Optimal Taxation when Children's Abilities Depend on Parents' Resources

Alexander Gelber and Matthew Weinzierl*

September 4, 2015

Abstract

Empirical research suggests that parents, and therefore tax policy that affects them, can have a significant effect on their children's future earnings abilities. We use the standard model of optimal taxation and U.S. microdata to take a first step toward characterizing, both qualitatively and quantitatively, how this intergenerational link matters for tax policy design. In our baseline case, we find that the utilitarian welfare-maximizing policy in this context would be more redistributive toward low-income parents than is the case under current U.S. tax policy. The additional income under such a policy would increase the probability that low-income children move up the economic ladder, and we estimate that it would thereby generate an aggregate welfare gain equivalent to 1.75 percent of lifetime consumption.

Keywords: Optimal taxation, family economics, intergenerational transmission of ability

*Gelber (corresponding author): UC Berkeley and NBER, agelber@nber.org; Weinzierl: Harvard Business School and NBER, mweinzierl@hbs.edu. We thank participants at the Cowles Summer Conference on Advances in Dynamic Taxation, the National Tax Association annual meeting, and seminars at IFS/LSE, Harvard, Hebrew University, the University of Pennsylvania, and Stanford. We thank Jillian Popadak, Jerry Yeh, and Roger Ou for research assistance and Benjamin B. Lockwood for exceptional work on the computational simulations. We thank the Entrepreneurship and Family Business Research Centre, Center for Human Resources, Mack Center, Zicklin Center, Risk and Decision Processes Center, and Global Initiatives Center, all at Wharton, for generous research support. We would especially like to thank Doug Bernheim, Caroline Hoxby, Marek Kapicka, Robert E. Lucas, Luigi Pistaferri, Monica Singhal, Michele Tertilt, Aleh Tsyvinski, and several anonymous referees for comments and suggestions. We thank Gordon Dahl, Lance Lochner, Dave Rapson, and Larry Kotlikoff for sharing data.

Introduction

Economists have long recognized that parents' resources may be a key determinant of their children's outcomes. Recent evidence shows, in particular, that increasing the disposable incomes of poor parents raises the performance of their children on tests of cognitive ability (Dahl and Lochner 2012; Milligan and Stabile forthcoming; Paxson and Schady 2007; Akee et al. 2010; Løken, Mogstad, and Wiswall 2012; Macours, Schady, and Vakis 2012).[1] Moreover, this literature generally finds larger positive impacts of parents' resources on children's ability among lower-income families. For example, as our paper will explore in detail, Dahl and Lochner (2012) find that a $1,000 increase in family income raises math and reading test scores by about 6% of a standard deviation, with larger impacts in more disadvantaged families. How should tax policy respond these findings? In this paper, we take a first step toward answering this question both qualitatively and quantitatively.[2]

We start with a formal model of the implications of this intergenerational link implications for tax policy design. We modify the standard modern optimal tax model to allow a child's income-earning ability to depend on three components: parental ability, parental disposable income, and a random shock. This mix of factors helps us to capture the complex process by which society's ability distribution is determined. Using that model, we derive analytical expressions for tax policy both within and across generations that would maximize utilitarian social welfare, that is, an unweighted aggregation of individual utility levels. These expressions make clear the new forces acting on tax design in this context. Within a generation, lower marginal tax rates at low incomes will yield gains to the extent that the marginal effect of resources on children's ability is relatively large for families with low incomes or low parental earnings-ability, but they will yield losses to the extent that they reduce the incentives for individuals to act to increase the likelihood that their descendants will have high ability. Across generations, resources should be allocated to equalize the cost of raising welfare, taking into account not only the marginal utilities of individuals in each generation (as in a conventional setting) but also the effects of resources on prior generations' incentives and future generations' tax payments and utility levels.

We then take the model to data and use it quantify the benefits society can gain by having a tax policy that takes advantage of this intergenerational link. We choose values for the model's

[1] Early work includes Leibowitz 1974; Becker and Tomes 1976; Becker 1981; and Becker and Tomes 1986.

[2] Emmanuel Farhi and Iván Werning (2010) characterize optimal taxation across generations, noting in their opening sentences that "One of the biggest risks in life is the family one is born into. We partly inherit the luck, good or bad, of our parents through the wealth they accumulate." Their important analysis largely assumed, however, that children's skills are independent of their parents' abilities and their parents' economic resources. We take up the complementary analysis.

key parameters by having the model's output (under existing U.S. tax policy) approximate new estimates of the transmission of ability across generations in the United States that we generate using data on parents and their children from the National Longitudinal Survey of Youth (NLSY) and Children of the National Longitudinal Survey of Youth (CNLSY). Then, we compute the utilitarian welfare-maximizing policy and find that it redistributes substantially more toward low-ability parents and earlier generations than does current policy. These changes cause average ability to rise across generations, yielding a welfare gain that would be equivalent to a 1.75 percent permanent increase in aggregate disposable income in our baseline case. Gains are larger for lower-income families than for higher-income families.

The paper proceeds as follows. Section 1 lays out the model of the tax design problem and derives analytical conditions on utilitarian welfare-maximizing policy both within and across generations. Section 2 generates new empirical estimates of the transmission of ability across generations and uses them to choose parameter values for the model. Section 3 uses that model to simulate and describe the effects of the utilitarian welfare-maximizing tax policy on the evolution of the ability distribution, and it calculates the welfare implications of those effects. Section 4 concludes. A lengthy online Appendix, which follows the working paper version of this paper (Gelber and Weinzierl 2012), contains details of the results, the robustness checks, the empirical strategy, and the relationship to prior literature.

1 Model

Our formal analysis starts with the standard dynamic optimal tax model based on Mirrlees (1971) and developed in, for example, Golosov et al., (2003).[3] Individuals obtain utility U from consumption c–equivalently, disposable income–and disutility from exerting work effort l. Each individual has an unobservable "type" w, which measures the ability to earn pre-tax income y and is taken from a fixed set of possible values indexed by $i \in \{1, 2, ..., I\}$. The product of individual ability and work effort, both of which are unobservable, is the observable pre-tax income. The tax authority levies taxes as a function of pre-tax income to maximize a utilitarian social welfare function, which we define as the total present-value utility of the population of families starting from the first generation and ending in generation T. The tax authority is constrained in two ways: disposable income

[3] For development of the dynamic model, see also Kocherlakota (2005), Golosov and Tsyvinski (2006), Albanesi and Sleet (2006), Golosov, Tsyvinski, and Werning (2007), Farhi and Werning (2010, 2013), and Weinzierl (2011). For work incorporating human capital accumulation, see Marek Kapicka (2006a, 2006b), Borys Grochulski and Tomasz Piskorski (2010), Dan Anderberg (2009), and Stefanie Stantcheva (2014).

must be funded by output (feasibility), and individuals maximize their own utility taking the tax system as given (incentive compatibility). We also impose the constraint that taxes may depend on only the *current* generation's characteristics and choices, a restriction we believe is appropriate given the strong aversion in real-world policy to having children's taxes depend on their parents' outcomes.

The intergenerational focus of this paper requires some additional structure. Individuals are linked in families, with one individual per generation in each family. Generations are indexed by $t \in \{1, 2, ..., T\}$, where the choice of T is immaterial to the results below. The distribution of individuals across types is taken as given in the first generation, but in subsequent generations it is a function of the distribution of disposable income in the previous generation as well as of the inheritance of type. Formally, denote with $p^j\left(w_t^i, c_t^{i'}\right)$ the probability that an individual of generation $t+1$ is of type j given that her parent (in generation t) was of type i and had disposable income of type i'. We assume no intergenerational transfers, so all disposable income is consumed within each generation. While bequests have interesting implications for the optimal taxation of bequestors and inheritors (see Piketty and Saez 2013 and Kopczuk 2013), they are highly concentrated among the very wealthy. Thus, individuals from most of the ability distribution will be unaffected by our assumption that there is no bequest motive, and even those who would be affected are already wealthy enough so that marginal changes to parental resources are immaterial.

As is common in optimal tax analyses, we translate the tax system set up by the tax authority into a menu of allocations of pre-tax and disposable income. Formally, the tax authority's problem is as follows:

Problem 1 *Tax Authority's Problem*

$$\max_{\{c_t^i, y_t^i\}_{t=1,i=1}^{T,I}} \sum_i p^i U_1^i \tag{1}$$

subject to feasibility:

$$\sum_i p^i R_1^i \geq \bar{R} \tag{2}$$

and incentive compatibility:

$$U_t^i \geq U_t^{i'|i} \quad \textit{for all } t \textit{ and } i, i', \tag{3}$$

where $U_t^{i'|i} = u\left(c_t^{i'}\right) - v\left(\frac{y_t^{i'}}{w_t^i}\right) + \beta \sum_{j=1}^I p^j\left(w_t^i, c_t^{i'}\right) U_{t+1}^j$, $U_t^i = U_t^{i|i}$, $R_t^i = \left(y_t^i - c_t^i\right) + \beta \sum_j p^j\left(w_t^i, c_t^i\right) R_{t+1}^j$. *$u$ represents the dependence of utility U on consumption c, v represents the dependence of utility*

U on hours worked l (equal to income y divided by ability w), β is a discount factor, and $\bar{R}$ is an exogenous revenue requirement.

We make a number of simplifying assumptions in this setup. First, only tax policy is modeled in this paper, but that does not imply that other policies play no role in reality. Our empirical estimates take as given the existing set of non-tax policies and institutions, such as public education or related policies like education grants, vouchers, loans, or subsidies, that have effects on children's abilities (including effects that may interact with the tax system). Our model implicitly assumes that these policies and institutions are held constant as taxes vary.

We restrict our model's instruments in this way for two reasons. First, it matches the empirical evidence we use to calibrate the model, which also is based on variation in tax policy holding other policies fixed. Second, the question of how tax policy can affect the ability distribution is of independent interest, both because in practice partial reforms to policy (e.g., to taxes only) are common and because a household's after-tax and transfer resources may have a direct effect on children.

At the least we should view the tax policy component as a complement to the education policy component. Of course, broader policy reforms could achieve even greater gains, so our estimates of the welfare gains from changes to taxes can be considered a lower bound on the potential for a holistic optimal policy. Such an optimal policy would also take into account the potentially complex interactions between optimal tax policy, optimal educational policy, and other programs, and we hope the results of this paper encourage future work on these important questions.

The importance of the tax policy component relative to the education policy component—or other complementary components like commodity taxes—may depend on the channels that link parental resources and children's abilities. For example, if parental resources matter for children's abilities primarily because the income elasticity of parental investments in schooling is positive, then targeted policies such as direct government expenditures on schooling could potentially be more effective than the more indirect route of tax policy, as some of the decreased tax burden could be spent on items other than schooling investment. However, if greater parental income leads to better outcomes for children through a route that is less similar to feasible government policies—e.g. parental income leads to better outcomes for the children because it reduces stress in the household—then tax policy could be a preferable approach for shaping children's outcomes. Future empirical work that can observe such channels could help in clarifying these issues.

A second simplifying assumption is that the formulation above has the same measure of parental resources serve as the quantity of consumption in the parent's utility function and the input to the child's ability production function. However, we are not asserting that the way in which parental disposable income is used is irrelevant to their child's ability. Rather, we are guided not only by tractability but also by the data. Our empirical evidence concerns the effect on a child's ability of transfers to her parents; we have no data on how those transfers were allocated. In order to calibrate to this evidence, our model must also leave the allocation of these transfers unspecified. We use the term disposable income, rather than consumption, to make this aspect of our analysis clear.

In principle, one could attempt to use the limited available data on the division of parental expenditure into consumption and investment in children's abilities to model more subtle optimal policies. Identifying the separate effects of these categories of expenditure on children's ability would not be possible using our data and identification strategy (as studying the causal effect of the EITC on different categories of consumption is severely limited by empirical power issues (Gelber and Mitchell 2012)). Moreover, the appeal of more subtle policy that distinguished between these categories would be diminished by incentives for (largely unobservable) misreporting of spending across categories. Similarly, in a dynamic framework one could model how consumption and savings over different periods are affected by taxes across periods. The empirical implementation of this would be difficult, however, as the intertemporal substitutability of consumption (i.e. the intertemporal elasticity of substitution) is the topic of much literature that is not considered settled, with elasticity estimates ranging from close to zero (Hall 1988, Dynan 1993) to over two (Gruber 2006). In general, it is possible to argue that the welfare gains we calculate reflect a lower bound on the gains if separate taxation of different types of consumption were modeled.

A third assumption in the problem above is that the allocation of parental time has no effect on children's abilities. Later in the paper, we extensively explore the case in which children's ability depends on both parental income and parents' hours worked. If parents work more, they are likely to spend less time with their children. This in principle could either worsen children's outcomes (if, say, parents teach children skills in their non-work time) or could improve children's outcomes (if, say, parents' increased work serves as a role model for children's work in school). While our empirical strategy strains the available data in this case, such that the results are best seen as suggestive, we find that on net allowing for a role of parental time allocation has only a small effect on our baseline results.

Finally, we do not constrain parent and child distributions of ability to be the same, as they might be in some steady state, to acknowledge that wage distributions and test scores have shown secular time trends in the data (e.g. Goldin and Katz 2007; Flynn 1987).

2 Analysis of utilitarian optimal policy

Our analysis of the tax authority's problem in expressions (1) through (3) generates two analytical results on utilitarian welfare-maximizing policy. First, we characterize the distortion to an individual's choice of how much to earn, that is the marginal tax rate on income. Second, we derive a necessary condition on optimal allocations across generations that shows the effects of this intergenerational link in an intuitive but powerful way.

2.1 Marginal tax rates

In the absence of taxes, parent i in generation t would choose how much income to earn to maximize her own utility subject to a personal feasibility constraint and given the expectation that her descendants, whose abilities are determined by the production function $p^j\left(w_t^i, c_t^i\right)$, would also choose optimally for themselves. That parent's optimal private choice would satisfy the following condition:

$$\frac{v'\left(y_t^i/w_t^i\right)}{w_t^i u'\left(c_t^i\right)} = \left(1 + \beta \sum_j \frac{\partial p^j\left(w_t^i, c_t^i\right)}{\partial c_t^i} \frac{U_{t+1}^j}{u'\left(c_t^i\right)}\right). \tag{4}$$

To interpret result (4), note that $v'\left(y_t^i/w_t^i\right)/[w_t^i u'\left(c_t^i\right)]$ is the ratio of the marginal disutility of labor to the marginal utility of consuming the income that labor earns. In a setting without endogenous child ability, this ratio equals one in the absence of taxes, and is less than one if the individual faces a positive marginal tax rate. In contrast, result (4) implies that in this no-tax setting parents set this ratio *greater* than one. That is, parents take into account the effect of their disposable income on their child's ability, so they will appear to choose labor supply as though there were a marginal subsidy equal to the second term on the right-hand side of expression (4), relative to a model in which they took only their own disposable income into account.

