



Universität St.Gallen

Opportunity and Inequality across Generations

Winfried Koeniger and Carlo Zanella

February 2020 Discussion Paper no. 2020-03

Editor: Vanessa Pischulti
University of St.Gallen
School of Economics and Political Science
Department of Economics
Müller-Friedberg-Strasse 6/8
CH-9000 St.Gallen
Phone +41 71 224 23 07
Email seps@unisg.ch

Publisher: School of Economics and Political Science
Department of Economics
University of St.Gallen
Müller-Friedberg-Strasse 6/8
CH-9000 St.Gallen
Phone +41 71 224 23 07

Electronic Publication: <http://www.seps.unisg.ch>

Opportunity and Inequality across Generations¹

Winfried Koeniger and Carlo Zanella

Author's address:

Prof. Dr. Winfried Koeniger
Swiss Institute for Empirical Economic Research (SEW)
University of St.Gallen
Varnbuelstrasse 14
CH-9000 St.Gallen
Phone +41 71 224 2309
Email winfried.koeniger@unisg.ch
Website www.sew.unisg.ch/en

¹ Winfried Koeniger: University of St.Gallen (Swiss Institute for Empirical Economic Research), CESifo, Center for Financial Studies and IZA, winfried.koeniger@unisg.ch. Carlo Zanella: Department of Economics, University of Zurich, Schönberggasse 1, 8001 Zurich, Switzerland, carlo.zanella@econ.uzh.ch.

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

Abstract

We analyze inequality and mobility across generations in a dynastic economy. Nurture, in terms of bequests and the schooling investment into the next generation, is observable but the draw of nature in terms of ability is hidden, stochastic and persistent across generations. We calibrate the model to U.S. data to illustrate mechanisms through which nurture and nature affect mobility and the transmission of income inequality across generations, thus complementing the vast empirical literature. To provide a benchmark for the observed status quo, we solve for the social optimum in which the planner weighs dynasties equally and chooses optimal tax schedules subject to incentive compatibility. Analyzing the transition from the calibrated steady state to this social optimum, we find that insurance against intergenerational ability risk increases on the transition path by making welfare of family dynasties more dependent on nurture relative to nature. The insurance comes at the cost of less social mobility. We compare welfare in the social optimum and economies with a simple history-independent tax and subsidy system.

Keywords

Human capital, schooling, bequests, asymmetric information, intergenerational mobility, inequality

JEL Classification

E24, H21, I24, J24, J62

1 Introduction

Inequality across generations is transmitted through parents' nurture and nature. A central question is how much the opportunities of each generation depend on nurture and nature. A tightly related question is whether, and how, optimal policy should change the observed patterns.

To shed light on these issues, we analyze inequality and opportunity across generations in an economy with family dynasties calibrated to the U.S. The transmission of inequality across generations in the model is influenced by nurture, in terms of bequests and schooling investment, and nature, in terms of ability that is private information, stochastic and persistent across generations. Mobility within a generation, resulting from the stochastic changes in ability, indicates the extent of opportunities available to each generation.

Based on the model, we make three contributions. Firstly, we use the calibrated model to complement the vast empirical literature on inequality and mobility by illustrating mechanisms through which nurture and nature affect mobility and the intergenerational transmission of inequality. Secondly, we analyze how mobility and the transmission of inequality across generations would change if, starting from the calibrated steady state, a reform implemented the social optimum. Thirdly, we compare welfare of economies with simple tax and subsidy systems to welfare in the social optimum, which, in theory, requires complex history-dependent tax schedules for its implementation as in Farhi and Werning (2010). To the best of our knowledge, such analysis of simple tax systems has not been provided yet in the context of intergenerational models with persistence in unobservable ability. We now elaborate further on each of the three contributions.

Concerning the intergenerational transmission of inequality, we find important differences in how bequests and schooling affect this transmission in our calibrated model. Bequests or inheritances decrease the incentive to exert labor effort through a negative wealth effect and thus induce mean reversion in labor earnings.¹ Investment into schooling instead increases the persistence in labor earnings across 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. Quantitatively, we find that the transmission of inequality is mostly determined by nature (ability) rather than nurture (bequests or human capital investment),

¹Such 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. (2018) on the consequences of this wealth effect for labor income taxation. The wealth effect also implies less investment into schooling because schooling results in more resources only if combined with labor effort. Although the negative wealth effect dominates on average, we find that more wealth relaxes borrowing constraints for some dynasties and thus increases their investment into schooling.

in line with recent empirical findings of Bingley et al. (2018) who identify the effects of nature and nurture with a credible “Children of Twins” design.²

Our analysis of the intergenerational transmission of inequality and mobility in the model calibrated to the U.S. complements the empirical evidence provided by Chetty et al. (2014) for the U.S.³ We highlight economic mechanisms, as discussed above, that generate the empirically observed patterns and we quantify the trade-off between intergenerational insurance and mobility. In the analysis of the reform that implements the social optimum, more income inequality does not make dynasties worse off *ex ante* because the reform also increases intergenerational insurance.

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 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).

In their intergenerational model with estate taxation calibrated to the U.S., De Nardi and Yang (2016) 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. We complement this evidence by showing 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 welfare terms instead.

We characterize the social optimum as the solution of a dynamic Mirrleesian problem, in which asymmetric information constrains the insurance that the planner can provide against the ability risk that dynasties face. This makes the quantitative comparison be-

²The 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.

³Güell et al. (2018) provide evidence for Italy, Markussen and Røed (2020) for Norway and Adermon et al. (2018) for Sweden. See also their references for further literature.

tween the calibrated economy and the social optimum non-trivial. Inequality and less than full insurance are also a feature of the social optimum, which provides a useful benchmark to put the observed mobility and transmission of inequality in perspective.

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, allows for a meaningful analysis of the transition from the steady state of the calibrated economy to the social optimum. Because the wedge between the discount rates is an important determinant of the distribution in the social optimum, we explain in Section 5 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.

In our quantitative analysis of the reform and the subsequent transition to the social optimum, we find that the importance of nurture (in terms of bequests and human capital investment) increases relative to nature (in terms of ability) compared with the steady state of the calibrated economy. We show that this implies more intergenerational insurance against ability risk at the cost of less social mobility, in the sense that ability is less important for the rank in the welfare distribution within each generation. We quantify this trade-off between insurance across generations and social mobility in the transition to the social optimum, in which the extent of intergenerational insurance and mobility is determined by optimal incentive provision as pointed out by Phelan (2006).

We find that production is decoupled more from consumption after the reform on the transition to the social optimum. The pass-through coefficient of an unexpected ability shock to consumption decreases implying more intergenerational insurance. At the same time, the correlation of ability and labor effort increases in the social optimum compared to the steady state calibrated to the U.S., and income mobility across and within generations remains quite stable after the reform on the transition to the social optimum.

At the time of the reform, we hold constant the present discounted value of the expected net costs for the allocation of each dynasty and give equal weight to dynasties in

the planner's objective function. We thus focus on insurance and refrain from adding motives for redistribution across dynasties by applying different weights in the 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.

Our analysis illustrates the general point, made in Lee and Seshadri (2018) for example, that trade-offs arise depending on whether equality of opportunity is considered from a dynastic point of view, tilting the balance towards more insurance across generations at the cost of less social mobility within generations, or from the point of view of individual families within a generation, tilting the balance towards more mobility within a generation and thus less intergenerational insurance.⁴ Interpretation of the empirical evidence on mobility patterns and inequality provided in Chetty et al. (2014) for the U.S., Güell et al. (2018) for Italy, Markussen and Røed (2020) for Norway or Adermon et al. (2018) for Sweden, for example, thus requires assumptions about the social welfare function. The welfare function used in this paper contains a dynastic motive which is disciplined by calibrating the discount factor of dynasties to match the empirically observed size of bequests.

Concerning the comparison of economies with simple tax and subsidy systems to the social optimum, we find that about half of the welfare gains of moving from the *laissez faire* to the social optimum can be achieved in economies with linear taxes and subsidies that condition on current labor income, bequests and schooling. We find that bequests and schooling are both subsidized but at different rates. 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.⁵ In the tax and subsidy system with optimal linear rates, 18% of labor income is taxed and bequests and schooling are subsidized at rates of 36% and 30%, respectively.

Our results show to which extent the endogenous state variables bequests and human capital together with labor income, which depends on the ability draw of the current generation, can capture the effect of history in our calibrated economy with persistent ability shocks. Albanesi and Sleet (2006) have shown that history can be fully summarized by conditioning optimal taxes on the endogenous state variable if shocks to unobserved

⁴See Arneson (2018) for a recent discussion of different interpretations of equal opportunity.

⁵Farhi 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.

ability are i.i.d. and thus not persistent.⁶ 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.⁷

Our analysis also relates to the large strand of literature on optimal taxation of 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. 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) and Stantcheva (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. 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 approach in our paper is to compare the status quo in the U.S., using approximations of existing tax schedules, with a social optimum that does not imply immiseration in the long run, and to analyze approximations of the social optimum with simple history-independent tax schedules. The focus of this paper is thus different from Koeniger and Prat (2018), who characterize socially optimal taxation of bequests and human capital if the planner and the dynasties apply the same weight to future generations, implying immiseration in the long run as in Atkeson and Lucas (1992).

