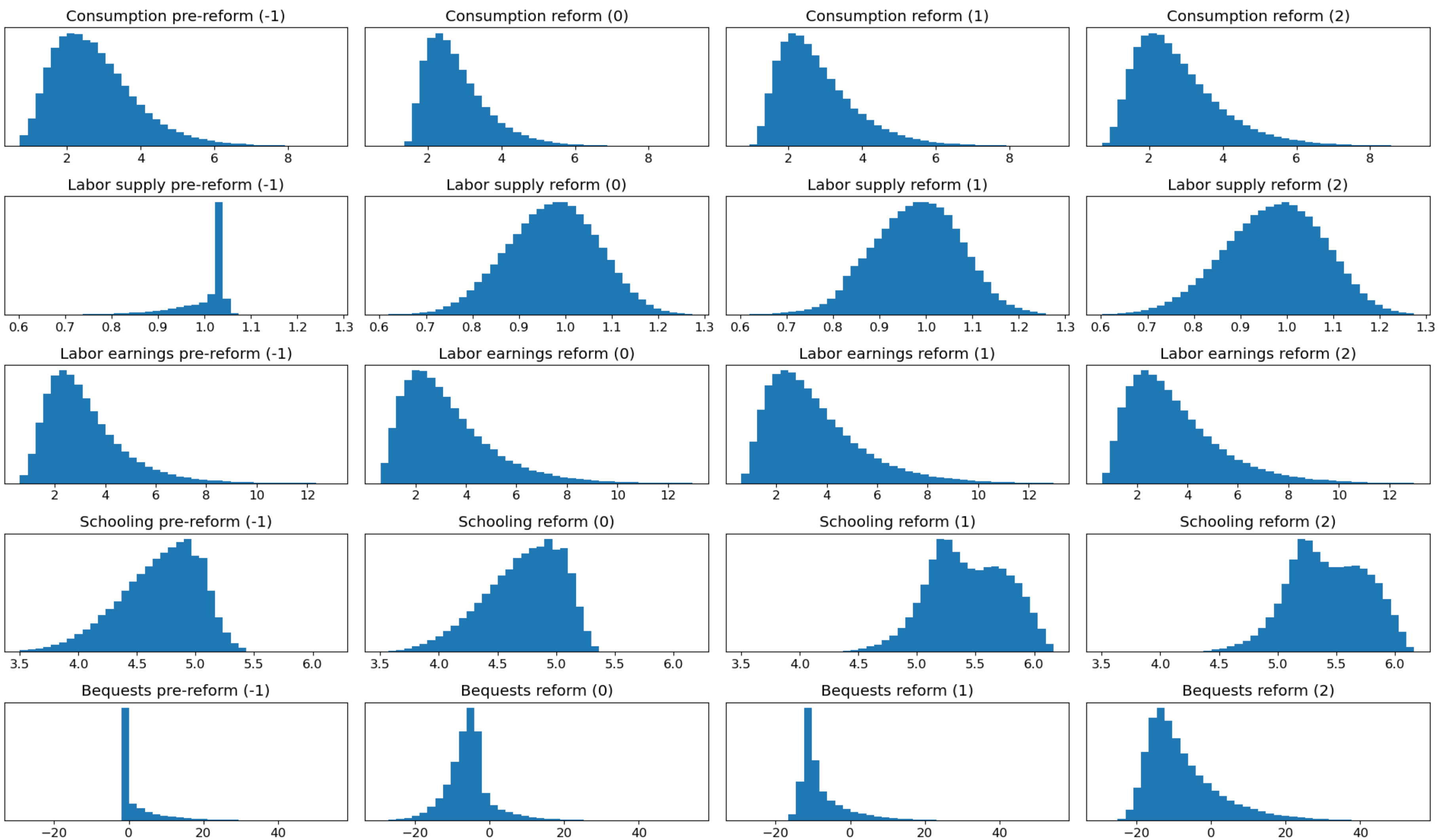

Figure G.2: Transition to the social optimum: evolution of distributions. *Notes:* The reform is implemented before generation 0 makes its choices. Units of monetary variables are in 10,000 US-$ and labor supply is normalized by the average level in the calibrated steady state. The changes of the means of the distributions are discussed in Section 5.