Lemma 1 shows that the tax authority distorts a parent's private choice.

Lemma 1 *Intratemporal Distortion: Let* $\mu_t^{i'|i}$ *denote the multiplier on* (3) *. The solution to the Tax*

Authority's Problem satisfies, for all $t \in [1, 2, ...T]$ and all $j \in [1, 2, ...I]$,

$$\frac{v'\left(y_t^j/w_t^j\right)}{w_t^j u'\left(c_t^j\right)} = \mathbb{A}_t^j \frac{\left(\mathbb{B}_t^j + \mathbb{C}_t^j\right)}{\left(\mathbb{B}_t^j + \mathbb{D}_t^j\right)} \left(1 + \beta \sum_k \frac{\partial p^k\left(w_t^j, c_t^j\right)}{\partial c_t^j} \frac{U_{t+1}^k}{u'\left(c_t^j\right)}\right) \tag{5}$$

where the expressions for $\mathbb{A}_t^j$, $\mathbb{B}_t^j$,$\mathbb{C}_t^j$, and $\mathbb{D}_t^j$ are as follows:

$$\mathbb{A}_t^j = \frac{1}{1 - \beta \sum_k \frac{\partial p^k\left(w_t^j, c_t^j\right)}{\partial c_t^j} R_{t+1}^k}, \quad \mathbb{B}_t^j = \beta^t \pi_t^j + \sum_{\tau=1}^{t-1} \beta^{t-\tau} \sum_i \sum_{i'} \mu_\tau^{i'|i} \left(\pi_t^j|_{c_\tau^i} - \pi_t^j|_{c_\tau^{i'}}\right), \tag{6}$$

$$\mathbb{C}_t^j = \sum_{j'} \mu_t^{j'|j} - \sum_{j'} \mu_t^{j|j'} \frac{1 + \beta \sum_k \frac{\partial p^k\left(w_t^{j'}, c_t^j\right)}{\partial c_t^j} \frac{U_{t+1}^k}{u'\left(c_t^j\right)}}{1 + \beta \sum_k \frac{\partial p^k\left(w_t^j, c_t^j\right)}{\partial c_t^j} \frac{U_{t+1}^k}{u'\left(c_t^j\right)}}, \quad \mathbb{D}_t^j = \sum_{j'} \mu_t^{j'|j} - \sum_{j'} \mu_t^{j|j'} \frac{\frac{1}{w_t^{j'}} v'\left(\frac{y_t^j}{w_t^{j'}}\right)}{\frac{1}{w_t^j} v'\left(\frac{y_t^j}{w_t^j}\right)} \tag{7}$$

where $\pi_t^j|_{c_\tau^i}$ is the probability that a generation t descendant of parent type i from generation τ is of type j and $\sum_i \pi_t^j|_{c_\tau^i}$ is denoted by the unconditional probability π_t^j.

Lemma 1 shows that the product $\mathbb{A}_t^j \frac{\left(\mathbb{B}_t^j + \mathbb{C}_t^j\right)}{\left(\mathbb{B}_t^j + \mathbb{D}_t^j\right)}$ is the optimal wedge distorting the parent's choice of earned income. To understand the determinants of this optimal wedge, we decompose it into three effects.

The first is the "revenue effect." Suppose the tax authority had full information on individuals' (endogenous) abilities. In that case, $\mathbb{B}_t^j = \beta^t \pi_t^j$ and $\mathbb{C}_t^j = \mathbb{D}_t^j = 0$. Then, the wedge would be simply $\mathbb{A}_t^j$, the value of which depends on the present value of the expected net revenue gain from raising the disposable income of parent j. The tax authority values that revenue gain, while the parent does not, so the revenue effect implies that a smaller downward distortion (or larger upward distortion) to parent j's effort is optimal. To the extent that the marginal effect of additional parental disposable income on children's abilities is larger for lower-ability parents, this revenue effect will be greater at low incomes, and smaller marginal taxes at low incomes will be optimal.

The second is the "relative return effect." Suppose the tax authority cannot observe ability, but parental resources have *no effect* on children's abilities. In that case, $\mathbb{A}_t^j = 1$, $\mathbb{B}_t^j = \beta^t \pi_t^j$, and $\mathbb{C}_t^j = \sum_{j'} \mu_t^{j'|j} - \sum_{j'} \mu_t^{j|j'}$, and the optimal distortion is driven by binding incentive constraints in the current generation. Introducing *endogenous* ability changes $\mathbb{C}_t^j$ because its value depends on how the inheritance of ability and parental resources interact in the production of child ability. In particular, including endogenous ability will increase $\mathbb{C}_t^j$ if the marginal effect of additional parental disposable income on children's abilities is larger for lower-*ability* parents. Meanwhile, $\mathbb{D}_t^j$

is unchanged, so $\mathbb{C}_t^j$ and $\mathbb{D}_t^j$ have more similar values and the optimal marginal taxes on low-skilled parents are smaller (note that $\mathbb{C}_t^j < \mathbb{D}_t^j$ when higher-skilled types are tempted to mimic lower-skilled types). Intuitively, if the extra resources redistributed to low-income families are of less value to high-ability parents because the intergenerational effect on abilities is smaller, high-income parents will be less tempted in this model to claim low ability and smaller marginal tax rates at low incomes will be optimal.

The third is the "ancestor incentive effect." Introducing endogenous ability also changes $\mathbb{B}_t^j$, which measures how an increase in c_t^j affects the incentives of earlier generations to increase the likelihood that their descendants have the type j. All else equal, a smaller marginal tax rate for a low-skilled type raises the temptation for previous generations to work less and accept a higher probability of low-skilled descendants. In this way, the ancestor incentive effect acts to offset the revenue and relative return effects, providing a rationale for larger marginal taxes at low incomes.

In the end, the net effect on utilitarian welfare-maximizing marginal tax rates depends on the relative strength of these three effects.

2.2 Allocations across generations

In the absence of the intergenerational link studied in this paper, the standard condition on utilitarian welfare-maximizing allocations across generations is

$$\sum_j \frac{\pi_t^j}{u'\left(c_t^j\right)} = \sum_k \frac{\pi_{t+1}^k}{u'\left(c_{t+1}^k\right)}. \tag{8}$$

This condition, parallel to the "Symmetric Inverse Euler Equation" in Weinzierl (2011), shows that the optimal allocation equalizes the cost, in disposable income units, of raising social welfare across generations. A version of it also applies to optimal tagging (Mankiw and Weinzierl 2010).

With endogenous ability, a modified version of (8) applies, which we state in the following proposition.

Proposition 1 *The solution to the Tax Authority's Problem satisfies*

$$\frac{1}{\Lambda_t}\sum_j \frac{\pi_t^j}{u'\left(c_t^j\right)} \frac{1-\beta\sum_k \frac{\partial p^k\left(w_t^j,c_t^j\right)}{\partial c_t^j} R_{t+1}^k}{1+\beta\sum_k \frac{\partial p^k\left(w_t^j,c_t^j\right)}{\partial c_t^j}\frac{U_{t+1}^k}{u'\left(c_t^j\right)}} = \frac{1}{\Lambda_{t+1}}\sum_k \frac{\pi_{t+1}^k}{u'\left(c_{t+1}^k\right)} \frac{1-\beta\sum_l \frac{\partial p^l\left(w_{t+1}^k,c_{t+1}^k\right)}{\partial c_{t+1}^k} R_{t+2}^l}{1+\beta\sum_l \frac{\partial p^l\left(w_{t+1}^k,c_{t+1}^k\right)}{\partial c_{t+1}^k}\frac{U_{t+2}^l}{u'\left(c_{t+1}^k\right)}} \tag{9}$$

where Λ_t is defined as:

$$\Lambda_t = 1+\sum_{\tau=1}^{t-1} \frac{\beta^{t-\tau}}{\beta^t} \sum_i \sum_{i'} \mu_\tau^{i'|i} \sum_k \left(\pi_t^k|_{c_\tau^i} - \pi_t^k|_{c_\tau^{i'}} \right) + \frac{\beta}{\beta^t} \sum_k \sum_{k'} \mu_t^{k|k'} \frac{\sum_j \frac{\partial p^j\left(w_t^k, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)} - \sum_j \frac{\partial p^j\left(w_t^{k'}, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)}}{1+\beta \sum_j \frac{\partial p^j\left(w_t^k, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)}}.$$

To understand Proposition 1, recall that (8) implies that the cost of raising social welfare through transfers to one generation must be the same for all generations. Proposition 1 is the same condition, but in the more complicated context of this model. To help with its interpretation, we decompose it into three effects of making a transfer to a given generation t.

First, if the transfer to generation t raises individual $j's$ investment in her children's abilities resulting in increased tax revenue from future generations, these revenue gains act to offset the costs of the transfer. Formally, this factor is captured in the expression $\beta \sum_k \left[\partial p^k \left(w_t^j, c_t^j\right) / \partial c_t^j\right] R_{t+1}^k$, which is closely related to the revenue effect identified in the discussion of Lemma 1. This expression is the present value of the net change in future taxes paid by individual $j's$ children when c_t^j increases.

Second, under the assumption that the transfer to individual j raises her investment in her children's abilities resulting in increased utilities for future generations, these welfare gains augment any direct changes in utility from the transfer. Formally, this factor is captured in the expression $\beta \sum_k \left[\partial p^k \left(w_t^j, c_t^j\right) / \partial c_t^j\right] \left[U_{t+1}^k / u'\left(c_t^j\right)\right]$. This expression is the present value, per additional unit of utility for individual j, of the increase in utility enjoyed by individual $j's$ children when c_t^j increases.

Third, both the relative return and ancestor incentive effects from the discussion of Lemma 1 also matter for intertemporal allocations as well, through their effects on Λ_t. The parameter $\Lambda_t = 1$ when two conditions hold: the marginal effect of parental financial resources on child ability is independent of parental ability (i.e., $\partial p^k \left(w_t^j, c_t^j\right) / \partial c_t^j = \partial p^k \left(w_t^{j'}, c_t^j\right) / \partial c_t^j$ for all j, j'), and incentive constraints do not bind in preceding generations (i.e., $\mu_\tau^{i'|i} = 0$ for all τ, i, i'). If, instead, transfers to a generation relax incentive constraints and increase the investment of low-income parents in their children, the expression Λ_t is less than one. Similarly, if transfers to a generation relax incentive constraints that bind on ancestors whose offspring are relatively common in the recipient generation, the expression Λ_t is less than one. The smaller is Λ_t, the larger is the optimal transfer to generation t.

To build intuition for how these effects combine, suppose that average ability is stable over time and the effects of parental resources on a child's ability are largest at lower skill levels (formally, $\sum_{k=1}^{K} \pi_t^k = \sum_{k=1}^{K} \pi_{t+1}^k$ for all $K \leq I$ and $\left|\partial p^j \left(w_t^{k+1}, c_t^{k+1}\right) / \partial c_t^\kappa\right| \leq \left|\partial p^j \left(w_t^k, c_t^k\right) / \partial c_t^k\right|$ for $\kappa > k$). A policy neglecting endogenous child ability would satisfy (8) and treat generations symmetrically. In contrast, an alternative policy that would satisfy (9) would transfer resources from generation $t+1$ to low-ability workers in generation t. Such a policy would raise the average inverse marginal utility of consumption in generation t but also increase the population proportion of higher-ability workers in generation $t+1$, putting greater weight on workers with larger inverse marginal utilities of disposable income and smaller gains in future revenue and utility for their descendants from marginal resources. As a result, the cost of raising social welfare in generations t and $t+1$ would rise together. In other words, transfers from future to earlier generations may generate gains for all generations: early generations from having higher disposable incomes, and future generations from having improved ability distributions. While result (9) does not prove that a policy improvement such as in this example exists, our data and simulations below show that this is a realistic scenario for the United States.

2.3 Labor as an input to children's ability

As we discuss above, it is also possible that children's ability could depend on parents' hours worked. In this case, the planner's problem includes the dependence of $p(\cdot)$ on labor effort (or, equivalently, time not devoted to labor effort).

Problem 2 *Planner's Problem with Parental Labor as an Input to Child Ability*

$$\max_{\left\{c_t^i, y_t^i\right\}_{t=1, i=1}^{T, I}} \sum_i p^i U_1^i \tag{10}$$

where

$$U_t^i = u\left(c_t^i\right) - v\left(\frac{y_t^i}{w_t^i}\right) + \beta \sum_{j=1}^{I} p^j \left(w_t^i, c_t^i, \frac{y_t^i}{w_t^i}\right) U_{t+1}^j \tag{11}$$

This is maximized subject to feasibility:

$$\sum_i p^i R_1^i \geq \bar{R}, \tag{12}$$

where $\bar{R}$ is an exogenous revenue requirement, and

$$\sum_i p^i \left[\left(y_t^i - c_t^i\right) + \beta \sum_j p^j \left(w_t^i, c_t^i, \frac{y_t^i}{w_t^i} \right) R_{t+1}^j \right] \geq \bar{R}, \tag{13}$$

and incentive compatibility for each generation:

$$U_t^i \geq U_t^{i'|i} \text{ for all generations } t \text{ and types } i, i', \tag{14}$$

where $U_t^{i'|i}$ denotes the utility obtained by an individual of type i when claiming to be type i' :

$$U_t^{i'|i} = u\left(c_t^{i'}\right) - v\left(\frac{y_t^{i'}}{w_t^i}\right) + \beta \sum_{j=1}^{I} p^j \left(w_t^i, c_t^{i'}, \frac{y_t^{i'}}{w_t^i} \right) U_{t+1}^j. \tag{15}$$

Simplifying as in the model without labor effort in the ability production function, we obtain the following Lemma, which is analogous to Lemma 1.