Our analysis proceeds in the following steps. In Section 2, we model the decision problem of family dynasties. We then calibrate the model to U.S. data in Section 3. In Section 4, we provide results for inequality, mobility and the importance of parental background in this economy. To benchmark these results, we characterize the planner problem in Section 5. We compute the constrained-efficient social optimum and analyze in Section 6 how a tax reform changes inequality and mobility on the transition to the socially optimal steady state. In Section 7, we compare economies with simple history-independent tax schedules to the social optimum. We discuss the properties of these schedules and conclude in Section 8. 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.

⁶Stantcheva (2017) has extended this result in a life-cycle model with assets and human capital.

⁷Given 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).

2 The model

We build on the dynasty model of Koeniger and Prat (2018) to understand 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 is deliberately simpler than the model by Lee and Seshadri (2019) who analyze the sequence of decisions over the life cycle for each generation in more detail. The simplicity of our model keeps the problem tractable when we solve for the social optimum. This allows us make two of our contributions in this paper: to analyze how the transmission of intergenerational inequality changes on the transition from the calibrated steady state to the social optimum; and to characterize simple 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 :

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

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 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.⁸

⁸We 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

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 $\Theta \equiv [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}_+$, where we assume a continuously differentiable distribution $F : \Theta \rightarrow [0, 1]$ 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 $|\partial f(\theta_t | \theta_{t-1}) / \partial \theta_{t-1}| \leq B$ 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. When we quantify the effect of nature in this paper, the effect 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

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

where β is the discount factor of the family measuring the strength of the altruism towards future generations, $0 \leq \rho \leq 1$ 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.

human capital investments does not provide information about ability.

3 Calibration

In this section, we calibrate the model of family dynasties to the U.S. economy. We will use the calibrated economy to characterize inequality, mobility and the effect of parental background in Section 4. In Section 6 we then use the steady state of the calibrated economy as a starting point to analyze the effect of a tax reform on inequality and mobility, on the transition to the socially-optimal steady state.

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 are going to see in Section 5, 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.⁹

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.¹⁰ We adjust this figure for household size dividing it by 1.4.¹¹

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 $U(c, l) = \ln(c) - l^\alpha/\alpha$, 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 $\alpha = \varepsilon^{-1} + 1 = 3$.

⁹In the robustness analysis discussed in Section 6 we report results if we match the conditional *mean* of bequests instead.

¹⁰The 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.

¹¹This 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.

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

$$\ln w(\theta, h) = \ln A(\theta, h) = (1-\xi) \ln h + \xi \ln \theta, \quad (2)$$

so that it is straightforward to use the variance of residual wages as target to calibrate the variance of unobserved ability θ .¹² 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 estimates by Heathcote et al. (2010) for the U.S. in 2005, for the part of the variance of residual log-wages that is generated by persistent shocks.¹³ We use the variance resulting from persistent shocks as target 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.

We refer to the large 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 (2) thus implies $1-\xi = 0.1$, given that years of schooling S correspond to $\ln h$ in

¹²Given (2), 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 school years correspond to $\ln h$.

¹³See panel C of Figure 3 in Heathcote et al. (2008).

Parameters	Target	Source
Preferences		
Discount factor (annualized): $\beta = 0.966$	Median bequest (conditional on receiving one, equalized): \$52,571	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	Standard
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: $\ln(\theta') = \rho \ln(\theta) + \epsilon, \ln \theta' \sim \mathcal{N}\left(-\frac{\sigma_\epsilon^2}{2(1-\rho^2)}, \frac{\sigma_\epsilon^2}{1-\rho^2}\right)$		
$\rho = 0.448$	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	Heathcote et al. (2008)
Education cost: $g(h', h) = \kappa(h')^{c_1} h^{c_2}$		
$\kappa = 0.0014$	Average years of schooling: 12.86	Barro and Lee (2013)
$c_1 = 0.7465$	Average net cost for an additional year of education: \$13,845	OECD (2011), Stantcheva (2017)
$c_2 = -0.0005$	Intergenerational correlation of years of schooling: 0.46	Hertz et al. (2008)

Table 1: Calibrated parameter values

our model, where compulsory schooling is defined as $h = 1$ in which case the chosen years of schooling $S = \ln 1 = 0$.

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 $\delta = 0.9276$.¹⁴ 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 now discuss the calibration of its parameters.

Stochastic process for ability and education costs.—The parameters for the persistence

¹⁴ $T^y(y)$ 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.

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, 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.

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 c_1 of the cost function $g(h', h)$ in Table 1 allow us to match average years of schooling and average net cost for schooling. 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 education 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.¹⁵ Expressed in dollars, this amount equals the mean annual earnings of high-school dropouts of \$20,241 in 2010 which have a present value for a 30-year

¹⁵A 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 $l^*(\theta, h) \equiv \arg \max \{ \ln(\delta [A(\theta, h)l]^{1-t_y}) - v(l) \}$. For $v(l) = l^\alpha/\alpha$, we obtain $l^*(\theta, h) = (1 - t_y)^{1/\alpha}$. $A(1, 1) = 1$ then implies that the period income of the average worker without any non-compulsory education is $y^*(1, 1) = (1 - t_y)^{1/\alpha} = 0.9356$, once we insert the parameter values documented in Table 1.

<i>Variable</i>	<i>Data</i> (1)	<i>Model</i> (2)
Median bequests, conditional on $b > 0$	\$52,571	\$52,717
Average years of schooling S	12.86	12.75
Correlation(S', S)	0.46	0.48
Intergenerational earnings elasticity	0.45	0.45
Average net cost of an additional year of schooling	\$13,845	\$13,674

Table 2: Target statistics in the data and model predictions

period of \$436,762.¹⁶

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 5 and 6 are concave.¹⁷ 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 four years of non-compulsory schooling, this reduces the cost of educating their children only by 2 per mille.

Simulation.—We solve the problem by applying the endogenous gridpoint method proposed in Hintermaier and Koeniger (2010). 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. Further implementation details on the numerical solution and calibration are provided in Appendix A.2.

Table 2 shows that the calibrated model matches the data targets quite closely. In Appendix A.5 we provide a set of 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.

¹⁶See Table 232 in the *Statistical Abstract of the United States 2012* available at <https://www.census.gov>.

¹⁷In 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.

4 Opportunity and inequality in the calibrated economy

We investigate the transmission of inequality across generations in the calibrated economy. In doing so, we highlight the mechanisms through which parental background affects this transmission, both in terms of nurture through bequests and human capital, and in terms of nature through innate ability.

<i>Rank-rank correlations</i>		
<i>Across generations</i>		
	Ability	0.43
	Welfare	0.67
	Labor earnings	0.43
	Consumption	0.64
	Bequests	0.59
	Years of schooling	0.46
	Labor effort	0.34
<i>Within generation</i>		
	Ability and Welfare	0.90
<i>Elasticity of labor earnings with respect to...</i>		
<i>Nurture</i>		
	Bequests	-0.033
	Schooling	0.50
<i>Nature</i>		
	Ability	0.93

Table 3: Mobility and the effect of nurture and nature on earnings

Notes: As discussed in Section 3, the empirical literature on the returns to education implies that the elasticities of wages (or productivity) with respect to ability and human capital in our model are 0.9 and 0.1, respectively. This gives empirical content to the changes in schooling and ability for the interpretation of the labor earnings elasticities.

The top panel of Table 3 shows rank-rank correlations across generations for all variables of interest in the steady state of the calibrated economy. For earnings, we can compare this with empirical results for the U.S. in Chetty et al. (2014) who report a rank-rank correlation of 0.34 in their table 1.¹⁸ In our 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. This is in line with results in OECD (2018) for the U.S.

¹⁸The estimated transition matrix reported in Table 2 of Chetty et al. (2014) is remarkably similar to Table 11 in Appendix A.4, generated by our calibrated model. The matrix in Chetty et al. (2014) predicts somewhat less persistence at the bottom and top of the income distribution. As shown in Table 16 of Appendix A.5, we match this matrix more closely in the robustness check for a calibration with less persistence that targets the lower end of estimates for the intergenerational earnings elasticity reported in Table 1 of Chetty et al. (2014). The model-implied rank-rank correlation of earnings is 0.29 in this case.

that it takes on average four to five generations for the offspring of a low-income family to reach the average income.

The rank-rank correlations in the top panel of Table 3 reveal that all variables, but for labor effort, are at least as persistent across generations as ability, where the persistence of ability is given by the calibrated stochastic process. The difference between the rank-rank correlation of labor earnings and the respective correlations for consumption or welfare in Table 3 shows that labor earnings are more mobile (and hence less persistent) across generations than consumption or welfare. This is because dynasties partially self insure ability shocks and thus smooth consumption across generations, implying a less than perfect rank-rank correlation of 0.90 between ability and welfare within generations. The insurance against ability shocks is achieved by nurture which implies persistence in bequests and schooling. The relatively lower persistence of labor effort than ability is the result of different effects: more bequests result in a negative wealth effect on labor effort but more schooling or ability make labor effort more productive so that the substitution effect would increase labor effort, *ceteris paribus*. These effects also explain the different sign of the average steady-state elasticities of labor earnings with respect to changes in bequests, schooling and ability, respectively, reported in the bottom panel of Table 3.