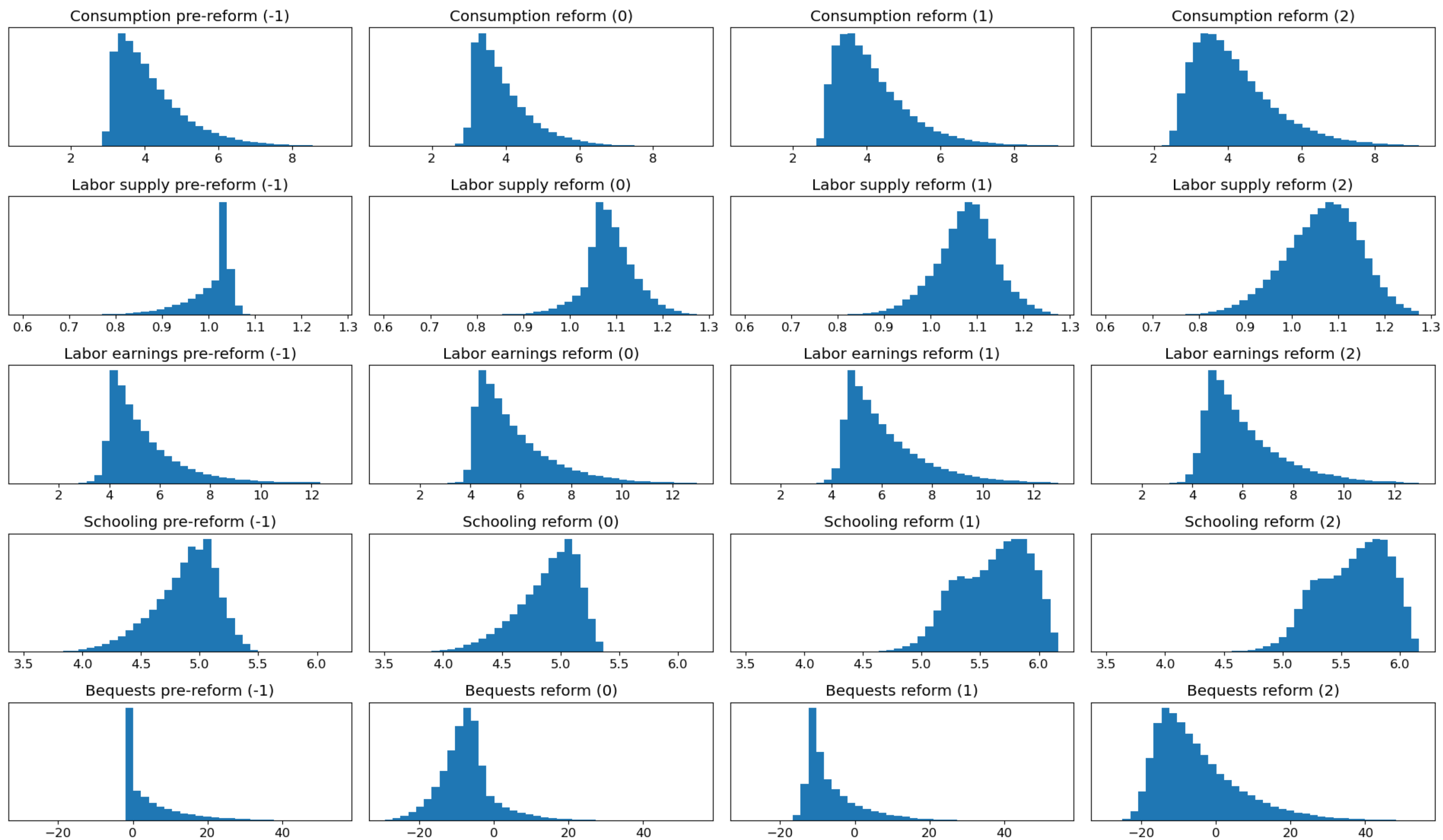

Figure G.3: Transition to the social optimum: evolution of distributions for top quartile of ability distribution in each generation.
Notes: See Figure G.2.