Lemma 2 *Intratemporal Distortion with Parental Labor as an Input to Child Ability: Let $\mu_t^{i'|i}$ denote the multiplier on* (14)*. The solution to the Planner's Problem with Parental Labor as an Input to Child Ability satisfies, for all $t \in [1, 2, ...T]$ and all $j \in [1, 2, ...I]$,*

$$\frac{v'\left(\frac{y_t^k}{w_t^k}\right)}{w_t^k u'\left(c_t^k\right)} = \mathbb{A}_t^j \frac{\left(\mathbb{B}_t^j + \mathbb{C}_t^j\right)}{\left(\mathbb{B}_t^j + \mathbb{D}_t^j\right)} \frac{\left(1 + \beta \sum_j \frac{\partial p^j\left(w_t^k, c_t^k, \frac{y_t^k}{w_t^k}\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)}\right)}{\left(1 + \beta \sum_j \frac{\frac{1}{w_t^k} \partial p^j\left(w_t^k, c_t^k, \frac{y_t^k}{w_t^k}\right)}{\partial y_t^k} \frac{U_{t+1}^j}{\frac{1}{w_t^k} v'\left(\frac{y_t^k}{w_t^k}\right)}\right)}$$

where

$$\mathbb{A}_t^j = \frac{1 + \beta \sum_j \frac{\frac{1}{w_t^k} \partial p^j\left(w_t^k, c_t^k, \frac{y_t^k}{w_t^k}\right)}{\partial y_t^k} R_{t+1}^j}{1 - \beta \sum_k \frac{\partial p^k\left(w_t^j, c_t^j, \frac{y_t^j}{w_t^j}\right)}{\partial c_t^j} R_{t+1}^k},$$

$$\mathbb{B}_t^j = \beta^t \pi_t^j + \sum_{\tau=1}^{t-1} \beta^{t-\tau} \sum_i \sum_{i'} \mu_\tau^{i'|i} \left(\pi_t^j|_{c_\tau^i} - \pi_t^j|_{c_\tau^{i'}} \right),$$

$$\mathbb{C}_t^j = \sum_{j'} \mu_t^{j'|j} - \frac{1 + \beta \sum_k \frac{\partial p^k\left(w_t^{j'}, c_t^j, \frac{y_t^j}{w_t^j}\right)}{\partial c_t^j} \frac{U_{t+1}^k}{u'\left(c_t^j\right)}}{1 + \beta \sum_k \frac{\partial p^k\left(w_t^{j}, c_t^j, \frac{y_t^j}{w_t^j}\right)}{\partial c_t^j} \frac{U_{t+1}^k}{u'\left(c_t^j\right)}} \sum_{j'} \mu_t^{j|j'},$$

$$\mathbb{D}_t^j = \sum_{j'} \mu_t^{j'|j} - \frac{\frac{1}{w_t^{j'}} v'\left(\frac{y_t^j}{w_t^{j'}}\right)}{\frac{1}{w_t^{j}} v'\left(\frac{y_t^j}{w_t^{j}}\right)} \frac{\left(1 + \beta \sum_k \frac{\frac{1}{w_t^{j'}} \partial p^k\left(w_t^{j'}, c_t^j, \frac{y_t^j}{w_t^{j'}}\right)}{\partial y_t^j} \frac{U_{t+1}^k}{\frac{1}{w_t^{j'}} v'\left(\frac{y_t^j}{w_t^{j'}}\right)}\right)}{\left(1 + \beta \sum_k \frac{\frac{1}{w_t^{j}} \partial p^k\left(w_t^{j}, c_t^j, \frac{y_t^j}{w_t^{j}}\right)}{\partial y_t^j} \frac{U_{t+1}^k}{\frac{1}{w_t^{j}} v'\left(\frac{y_t^j}{w_t^{j}}\right)}\right)} \sum_{j'} \mu_t^{j|j'}.$$

Note that the division of the intratemporal result into a wedge and an expression equal to what the parent would choose continues to hold in this setting. The parent's optimum would now be:

$$\frac{\frac{1}{w_t^i} v'\left(\frac{y_t^i}{w_t^i}\right)}{u'\left(c_t^i\right)} = \frac{1 + \beta \sum_{j=1}^I \frac{\partial p^j\left(w_t^i, c_t^i, \frac{y_t^j}{w_t^j}\right)}{\partial c_t^i} \frac{U_{t+1}^j}{u'\left(c_t^i\right)}}{1 + \beta \sum_{j=1}^I \frac{1}{w_t^j} \frac{\partial p^j\left(w_t^i, c_t^i, \frac{y_t^j}{w_t^j}\right)}{\partial y_t^j} \frac{U_{t+1}^j}{\frac{1}{w_t^i} v'\left(\frac{y_t^i}{w_t^i}\right)}}.$$

The main differences once parental labor effort enters the child ability production function are as follows (recall that parents internalize the direct effect of their time allocation on their children's abilities, as in the expression for the parental optimum above). Assume extra parental time at work is detrimental to child ability, so that $\frac{\frac{1}{w_t^{j'}} \partial p^k\left(w_t^{j'}, c_t^j, \frac{y_t^j}{w_t^{j'}}\right)}{\partial y_t^j} < 0$. First, the term $\mathbb{A}_t^j$ now captures that extra parental effort will lower future revenues, so the optimal distortion discouraging labor effort is larger (*i.e.* the term $\mathbb{A}_t^j$ is smaller). Second, if extra parental time at work is more detrimental to child ability for low-ability parents, then $\mathbb{D}_t^j$ will be smaller. This reduces the optimal distortion of labor suply for lower skill types. Intuitively, incentive constraints are less binding with this effect, because high-skilled parents gain relatively less from the lower labor effort requirements they would enjoy if they claimed a low income.

3 Model calibration under existing U.S. tax policy

In this section, we generate new estimates of the effect of parental resources on children's ability under existing U.S. tax policy and then use those estimates to parameterize the model described in Section 1.

3.1 Empirical estimates of the target statistics

We adapt to our framework the empirical work from a recent major study of parents' taxes and children's outcomes. Dahl and Lochner (2012) study the effect of expansions of the EITC in the 1990s and other variation over this period in tax and transfer programs on children's test score outcomes.[4] Rather than calibrating our model using a cross-section of data, we use a modified version of the Dahl and Lochner empirical strategy in order to generate more credible estimates of the causal effect of parental income on child ability. We briefly describe their empirical strategy here, often borrowing from their description, but we refer readers to their paper for a full description of their empirical strategy and its motivation.[5] We note the caveat that not all papers have found important effects of transfer programs on children; for example, Kapustin and Ludwig (2014) find that randomly receiving a housing voucher has mostly statistically insignificant effects on children's outcomes, which are always smaller than those in recent studies of cash transfers.

Dahl and Lochner investigate how increases in parental income due to the EITC and other tax and transfer programs affect the cognitive achievement of disadvantaged children. Their estimation strategy is based on the observation that low- to middle-income families received large increases in payments from expansions of the EITC in the late-1980s and mid-1990s but higher-income families did not. If parental disposable income affects child ability, this disparity in the changes to disposable income should have caused an increase over time in the test scores of children from low-to-middle income families relative to those from higher income families. Dahl and Lochner use the CNLSY, which contain data on several thousand children matched to their mothers (from the main NLSY sample). Income and demographic measures are included in the data, in addition to as many as four repeated measures of cognitive test scores per child taken every other year. They use measures of child ability based on standardized scores on the Peabody Individual Achievement Test (PIAT), which measures oral reading ability, mathematics ability, word recognition ability, and reading comprehension. Dahl and Lochner's instrumental variables estimates suggest that a \$1,000 increase in family income raises math and reading test scores by about 6% of a standard deviation.

[4] See Hotz and Scholz (2003) or Nada and Hoynes (2005) for detailed descriptions of the EITC program and a summary of related research.

[5] Several papers estimate the effect of parents' income on their children's achievement levels (see the citations in Section I). The marginal effect of parental resources on child ability is more elusive, in part because it is difficult to find exogenous variation in parents' disposable income and linked data on parents' and their children's outcomes. Our paper suggests that the effect of parents' disposable income on children's wages is an important topic ofr future research.

We estimate a model similar to Dahl and Lochner's, designed to attain a slightly different goal. Motivated by our model above and the simulation below, we estimate the effect that income has on the probability that a parent of given ability type produces a child of a given ability type (controlling for the child's lagged ability type); details are provided in the online appendix.

One potential limitation of this approach is that EITC variation affects only lower and lower-middle income families. However, it is important to note that our measure of tax liability is based on the NBER Taxsim calculator, which incorporates all tax liabilities, not just those due to the EITC. At the same time, the biggest variation in income taxes over this period related to the EITC, as well as to increases in marginal tax rates for top incomes from the 1990 and 1993 reforms.[6]

In our main specification, we divide parents into four wage (ability) categories $\{P_i\}_{i=1}^4$ and divide children into four test score categories $\{C_i\}_{i=1}^4$. Each category comprises one quartile of the sample distribution of wages or test scores, respectively, with subscript i indicating the quartile of the distribution, where $i = 1$ is the lowest quartile. Because there are four parent types, we estimate four separate regressions, in each of which the dependent variable is a dummy that equals one when the child has ability in the i-th category. We classify parents into wage types by ranking them according to their average wage over the full sample period.

In choosing the number of categories, we take into account competing technical and conceptual considerations: more categories will give the calibration more targets as well as better describe the true heterogeneity of the population and therefore the potential gains from optimal policy, but too many categories will prevent the regressions in the empirical estimation from having enough positive values of the dependent variable to yield meaningful results. Using too few categories fails to provide enough empirical targets for the calibration exercise to converge on a best set of parameter values: this factor requires us to use at least four categories.[7]

We construct the simulated model to match (i.e., minimize the sum of squared deviations from) three sets of statistics obtained from this analysis: 1) the marginal effects of parental resources on their children's abilities; 2) the ability transition matrix between generations; and 3) the expected log wage within generations. The signs of these estimates generally conform to expectations: higher parental income predominantly increases the probability that a child is high-ability (i.e., in the third or fourth quartile) and decreases the probability that a child is low-ability (i.e., in the first or second

[6] To address this issue further, we examine the effects of analyzing higher income levels in the Appendix, where we show the robustness of the results to including redistribution from still higher incomes in the calibration.

[7] In the Appendix, we also show results with five and ten categories. Those results show heterogeneity in tax rates at a finer level of disaggregation at the cost of a substantial loss of power in the empirical estimates. Our analyses with five and ten types yield similar results to those from our benchmark four-type model.

quartile), with the "correct" sign of the relationship occurring in 13 out of 16 regressions. Using the same dataset and definition of types, we next generate this transition matrix by calculating the fraction of the sample from each parent wage category that began the sample period with the child test score in each category. Finally, the expected log wage within a generation is readily calculated as the average of the log of the four ability levels. Details of these calculations are provided in the online appendix.

3.2 The model's parameter values

We need to specify some features of the general model from Section 1 to generate output that can be matched to the targeted statistics.

First, we assume that the ability production function is

$$E\left[\ln w_{t+1}|w_t^j, c_t^{j'}\right] = \alpha_a\left(\rho \ln w_t^j + (1-\rho)\ln \bar{w}\right) + \alpha_c^j \ln c_t^{j'}, \tag{16}$$

which implies that the child's expected ability is a function of the parent's ability, a fixed "reference" ability level, and the parent's disposable income. The child's expected ability is influenced by the parent's ability w_t^j relative to the fixed ability level $\bar{w}$, indicating mean reversion in characteristics transmitted across generations, consistent with, e.g., Haider and Solon (2006).

This log-linear functional form concisely captures the basic forces at work in this model. It allows us to adjust the relative importance of parental ability and mean ability (through ρ) and the relative importance of parental ability and parental resources (through α_a and α_c^j). Note that our measure of parental ability (*i.e.*, the wage) implicitly includes not just innate talent but also the education and other factors that contribute to the parent's wage. A further step in the analysis would be to separate out the determinants of parental ability, including parents' own education decisions, and independently assess their importance for determining children's ability. Note that the importance of parental resources depends on parental ability through the dependence of α_c^j on j, establishing a direct connection between the exogenous and endogenous components of the ability production function. Importantly, we impose no restrictions on the form this connection takes—the marginal value of parental resources may increase, be constant, decrease, or exhibit complex nonlinearities as innate ability increases.[8]

[8] We do not estimate this production function directly using our empirical approach because our empirical approach relies on a fixed effects specification, which would difference out parent ability. Our regression specification estimates a coefficient on parental income that is comparable to the coefficient on parental income in (16).

Second, we translate the expected ability in (16) into an ability distribution for the population of children of parents of type j with disposable income $c_t^{j'}$ by assuming that ability is distributed lognormally with variance σ^2 :

$$\ln w_{t+1} \sim N\left(E\left[\ln w_{t+1}|w_t^j, c_t^{j'}\right], \sigma^2\right). \tag{17}$$

The ability distribution over the income range relevant to this paper is commonly calibrated as lognormal (e.g. Tuomala 1990). The variance σ^2 represents an exogenous, stochastic shock to child ability common across parent types.

The simulations of this model will use a discrete distribution of abilities, consistent with the model described in Section 1, whereas (16) and (17) appear to produce continuous ability distributions. To classify individuals into I discrete types, we define fixed ranges of w that correspond to each type $i \in I$. By "fixed," we mean that the boundary values of w that determine whether an individual is assigned wage w^i or w^{i+1} are exogenously given. With these fixed ranges, we can translate the distribution of ability for a given child implied by (17) into transition probabilities among types across generations.[9]

Finally, before proceeding with the calibration, we specify the tax system facing individuals, the utility function those individuals maximize, and the set of ability types.[10] For the status quo tax system, we assume that the Kotlikoff and Rapson (2007) calculations of marginal effective tax rates on income for 30-year-old couples in the United States in 2005 are a good approximation of the status quo tax policy facing parents of young children.[11] These authors' detailed calculations go well beyond statutory personal income tax schedules and include a wide array of transfer programs as well as corporate, payroll, and state and local income and sales taxes. Our computations require a smooth tax function, so we take a fifth-order polynomial approximation of the Kotlikoff-Rapson schedule up to $75,000, well above the EITC phase-out. The government's tax system also includes a grant to all individuals, which we determined by enforcing the government's budget constraint in

[9] An example may help clarify the procedure. Suppose $I = 2$, so that there are two ability types. Denote the fixed wage level that separates types 1 and 2 as w^*. A mother of type j expects her child to, on average, have the ability $E\left[\ln w_2|w_1^j, c_1^j\right]$ as defined by expression (16). In reality, her child's ability is a random variable distributed according to $N\left(E\left[\ln w_2|w_1^j, c_1^j\right], \sigma^2\right)$. The probability that her child's ability ends up in the lower half of the full distribution of wages across all children is, therefore, the value at w^* of the cumulative density function implied by this normal distribution.

[10] We demonstrate the robustness of our results to alternative specifications in the Appendix.

[11] The biggest subsequent changes in tax policy not captured in these calculations are the increases in marginal tax rates among the highest income earners that have occurred through the Patient Protection and Affordable Care Act (ACA) of 2010 and the American Taxpayer Relief Act of 2012, as well as the phaseouts of subsidies for health insurance for low-to-middle income earners under the ACA.

each generation. The individual's current-generation utility takes a separable, isoelastic form

$$u\left(c_t^i\right) - v\left(\frac{y_t^i}{w_t^i}\right) = \frac{\left(c_t^i\right)^{1-\gamma} - 1}{1-\gamma} - \frac{\theta}{\sigma}\left(\frac{y_t^i}{w_t^i}\right)^{\sigma},$$

and we set $\gamma = 2$ and $\sigma = 3$ to be consistent with mainstream estimates of these parameters (see, for example, the range of estimates for the former in Barsky et al. 1996 and Chetty 2006, and for the latter the survey of Chetty 2012). We choose $\theta = 2.5$ so that hours worked in the simulation approximately match the average labor supply in the population. We initially set $\beta = 1$, reflecting no discounting of utility across generations, the assumption preferred by Ramsey (1928). Guided by the empirical analysis discussed above, we assume ability comes in $I = 4$ fixed types (roughly interpretable as the hourly wage): $w_t^i \in \{3.44, 6.30, 9.42, 19.57\}$ for all $t = \{1, 2, ..., T\}$. The probability distribution across those types is uniform in the first generation but is endogenously determined in the model for subsequent generations.