Ability / Welfare	Quintiles				
	1	2	3	4	5
1	0.86 (-6.03,4.4)	0.08 (37.64,4.8)	0.03 (93.7,4.8)	0.02 (161.32,4.8)	0.01 (283.21,4.8)
2	0.14 (-8.1,4.4)	0.66 (-5.0,4.6)	0.11 (45.71,4.9)	0.06 (116.35,4.9)	0.03 (245.21,4.9)
3	0.0 (.)	0.26 (-8.4,4.5)	0.53 (0.01,4.7)	0.14 (72.03,5.0)	0.07 (205.73,5.0)
4	0.0 (.)	0.0 (.)	0.33 (-7.88,4.6)	0.5 (13.55,4.8)	0.17 (157.32,5.0)
5	0.0 (.)	0.0 (.)	0.0 (.)	0.28 (-4.6,4.7)	0.72 (76.71,5.0)

Table 4: Social mobility matrix in the calibrated economy

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

To investigate the extent of social mobility in the calibrated economy further, we show in Table 4 how shocks to ability translate into differences in welfare. The matrix in Table 4 displays the probability that a family in a quintile of the ability distribution is in a

specific quintile of the welfare distribution. The welfare measure is dynastic and includes the discounted welfare of future generations. We define perfect social mobility within a generation as the case in which nature, in terms of ability θ , fully determines the position in the welfare distribution. In other words, nurture, in terms of received bequests b and obtained human-capital investment h , is then irrelevant for the position in the welfare distribution. In this case, the matrix would be an identity matrix. 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.¹⁹

For example, the cell in the first row and fifth column of the matrix 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. These families have received \$283,210 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 86% of families currently in the lowest ability quintile are also in the lowest quintile of the welfare distribution. As shown in Table 3, the rank-rank correlation between ability and welfare within a generation is 0.90. More insurance across generations and less social mobility within a generation are two sides of the same coin.

We investigate further the mechanisms through which nurture affects the transmission of inequality. As mentioned above, bequests reduce labor effort through a wealth effect. This introduces an endogenous 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. The bottom panel of Table 3 shows 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 in-

¹⁹We find that labor supply is approximately uncorrelated with ability in our calibrated model so that changes in welfare result mostly from changes in consumption.

come. 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 1.5 percent.

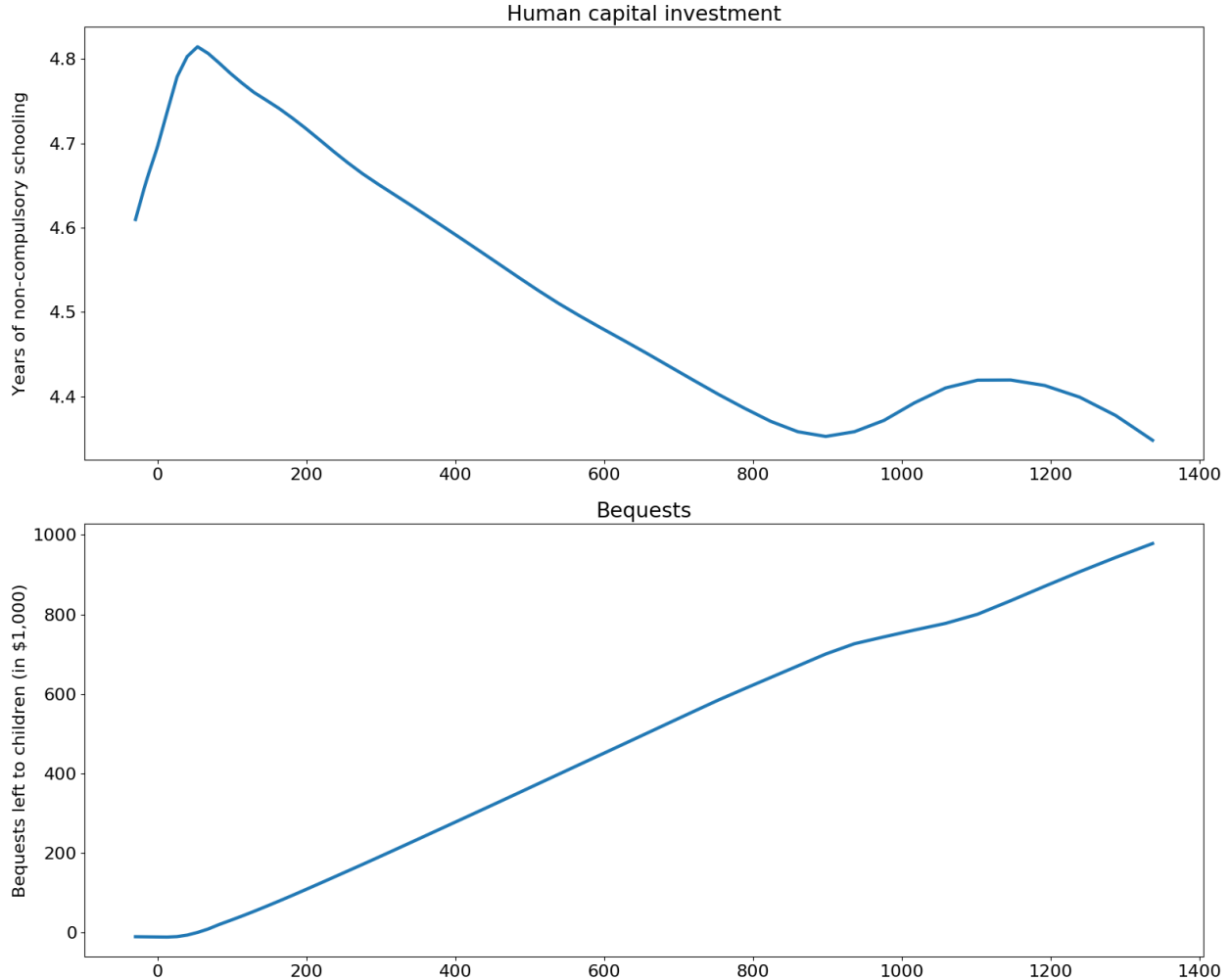


Figure 1: Human capital investment and bequests as a function of received bequests. *Notes:* We condition on parents with 12 years of schooling and median ability. Bequests are in units of \$1,000.

Another channel through which bequests affect mobility, 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.²⁰

²⁰In 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). An important difference in our model is that parents do not perfectly know the ability of their children when they invest into their schooling so that the effect of borrowing constraints on schooling investment and labor productivity is less strong. Furthermore, as explained

To illustrate the consequences for human capital investment, Figure 1 plots human capital investment 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 human capital investment is a very non-monotonic function of bequests because there are various effects at work. At a low level of bequests, human capital investment 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, less mobile and 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.²¹ The bottom panel of Table 3 shows that this is the case in the calibrated economy. The average elasticity of earnings with respect to schooling investment is 0.50, 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. The elasticity of labor earnings with respect to nature (ability) is even 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.

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 from an intergenerational insurance perspective. To answer this question, we report results of the following experiment in Table 5. Consider a family characterized by the initial

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 A.5 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 6, this increases the amount of insurance in the calibrated steady state relative to the benchmark calibration.

²¹The 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.

conditions (b, h, θ) . Then compute the welfare increase if that family receives additional \$10,000, for example as bequest. Table 5 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 12 in Appendix A.4.²²

Earnings quintile	Increase in years of schooling	Cost of additional schooling	Increase in ability (in units of standard deviation)
1	0.97	19,393	0.08
2	0.70	13,772	0.09
3	0.57	11,429	0.09
4	0.47	9,489	0.10
5	0.34	7,106	0.11

Table 5: 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 5 illustrate the effectiveness of nurture (b, h) relative to nature θ in generating the same amount of welfare. In terms of efficiency, 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, 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. Thus, schooling is less effective for these families in generating welfare than bequests. For families in the top of the earnings distribution instead, schooling is more efficient than bequests for generating additional welfare, given the complementarity of ability and human capital in generating 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 5 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 abil-

²²The correlation between earnings and welfare is very high at 0.81 so that the values in the last column of Tables 5 and 12 do not differ up until the second decimal place.

²³For 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 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.

ity and shows the extent to which a given shock to ability has a stronger welfare impact at the bottom of the earnings distribution.

After illustrating the mechanisms through which nurture and nature affect the transmission of inequality in the calibrated model, we compute the quantitative effects of nature and nurture on labor earnings y and the intergenerational transmission of earnings. We find that the effect of nurture on earnings through schooling and bequests is modest in the calibrated model.²⁴ The variation in bequests and schooling given to a generation explains 0.9 – 1.2 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 4.8 percent of the part of the variation of children’s earnings that can be attributed to parents’ nature and nurture. These results are consistent 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 using unique Danish data. They find that two thirds of the intergenerational earnings elasticity can be attributed to nature. Our model attributes an even larger role to ability but this has to be interpreted as an upper bound when assessing the effect of nature, given that ability at labor market entry in our model may also contain some nurture component.

5 Social optimum

To put inequality and mobility in the calibrated economy into perspective, we will compare them to their respective counterparts in the social optimum. We assume that asymmetric information constrains the insurance or redistribution provided by a utilitarian planner who discounts the future and weighs family dynasties equally. This implies non-degenerate inequality and mobility in the social optimum so that comparison between the calibrated economy and the social optimum is non-trivial.²⁵

We focus 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 , they considered the case in which $\psi > \beta$. They showed that this assumption breaks the immiseration result of Atkeson and

²⁴While the effect of nurture on earnings is small, nurture insures generations against ability risk as shown in Table 4 above. Nurture thus allows to decouple production from consumption. We elaborate on the extent of consumption insurance through nurture in Section 6.

²⁵In the first best with full insurance all inequality in consumption among ex-ante identical households would be eliminated.

Lucas (1992) and implies a non-degenerate stationary distribution of consumption and welfare in the planner problem.