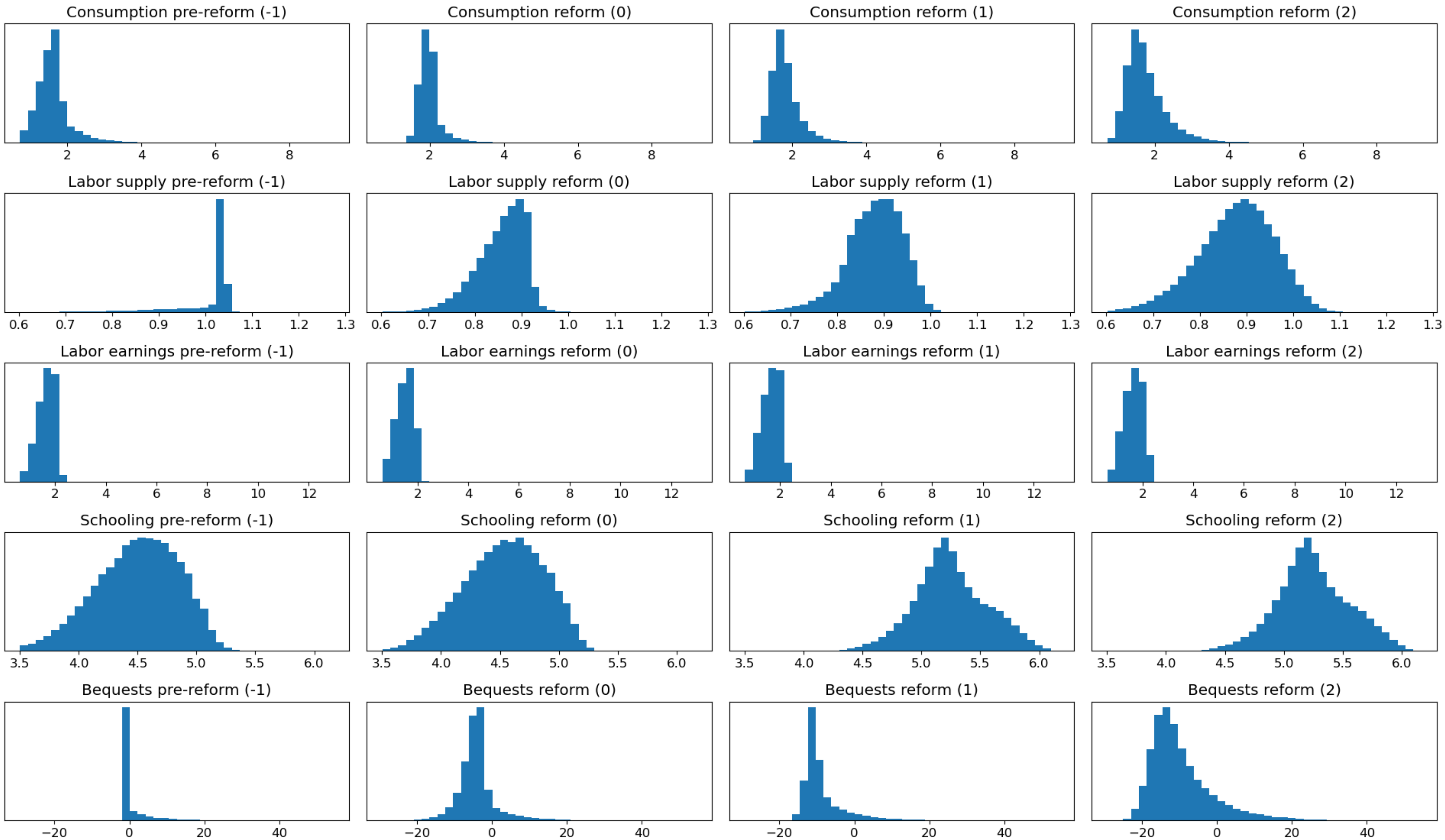

Figure G.4: Transition to the social optimum: evolution of distributions for bottom quartile of ability distribution in each generation. *Notes:* See Figure G.2.