3.3 Calibration Results

Expressions (16) and (17) indicate that the model calibration will search over values of the following seven parameters: $\left\{\rho, \alpha_a, \left\{\alpha_c^j\right\}_{j=1}^{I}, \sigma\right\}$. We impose values for two parameters based on prior research. We set $\rho = 0.5$ for the parameter controlling the transmission of ability across generations. This assumption is based on the voluminous evidence surveyed in Feldman, Otto, and Christiansen 2000, which has produced a range of plausible estimates for ρ—generally in the range of 0.3 to 0.7. This evidence enables us to impose a value of ρ (rather than estimate it), thereby increasing the model's ability to estimate the parameters on which we have more limited evidence.[12] We also impose the value of $\sigma = 0.76$, which we calculate using data on wages from the NLSY sample. This leaves five parameters to be chosen.

To calibrate the five remaining parameters, we minimize a weighted sum of squared errors between our model's output and the three sets of target statistics: the marginal effects and transition matrix and the mean log wage. For reference, we calculate the marginal effects of parental disposable income from the model output as the increase in the probability of a given child type caused by an increase of one percent in a given parent type's disposable income, while the model output directly implies transition probabilities for all parent and child types and mean log wages. We weight the squared errors by the inverse of the targets' standard errors, which has the effect of putting much

[12] In the Appendix, we show that our qualitative conclusions are robust to ρ ranging from 0.3 to 0.7.

greater weight on the more-precisely-estimated transition matrix elements and the mean log wage. We use ten generations ($T = 10$) in the simulations, using the middle (fifth-to-sixth) generation as the target for the calibration exercise.[13]

Table 1 shows the parameter values chosen by the simulation.

Table 1: Parameter values chosen in calibration

	α_a	α_c^1	α_c^2	α_c^3	α_c^4
Value under status quo policy	0.88	0.57	0.23	0.17	0.00

Recall that α_a and α_c^j are the weights on the two channels, ability and economic resources, through which parents affect their child's ability. The product of ρ and α_a gives the weight on parental ability in expected child ability, while α_c^j gives the (parental type-specific) weight on parental resources. The monotonically declining values of α_c^j in Table 1 suggest that parental resources play a greater role among lower-ability parents, consistent with the empirical evidence. Key moments determining the estimates of the α_c^j are the coefficients on parent income in determining child ability. Key moments determining both the estimates of the α_c^j and the estimate of α_a are the elements of the transition matrix of parent ability to child ability, as these determine the combined role that parent ability and parent resources play in determining child ability.

The simulation performs well in matching the empirical targets for which the data are most informative, the transition matrix and mean log wage. The simulation yields marginal effects of parental resources that differ substantially from the data.[14] The calibrated status quo marginal effects exhibit a pattern much closer to what the empirical literature described previously would suggest—negative for lower child types and positive for higher child types—than do the estimated effects in the data. This divergence is not surprising, however, given the statistical insignificance of the empirical estimates and their often-unexpected signs.

Table 2 shows that the simulation closely matches the data for the parent-child type transition matrix between generations (we show the transition between the fifth and sixth generations of the

[13] We show robustness to this choice in the Appendix.

[14] Appendix Table 7 shows this for the fifth (middle) generation of the simulation.

simulation as an illustrative example).

Table 2. Parent-child type transition matrix

	Data					Calibrated status quo policy			
	Child type					Child type			
Parent type	1	2	3	4	Parent type	1	2	3	4
1	0.31	0.24	0.22	0.22	1	0.30	0.26	0.26	0.19
2	0.29	0.25	0.23	0.23	2	0.28	0.25	0.26	0.20
3	0.20	0.27	0.29	0.25	3	0.22	0.24	0.28	0.26
4	0.20	0.23	0.26	0.31	4	0.19	0.22	0.29	0.30

Finally, the simulation matches the mean log wage, 2.07.

4 Effects of Optimal Policy on Ability Distribution and Welfare

In this section, we simulate a many-period version of the tax authority's problem using the parameter values and the model from the previous sections. We characterize utilitarian welfare-maximizing policy by comparing it to the status quo policy used in that calibration, and we calculate the welfare gains from the optimal policy's effects on the evolution of the ability distribution.

4.1 Utilitarian optimal policy compared to the status quo

Table 3 shows average and marginal tax rates for each type under the utilitarian welfare-maximizing and status quo policies. Average tax rates are calculated as the ratio $(y - c)/y$. For marginal tax rates, we compare the marginal tax rates imposed by the status quo policy to the marginal tax rates that would implement the utilitarian welfare-maximizing allocation.

Table 3. Marginal and average tax rates

	Marginal tax rate		Average tax rate	
Type	W-max	Status Quo	W-max	Status Quo
1 (lowest)	32%	21%	-192%	-42%
2	44%	31%	-60%	-14%
3	46%	35%	-13%	0%
4 (highest)	0%	32%	37%	16%

The utilitarian welfare-maximizing policy is substantially more redistributive than the status quo, generating large transfers to low-skilled parents. This result is due entirely to the redistributive preferences in the assumed social welfare function. Nevertheless, these redistributive transfers generate an improved ability distribution by taking advantage of the gap between the impact of increased disposable income on the children of low-ability parents and high-ability parents. The larger marginal distortions (other than on $I = 4$) under the utilitarian welfare-maximizing policy make the allocations enjoyed by lower types less attractive to those with higher ability, as is common in optimal tax analyses.[15]

One might be concerned that these results could be driven in important ways by having only four types. For example, in a Mirrlees framework, having a finite number of types yields a zero marginal tax rate at the top, as we see above. However, the differences in welfare between the status quo and the welfare-maximizing policy will stem largely from the much lower average tax rates at the bottom of the distribution under the welfare-maximizing policy, rather than from differences in marginal distortions at the top.[16]

The utilitarian welfare-maximizing policy adjusts intertemporal allocations to capitalize on the endogeneity of ability, as was suggested in the discussion of Proposition 1. Table 4 reports the difference between the tax authority's "budget balance" as a share of aggregate income in each generation under the welfare-maximizing policy and under the status quo. In other words, it is the additional average tax rate assessed on each generation relative to a balanced budget (assumed for the status quo).

Table 4. Intertemporal allocations

	Difference in government budget balance (as pct of output)									
	1	2	3	4	5	6	7	8	9	10
W-max - Status Quo	-5.1%	-0.7	-0.9	-0.9	-0.9	-0.9	-0.9	-0.8	-1.0	11.3

The utilitarian welfare-maximizing policy borrows from future generations to fund greater invest-

[15] If, as in a standard Mirrleesian analysis, the planner believed the starting ability distribution was fixed, the optimal policy would generate average tax rates of [-253%, -73%, -28%, -2%, and 42%]. Therefore, the presence of endogenous ability reduces the extent of redistribution that is optimal in a standard Utilitarian model. The intuition for this result is that low earners in the conventionally optimal policy face high marginal tax rates, with much of their welfare coming through increased leisure time. With endogenous ability, it can be optimal to have these workers exert more effort, earn more income, and therefore enjoy more disposable income than in a static model. Note that the status quo policy includes much less redistribution, so that the increased redistribution under the utilitarian-optimal policy generates large welfare gains through its effects on ability alone.

[16] Consistent with this logic, we show in the Appendix that our results are similar with other numbers of types (i.e., 5 or 10 types).

ment in the skills of the current generation relative to the status quo. Of course, our model abstracts from many features of the economy, notably capital as a factor of production, some of which may make deficit-financed investment in children less appealing. However, the key point illustrated by Table 4 is that society can benefit by having later generations contribute, through higher taxes, to improving the ability distribution generated by earlier generations.

These differences in tax policy affect the evolution of the ability distribution. The ability transition matrices across generations are shown in Table 5.

Table 5. Transition matrix

	W-max					Status Quo			
	Child type					Child type			
Parent type	1	2	3	4	Parent type	1	2	3	4
1	0.24	0.24	0.28	0.24	1	0.30	0.26	0.26	0.19
2	0.27	0.25	0.27	0.21	2	0.28	0.25	0.26	0.20
3	0.22	0.24	0.28	0.26	3	0.22	0.24	0.28	0.26
4	0.19	0.22	0.29	0.30	4	0.19	0.22	0.29	0.30

The utilitarian welfare-maximizing policy enables a substantially greater share of the children of lower-skilled parents to move up the skill ladder than does the status quo policy, while it has much smaller–though non-negative–effects on the prospects of the children of higher-skilled parents. Figure 1 shows the result: the welfare-maximizing policy leads to 1.8% fewer individuals of the lowest type and 1.6% more of the highest type than does the status quo policy.

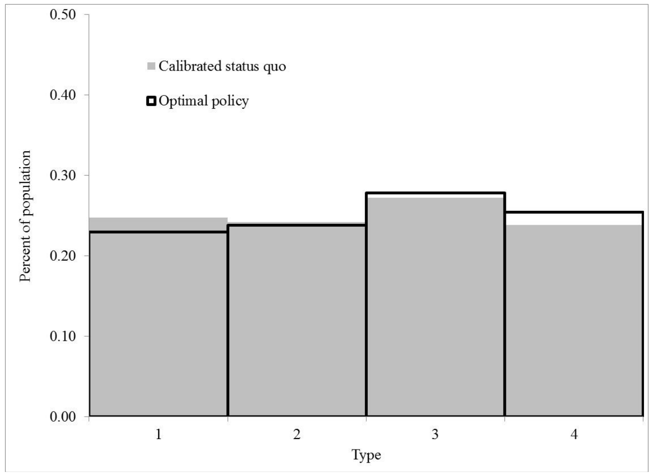


Figure 1: Ability distribution under two policies.

4.2 Welfare gains from improved ability distribution

Finally, we turn to welfare impacts. While utilitarian welfare is much higher under the welfare-maximizing policy than under the status quo, that difference is driven by the model's assumed social preference for income equality. For the purposes of this paper, we are more interested in the welfare gains due to endogenous ability by itself, so we consider the following thought experiment. We define a setting called the "adjusted status quo" in which the distribution of abilities generated by the utilitarian welfare-maximizing policy prevails but the within-period utility levels of individuals of each ability type are those from the status quo policy outcome. We then calculate the factor by which disposable income would have to rise in the status quo to reach the welfare of the adjusted status quo as our measure of the welfare gain due solely to the utilitarian welfare-maximizing policy's effects on the ability distribution over time. Similar factors can be calculated for each type of first-generation parent, indicating how the welfare gains through this channel are shared. Table

6 shows the results for the baseline case of ten generations.

Table 6. Welfare

	Welfare level (in utils)		Welfare gain
	Status quo policy	Adjusted status quo	(Percent of disposable income)
Overall	6.44	6.49	1.75%
Type-specific			
1 (Lowest)	6.25	6.30	2.2%
2	6.41	6.46	1.6%
3	6.49	6.54	1.6%
4 (Highest)	6.61	6.66	1.6%

As these results show, the utilitarian welfare-maximizing policy has the potential to generate a welfare gain equivalent to 1.75 percent of aggregate disposable income simply by shifting the ability distribution over time. The gains are somewhat larger among low-skilled parents, as would be expected. Nevertheless, high-skilled parents gain substantially, as the gains accruing to future generations raise the current generation's present-value welfare. Gains for future generations follow the same patterns.[17]

4.3 Effects of Parental Time Allocation

In this brief subsection, we incorporate parental time allocation into the quantitative analysis by estimating the welfare gain from the version of the model in which parental time working can affect child ability.[18] In particular, we target these marginal effects as well as the transition matrix and mean log wage as in Section 3, calibrating the following reduced-form equation, which is analogous to (18):

$$E\left[\ln w_2 | w_1^j, c_1^{j'}\right] = \alpha_a \left(\rho \ln w_1^j + (1-\rho)\ln \bar{w}\right) + \alpha_c^j \ln c_1^{j'} + \alpha_l^j l_1^{j'}, \tag{18}$$

[17] In the Appendix, we explore the robustness of these baseline results to variation in time discounting β, the number of generations T, the assumed persistence of type across generations ρ, and the number of types I, and other details of our specification. The qualitative and quantitative lessons of the baseline analysis prove to be robust, consistent with our message that a utilitarian welfare-maximizing policy that redistributes more than the status quo and shifts resources to earlier generations raises average ability over time and yields a sizeable welfare gain.

[18] In Appendix Table 5 we estimate the marginal effects of both parental disposable income and parental work effort on child ability, which we use as additional targets in a calibration exercise similar to that described in the main text of the paper.

where $l_1^{j'}$ is the labor effort of the parent of type j'. The calibration therefore searches for values of the parameters $\left\{\rho, \alpha_a, \left\{\alpha_c^j\right\}_{j=1}^I, \left\{\alpha_l^j\right\}_{j=1}^I, \sigma\right\}$ to match the empirical targets; with $I = 4$ there are 11 values to estimate. As before, we assume baseline values for $\rho = 0.50$ and $\sigma = 0.76$.

The results of the calibration exercise, indicate that the values for α_a and α_c^j (the weights on the parental ability and parental economic resources) are only slightly changed by the inclusion of parental effort. Most of the α_l parameters are near zero, though the substantially larger negative value for the highest ability type suggests that creating greater incentives for effort among the higher earners in this group may have some costs in terms of child ability; see the online appendix for further details.

As in the baseline case, the simulation does well in matching the empirical targets for which the data are most informative, the transition matrix and mean log wage. The calibrated marginal effects of parental resources and effort imply that child ability is increasing in parental resources and leisure (i.e., decreasing in parental labor effort), patterns not significantly apparent in the data. For brevity, we omit these results, which are very similar to the relevant analogues shown previously.

The optimal policy given the calibrated model also strongly resembles that in the baseline case (where parental effort did not affect children's abilities). In particular, the optimal policy is substantially more redistributive across incomes than the status quo policy and borrows from future generations. The ability distribution improves over time under the optimal policy, so that 1.3% more of the population is of the highest type and 1.5% less is of the lowest type in the fifth (middle) generation than under the status quo policy. The welfare gain from this improvement in the ability distribution, calculated just as in the main text, is 1.51% (compared to 1.75% in the baseline case). The slightly smaller values for the shifts in the ability distribution and welfare gain are due to the negative effects of increased parental work effort on child ability (the α_l coefficients), which offset in part the gains from having greater output and thus parental disposable income.

5 Conclusion

This paper starts with the recent empirical finding that the cognitive performance of children in poor families improves when their parents have more disposable income. How might tax policy respond to this relationship?