We refer to Appendix A.1 for details about the planner problem and its 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 truthfully revealed by families to the planner in equilibrium. In the social optimum,

$$\mathbb{E} \left[\frac{1}{u'(c_t(\theta^t))} \right] = \frac{1/(1+r)}{\psi} \mathbb{E} \left[\frac{1}{u'(c_{t+1}(\theta^t))} \right]. \quad (3)$$

Stationarity of (average) consumption thus requires $\psi = 1/(1+r)$, 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 :

$$\mathbb{E} \left[\frac{1}{u'(c_{t+1}(\theta^{t+1}))} \middle| \theta^t \right] = \frac{\beta}{\psi} \frac{1}{u'(c_t(\theta^t))} + \eta_t \left(1 - \frac{\beta}{\psi} \right), \quad (4)$$

where η is the multiplier attached to the constraint that captures the differences in the promises made by the planner compared to the promises which the planner would make if the same discounting as the family were applied to the welfare of future generations. See the recursive problem and its constraints in Appendix A.1.

For $\beta/\psi < 1$ and $\eta > 0$, equation (4) shows that $1/u'(c(\theta^t))$ is mean-reverting. Because the planner cares more about providing equal opportunities for new generations in the future, shocks to ability are not fully passed on to future generations. The mean reversion implies a form of 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 calibrated to the U.S.

6 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. Appendices A.2, A.3 together with equation (A.21) and the subsequent discussion in Appendix A.1, provide further details on how we implement the numerical solution and the reform.

We focus 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.

The evolution of key variables of interest after the reform is intuitive. Average consumption increases by 1.2% after the reform illustrating the efficiency gains of the social optimum compared to the calibrated economy. Expenditures for human capital and average labor earnings increase although average labor effort decreases.

The distributions in Figure 2, Appendix A.4, show that the averages hide substantial heterogeneity. The figure plots distributions up to two generations after the reform and reveals that labor supply and human capital become much more dispersed after the reform as the planner decouples production of families from their consumption. Figures 3 and 4 show the evolution of the 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 generation 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 generation 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 resources implied by promised utility (which we continue to call bequests after the reform in the figures). Thus, labor supply and human capital expenditures become more dispersed within the top quartile of the ability distri-

bution, and the dispersion of labor supply increases also within the bottom quartile. 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.

Beyond the effects on inequality, we find that income mobility (or persistence of labor earnings) remains quite stable on the transition to the social optimum. Two generations after the reform (corresponding to 30 – 60 years), for example, the rank-rank correlation of earnings is 0.39 compared to 0.43 in the calibrated economy and the rank-rank correlation of consumption is 0.77 compared to 0.64. The similar persistence of earnings across generations, accompanied by the increase of the persistence of consumption, suggests that more intergenerational insurance against ability shocks is provided after the reform, reducing social mobility within generations.

Ability / Welfare	Quintiles				
	1	2	3	4	5
1	0.38 (-264.19,5.0)	0.24 (-256.43,5.2)	0.17 (-222.59,5.4)	0.13 (-122.23,5.5)	0.07 (111.82,5.5)
2	0.25 (-272.46,5.1)	0.23 (-265.87,5.3)	0.21 (-240.52,5.5)	0.18 (-158.71,5.6)	0.13 (64.26,5.7)
3	0.18 (-276.69,5.1)	0.21 (-270.64,5.3)	0.22 (-247.81,5.5)	0.21 (-174.55,5.6)	0.18 (43.01,5.7)
4	0.13 (-278.87,5.1)	0.18 (-274.97,5.3)	0.22 (-254.32,5.5)	0.23 (-187.46,5.6)	0.24 (31.15,5.8)
5	0.06 (-285.55,5.1)	0.13 (-281.24,5.3)	0.19 (-262.42,5.5)	0.25 (-201.38,5.7)	0.37 (28.66,5.9)

Table 6: Social mobility matrix on the transition to the social optimum (in $t=1$, i.e., the second generation after the reform)

Notes: See Table 4.

This suggestive evidence is confirmed comparing the social mobility matrix two generations after the reform, presented in Table 6, with the mobility matrix of the calibrated economy in Table 4. The mobility matrix after the reform has less weight on the diagonal, implying more insurance and less mobility. The correlation between the rank in the ability distribution and the rank in the welfare distribution is indeed much smaller at 0.38, compared with 0.90 in the calibrated economy.²⁶ 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,²⁷ and the correlation between

²⁶Given that we simulate the economy for a sample of 500,000 families, this difference is statistically significant.

²⁷In the social optimum, families do not face borrowing constraints as in calibrated economy so that wealth (or promises) can take more negative values.

wealth (or promises) and human capital falls after the reform: in the second generation after the reform ($t = 1$), $cor(b, \log(h)) = 0.12$ compared with 0.44 in the calibrated economy.

To further compare the role of parental background in the calibrated economy and on the transition path to the social optimum, 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.

<i>Rank in distribution of...</i>	<i>Welfare rank</i>		
	(1) Calibrated economy	(2) 2nd generation after reform	(3) steady state after reform
b	0.29	0.55	0.99
S	0.14	0.68	0.36
θ	0.77	0.02	0.01
N	500,000	500,000	500,000
R^2	0.92	0.89	0.97

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.

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 5.5 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 (or promises in the terminology of the planner problem) become more important than schooling for the position in the welfare distribution.

	(1)	(2)	(3)	(4)
	Pass-through coefficient	Rank-rank correlation of ability and welfare	Rank-rank correlation of ability and labor earnings	IGE
Calibrated economy	0.67	0.90	0.99	0.45
Social optimum				
2 nd generation after reform	0.59	0.38	0.99	0.42
Steady state after reform	0.63	0.24	0.96	0.44

Table 8: Insurance and Mobility

Notes: The pass-through coefficient captures the effect of unexpected changes in productivity on consumption, obtained from a linear regression of log consumption on log productivity. IGE denotes the intergenerational earnings elasticity.

Table 8 summarizes the analysis by showing how insurance and mobility evolve after the reform, starting from the steady state of the calibrated economy. The pass-through of an ability shock to consumption falls after the reform, illustrating the increase of consumption insurance. Column (1) in Table 8 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 0.59 in the second generation after the reform and 0.63 in the steady state.²⁸ That is, in the steady state after the reform 37 percent of ability shocks are insured compared to 33 percent in the calibrated economy.²⁹ Similarly, the rank-rank correlation between ability and welfare within generations decreases in column (2). Instead, columns (3) and (4) show that mobility of income across and within generations changes very little. In particular, the high rank-rank correlation between ability and labor earnings within generations, that remains close to one after the reform, shows

²⁸Given the process for ability specified in Table 1 and the assumption of perfect competition, equation (2) shows how the ability shock translates into unexpected wage changes.

²⁹Although 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 US 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 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.

that the additional insurance after the reform can be decoupled from efficiently higher production of more able families. At the same time, the transition after the reform is associated with an increase in inequality of earnings, labor effort, consumption and bequests, as measured by the respective standard deviation reported in Table 13 in Appendix A.4.

Table 14 in Appendix A.5 shows that the results on insurance and mobility reported in Table 8 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 a higher Frisch elasticity of 0.86 instead of 0.5, 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 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 14 in Appendix A.5 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 0.45 to 0.62. As shown in column (2) of Table 14 in Appendix A.5, a higher target level of bequests reduces social mobility, in terms of the correlation between ability and welfare, and increases insurance in the calibrated economy compared to the benchmark case. Qualitatively as in the benchmark case, social mobility decreases after the reform. The pass-through coefficient increases instead so that it is of similar size in the social optimum as in the benchmark case.

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 trade less mobility within a generation for 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.³⁰

³⁰Lack 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 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

It is worth noting that compared to the transition to a social optimum with immiseration, there is more social mobility 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.

7 Simple tax and subsidy systems and the social optimum

Wedges, as defined in Appendix A.6, capture the differences 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. We provide simple history-independent approximations of these taxes and subsidies and discuss to which extent they allow to achieve the welfare gains of the reform.

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 (A.27) specified in Appendix A.6 and modified to include linear, constant rates for taxes and subsidies. As an alternative, we approximate linear schedules that may vary across generations analogous 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 further explained in Appendix A.6.³¹

Table 9 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. The planner has a lower discount rate than the dynasties and thus wants to provide further incentives so that dynasties shift resources to the future. The Pigouvian subsidy of bequests thus aligns the dynasties' incentives with the social opti-

intergenerational transfers.

³¹Note that by design of the reform, further explained in Appendix A.3, each dynasty in state $s = (b, h, \theta)$ at the time of the reform has to contribute the same amount of net-taxes as in the calibrated economy. Hence, all tax systems we present include a lump-sum component to ensure this is the case. Also 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-Mead 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.

mum. Human capital accumulation is subsidized as well. As shown in Bovenberg and Jacobs (2005) and illustrated in Appendix A.6, this is done also to offset the distortion of the schooling decision resulting from the taxation of labor income. The size of the subsidy further depends on the riskiness of human capital and possible distortions of the accumulation motive or incentives as emphasized by Stantcheva (2017) and Koeniger and Prat (2018).

	Bequest	Schooling	Labor Income
Optimal linear taxes / subsidies, constant across generations	-0.36	-0.30	0.18
Approximated linear taxes / subsidies, varying across generations:			
2 nd generation after reform	-0.21	-0.45	0.31
Steady state after reform	-0.23	-0.50	0.41

Table 9: Simple linear taxes and subsidies

Notes: The taxes and subsidies varying across generations are cross-sectional averages derived in Appendix A.6. A positive value implies a tax, a negative value implies a subsidy.