Figure G.5 shows that the path of intergenerational earnings mobility is similar across three commonly used measures of earnings mobility: the intergenerational earnings elasticity (IGE), the intergenerational correlation of earnings (IGC) and the rank-rank correlation of earnings (RRC). As discussed in Mazumder (2016), for example, the mobility measured by the IGE and the IGC or the rank-rank correlation differ if the standard deviation of log earnings σ_y changes. In particular, the $\text{IGE}_t = \sigma_{y_t}/\sigma_{y_{t-1}}\ \text{IGC}_t$. Analogous to the empirical literature, we define (relative) mobility at time t as the earnings mobility of children in t relative to earnings of parents in $t-1$.

Figure G.5 shows that the initial spike of earnings mobility on impact ($t = 0$) of the reform is largest for the IGE but the change between mobility in the steady state of the calibrated economy ($t = -1$) and in the second generation after the reform ($t = 1$) is quantitatively small and similar across all three measures.

Let us elaborate on the non-monotonic path of earnings mobility, for concreteness based on the IGE measure. Recall that the $\text{IGE}_t = \sigma_{y_t,y_{t-1}}/\sigma^2_{y_{t-1}}$ where $\sigma_{y_t,y_{t-1}}$ is the intergenerational covariance of log earnings. On impact of the reform ($t = 0$), the earnings mobility increases because $\sigma^2_{y_{t-1}}$ is given whereas $\sigma_{y_t,y_{t-1}}$ increases. As visible in Figure G.2 above, the increase in $\sigma_{y_t,y_{t-1}}$ is driven by the higher dispersion of labor supplied which is correlated with the ability of parents in the previous generation. In the following periods, once both the variance of parents' earnings and the intergenerational covariance of earnings have increased after the reform, the IGE falls slightly relative to its value before the reform. As the economy approaches the new steady state, the IGE then increases to a similar value as prior to the reform.

In the tables in the main text, we report the IGE in the second generation after the reform ($t = 1$) because for this generation also the earnings of the parents have been affected by the reform. Thus, both the earnings variance of the parents and the intergenerational covariance between parents and children, used for computing the IGE, are then measured after the reform. We also report the values of the IGE in the steady state. As Figure G.5 shows, the IGE at both points in time summarizes the transition path quite well noting that changes in the mobility measures after the reform are quantitatively small if one abstracts from the mechanical spike in the mobility measures on impact.

The decomposition of the IGE presented in Table 6 in the main text shows that the changes in the IGE, after the initial mechanical effect, quantitatively depend on how much the higher dispersion of labor effort increases the variance of labor earnings relative to the intergenerational covariance of earnings.

The variance of labor earnings depends on the correlation between ability and labor earnings within a generation which is higher in the second generation after the reform relative to the steady state (see column (3) of Table 5 in the main text). Table 6 shows