We take an initial step toward answering this question, providing both qualitative and quantitative results on tax policy in this context. First, we characterize conditions describing welfare-maximizing tax policy both within and across generations when children's abilities are a potentially complex function of inherited characteristics and parental (financial) resources. We use these conditions to better understand the ways in which this dependence changes the proper design of tax policy. Second, we use the model to quantify the potential benefits of well-designed policy. We specify the model's parameters to match new estimates of the intergenerational transmission of skills that we obtain by analyzing panel data from the NLSY. We then simulate the effects of the utilitarian welfare-maximizing policy in this model and show that its substantially greater progressivity shifts the ability distribution up over time. This shift generates an aggregate welfare gain equivalent to 1.75 percent of total disposable income in perpetuity, with larger gains for the poor. Even higher-skilled members of the current generation benefit substantially, however, as the welfare gains experienced by future generations increase the present value of the welfare of the current generation.

Our paper abstracts from an important issue: some individuals in the population do not have children. Thus, we must be careful to interpret our results, for example the welfare gains from policy change, as applying to *families with children*, not in the full population. Aside from this interpretive point, however, three factors suggest that omitting individuals without children will have at most minor effects on our results. First, the U.S. Department of Health and Human Services (2002) reports that 84 percent of men and 86 percent of women age 45 and older have ever had a biological child. In a benchmark lifecycle model, the lifetime tax burden is relevant for expenditure decisions, so our analysis of mothers with children captures the effects of our policy for the great majority of the population. Second, our calculations of the welfare gains from tax reform are restricted to its effects on the ability transition matrix. Those effects are relevant only for families with children, by definition. Finally, to the extent that tax policy can be (and often is, as with the Child Tax Credit, the EITC, or Temporary Assistance for Needy Families) made to depend on the presence of children, implementation of the policy change we consider may apply only to adults with children.

Nonetheless, it may be valuable to consider how including adults without children would affect our analysis. If the policy changes were not restricted to households with children, the additional revenue would be raised from and distributed to individuals with and without children. Because childless households have no effect on the ability transition matrix, the benefits and costs of redistribution—in terms of its effects on the ability transition matrix—would be reduced, and the welfare gains from reform may change.[19] In future work, it could be possible to model taxing or spending differentially on those with and without children, though in principle this could create distortions to the choice of whether to have children.

We take a "welfarist" approach to defining optimal policy by using the standard utilitarian social welfare function equal to the sum of utilities across the population. Philosophers and social choice theorists have suggested alternative goals for policy, including one of particular relevance to this paper: namely, equality of opportunity.[20] Roemer's (2000) leading treatment of that principle emphasizes that it requires a distinction between "before" and "after," in that opportunities should be equalized "before" competition among individuals begins but outcomes "after" competition begins should not. A classic tension in following this principle arises when parents' resources affect children's disutility of effort, so that what is "before" for children is "after" for their parents (see Fishkin 1987). A related point is that our use of a utilitarian social welfare function means that we do not fully insure children against "before" factors (those over which they have no control, such as their parents' financial resources) because the benefits from doing so must be weighed against the costs of dampening parents' incentives to earn more income. Future research could explore such issues.

Of course, future research may be able to improve our understanding of the tax policy studied in this paper. For example, when a panel dataset of sufficient duration allows us to link data on parents' and children's wages, this will allow estimates of the intergenerational effect of parental income on parent-child wage transitions. Exploring the potential for interactions between optimal tax policy and optimal government policy in other domains, including education, also seems to be a promising area for future research. Incorporating other dimensions of parental influence is another natural next step. We have shown that parental leisure versus work time does not seem to exert an important influence in this case, but one might study how the composition of parents' available resources (i.e., as disposable income or in-kind, such as education) affects the results. Such analyses may have implications for a broader class of policies that, like the taxes in this paper, could be used to affect–rather than merely respond to–the dynamics of the ability distribution.

[19] This discussion also raises important issues about policy's design toward the number of children in a household, which we leave to future work.

[20] Economists have begun including unconventional objectives in optimal tax research: see Weinzierl (2014, 2015), Saez and Stantcheva (2015), Lockwood, Nathanson, and Weyl (2015).

6 Disclosures

Gelber served as Deputy Assistant Secretary for Economic Policy at the U.S. Department of the Treasury from 2012 to 2013, although he had a minimal role in the development of tax policy.

We have no financial arrangements that might give rise to conflicts of interest with respect to the research reported in this paper.

References

[1] Albanesi, Stefania, and Christpher Sleet. 2006. "Dynamic Optimal Taxation with Private Information." *Review of Economic Studies* 73(1): 1-30.

[2] Akee, Randall K. Q., William E. Copeland, Gordon Keeler, Adrian Angold, and E. Jane Costello. 2010. "Parents' Income and Children's Outcomes: A Quasi-experiment." *American Economic Journal: Applied Economics* 2(1): 86-115.

[3] Anderberg, Dan. 2009. "Optimal Policy and the Risk Properties of Human Capital Reconsidered." *Journal of Public Economics* 93: 1017-1026.

[4] Barsky, Robert B., Thomas F. Juster, Miles S. Kimball, and Matthew D. Shapiro. 1997. "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study." *Quarterly Journal of Economics*, 112: 537-580.

[5] Becker, Gary S. 1981. "Altruism in the Family and Selfishness in the Market Place." *Economica* 48 (189): 1-15.

[6] Becker, Gary S., and H. Gregg Lewis. 1973. "On the Interaction between the Quantity and Quality of Children." *Journal of Political Economy* 81(2): 279-288.

[7] Becker, Gary S., and Nigel Tomes. 1976. "Part II: Labor Supply and the Family: Child Endowments and the Quantity and Quality of Children." *Journal of Political Economy* 84 (4): 143-162.

[8] Becker, Gary S., and Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics* 4(3): 1-39.

[9] Bernal, Raquel, and Michael P. Keane. 2011. "Child care choices and children's cognitive achievement: The case of single mothers." *Journal of Labor Economics* 29(3): 459-512.

[10] Best, Michael and Henrik Kleven. 2013. "Optimal Taxation with Career Effects of Work Effort." Working paper.

[11] Blau, David M. 1999. "The effect of child care characteristics on child development." *Journal of Human Resources* 34(4): 786-822.

[12] Blau, David M. and Janet Currie. 2004. "Preschool, Day Care, and Afterschool Care: Who's Minding the Kids," in *Handbook of the Economics of Education*, Hanushek, Eric and Finis Welch eds. Elsevier.

[13] Blau, Francine and Adam Grossberg, 1992, "Maternal Labor Supply and Children's Cognitive Development," *Review of Economics ad Statistics* 74(3): 474-481.

[14] Chetty, Raj. 2006. "A New Method of Estimating Risk Aversion." *American Economic Review* 96(5): 1821-1834.

[15] Chetty, Raj. 2012. "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply." *Econometrica* 80(3): 969-1018.

[16] Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, 2014, "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," Harvard University working paper.

[17] Dahl, Gordon, and Lance Lochner. 2012. "The Impact of Family Income on Child Achievement: Evidence from Changes in the Earned Income Tax Credit." *American Economic Review* 102(5): 1927-56).

[18] Dynan, Karen. 1993. "How Prudent Are Consumers?," *Journal of Political Economy*, 101, 1104-1113.

[19] Eissa, Nada, and Hilary Hoynes. 2006. "Behavioral Responses to Taxes: Lessons from the EITC and Labor Supply." *Tax Policy and the Economy* 20: 73-110.

[20] Farhi, Emmanuel, and Ivan Werning. 2010. "Progressive Estate Taxation." *Quarterly Journal of Economics* 125(2)..

[21] Farhi, Emmanuel, and Ivan Werning. 2013. "Insurance and Taxation over the Life Cycle." *Review of Economic Studies* 810, no. 2: 596–635.

[22] Feenberg, Daniel, and Elisabeth Coutts. 1993. "An introduction to the TAXSIM model." *Journal of Policy Analysis and Management* 12(1): 189-194.

[23] Feldman, Marcus W., Sarah P. Otto, and Freddy B. Christiansen. 2000. "Genes, Culture, and Inequality." In *Meritocracy and Inequality*, edited by Kenneth Arrow, Samuel Bowles, and Stephen Durlauf, 61-85. New Jersey: Princeton University Press.

[24] Fishkin, James. 1987. "Liberty Versus Equal Opportunity." *Social Philosophy and Policy Bulletin* 5(1): 32-48.

[25] Flynn, James R. 1987. "Massive IQ gains in 14 Nations: What IQ tests Really Measure." *Psychological Bulletin* 101(3): 171-191.

[26] Gelber, Alexander. 2011. "How do 401(k)s Affect Saving? Evidence from Changes in 401(k) Eligibility." *American Economic Journal: Economic Policy* 3(4):103-122.

[27] Gelber, Alexander, and Joshua Mitchell. "Taxes and Time Allocation: Evidence from Single Women and Men." *Review of Economic Studies* 79(3): 863-897.

[28] Gelber, Alexander, and Matthew Weinzierl. "Equalizing Abilities and Equalizing Opportunities: Optimal Taxation when Children's Abilities Depend on Parents' Resources." NBER Working Paper 18332.

[29] Goldin, Claudia, and Lawrence Katz. 2007. "Long-run Changes in the U.S. Wage Structure: Narrowing, Widening, Polarizing." In *Brookings Papers on Economic Activity* 2:2007, edited by Douglas Elmendorf and William Gale, 135-165. Washington: Brookings Institution Press.

[30] Golosov, Mikhail, and Aleh Tsyvinski. 2006. "Designing Optimal Disability Insurance: A Case for Asset Testing." *Journal of Political Economy* 114(2): 257-279.

[31] Golosov, Mikhail, Narayana Kocherlakota, and Aleh Tsyvinski. 2003. "Optimal Indirect and Capital Taxation." *Review of Economic Studies* 70(3): 569-587.

[32] Grochulski, Borys, and Tomasz Piskorski. 2010. "Risky Human Capital and Deferred Capital Income Taxation." *Journal of Economic Theory* 145(3): 908-943.

[33] Gruber, Jonathan. 2006. "A Tax-Based Estimate of the Intertemporal Elasticity of Substitution." NBER Working Paper 11945.

[34] Gruber, Jonathan, and Emmanuel Saez. 2002. “The Elasticity of Taxable Income: Evidence and Implications.” *Journal of Public Economics* 84(1): 1-33.

[35] Haider, Steven, and Gary Solon. 2006. "Life-Cycle Variation in the Association between Current and Lifetime Earnings." *American Economic Review* 96(4): 1308-1320.

[36] Hall, Robert. 1988. “Intertemporal Substitution in Consumption,” *Journal of Political Economy*, 96, 339-57.

[37] Hotz, V. Joseph, and John Karl Scholz. 2003. “The Earned Income Tax Credit.” In *Means-Tested Transfer Programs in the United States*, edited by Robert Moffitt, 141-198. Chicago: Chicago Press.

[38] Kapicka, Marek. 2006. “The Dynamics of Optimal Taxation when Human Capital is Endogenous.” UCSB working paper.

[39] Kapicka, Marek. 2006. “Optimal Income Taxation with Human Capital Accumulation and Limited Recordkeeping.” *Review of Economic Dynamics* 9(4): 612-639.

[40] Jacob, Brian, Max Kapustin, and Jens Ludwig. 2014. "Human Capital Effects of Anti-Poverty Programs: Evidence from a Randomized Housing Voucher Lottery." NBER Working Paper 20164.

[41] Kopczuk, Wojciech. 2013. "Incentive Effects of Inheritances and Optimal Estate Taxation," *American Economic Review: Papers and Proceedings*, 103(3), 472-77.

[42] Kotlikoff, Laurence, and David Rapson. 2007. "Does It Pay, at the Margin, to Work and Save? Measuring Effective Marginal Taxes on Americans’ Labor Supply and Savings.” *Tax Policy and the Economy* 21: 83-144.

[43] Leibowitz, Arleen. 1974. “Home investments in children.” *Journal of Political Economy* 82(2): 111-131.

[44] Løken, Katrine, Magne Mogstad, and Matthew Wiswall. 2012. “What Linear Estimators Miss: The Effects of Family Income on Child Outcomes.” *American Economic Journal: Applied Economics* 4(2): 1-35.

[45] Macours, Karen, Norbert Schady, and Renos Vakis. 2012. “Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment.” *American Economic Journal: Applied Economics* 4(2): 247-273.

[46] Mankiw, N. Gregory, and Matthew Weinzierl. 2010. "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution.” *American Economic Journal: Economic Policy* 2(1): 155-176.

[47] Milligan, Kevin, and Mark Stabile. "Do Child Tax Benefits Affect the Wellbeing of Children? Evidence from Canadian Child Benefit Expansions.” *American Economic Journal: Economic Policy* 3(3): 175–205.

[48] Mirrlees, James A. 1971. "An Exploration in the Theory of Optimal Income Taxation.” *Review of Economic Studies* 38(2): 175-208.

[49] Paxson, Christine, and Norbert Schady. 2007. "Cognitive Development Among Young Children in Ecuador: The Roles of Wealth, Health and Parenting." *Journal of Human Resources* 42(1): 49-84.

[50] Pence, Karen. 2006. "The Role of Wealth Transformations: An Application to Estimating the Effect of Tax Incentives on Saving." *B.E. Journal of Economic Analysis and Policy* 5(1): 1-24.

[51] Piketty, Thomas, and Emmanuel Saez. 2013. "A Theory of Optimal Inheritance Taxation." *Econometrica* 81(5): 1851-1886.

[52] Ramsey, Frank P. 1928. "A Mathematical Theory of Saving," *Economic Journal,* 38(152), December.

[53] Roemer, John. 2000. "Equality of Opportunity." Cambridge, MA: Harvard University Press.

[54] Rothschild, Casey and Florian Scheuer. 2011. "Optimal Taxation with Rent-Seeking," *NBER Working Paper* 17035.

[55] Ruhm, Christopher J. 2004. "Parental employment and child cognitive development." *Journal of Human Resources* 39(1): 155-192.

[56] Stantcheva, Stefanie. 2014. "Optimal Taxation and Human Capital Policies over the Life Cycle", Working paper.

[57] Tuomala, Matti. 1990. *Optimal Income Tax and Redistribution.* New York: Oxford University Press.

[58] U.S. Department of Health and Human Services. 2002. "Charting Parenthood: A Statistical Portrait of Fathers and Mothers in America." Washington, D.C.: Government Printing Office.

[59] Weinzierl, Matthew. 2011. "The Surprising Power of Age-Dependent Taxes." *Review of Economic Studies* 78(4): 1490-1518.

Appendix to: Optimal Taxation when Children's Abilities Depend on Parents' Resources

Alexander Gelber and Matthew Weinzierl

September 18, 2015

1 Related theoretical literature

This paper introduces a new element to the active literature in dynamic optimal taxation. Following the original contribution of Mikhail Golosov, Narayana Kocherlakota, and Aleh Tsyvinski (2003), most work in this area has considered the impact of stochastic and exogenous skill processes on the optimal taxation of an individual over his lifetime. Contributions include Kocherlakota (2005), Golosov and Tsyvinski (2006), Albanesi and Sleet (2006), Golosov, Tsyvinski, and Werning (2007), Farhi and Werning (2010, as discussed in the paper), and Weinzierl (2011).