Table 9 shows that bequests and schooling are subsidized more compared to 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 9 corresponds to the quantitative difference of the rates to the calibrated economy. The first row of Table 9 further shows that the optimal subsidy rate for schooling is six percentage points smaller than for bequests.

Concerning labor income taxation, the optimal linear tax rate reported in the first row of Table 9 is eleven percentage points smaller than the average marginal income tax in the calibrated economy. The bottom part of Table 9 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 9 reveals that the level of the approximated rates differs substantially from the optimal but constant rates. We now show that the latter imply a better approximation of the social optimum in our application.

We assess to which extent the linear tax and subsidies achieve the welfare gains of the

second-best allocation relative to the laissez-faire allocation.³² 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.

	<i>Welfare gain in percent of consumption equivalents</i>
Second best	4.8
Optimal linear taxes / subsidies, constant across generations	2.2
Approximated linear taxes / subsidies, varying across generations	1.5

Table 10: Welfare gains

Notes: Welfare gains compared to the laissez faire, 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 10 shows the welfare gains 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 .³³ The welfare gains of 4.8% are larger than the gains 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 cannot explain the difference, as we discuss further below. If compared to the calibrated economy, the welfare gains of the reform are a bit higher at 5.7% because of the borrowing constraint that

³²We compare the welfare gains relative to the laissez faire without distortionary taxes, starting from the steady state of the calibrated economy. In the laissez-faire economy we add a lump-sum tax so that each dynasty of type (b, h, θ) contributes the same present value of net taxes as in the calibrated economy and thus, by design, also as in the second best. This prevents a wealth effect from biasing the welfare comparison. Appendix A.3 contains further details.

³³If we evaluate the welfare gains from the perspective of the family, applying the discount factor β in the objective function, the welfare gains are smaller by a factor 0.31 as gains accruing in the future then receive less weight.

is present in the calibrated but not in the laissez-faire economy.

The second and third row of Table 10 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 10, achieves 31% of the welfare gains obtained by moving from the laissez faire to the second best after the reform. The economy with optimized but constant linear taxes and subsidies, in the second row of Table 10, achieves 46% 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.³⁴

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.³⁵ As 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 to the laissez faire are 6.7%. Approximated linear taxes and subsidies, which vary across generations, imply welfare gains of 1.2%, thus achieving 18% of the welfare gains from the laissez faire to the second best after the reform. These results are similar in terms of orders of magnitude to those reported in Table 10 and illustrate that our conclusions for the benchmark case do not depend on the wedge between the discount factor of the planner and the dynasties in a critical way.

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.

³⁴We 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 A.6 provides further details on the non-linear approximation.

³⁵For a stationary distribution in the calibrated economy, we need to maintain the assumption that $\beta < 1/(1+r)$.

8 Conclusion

We have analyzed inequality and mobility in a dynasty model that we calibrated to match empirical evidence for the U.S. We have compared mobility and the intergenerational transmission of inequality in the steady state of the calibrated economy and on the transition to the social optimum, based on a social welfare function in which dynasties are weighed equally and, as a result, future generations receive a larger weight than in the welfare function of dynasties. The wedge between the weight of future generations in the dynasties' problem and the planner's problem has been disciplined by observable data on bequests in our calibration.

We have found that, compared with the calibrated U.S. economy, social optimality implies less opportunity within a generation, measured in terms of less mobility of agents with different ability across the welfare distribution, and more intergenerational insurance against ability risk. This is achieved with a stronger influence of nurture, in terms of schooling and bequests, on the family's position in the welfare distribution. We find that income mobility across and within generations remains quite stable, indicating that consumption can be decoupled more from labor earnings in the social optimum compared to the calibrated economy.

We have shown that simple linear schedules for taxes and subsidies achieve about half of the welfare gains of the socially optimal tax system compared to the *laissez faire*. On average, bequests and schooling are subsidized in the social optimum to provide insurance of future generations against ability risk. The optimal linear rates of the simple tax and subsidy system are 18% for the labor income tax, 36% for the bequest subsidy and 30% for the schooling subsidy.

Our analysis of the transition to the social optimum shows that less social mobility 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 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 one plausible set of these assumptions, we have shown how intergenerational insurance and mobility may be shaped by the tax and subsidy system.

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.
- Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy*, 75(4):352–365.
- 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.
- 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.
- 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.
- De Nardi, M. and Yang, F. (2016). Wealth Inequality, Family Background, and Estate Taxation. *Journal of Monetary Economics*, 77(C):130–145.
- 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.

- 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.
- 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.
- 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 Mirrleesian 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. (2018). Inheritance Taxation and Wealth Effects on the Labor Supply of Heirs. Manuscript.
- 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.
- Markussen, S. and Røed, K. (2020). Economic Mobility Under Pressure. *Journal of the European Economic Association*.
- 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 Appendix

A.1 Planner problem: recursive formulation and results

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.³⁶ We use the first-order conditions of the relaxed problem to derive the key equations (3) and (4) discussed in Section 5 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 $\theta^t \equiv \{\theta_0, \theta_1, \dots, \theta_t\}$ 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 $x_t : \Theta^t \rightarrow \mathbb{R}_+^3$. 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

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

where \mathbb{E}_0 is the expectation operator conditional on information available at time 0 and β is the discount factor 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 $r_t : \Theta^t \rightarrow \Theta$. 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

³⁶This 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.

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

$$W = \max_{\{c_t, y_t, h_{t+1}\}} \mathbb{E}_0 \sum_{t=1}^{\infty} \psi^{t-1} U(c_t, y_t, h_t, \theta_t) \quad (\text{A.2})$$

$$\text{s.t. } \mathbb{E}_0 \sum_{t=1}^{\infty} \beta^{t-1} U(\cdot) \geq V, \quad (\text{A.3})$$

(IC),

$$\mathbb{E}_0 \sum_{t=1}^{\infty} q^{t-1} z_t \leq \Gamma_0,$$

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 $t \rightarrow \infty$. 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 6 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

$$\Gamma_0 = \min_{\{c_t, y_t, h_{t+1}\}} \mathbb{E}_0 \sum_{t=1}^{\infty} q^{t-1} z_t \quad (\text{A.4})$$

$$\text{s.t. } \mathbb{E}_0 \sum_{t=1}^{\infty} \beta^{t-1} U(\cdot) \geq V$$

(IC),

$$\mathbb{E}_0 \sum_{t=1}^{\infty} \psi^{t-1} U(\cdot) \geq W.$$

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.³⁷ For this purpose, we define the equilibrium continuation utility $\omega(\theta^t)$ for a given history θ^t as

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

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

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

Incentives are compatible ex-post if

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

We use \mathbf{x}^{IC} to denote the set of all allocations \mathbf{x} satisfying (A.6).³⁸

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. As in Koeniger and Prat (2018), the envelope condition for the considered problem is

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

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-

³⁷In this part we draw heavily on material in Koeniger and Prat (2018) which we present here for completeness.

³⁸Note that allocations in \mathbf{x}^{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}^{FOA} as the set of allocations so that condition (A.7) holds for all θ^t , we note that $\mathbf{x}^{\text{IC}} \subseteq \mathbf{x}^{\text{FOA}}$. Replacing the incentive constraint in problem (A.4) by $\mathbf{x} \in \mathbf{x}^{\text{FOA}}$ 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 “ \prime ” values of variables one period in the future and with “ $-$ ” values of variables with a one-period lag, the stationary recursive problem is:

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

$$\text{s.t.} \quad \omega(\theta) = U(c(\theta), y(\theta), h, \theta) + \beta V'(\theta), \quad (\text{A.8})$$

$$\tilde{\omega}(\theta) = U(c(\theta), y(\theta), h, \theta) + \psi W'(\theta) \quad (\text{A.9})$$

$$V = \int_{\Theta} \omega(\theta) dF(\theta|\theta_-), \quad (\text{A.10})$$

$$W = \int_{\Theta} \tilde{\omega}(\theta) dF(\theta|\theta_-), \quad (\text{A.11})$$

$$\Phi = \int_{\Theta} \omega(\theta) \frac{\partial f(\theta|\theta_-)}{\partial \theta_-} d\theta, \quad (\text{A.12})$$

$$\frac{\partial \omega(\theta)}{\partial \theta} = \frac{\partial U(c, y, h, \theta)}{\partial \theta} + \beta \Phi'(\theta). \quad (\text{A.13})$$

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

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

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

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

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

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

and the first-order condition for W' is

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

We now use these two equations to derive a modified reciprocal Euler equation and the key equations (3) and (4) discussed in Section 5 of the main text. As a first step, we substitute the envelope condition $\partial \Gamma / \partial W = \eta$ into equation (A.17) which implies

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

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

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

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

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

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 $\hat{\theta}$.