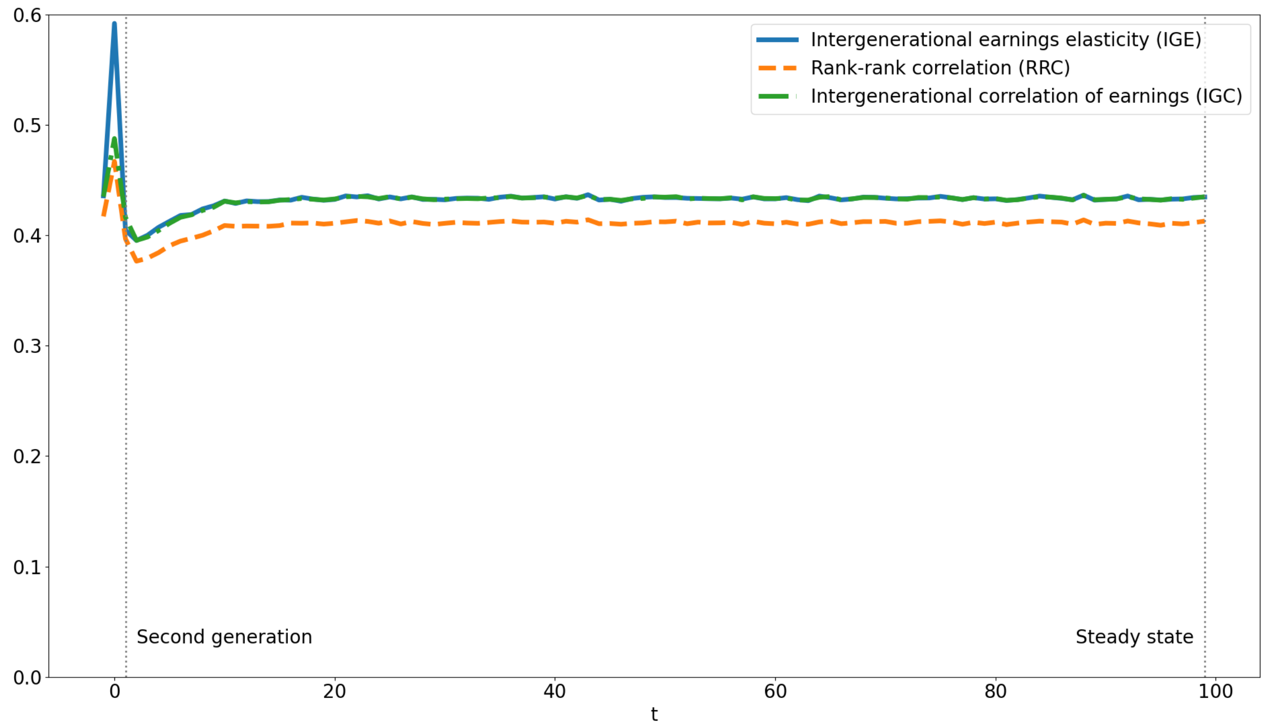

Figure G.5: Transition paths for different measures of intergenerational mobility after the reform. *Notes:* The figure displays the transition path for three measures of intergenerational mobility: the intergenerational earnings elasticity (IGE), the intergenerational correlation of earnings (IGC) and the rank-rank correlation of earnings (RRC).

that the larger covariance of ability with labor effort within a generation (determining $\mathcal{V}_3$) indeed contributes more to the higher variance of earnings after the reform.

At the same time, the higher intergenerational covariances between labor effort and ability (determining $\mathcal{C}_3$ and $\mathcal{C}_6$) contribute more to the higher intergenerational covariance of earnings. These contributions are smaller in the steady state than in the second generation after the reform.

In the new steady state, nurture provides relatively more insurance against ability risk so that also the intergenerational covariance of labor effort ($\mathcal{C}_3$) is higher. Quantitatively, it turns out that the IGE thus increases slightly after the third generation after the reform as the economy approaches the new steady state.

G.1 Robustness analysis

This subsection presents the results of the robustness analysis. Table G.3 shows that the results on the evolution of insurance and mobility for our benchmark calibration, presented in Table 5 in the main text, are robust across most alternative calibrations. To be concise, we focus on the pass-through coefficient of ability shocks to consumption and the rank-rank correlation between ability and welfare. For convenience, we repeat the results of the benchmark case in the first column of Table G.3.

Worth noting is that the dependence of welfare on nurture, in terms of the correlation between ability and welfare, decreases and insurance increases in the calibrated economy if we target the conditional mean of bequests in column (2). This calibration implies more bequests than in the benchmark calibration so that the ability shock can be smoothed more by dynasties.