We analyze optimal tax policy when the skill distribution of one generation depends on the skill distribution and the choices of the previous generation (subject to stochastic shocks). Because we allow the skill distribution to be endogenously determined, our paper is closely related to another body of work that extends the original dynamic optimal tax literature by allowing individuals' choices to affect their own ability levels (see Casey Rothschild and Florian Scheuer 2011 or Michael Best and Henrik Kleven 2013, for example). Other examples include the following. Marek Kapicka (2006a, 2006b) allows a deterministic skill process to be endogenous and has heterogeneity in natural ability, but each type is fixed for life, and all types share the same human capital production function. Borys Grochulski and Tomasz Piskorski (2010) allow a population of identical agents to choose a human capital investment, the output and depreciation of which are stochastic, thus combining stochasticity with a form of endogeneity. Grochulski and Piskorski (2010) have no heterogeneity outside of shocks to the human capital production function, the returns to which are therefore not dependent on natural ability. Dan Anderberg (2009) extends that approach by allowing for heterogeneous ability shocks, the effects of which on earnings can be magnified or reduced by human capital investment undertaken by identical agents before the ability shocks are realized, but human capital investment decisions are made by agents before their ability heterogeneity is realized.

2 Proof of Lemma 1

The planner's problem yields these first-order conditions for c_t^k and y_t^k:

$$\left[\begin{array}{c} u'\left(c_t^k\right)\left(1+\beta\sum_j \frac{\partial p^j\left(w_t^k,c_t^k\right)}{\partial c_t^k}\frac{U_{t+1}^j}{u'\left(c_t^k\right)}\right) \\ *\left(\begin{array}{c} \beta^t\pi_t^k+\sum_{\tau=1}^{t-1}\beta^{t-\tau}\sum_i\sum_{i'}\mu_\tau^{i'|i}\left(\pi_t^k|_{c_\tau^i}-\pi_t^k|_{c_\tau^{i'}}\right) \\ +\sum_{k'}\mu_t^{k'|k}-\sum_{k'}\mu_t^{k|k'}\frac{1+\beta\sum_j\frac{\partial p^j\left(w_t^{k'},c_t^k\right)}{\partial c_t^k}\frac{U_{t+1}^j}{u'\left(c_t^k\right)}}{1+\beta\sum_j\frac{\partial p^j\left(w_t^k,c_t^k\right)}{\partial c_t^k}\frac{U_{t+1}^j}{u'\left(c_t^k\right)}} \end{array}\right) \end{array}\right]$$
$$= \lambda\beta^t\pi_t^k\left(1-\beta\sum_j\frac{\partial p^j\left(w_t^k,c_t^k\right)}{\partial c_t^k}R_{t+1}^j\right),$$

$$\frac{1}{w_t^k}v'\left(\frac{y_t^k}{w_t^k}\right)\left(\begin{array}{c} \beta^t\pi_t^k+\sum_{\tau=1}^{t-1}\beta^{t-\tau}\sum_i\sum_{i'}\mu_\tau^{i'|i}\left(\pi_t^k|_{c_\tau^i}-\pi_t^k|_{c_\tau^{i'}}\right) \\ +\sum_{k'}\mu_t^{k'|k}-\sum_{k'}\mu_t^{k|k'}\frac{\frac{1}{w_t^{k'}}v'\left(\frac{y_t^k}{w_t^{k'}}\right)}{\frac{1}{w_t^k}v'\left(\frac{y_t^k}{w_t^k}\right)} \end{array}\right)=\lambda\beta^t\pi_t^k$$

Simplifying by eliminating λ and denoting terms as in the text yields the Lemma, where

$$\mathbb{A}_t^j=\frac{1}{1-\beta\sum_k\frac{\partial p^k\left(w_t^j,c_t^j\right)}{\partial c_t^j}R_{t+1}^k},\quad \mathbb{B}_t^j=\beta^t\pi_t^j+\sum_{\tau=1}^{t-1}\beta^{t-\tau}\sum_i\sum_{i'}\mu_\tau^{i'|i}\left(\pi_t^j|_{c_\tau^i}-\pi_t^j|_{c_\tau^{i'}}\right), \tag{1}$$

$$\mathbb{C}_t^j=\sum_{j'}\mu_t^{j'|j}-\sum_{j'}\mu_t^{j|j'}\frac{1+\beta\sum_k\frac{\partial p^k\left(w_t^{j'},c_t^j\right)}{\partial c_t^j}\frac{U_{t+1}^k}{u'\left(c_t^j\right)}}{1+\beta\sum_k\frac{\partial p^k\left(w_t^j,c_t^j\right)}{\partial c_t^j}\frac{U_{t+1}^k}{u'\left(c_t^j\right)}},\quad \mathbb{D}_t^j=\sum_{j'}\mu_t^{j'|j}-\sum_{j'}\mu_t^{j|j'}\frac{\frac{1}{w_t^{j'}}v'\left(\frac{y_t^j}{w_t^{j'}}\right)}{\frac{1}{w_t^j}v'\left(\frac{y_t^j}{w_t^j}\right)} \tag{2}$$

where $\pi_t^j|_{c_\tau^i}$ is the probability that a generation t descendant of parent type i from generation τ is of type j and $\sum_i \pi_t^j|_{c_\tau^i}$ is denoted by the unconditional probability π_t^j. In words, A_t^j depends on $\beta\sum_k\frac{\partial p^k\left(w_t^j,c_t^j\right)}{\partial c_t^j}R_{t+1}^k$, the weighted present value sum of net revenues obtained across types over time, with the weight on type k in generation $t+1$ representing the increase in probability that children of parent type j will be type k when c_t^j is increased slightly.

2.1 Optimal condition with two ability types

We assume that the incentive constraints bind "downward," as is the standard case in Mirrleesian optimal tax models. Formally, we assume that $w^j > w^i$ and that $\mu_t^{i|j} > 0$ but $\mu_t^{j|i} = 0$ for all generations t. Then, the result for each ability type in generation t is as follows:

$$\frac{v'\left(y_t^i/w_t^i\right)}{w_t^i u'\left(c_t^i\right)} =$$

$$\left(\frac{1+\beta\sum_k \frac{\partial p^k\left(w_t^j,c_t^j\right)}{\partial c_t^j}\frac{U_{t+1}^k}{u'\left(c_t^j\right)}}{1-\beta\sum_k \frac{\partial p^k\left(w_t^i,c_t^i\right)}{\partial c_t^i}R_{t+1}^k}\right)$$

$$*\left(\frac{\beta^t\pi_t^i+\sum_{\tau=1}^{t-1}\beta^{t-\tau}\mu_\tau^{i|j}\left(\pi_t^i|_{c_\tau^j}-\pi_t^i|_{c_\tau^i}\right)-\mu_t^{i|j}\frac{1+\beta\sum\limits_k \frac{\partial p^k\left(w_t^j,c_t^i\right)}{\partial c_t^i}\frac{U_{t+1}^k}{u'\left(c_t^i\right)}}{1+\beta\sum\limits_k \frac{\partial p^k\left(w_t^i,c_t^i\right)}{\partial c_t^i}\frac{U_{t+1}^k}{u'\left(c_t^i\right)}}}{\beta^t\pi_t^i+\sum_{\tau=1}^{t-1}\beta^{t-\tau}\mu_\tau^{i|j}\left(\pi_t^i|_{c_\tau^j}-\pi_t^i|_{c_\tau^i}\right)-\mu_t^{i|j}\frac{\frac{1}{w_t^j}v'\left(\frac{y_t^i}{w_t^j}\right)}{\frac{1}{w_t^i}v'\left(\frac{y_t^i}{w_t^i}\right)}}\right) \tag{3}$$

$$\frac{v'\left(y_t^j/w_t^j\right)}{w_t^j u'\left(c_t^j\right)} = \frac{1}{1-\beta\sum_k \frac{\partial p^k\left(w_t^j,c_t^j\right)}{\partial c_t^j}R_{t+1}^k}\left(1+\beta\sum_k \frac{\partial p^k\left(w_t^j,c_t^j\right)}{\partial c_t^j}\frac{U_{t+1}^k}{u'\left(c_t^j\right)}\right) \tag{4}$$

3 Proof of Proposition 1

Rewrite the first-order condition for disposable income from the proof of Lemma 1 as

$$\beta^t\pi_t^k+\sum_{\tau=1}^{t-1}\beta^{t-\tau}\sum_i\sum_{i'}\mu_\tau^{i'|i}\left(\pi_t^k|_{c_\tau^i}-\pi_t^k|_{c_\tau^{i'}}\right)+\sum_{k'}\mu_t^{k'|k}-\sum_{k'}\mu_t^{k|k'}\frac{1+\beta\sum\limits_j \frac{\partial p^j\left(w_t^{k'},c_t^k\right)}{\partial c_t^k}\frac{U_{t+1}^j}{u'\left(c_t^k\right)}}{1+\beta\sum\limits_j \frac{\partial p^j\left(w_t^k,c_t^k\right)}{\partial c_t^k}\frac{U_{t+1}^j}{u'\left(c_t^k\right)}}$$

$$= \lambda\beta^t\frac{\pi_t^k}{u'\left(c_t^k\right)}\frac{1-\beta\sum\limits_j \frac{\partial p^j\left(w_t^k,c_t^k\right)}{\partial c_t^k}R_{t+1}^j}{1+\beta\sum\limits_j \frac{\partial p^j\left(w_t^k,c_t^k\right)}{\partial c_t^k}\frac{U_{t+1}^j}{u'\left(c_t^k\right)}}.$$

Then, sum each side over k:

$$\sum_k \left(\begin{array}{c} \beta^t \pi_t^k + \sum_{\tau=1}^{t-1} \beta^{t-\tau} \sum_i \sum_{i'} \mu_\tau^{i'|i} \left(\pi_t^k|_{c_\tau^i} - \pi_t^k|_{c_\tau^{i'}} \right) \\ + \sum_{k'} \mu_t^{k'|k} - \sum_{k'} \mu_t^{k|k'} \frac{1+\beta \sum\limits_j \frac{\partial p^j\left(w_t^{k'}, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)}}{1+\beta \sum\limits_j \frac{\partial p^j\left(w_t^{k}, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)}} \end{array} \right)$$

$$= \sum_k \lambda \beta^t \frac{\pi_t^k}{u'\left(c_t^k\right)} \frac{1-\beta \sum\limits_j \frac{\partial p^j\left(w_t^k, c_t^k\right)}{\partial c_t^k} R_{t+1}^j}{1+\beta \sum\limits_j \frac{\partial p^j\left(w_t^k, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)}}.$$

We can rewrite this as:

$$\beta^t + \sum_k \left(\sum_{\tau=1}^{t-1} \beta^{t-\tau} \sum_i \sum_{i'} \mu_\tau^{i'|i} \left(\pi_t^k|_{c_\tau^i} - \pi_t^k|_{c_\tau^{i'}} \right) + \sum_{k'} \mu_t^{k'|k} - \sum_{k'} \mu_t^{k|k'} \frac{1+\beta \sum\limits_j \frac{\partial p^j\left(w_t^{k'}, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)}}{1+\beta \sum\limits_j \frac{\partial p^j\left(w_t^{k}, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)}} \right)$$

$$= \lambda \beta^t \sum_k \frac{\pi_t^k}{u'\left(c_t^k\right)} \frac{1-\beta \sum\limits_j \frac{\partial p^j\left(w_t^k, c_t^k\right)}{\partial c_t^k} R_{t+1}^j}{1+\beta \sum\limits_j \frac{\partial p^j\left(w_t^k, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)}}.$$

Rearranging, we obtain

$$\frac{1}{\lambda} = \frac{\sum_k \frac{\pi_t^k}{u'\left(c_t^k\right)} \frac{1-\beta \sum\limits_j \frac{\partial p^j\left(w_t^k, c_t^k\right)}{\partial c_t^k} R_{t+1}^j}{1+\beta \sum\limits_j \frac{\partial p^j\left(w_t^k, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)}}}{1 + \sum_{\tau=1}^{t-1} \frac{\beta^{t-\tau}}{\beta^t} \sum_i \sum_{i'} \mu_\tau^{i'|i} \sum_k \left(\pi_t^k|_{c_\tau^i} - \pi_t^k|_{c_\tau^{i'}} \right) + \frac{\beta}{\beta^t} \sum_k \sum_{k'} \mu_t^{k|k'} \frac{\sum\limits_j \frac{\partial p^j\left(w_t^k, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)} - \sum\limits_j \frac{\partial p^j\left(w_t^{k'}, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)}}{1+\beta \sum\limits_j \frac{\partial p^j\left(w_t^k, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)}}}.$$

This holds for all t, yielding the Proposition.

$$\frac{1}{\Lambda_t} \sum_j \left[\frac{\pi_t^j}{u'\left(c_t^j\right)} \frac{1-\beta \sum_k \frac{\partial p^k\left(w_t^j, c_t^j\right)}{\partial c_t^j} R_{t+1}^k}{1+\beta \sum_k \frac{\partial p^k\left(w_t^j, c_t^j\right)}{\partial c_t^j} \frac{U_{t+1}^k}{u'\left(c_t^j\right)}} \right]$$

$$= \frac{1}{\Lambda_{t+1}} \sum_k \left[\frac{\pi_{t+1}^k}{u'\left(c_{t+1}^k\right)} \frac{1-\beta \sum_l \frac{\partial p^l\left(w_{t+1}^k, c_{t+1}^k\right)}{\partial c_{t+1}^k} R_{t+2}^l}{1+\beta \sum_l \frac{\partial p^l\left(w_{t+1}^k, c_{t+1}^k\right)}{\partial c_{t+1}^k} \frac{U_{t+2}^l}{u'\left(c_{t+1}^k\right)}} \right],$$

where

$$\Lambda_t = 1 + \sum_{\tau=1}^{t-1} \frac{\beta^{t-\tau}}{\beta^t} \sum_i \sum_{i'} \mu_\tau^{i'|i} \sum_k \left(\pi_t^k|_{c_\tau^i} - \pi_t^k|_{c_\tau^{i'}} \right) + \frac{\beta}{\beta^t} \sum_k \sum_{k'} \mu_t^{k|k'} \frac{\sum_j \frac{\partial p^j\left(w_t^k, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)} - \sum_j \frac{\partial p^j\left(w_t^{k'}, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)}}{1 + \beta \sum_j \frac{\partial p^j\left(w_t^k, c_t^k\right)}{\partial c_t^k} \frac{U_{t+1}^j}{u'\left(c_t^k\right)}}.$$

4 Empirical Strategy

This section describes the specification that we use in our empirical estimates, our data, and our empirical results.