As a further step, we take unconditional expectations in equation (A.20) and rearrange to obtain

$$\mathbb{E} \left[\frac{1}{\partial u(c'(\theta'))/\partial c'(\theta')} \right] = \frac{\beta}{q} \frac{1}{\partial u(c(\theta))/\partial c(\theta)} + \eta \left(\frac{\psi}{q} - \frac{\beta}{q} \right). \quad (\text{A.21})$$

For the implementation of the reform, as discussed in further detail in Section A.3, we set $\eta = \lambda$ at the time of the reform. We now show that this choice implies that average consumption changes at the time of the reform and remains constant afterwards at its steady state level if $\psi = q$, i.e., if the planner's discount rate equals the real interest rate r . Equation (A.21) implies that for $\eta = \lambda$ and $\psi = q$,

$$\mathbb{E}[(\partial u(c'(\theta'))/\partial c'(\theta'))^{-1}] = \mathbb{E}[\lambda'] = \eta. \quad (\text{A.22})$$

Iterating forward, using (A.21) with a one-period lead, $\eta = \mathbb{E}[\lambda']$ and $\psi = q$, it follows that average consumption remains constant after an initial adjustment at the time of the reform.

Long-run properties.— If we take unconditional expectations on both sides of (A.20) and rearrange using that $\mathbb{E}[\eta] = \mathbb{E}[\lambda]$, we obtain key equation (3) in Section 5 of the main text, denoting the marginal utility of consumption as $u'(\cdot)$:

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

Clearly, stationarity of consumption requires $\psi = q$. Substituting this into the modified

reciprocal Euler equation (A.20), we obtain key equation (4) in Section 5 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 get 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 A.3.

A.2 Numerical solution

We solve 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-Mead 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 . 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.

A.3 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 which in the calibrated economy amount to 27 percent of average labor earnings. 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

$$\widehat{\Gamma}_\eta(\lambda(s), 0, h(s), \theta) = (1 + r)b(s) + t(s), \quad (\text{A.23})$$

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.1. 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 (A.23) 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.1, 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.

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 frequently 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.

A.4 Further results

y_t / y_{t+1}	Quintiles				
	1	2	3	4	5
1	0.42	0.25	0.17	0.11	0.05
2	0.24	0.25	0.22	0.18	0.11
3	0.16	0.21	0.22	0.22	0.18
4	0.11	0.17	0.21	0.25	0.27
5	0.07	0.12	0.17	0.24	0.40

Table 11: 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	0.94	17,465	0.08
2	0.68	12,977	0.09
3	0.57	11,317	0.09
4	0.48	10,251	0.10
5	0.40	9,177	0.11

Table 12: Average increase in schooling and ability that is welfare equivalent to receiving an additional \$10,000 as bequests, by welfare quintile

	Standard deviations of logarithms				
	Bequests	Schooling	Labor effort	Earnings	Consumption
Calibrated economy	0.16	0.08	0.06	0.46	0.41
Social optimum					
2 nd generation after reform	0.39	0.06	0.10	0.51	0.36
Steady state of reform	1.04	0.07	0.16	0.53	0.55

Table 13: Standard deviations in the calibrated and reformed economy

Notes: For bequests we report the standard deviation of the level and not the logarithm because the variable 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 A.3.

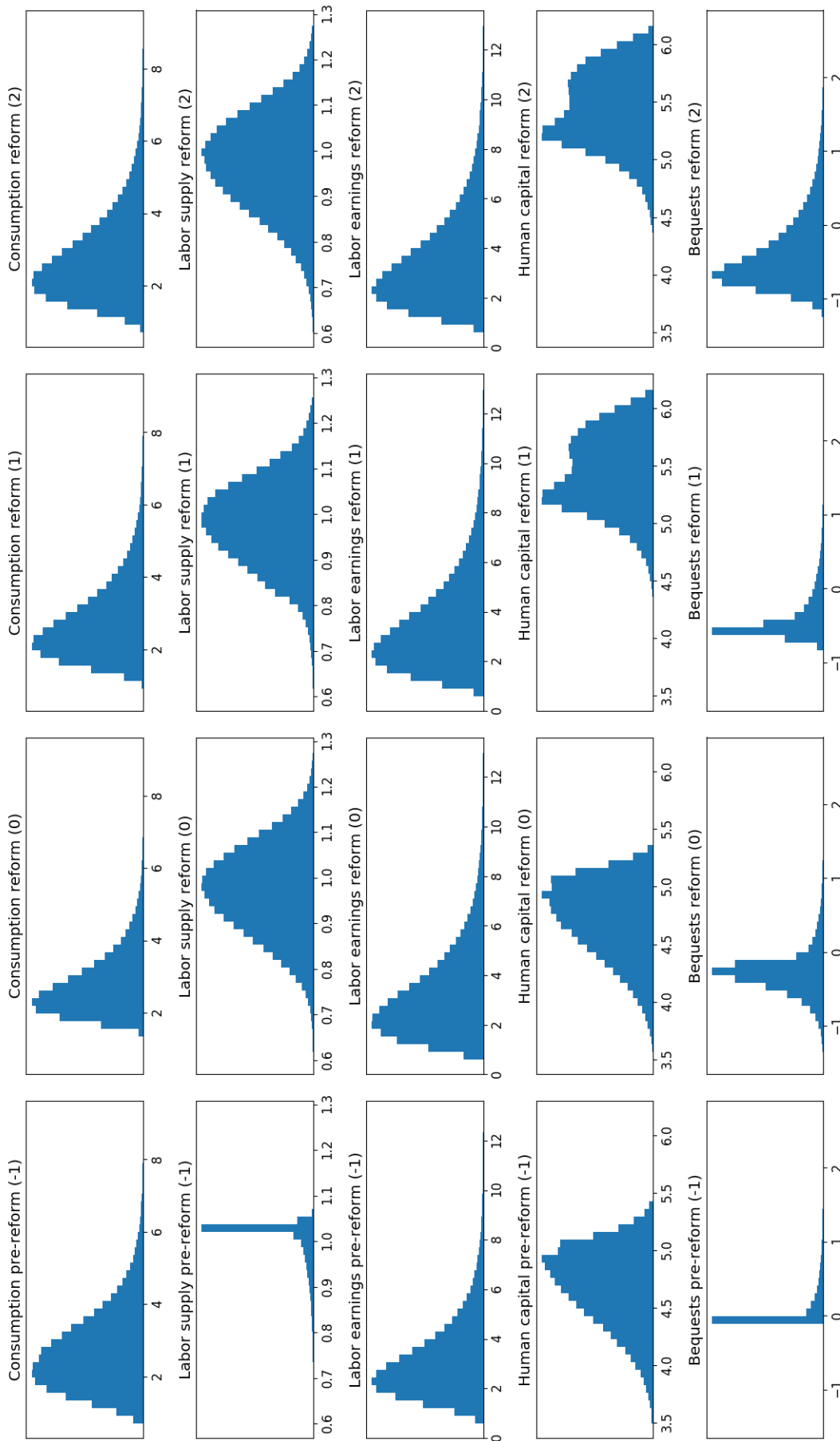


Figure 2: Transition to the social optimum: evolution of distributions. *Notes:* The reform is implemented before generation 0 makes its choices. Units of monetary variables are in 10,000 US-\$ and labor supply is normalized by the average level in the calibrated steady state.

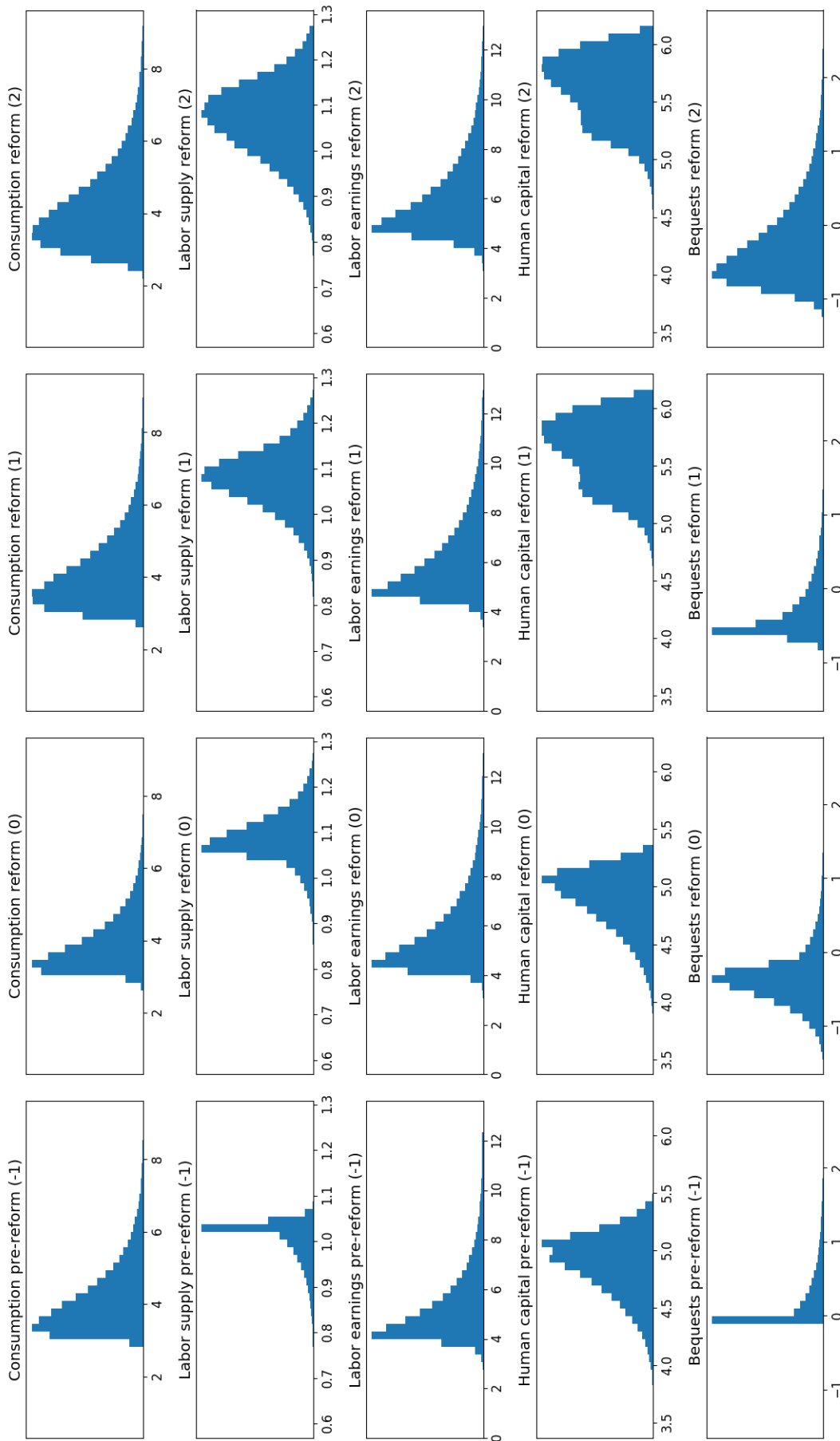


Figure 3: Transition to the social optimum: evolution of distributions for top quartile of ability distribution in each generation.
 Notes: See Figure 2.