When we recalibrate the Frisch elasticity of labor supply to a higher value in column (4), the pass-through coefficient first falls after the reform and then increases to a higher value in the steady state of the social optimum. This illustrates how a higher Frisch elasticity changes the trade-off between insurance and incentives in the social optimum.

Table G.4 provides details on the results of the calibration for each of the considered robustness checks, in terms of the recalibrated parameter values and the implied target statistics. We now provide further information for each of the robustness checks.

Mean bequests as target.— We target the mean bequests of households that received a bequest. As before, we convert the mean bequest of \$408,400 for households, reported in Table 2 of Wolff and Gittleman (2014), into adult equivalents dividing by 1.4 so that our target is \$291,714. As shown in the second column of Table G.4, the calibration matches this target quite closely.

Lower intergenerational elasticity of earnings as target.— We recalibrate the model if we target a lower intergenerational earnings elasticity $\iota = 0.3$, which is at the low end of estimates reported in Table 1 of Chetty et al. (2014). The third column of Table G.4 shows that the recalibrated model continues to match the target statistics closely. Furthermore, the model-implied mobility matrix, displayed in Table G.5, matches more closely the estimated transition matrix reported in Table 2 of Chetty et al. (2014). The match of the rank-rank correlation of income also improves, as mentioned in the main text of Section 3.

	(1) *Benchmark*	(2) *Mean bequests as target*	(3) *Lower IGE as target*	(4) *Higher Frisch elasticity*	(5) *Higher complementarity between θ and h*
Pass-through coefficient					
Calibrated economy	0.67	0.41	0.65	0.68	0.72
Social optimum					
2^{nd} generation after reform	0.58	0.40	0.57	0.66	0.58
Steady state after reform	0.63	0.74	0.63	0.74	0.61
Rank-rank corr(ability,welfare)					
Calibrated economy	0.90	0.48	0.88	0.91	0.95
Social optimum					
2^{nd} generation after reform	0.36	0.20	0.25	0.39	0.38
Steady state after reform	0.22	0.11	0.16	0.24	0.26

Table G.3: Robustness of results for insurance and mobility

	(1)	(2)	(3)	(4)	(5)
	Benchmark	*Mean bequests as target*	*Lower IGE as target*	*Higher Frisch elasticity*	*Higher complementarity between θ and h*
Recalibrated parameters					
Discount factor (annualized) β	0.966	0.970	0.966	0.966	0.963
Persistence ρ	0.44	0.44	0.32	0.47	0.44
Education cost parameter κ	0.0014	0.0015	0.0014	0.0017	0.0016
Education cost parameter ς_1	0.75	0.77	0.81	0.71	0.71
Education cost parameter ς_2	-0.0004	-0.0356	-0.0760	0.0132	0.0177
Predictions for target statistics					
Bequests	51,595	285,517	50,680	51,774	51,459
Average years of (non-compulsory) schooling S	4.76	4.73	4.79	4.74	4.91
Correlation(S', S)	0.47	0.46	0.46	0.46	0.46
Intergenerational earnings elasticity	0.44	0.44	0.30	0.45	0.45
Average net cost of an additional year of schooling	13,858	13,668	13,664	13,739	13,838

Table G.4: Calibration results for robustness checks

	Quintiles				
y_t / y_{t+1}	1	2	3	4	5
1	0.34	0.24	0.19	0.15	0.08
2	0.23	0.22	0.21	0.19	0.14
3	0.17	0.2	0.21	0.21	0.2
4	0.14	0.18	0.2	0.23	0.26
5	0.12	0.15	0.18	0.22	0.32

Table G.5: Earnings quintile transition matrix, in the steady state of the calibrated economy, targeting an intergenerational elasticity of earnings of 0.3

Notes: The probabilities across columns in each row may not add up to 1 because of rounding.