4.1 Regression Specification

Let x_i denote observable characteristics, η_{ia} denote time-varying unobserved shocks to the child or family, and c_{ia} denote total family disposable income for child i at age a (where income is measured net of any taxes and transfers, including EITC payments).[1] Child outcomes are denoted w_{ia}, which are a function of the child's and parents' characteristics and income. $\chi_a^{s_{ia}}(y_{ia})$ denotes EITC income, which is a function of pre-tax income, y_{ia}. Taxes other than the EITC are denoted $T_a^{s_{ia}}(P_{ia})$. The EITC schedules vary within a year based on income and number of children, and the EITC schedules also vary across years. The superscript s_{ia} on the EITC and tax functions denotes which schedule a child's family is on; the tax schedules may vary based upon the number of children in the household and marital status. Family disposable income is $c_{ia} = y_{ia} + \chi_a^{s_{ia}}(y_{ia}) - T_a^{s_{ia}}(y_{ia})$.

We use $\chi_a^{IV}(y_{i,a-1}) \equiv \chi_a^{s_{i,a-1}}(\hat{E}[y_{i.a}|y_{i.a-1}]) - \chi_{a-1}^{s_{i,a-1}}(y_{i,a-1})$ to instrument for the change in family disposable income from age $a-1$ to age a.[2] Here $\hat{E}[y_{i.a}|y_{i.a-1}]$ represents predicted pre-tax income at age a conditional on pre-tax income at age $a-1$. Following Dahl and Lochner, in order to calculate $\hat{E}[y_{i.a}|y_{i.a-1}]$, we regress pre-tax income on an indicator for positive lagged pre-tax income and a fifth-order polynomial in lagged pre-tax income, and then we obtain the fitted values. As in Jonathan Gruber and Emmanuel Saez (2002), we predict changes in EITC payments by applying the change in the EITC schedule to predicted current income, where the prediction is based on lagged pre-tax income. We exploit variation in predicted EITC income resulting only from policy changes in EITC schedules over time, as opposed to those resulting from changes in family structure, because we hold the type of EITC schedule (e.g. one versus two children) fixed over time.

As both Gruber and Saez and Dahl and Lochner note, the autoregressive process determining income is likely to include serially correlated income shocks. Using χ_a^{IV} as an instrument, without conditioning on lagged income, is therefore likely to yield biased and inconsistent estimates of the coefficient on parent income. This is because predicted changes in EITC payments depend on pre-tax family income at age $a-1$, namely $y_{i,a-1}$, which will be correlated with the subsequent change in income if, for example, mean reversion characterizes the evolution of income. Therefore, following Gruber and Saez and Dahl and Lochner, we control in the regression for a flexible function $\Phi(y_{i,a-1})$ of $y_{i,a-1}$. Like Dahl and Lochner, we specify this function $\Phi(y_{i,a-1})$ as an indicator for positive lagged pre-tax income and a fifth order polynomial in lagged pre-tax income.

Dahl and Lochner estimate the effect of parental after-tax income on children's ability, but in our calibration later, we will be interested in a related but different object: the effect of parental after-tax income on the probability that a child of a given ability level, conditional on the parent having a given (and

[1] The subscript i indexes individuals in this section; this should not be confused with the superscript i in the model in Section 1.

[2] We use "initial period" to refer to child age $a-1$ and "final period" to refer to child age a.

possibly different) ability level. We adapt the Dahl and Lochner empirical specification by estimating the following model:

$$D_{ia}^{l} = x_i'\alpha + \Delta c_{ia}\beta + W_{i,a-1}\delta + \Phi(y_{i,a-1}) + \eta_{ia} \tag{5}$$

using χ_a^{IV} as an instrument for Δc_{ia}. We relate a binary dummy D_{ia}^{l} equal to one when the child is in ability category l to observable characteristics x (which include child gender, age, and number of siblings), the change in parental income over the period in question Δc_{ia}, a vector of dummies $W_{i,a-1}$ for whether the child's lagged ability level (at age $a-1$) fell in each of the ability categories, and the flexible function $\Phi(y_{i,a-1})$ of lagged pre-tax income.[3] We run this regression separately for parents of different ability (i.e. wage) types, to investigate the separate effect of parental income on child ability among each type of parents. Intuitively, for a regression involving a given parent wage category, the coefficient β approximately tells us the effect of a 100% increase in parental disposable income on the fraction of children ending up in a given ability category, conditional on the parent being in the wage category in question, and given the child's initial ability level. By controlling for lagged child ability, we effectively remove permanent differences in child ability levels across families. Thus, our specification effectively relates changes in child ability to (instrumented) changes in parental income, using policy changes in EITC schedules to predict differential changes in after-tax family income across families. We run a linear probability model to estimate (4) because a logit or probit model would lead to an incidental parameters problem.[4]

To address the possibility that parents' hours worked could affect children's ability, we also explore specifications in which we additionally control for the first-difference of parents' hours worked from the initial period to the final period. However, it is worth bearing in mind that we have only one instrument (i.e. simulated changes in after-tax income) but two endogenous variables (i.e. the first-difference of parent after-tax income and the first-difference of parent hours worked). Thus, we are not able to instrument for both independent variables simultaneously; we simply control for the first-difference in hours worked while instrumenting for the change in after-tax income with the simulated change. As a result, the estimate of the effect of hours worked on children's ability should be considered suggestive. We therefore face a tradeoff: the estimate of this effect is more suggestive, but it allows us to simulate a richer model in which parents' time allocation may affect children's outcomes. We explore both possibilities in simulations, and they both yield similar results.[5]

As we discuss later, the formal model whose moments we will match to the data will be specified in

[3] This specification implicitly makes assumptions that mirror those made in Dahl and Lochner. First, parental income has an effect on child ability that is the same at all child ages. Second, conditional on lagged child ability, lagged changes in income have no effect on current income.

[4] Of course, a well-known limitation of linear probability models is that they may predict probabilities outside of the range [0,1]. We consider estimation of consistent effects to be the more important consideration, and thus we estimate a linear probability model rather than a logit or probit. Running a Chamberlain random effects ordered probit gives similar results to those shown but entails additional assumptions about the distribution of the random effect.

[5] Examining proxies for parents' time with children in the data—such as measures in the National Longitudinal Study of Youth like parents' participation in parent-teacher conferences—yielded unstable and typically insignificant point estimates. Other papers such as Bernal and Keane (2011), Blau and Grossberg (1992), Blau (1999), Ruhm (2004), and Blau and Currie (2004) have investigated the effect of parental employment and other home inputs on child outcomes.

terms of the effect of log parental income on child ability. Thus, it is useful for us to estimate the effect of log parental income on child ability, and ideally Δc_{ia} would represent the change in log parental income over the period in question. However, estimating exactly this specification would lead to a problem: the log of zero is undefined, but we would like to include individuals in the regressions whose parents may have had income of zero in the final period. Thus, we approximate log income using the inverse hyperbolic sine of income. The inverse hyperbolic sine approximates the log function but is defined at zero values (e.g. see similar work in Karen Pence 2006 or Alexander Gelber 2011).[6] It is important to emphasize that our results are similar when we use several alternative specifications: a linear specification (which is less compatible with our formal model but which allows us to include zero values of parental income); a specification in which we add 1 to income before logging it (which clearly allows us to log income, at the cost of adding an arbitrary value to income before logging it); and a specification in which we simply log income and discard observations in which income is zero (whose sample size is substantially reduced from the sample size we use in our regressions).

4.2 Data

The NLSY has not yet generally followed a sufficient number of children to an age when they can be observed participating in the labor force with their post-schooling wage, so we follow Dahl and Lochner in using child test scores as a measure of child ability.[7] We control for the child's initial test score category (i.e. by quartile), but the results are very similar when we instead control for linear or higher-order terms in the child's initial test score. We have measured parent wage category using their wages at the beginning of the sample period, so that their wages are not affected by subsequent EITC variation. To calculate the hourly wage, we divide earnings by hours worked for NLSY survey respondents. Over 99% of respondents are mothers.[8]

Our sample of children is constructed as Dahl and Lochner construct their sample, as described presently. The sample contains children observed in at least two consecutive even-numbered survey years between 1988 and 2000 with valid scores, family background characteristics, and family income measures. Our sample follows children over this period. We calculate each family's state and federal EITC payment and tax burden using the TAXSIM program (version 9) (Daniel Feenberg and Elizabeth Coutts, 1993). We also limit our sample to children whose mothers did not change marital status during two-year intervals when test scores are measured. Our main sample includes 3,714 interviewed children born to 2,108 interviewed mothers, with children observed 2.9 times on average. From 1986 to 2000, the tests were administered every two years to children ages five and older. Children took each individual test at most five times.

Appendix Table 1 shows summary statistics. Children's mean age is 11.31 years old. Nearly half of the

[6] The inverse hyperbolic sine of A is defined as $\sinh^{-1}(A) = \ln(A + \sqrt{1 + A^2})$. The change in parental income Δc_{ia} is therefore defined as $\Delta c_{ia} = \sinh^{-1}(c_{ia}) - \sinh^{-1}(c_{ia-1})$, where c_{ia-1} represents parent income when the child was age $a-1$. A more general form of the inverse hyperbolic sine function adds a scaling parameter; our results are similar when we use other scaling parameters.

[7] Our calibration therefore assumes that child test scores translate into hourly wages, as models of wage determination predict.

[8] The number of observations in our regressions is somewhat smaller than the number of observations in the baseline sample in Dahl and Lochner. Some respondents do not work in the initial period, implying that their hourly wage is unobserved. We drop these individuals from the sample, so that our sample consists only of working individuals.

children are male. Respondents work a mean of 1,692.68 hours per year. The mean calculated hourly wage is \$8.16/hour.

4.3 Results

As a preliminary step, we can consider simplified empirical exercises designed to test the viability of our approach by assessing whether increases in parental income increase the probability that children are high-ability. In Appendix Table 2, we show the results of a regression in which the dependent variable is binary, taking the value of 1 when the child has an above-median score in the final period and a value of 0 when the child has a below-median score in the final period. The right-hand-side of this regression is identical to the main regression specification (5) above, including a binary dummy measuring whether the lagged child test score is above or below the median. Increases in parental disposable income increase the dependent variable positively and significantly (at the 1% level). The point estimate shows that a 1% increase in parental income causes an increase in the probability that the child is in the high ability category of 0.67 percentage points. Evaluating this at the mean of parental income (\$34,679.13), this point estimate implies that a \$1,000 increase in parent income causes an increase in the probability that the child is in the high ability category of 1.93 percentage points, which represents a moderate-sized impact that makes sense in light of the moderate impacts that Dahl and Lochner found in their paper. Controlling for the first-difference of parent hours worked shows that parent hours worked has a negative effect on child ability that is significant at the 5% level. However, this effect is very small, implying that doubling parent hours would cause child ability to fall by only 0.1 standard deviation on average.

In Appendix Table 3, we illustrate the heterogeneity of the results across low- and high-wage parents, using only two ability types in order to increase the power of the estimates. This illustrates the viability of the approach and gives a sense of the variation in the data that underlies our later empirical estimates with more ability types. For both low- and high-ability parents, the point estimates of the coefficients are positive, as we would expect: higher parental income increases the probability that a child is high-ability. The point estimates are moderate-sized and reasonable. For low-ability parents, the coefficient is significantly different from zero at the 5% level: parental income has a positive, substantial, and statistically significant impact on the probability that a child is high-ability. However, it is worth noting that the coefficient is smaller and insignificantly different from zero among high-ability parents. The point estimates show that a 1% increase in parental income among low-ability parents causes a 0.75 percentage point increase in the probability that a child is high-ability, and that a 1% increase in parental income among high-ability parents causes a 0.57 percentage point inrease in the probability that a child is in the high-ability category. Evaluating these point estimates at the mean of parental income implies that a \$1,000 increase in parental income among low-ability parents causes a 2.80 percentage point increase in the probability that a child is high-ability, and that a \$1,000 increase in parental income among low-ability parents causes a 1.35 percentage point increase in the probability that a child is high-ability (substantially lower than the increase among low-ability parents). The regressions that additionally control for the first-difference of parent hours worked show a very small and insignificant impact of parent hours on child ability, as we discuss further in the Appendix.

The signs of the coefficients in our main results in Appendix Table 4 generally conform to expectations:

higher parental income usually increases the probability that a child is high-ability (i.e. in the third or fourth quartile) and decreases the probability that a child is low-ability (i.e. in the first or second quartile), with the "correct" sign of the relationship occurring in 13 out of 16 regressions.[9] In one case, the point estimate is statistically significantly different from zero (at the 10% level), but the other estimates are insignificant.[10] The insignificance of the other estimates should be unsurprising given that the sample size is limited when we use four parental types, and that there are a limited number of times that the dependent variable takes the value of 1 when we use four child types. Nonetheless, it is reassuring that the relationship predominantly takes the correct sign. Moreover, in the three cases in which the relationship takes the "wrong" sign, only one of these point estimates shows a non-negligible effect: only one shows that the effect of a 1% increase in parental income causes a change in the child's probability of being in a given a category larger than 0.1 percentage point.

[9] The point estimates of the coefficients in Table 1 sum to zero across child types (within a parent type). Tax schedule changes affecting middle-class individuals help to drive identification of the effects for our highest-income group.

[10] The point estimate that is significant also takes the "correct" sign, showing that increases in income for parents in the second quartile decrease the probability that children are in the first quartile of ability.

5 Appendix Tables 1-8

Appendix Table 1: Summary Statistics

(1) Variable	(2) Mean	(3) Standard deviation
Household Income	34,679.13	22,969.28
Hours worked of respondent	1,692.68	785.42
Hourly wage	8.16	3.25
Child age	11.31	2.02
Child male (dummy)	0.497	0.500

Notes: The table shows the means and standard deviations of the key variables used. The data are taken from the NLSY, with sample restrictions corresponding to the baseline specification in Column 1 of Table 3 of Dahl and Lochner (forthcoming). The variable in question is shown in each row in Column 1, the mean is shown in Column 2, and the standard deviation in Column 3. The hourly wage is calculated as a respondent's earnings divided by a respondent's yearly hours worked. The number of observations in the full sample is 6,902, corresponding to 2,108 mothers and 3,714 children. Income is measured in year 2000 dollars.

Appendix Table 2 shows the effect of parent after-tax income on child's probability of being high-ability. Two-stage least squares results.

Appendix Table 2:
Effect of parent after-tax income on child ability

Income	0.67	0.66
	(0.26)**	(0.27)**
Parent Hours Worked		-.000059
		(.000027)**
N	6,902	6,602
n	2,108	2,057

Notes: The dependent variable is a dummy variable that equals 1 if a child's measured ability is above the median. Child ability is measured by their test scores on math and reading components of the PIAT, as described in the text and in Dahl and Lochner (forthcoming). Child test scores are measured at the end of each sample period. This binary variable is regressed on the change in the parent's net-of-tax income (instrumented using the change in the parent's net-of-tax income predicted using lagged income), a fifth-order polynomial in lagged income, an indicator for positive lagged income, a dummy that equals one if the child's lagged test score is above the median (and zero otherwise), gender, age, and number of siblings. The number of children is 3,714. N refers to the number of observations, and n refers to the number of parents included in each regression. The sample size is slightly smaller when controlling for the first-difference of parent hours worked because hours worked is occasionally missing. The table shows the coefficient on income, with the standard error below in parentheses. Parent income is measured in 1,000's of year 2000 dollars. To approximate the log functional form, we take the inverse hyperbolic sine of income in each period before we first-difference it, so that we approximately estimate the effect of log income on child ability, as described in the text. Standard errors are clustered at the level of the mother. The regression controls for child gender, age, and number of siblings. *** denotes significance at the 1% level.