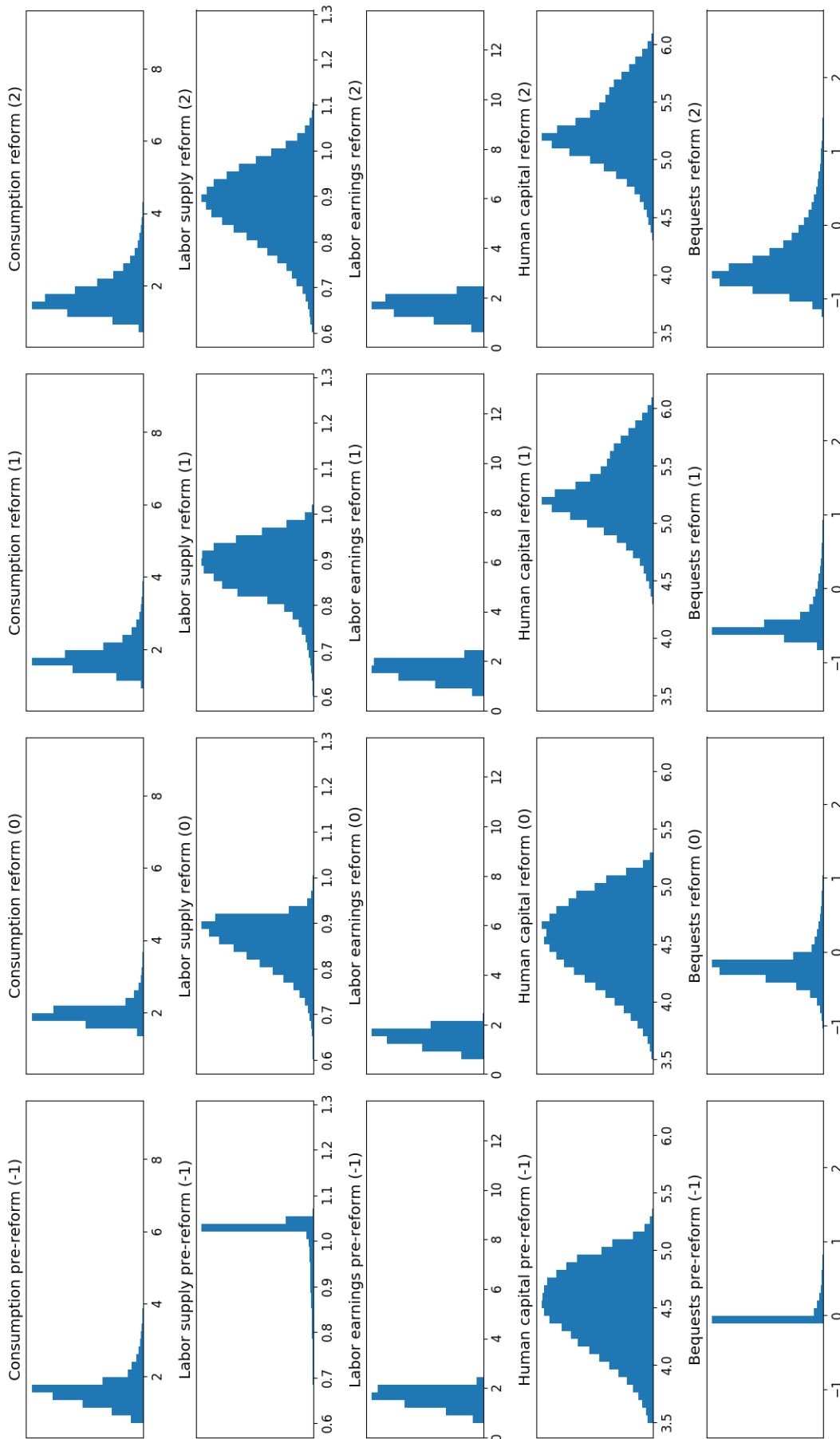


Figure 4: Transition to the social optimum: evolution of distributions for bottom quartile of ability distribution in each generation. *Notes:* See Figure 2.

A.5 Robustness analysis

This subsection presents the results of the robustness analysis. Table 14 shows that the results on the evolution of insurance and mobility for our benchmark calibration, presented in Table 8 in the main text, are robust across most alternative calibrations. For convenience, we repeat the results of the benchmark case in the first column of Table 14.

Worth noting is that social mobility, 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 and thus less dependance of welfare on current ability.

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 15 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 15, 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 Chetty et al. (2014), table 1. The third column of Table 15 shows that the recalibrated model continues to match the target statistics closely. Furthermore, the model-implied mobility matrix, displayed in Table 16, matches more closely the estimated

	(1)	(2)	(3)	(4)	(5)
	<i>Benchmark</i>	<i>Mean bequests as target</i>	<i>Lower IGE as target</i>	<i>Higher Frisch elasticity</i>	<i>Higher complementarity between θ and h</i>
<i>Pass-through coefficient</i>					
Calibrated economy	0.67	0.40	0.65	0.69	0.71
Social optimum					
2 nd generation after reform	0.59	0.40	0.58	0.67	0.60
Steady state after reform	0.63	0.65	0.62	0.73	0.63
<i>Rank-rank corr(ability,welfare)</i>					
Calibrated economy	0.90	0.49	0.87	0.92	0.95
Social optimum					
2 nd generation after reform	0.38	0.19	0.26	0.42	0.38
Steady state after reform	0.24	0.12	0.17	0.26	0.27

Table 14: Robustness of results for insurance and social mobility

	(1)	(2)	(3)	(4)	(5)
	Benchmark	Mean bequests as target	Lower IGE as target	Higher Frisch elasticity	Higher complementarity between θ and h
<i>Recalibrated parameters</i>					
Discount factor (annualized) β	0.966	0.970	0.965	0.965	0.963
Persistence ρ	0.45	0.44	0.33	0.49	0.44
Education cost parameter κ	0.0014	0.0015	0.0015	0.0017	0.0017
Education cost parameter ζ_1	0.75	0.78	0.78	0.68	0.70
Education cost parameter ζ_2	-0.0005	-0.0489	-0.0514	0.0443	0.0177
<i>Predictions for target statistics</i>					
Bequests	52,772	298,458	52,681	51,678	52,711
Average years of schooling S	12.75	12.73	12.77	12.71	12.88
Correlation(S', S)	0.48	0.47	0.44	0.44	0.46
Intergenerational earnings elasticity	0.45	0.43	0.30	0.48	0.44
Average net cost of an additional year of schooling	13,954	13,691	13,696	13,821	13,725

Table 15: Calibration results for robustness checks

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 footnote 18.

Higher Frisch elasticity.— We recalibrate the model for a larger Frisch elasticity of 0.86 for aggregate hours, reported in Chetty et al. (2013), table 2. 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 15 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 ζ_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.³⁹ The calibration of these parameters is done jointly with the other parameters reported in Table 15. Our recalibration results in $\chi = 0.786$, $\sigma_\epsilon^2 = 0.295$ and $\xi = 0.73$ compared with $\chi = 1$, $\sigma_\epsilon^2 = 0.217$ and $\xi = 0.9$ in the benchmark calibration. 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 15 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.

y_t / y_{t+1}	Quintiles				
	1	2	3	4	5
1	0.35	0.24	0.19	0.14	0.08
2	0.22	0.23	0.22	0.19	0.14
3	0.17	0.20	0.21	0.22	0.20
4	0.14	0.18	0.20	0.23	0.25
5	0.12	0.15	0.18	0.23	0.32

Table 16: 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.

³⁹Given 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 given that ability is not fully observable and correlated with schooling.