Higher Frisch elasticity.— We recalibrate the model for a larger Frisch elasticity of 0.86 for aggregate hours, reported in Table 2 of Chetty et al. (2013). This elasticity is based on micro-estimates from quasi-experimental studies and contains responses of hours at the intensive and extensive margin. We thus set $\alpha = 2.16$. The fourth column of Table G.4 shows that in the recalibrated model persistence of the ability shock slightly increases to match the intergenerational earnings elasticity. This quantitative result obtains because labor supply is not only a function of ability but also of bequests and human capital. If labor supply instead were a power function of ability, then one can show that a higher Frisch elasticity would only affect the variance of log income but not the persistence of log income across generations. Further note that ς_2 is calibrated to be positive, implying that families with lower human capital have a cost advantage for educating their children (which could be due to lower opportunity costs, for example).

Higher complementarity between human capital and ability.— We recalibrate our model to match the complementarity between years of schooling and ability, as suggested by findings in the second row of Table 3 in Cunha et al. (2006). Using test results of the AFQT as a measure of ability, they find that the return to one year of college in percent at the 95-th percentile of the ability distribution is 1.6 times higher than the return at the 5-th percentile of the ability distribution (and not constant across the ability distribution as implied by the Cobb-Douglas assumption for productivity). We use this target to calibrate the elasticity of substitution χ between ability θ and human capital h in the function for productivity. Because wages are no longer log-separable in ability and years of schooling if productivity is not Cobb-Douglas, we also calibrate the variance of the innovation of the ability process and the parameter ξ using the variance of residual wages of 0.2, obtained from a Mincer wage regression on the years of schooling, and the return to schooling of 0.1 as target statistics.[50] The calibration of these parameters is done jointly with the other parameters reported in Table G.4.

Our recalibration results in $\chi = 0.786$, $\sigma_\epsilon^2 = 0.276$ and $\xi = 0.73$ compared with $\chi = 1$, $\sigma_\epsilon^2 = 0.2$ and $\xi = 0.9$ in the benchmark calibration. Note that σ_ϵ^2 is approximately equal to the variance of residual wages in the benchmark case because the parameter values imply that $1-\rho^2 \approx \xi^2$. The additional targets are matched well by the recalibrated model: the return to the first year of college at the 95-th percentile of the ability distribution is 1.6 times higher than the return at the 5-th percentile of the ability distribution, the variance of residual wages is 0.2 (both statistics equal the respective target up to three digits of precision), and the return to schooling is 0.09. The fifth column of Table G.4 shows that the recalibrated model also continues to match the other targets reasonably well where ς_2 is calibrated to be positive, as in the robustness check with the higher Frisch elasticity.

Table G.6 displays the results for the average taxes and subsidies that incentivize the optimal insurance for the recalibrated economy with higher complementarity between human capital and ability. Compared with Table 9 for the benchmark, the level of tax and subsidy rates increases. The progressivity of bequest subsidies across the earnings distribution, which implies that bequest subsidies are targeted relatively more to the earnings poor, is robust for the recalibrated economy with a higher complementarity.

[50] Given that we observe ability in our simulated data based on the model, we obtain the model counterpart for the unbiased empirical estimate of the return to schooling by controlling for ability in the regression. In the empirical literature, identification of the return to schooling is based on a instrumental-variable regression because ability is not fully observable and correlated with schooling.

	Earnings quintile				
	1	2	3	4	5
Bequest subsidy	0.59	0.53	0.50	0.46	0.40
Schooling subsidy	0.63	0.60	0.60	0.60	0.59
Labor income tax	0.50	0.50	0.48	0.46	0.39
	Bequest quintile				
	1	2	3	4	5
Bequest subsidy	0.42	0.46	0.52	0.58	0.49
Schooling subsidy	0.60	0.60	0.61	0.61	0.59
Labor income tax	0.47	0.47	0.47	0.47	0.44

Table G.6: Average rates of subsidies or taxes for bequests, schooling, and labor per quintile, in the recalibrated economy with a higher complementarity between θ and h.

Note: Tax and subsidy rates are reported for the social optimum in the second generation after the reform.