A.6 Details on the wedges and the approximate implementation

In the laissez faire each dynasty solves the maximization problem⁴⁰

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

where the family chooses functions $b', h', l, c : \Theta \rightarrow \mathbb{R}$. 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:

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

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

$$\tau_{b,t}(\theta^t) \equiv 1 - \frac{q}{\beta} \frac{\partial u(c_t(\theta^t))/\partial c_t}{\mathbb{E} \left[\partial u(c_{t+1}(\theta^{t+1}))/\partial c_{t+1} \middle| \theta^t \right]}, \quad (\text{A.24})$$

$$\tau_{h,t}(\theta^t) \equiv \frac{\beta}{\frac{\partial g(h_{t+1}(\theta^t), h_t(\theta^{t-1}))}{\partial h_{t+1}}} \mathbb{E} \left[\frac{\frac{\partial u(c_{t+1}(\theta^{t+1}))}{\partial c_{t+1}}}{\frac{\partial u(c_t(\theta^t))}{\partial c_t}} \left(\frac{\partial y_{t+1}(\theta^{t+1})}{\partial h_{t+1}} - \frac{\partial g(h_{t+2}(\theta^{t+1}), h_{t+1}(\theta^t))}{\partial h_{t+1}} \right) \middle| \theta^t \right] - 1, \quad (\text{A.25})$$

$$\tau_{l,t}(\theta^t) \equiv 1 - \frac{\partial v(y_t(\theta^t), h_t(\theta^{t-1}), \theta_t)/\partial y_t}{\partial u(c_t(\theta^t))/\partial c_t}, \quad (\text{A.26})$$

⁴⁰The first-order conditions of this problem are equivalent for the problem $W^L(b(\theta), h(\theta), \theta) = \max_{\{b'(\theta), h'(\theta), l(\theta), c(\theta)\}} \left\{ [\mathbf{U}(c(\theta), l(\theta)) + \beta \int_{\Theta} W^L(b'(\theta), h'(\theta), \theta')] dF(\theta' | \theta) \right\}$, s.t. the constraints.

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

The social optimum can be decentralized with a general, history-dependent tax schedule as shown, for example, in Stantcheva (2017). We will 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

$$\begin{aligned} \max_{\{b_{t+1}, h_{t+1}, l_t, c_t\}} \mathbb{E}_t \left[\sum_{s=t}^{\infty} \beta^{s-t} \mathbf{U}(c_s, l_s) \right] & \quad (\text{A.27}) \\ \text{s.t. } b_{t+1} = (1+r) \left(1 - t_{b,t}(\theta^{t-1}) \right) b_t + (1 - t_{y,t}(\theta^t)) y_t - c_t - (1 + t_{h,t}(\theta^t)) g(h_{t+1}, h_t) - T(\theta^t), \\ y_t = Y(h_t, \theta_t, l_t), \\ \ln(\theta) = \rho \ln(\theta_-) + \epsilon, \end{aligned}$$

where the tax shifter $T(\theta^t)$ 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 $t_{b,t}(\theta^{t-1})$ 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 (A.26) 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 Section 7 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:

$$\bar{t}_{y,t} = \mathbb{E}[\tau_{l,t}(\theta^t)] \quad (\text{A.28})$$

so that every dynasty faces the same labor income tax.

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

$$\frac{\partial u(c_t)}{\partial c_t} = \beta(1+r)(1-t_{b,t+1}(\theta^t))\mathbb{E}_t\left[\frac{\partial u(c_{t+1})}{\partial c_{t+1}}\right]. \quad (\text{A.29})$$

Thus, comparison with equation (A.24) implies that bequests of a dynasty with a certain history θ^t should be taxed at rate $\tau_{b,t}(\theta^t)$.⁴¹ We approximate the bequest tax with the cross-sectional average of the bequest wedge, i.e.,

$$\bar{t}_{b,t+1} = \mathbb{E}[\tau_{b,t}(\theta^t)]. \quad (\text{A.30})$$

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

$$\begin{aligned} \frac{\partial g(h_{t+1}, h_t)}{\partial h_{t+1}}(1+t_{h,t}(\theta^t))\frac{\partial u(c_t)}{\partial c_t} = & \beta\mathbb{E}_t\left[\left(\frac{\partial y_{t+1}}{\partial h_{t+1}}(1-t_{y,t+1}(\theta^{t+1}))\right) \right. \\ & \left. - \frac{\partial g(h_{t+2}, h_{t+1})}{\partial h_{t+1}}(1+t_{h,t+1}(\theta^{t+1}))\right)\frac{\partial u(c_{t+1})}{\partial c_{t+1}} \end{aligned} \quad (\text{A.31})$$

with the definition of the wedge for human capital (A.25) and solve for $t_{h,t}(\theta^t)$:

$$t_{h,t}(\theta^t) = \tau_{h,t}(\theta^t) - \frac{\beta}{\frac{\partial g(h_{t+1}, h_t)}{\partial h_{t+1}}}\mathbb{E}_t\left[\left(\frac{\partial y_{t+1}}{\partial h_{t+1}}t_{y,t+1}(\theta^{t+1}) + \frac{\partial g(h_{t+2}, h_{t+1})}{\partial h_{t+1}}t_{h,t+1}(\theta^{t+1})\right)\frac{\frac{\partial u(c_{t+1})}{\partial c_{t+1}}}{\frac{\partial u(c_t)}{\partial c_t}}\right]. \quad (\text{A.32})$$

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-

⁴¹As 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 non-linear taxes below.

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.,

$$\bar{t}_{h,t} = \mathbb{E}\left[t_{h,t}(\theta^t)\right]. \quad (\text{A.33})$$

Given the recursive nature of equation (A.32), 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 (A.27) 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.

For the welfare comparisons of the decentralized economy with a simple non-linear approximation of the tax schedules, mentioned in footnote 34, 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 (A.27). 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 (A.27), with the modified tax schedules, and the definition of the wedges then imply that

$$t_{y,t}(y_t, \theta^t) + \frac{\partial t_{y,t}(y_t, \theta^t)}{\partial y_t} y_t = \tau_{l,t}(\theta^t), \quad (\text{A.34})$$

$$t_{b,t}(b_t, \theta^t) + \frac{\partial t_{b,t}(b_t, \theta^t)}{\partial b_t} b_t = 1 - \frac{q}{\beta} \frac{\partial u(c_{t-1}(\theta^{t-1})) / \partial c_{t-1}}{\partial u(c_t(\theta^t)) / \partial c_t}, \quad (\text{A.35})$$

and

$$\begin{aligned}
t_{h,t}(h_{t+1}, \theta^t) + \frac{\partial t_{h,t}(h_{t+1}, \theta^t)}{\partial h_{t+1}} \frac{1}{\epsilon_{g(\cdot), h_{t+1}}} h_{t+1} &= \tau_{h,t}(\theta^t) \\
- \frac{\beta}{\frac{\partial g(h_{t+1}, h_t)}{\partial h_{t+1}}} \mathbb{E}_t \left[\left(\frac{\partial y_{t+1}}{\partial h_{t+1}} \left(t_{y,t+1}(y_{t+1}, \theta^{t+1}) + \frac{\partial t_{y,t+1}(y_{t+1}, \theta^{t+1})}{\partial y_{t+1}} y_{t+1} \right) \right. \right. \\
&\quad \left. \left. + \frac{\partial g(h_{t+2}, h_{t+1})}{\partial h_{t+1}} t_{h,t+1}(h_{t+2}, \theta^{t+1}) \right) \frac{\frac{\partial u(c_{t+1})}{\partial c_{t+1}}}{\frac{\partial u(c_t)}{\partial c_t}} \right],
\end{aligned} \tag{A.36}$$

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 (A.35) differs from the right-hand side of (A.24) 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., $t(x) = \tilde{\alpha} + \tilde{\beta}x$. The left-hand side of equations (A.34), (A.35) and (A.36) then becomes

$$\tilde{\alpha}_{y,t} + 2\tilde{\beta}_{y,t}y_t, \tag{A.37}$$

$$\tilde{\alpha}_{b,t} + 2\tilde{\beta}_{b,t}b_{t+1}, \tag{A.38}$$

and

$$\tilde{\alpha}_{S,t} + \tilde{\beta}_{S,t} \frac{1}{c_1} + \tilde{\beta}_{S,t} S_{t+1}, \tag{A.39}$$

where (A.39) follows from the parametric assumption for the cost function $g(h', h) = \kappa(h')^{\zeta_1} h^{\zeta_2}$ and from the change of variable $S = \ln(h)$ to express the tax rate as a function of years of schooling S , implying $\frac{\partial t(h)}{\partial h} h = \frac{\partial t(S)}{\partial S}$.⁴²

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 (A.34), (A.35), (A.36) for each generation t on a constant and $x = y_t, b_{t+1}, S_{t+1}$, respectively.⁴³ Given the recursive nature of equation (A.36), we use the approximated taxes in $t + 1$ when approximating taxes in t , as before.

⁴²See Section 3 for an explanation why $S = \ln(h)$ in our model.

⁴³The approximation of the linear tax with the cross-sectional average obtains if we only regress the right-hand sides on a constant.

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:

$$\widehat{\alpha}_{y,t} = \widetilde{\alpha}_{y,t}, \quad (\text{A.40})$$

$$\widehat{\alpha}_{b,t} = \widetilde{\alpha}_{b,t}, \quad (\text{A.41})$$

$$\widehat{\alpha}_{S,t} = \widetilde{\alpha}_{S,t} + \widetilde{\beta}_{S,t} \frac{1}{G_1}, \quad (\text{A.42})$$

$$\widehat{\beta}_{y,t} = 2\widetilde{\beta}_{y,t}, \quad (\text{A.43})$$

$$\widehat{\beta}_{b,t} = 2\widetilde{\beta}_{b,t}, \quad (\text{A.44})$$

$$\widehat{\beta}_{S,t} = \widetilde{\beta}_{S,t}. \quad (\text{A.45})$$