Appendix Table 3: Empirical marginal effects of parental resources on child ability distribution (percentage points)

	Low-Ability Parents		High-Ability Parents	
Parent consumption	0.75	0.76	0.57	0.62
	(0.37)**	(0.39)*	(0.38)	(0.47)
Parent hours worked		-0.000052		-0.000076
		(0.000036)		(0.000055)
N	3,365	3,194	3,537	3,408
n	1,054	1,022	1,054	1,035

Notes: The dependent variable is a dummy variable that equals 1 if the child has above-median test scores on math and reading components of the PIAT, as described in the text and in Dahl and Lochner (forthcoming). The dependent variable is regressed on the change in the parent's net-of-tax income (instrumented using the change in the parent's net-of-tax income predicted using lagged income), a fifth-order polynomial in lagged income, an indicator for positive lagged income, gender, age, number of siblings, and dummies for each child's lagged test score category. This regression is run separately for parents who have below and above-median hourly wages, corresponding to the regressions in Columns A and B, respectively. The table shows the estimated coefficient, with the standard error below in parentheses. Parent wage is measured as the mean hourly wage over the sample period, and child test score is measured at the end of each sample period. N refers to the number of observations in the regression, and n refers to the number of mothers (which is the same as the number of clusters). N refers to the number of observations, and n refers to the number of parents. The sample size is slightly smaller when controlling for the first-difference of parent hours worked because hours worked is occasionally missing. The total number of children in the full sample (including both those in the high-ability and the low-ability parent groups) is 3,714. Parent income is measured in 1,000's of year 2000 dollars. To approximate the log functional form, we take the inverse hyperbolic sine of income in each period before we first-difference it, so that we approximately estimate the effect of log income on the dependent variable, as described in the text. Standard errors are clustered at the level of the mother.

Appendix Table 4: Empirical marginal effects of parental resources on child ability distribution, in percentage points						
	Child type					
Parent type	1	2	3	4	N	n
1 (lowest)	-0.10	-0.34	0.80	-0.36	1,544	527
	(1.22)	(1.54)	(1.97)	(1.09)		
2	-0.56	0.04	0.25	0.27	1,907	527
	(0.32)*	(0.30)	(0.24)	(0.24)		
3	-0.12	-1.04	1.26	-0.10	1,813	527
	(0.60)	(0.97)	(1.06)	(0.54)		
4 (highest)	-0.05	-0.28	0.07	0.26	1,638	527
	(0.39)	(0.56)	(0.47)	(0.49)		

Notes: The dependent variable is a dummy variable that equals 1 if the child is in a given ability quartile. The child's ability is measured by test scores on math and reading components of the PIAT, as described in the text and in Dahl and Lochner (forthcoming). The dependent variable is regressed on the change in the parent's net-of-tax income (instrumented using the change in the parent's net-of-tax income predicted using lagged income), a fifth-order polynomial in lagged income, an indicator for positive lagged income, gender, age, number of siblings, and dummies for each child's lagged test score category. This regression is run separately for parents who have wages in each quartile of the parent wage distribution. The table shows the estimated coefficient, with the standard error below in parentheses. Parent wage is measured as the mean hourly wage over the sample period, and child test score is measured at the end of each sample period. Standard errors are clustered at the level of the mother. N refers to the number of observations, and n refers to the number of mothers (which is the same as the number of clusters), included in each of four regressions estimated in a parent quartile. N differs across regressions because the number of missing observations differs across parents. The total number of children in the full sample (including both those in the high-ability and the low-ability parent groups) is 3,714. Parent income is measured in 1,000's of year 2000 dollars. To approximate the log functional form, we take the inverse hyperbolic sine of income in each period before we first-difference it, so that we approximately estimate the effect of log income on the dependent variable, as described in the text.

Appendix Table 5: Empirical marginal effects of parental resources on child ability distribution, in percentage points						
	Child type				N	n
Parent type	1	2	3	4		
1 (lowest)	-0.02, .000012	-0.29, .000030	0.26, -0.000040	0.05, 0.0000030	1,446	507
	(0.75, .000064)	(0.95, 0.000081)	(0.84, 0.000071)	(.6258181, .0000531)		
2	-0.55, 0.000076	0.05, -0.000024	0.19, -0.000023	0.31, -0.000029	1,852	519
	(0.36, 0.000047)	(0.34, 0.000044)	(0.26, 0.000034)	(0.28, 0.000037)		
3	-0.08, -0.000016	-1.05, 0.00011	1.18, -0.00010	-0.06, 0.0000067	1,756	520
	(0.60, 0.000054)	(0.90, 0.000083)	(0.87, 0.000079)	(0.53, 0.000049)		
4 (highest)	-0.12, -0.0000063	-0.16, 0.000041	0.042, 0.0000011	0.24, -0.000035	1,548	511
	(0.44, 0.000033)	(0.59, 0.000046)	(0.51, 0.000040)	(0.46, 0.000035)		

Notes: See notes to Appendix Table 4. Each cell of the table shows the estimated coefficient on the change in parent income, followed by the estimated coefficient on the change in parent hours worked after the comma. In parentheses below the coefficients, the standard error on the change in parent income is shown, followed by the standard error on the change in parent hours worked. n differs slightly across regressions because the total number of parents (for whom hours worked is observed) is not a multiple of four (and the results are not sensitive to the this allocation across quartiles).

Appendix Table 6: Empirical ability transitions						
	Child type					
Parent type	1	2	3	4	N	n
1	0.31	0.24	0.22	0.22	1,192	421
2	0.29	0.25	0.23	0.23	1,531	422
3	0.20	0.27	0.29	0.25	1,461	421
4	0.20	0.23	0.26	0.31	1,397	422

See notes to Appendix Table 4. The table shows the empirical probability that parents of a given ability type have children of a given (possibly different) ability type.

Appendix Table 7. Marginal effects of parental resources									
	Data					Calibrated status quo policy			
	Child type					Child type			
Parent type	1	2	3	4	Parent type	1	2	3	4
1	-0.10	-0.34	0.80	-0.36	1	-0.26	-0.04	0.09	0.20
2	-0.56	0.04	0.25	0.27	2	-0.10	-0.02	0.03	0.09
3	-0.12	-1.04	1.26	-0.10	3	-0.07	-0.02	0.02	0.07
4	-0.05	-0.20	0.07	0.26	4	-0.00	-0.00	0.00	0.00

The table shows a substantial mismatch between the data and the model in the marginal effects of parental resources on child ability. Given our parsimonious model, this should not be surprising. Moreover, the estimated effects from the data are very imprecise, such that they receive relatively little weight in the calibration exercise. Nevertheless, the calibrated results take on values that are reassuringly "intuitive," with more resources always increasing the probability that parents have high-ability children, especially for low-ability parents.

Appendix Table 8: Parameter values chosen in calibration					
	α_a	α_c^1	α_c^2	α_c^3	α_c^4
With effect of labor	0.90	0.53	0.19	0.13	0.00
Baseline	0.88	0.57	0.23	0.17	0.00
		α_l^1	α_l^2	α_l^3	α_l^4
With effect of labor		-2.9E-5	-2.3E-4	-7.9E-5	-0.17
Baseline		n/a	n/a	n/a	n/a

The table shows the parameter values chosen by the calibration in the version of the model in which parental time working can affect child outcomes.

6 Robustness of results to variation in ρ, T, β, and I

Here, we describe the robustness of our status quo calibration results to modifying four assumptions: the value of the parameter ρ, the number of generations T, the value of the parameter β, and the number of types of parent and child ability levels I.

6.1 Value of ρ

ρ indicates the role of parental ability, relative to a mean ability level, in determining a child's ability. Higher values for ρ indicate slower mean-reversion of ability across generations. In the baseline estimates above, we set $\rho = 0.50$ based on a large body of empirical research. That same research, however, acknowledges a potentially wide range of values for what ρ represents in our model: namely, the extent to which parents' abilities are passed to their children through both genetic and environmental channels not influenced by parents' financial resources. Here we show how our results vary with the value of ρ.

We consider values of ρ, from 0.30 through 0.70. Appendix Table 9 shows the parameter values chosen by the simulations (as in Table 1). Appendix Table 10 shows the ability distribution that obtains under the optimal policy (as in Figure 2) and the corresponding welfare gain from applying the optimal policy's intergenerational ability transition matrices to the status quo's utility levels. The results for the baseline case of $\rho = 0.50$ are given for reference.

Appendix Table 9: Parameter values with alternative ρ values

	α_a	α_c^1	α_c^2	α_c^3	α_c^4
$\rho = 0.30$	0.94	0.25	0.08	0.09	0.00
$\rho = 0.40$	0.91	0.42	0.16	0.13	0.00
$\rho = 0.50$	0.88	0.57	0.23	0.17	0.00
$\rho = 0.60$	0.85	0.70	0.30	0.20	0.00
$\rho = 0.70$	0.82	0.79	0.35	0.24	0.00

Appendix Table 10: Ability distribution and welfare gains with alternative ρ values

	Ability type				Welfare gain
	1	2	3	4	α_c^5
$\rho = 0.30$	0.24	0.24	0.28	0.25	0.80%
$\rho = 0.40$	0.23	0.24	0.28	0.25	1.31%
$\rho = 0.50$	0.23	0.24	0.28	0.25	1.75%
$\rho = 0.60$	0.23	0.24	0.28	0.26	2.15%
$\rho = 0.70$	0.22	0.24	0.28	0.26	2.51%

These tables show that the main lessons from the baseline analysis are robust to variation in ρ. Appendix Table 10 shows that as innate ability is more important for realized ability (as ρ increases), the welfare impact of optimal policy increases. To see why, note that the spread of weights on parental disposable income increases with ρ. Intuitively, to match the status quo empirical targets with higher heritability of ability, the simulation requires that the value of disposable income be even greater for the children of low-ability parents (and less for the children of high-ability parents) than in the benchmark case. Because optimal policy affects the allocation of disposable income, these higher $\left\{\alpha_c^j\right\}_{j=1}^I$ values make optimal policy more powerful.

6.2 Value of T

We also describe the robustness of our results to variation in T, the number of generations simulated. The results are virtually unchanged when we consider two alternative horizons, namely $T = 8$ and $T = 12$. In particular, the parameter values chosen by the calibration to the status quo policy are the same as those shown in Table 1 of the main paper. The optimal policies are essentially unchanged, with the only small difference being that adding generations lowers slightly the welfare gain from the optimal policy's improvement in the ability distribution (for example, the welfare gain is 1.70 percent of total income if $T = 8$ compared to 1.75 percent in the baseline case of $T = 10$ and 1.79 percent in the case of $T = 12$.

6.3 Value of β

Next, we vary the value of β (and thus $R = 1/\beta$). The appropriate value of β is far from clear, both normatively and positively. The benchmark analysis of Ramsey (1928) showed that the discount rate applied by society ought to equal the sum of the rate of pure time preference and the product of the consumption elasticity of marginal utility and the growth rate of income. In this model, there is no steady state growth (when the ability distribution is stable), so we are left with the rate of pure time preference. While that rate may be positive for households, a case can be made that society should not discount future utilities. Ramsey (1928) himself wrote: "it is assumed that we do not discount later enjoyments in comparison with earlier ones, a practice which is ethically indefensible and arises merely from the weakness of the imagination." This perspective is reflected in our baseline assumption of $\beta = 1$, which in the context of intergenerational optimization seems particularly appropriate.

Nevertheless, we consider a case in which $\beta = 0.95$, so that each generation's utility is worth five percent less than the previous generation's. This scale of discounting is far less than a standard annual discounting model would imply, in which a pure time preference rate of two percent would imply a 25-year generational discount factor of 0.60, but we view that as an implausible degree of discounting for this scenario. Estimating the model with $\beta = 0.95$ and the other baseline values yields almost identical results to the baseline case. In particular, the parameter estimates for the calibration to the status quo are the same as those shown in Table 1 of the main paper. As might be expected, the use of a discount factor less than one affects the optimal policy results very similarly to using a shorter horizon (smaller T), in that the welfare gain is slightly smaller when $\beta = 0.95$ than in the baseline case (i.e., 1.70 percent of total consumption compared to 1.75 percent).

6.4 Value of I

Finally, we allow for $I = 5$ or (alternatively) $I = 10$ parent wage and child ability categories, rather than the four categories in the baseline results. The $I = 5$ case shows the effects of a small increase in the number of types, whereas the $I = 10$ case allows substantially more ability types for comparison. The wage levels assigned to each category (calculated as in the benchmark case) in the five-type case are {3.01, 5.55, 7.74, 10.58, 21.38}. In the ten-type case, they are {1.77, 3.74, 5.04, 6.05, 7.14, 8.34, 9.7, 11.48, 14.35, 28.42}. In Appendix Table 11, we show the values of the parameters chosen by the simulation to match the empirical targets (i.e., the transition matrix, marginal effects of parental disposable income, and mean log wage) in the $I = 5$ and $I = 10$ cases, as well as in the $I = 4$ baseline case.[11] We omit these targets and the corresponding calibration results for brevity, but they resemble those of the baseline case. For reference, we repeat the chosen parameter values from that baseline case.

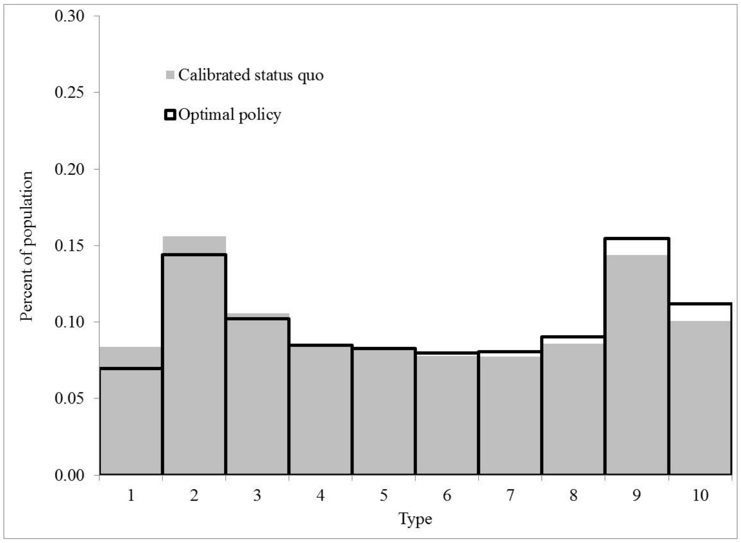
Appendix Table 11: Calibrated parameter estimates for $I = 4$, $I = 5$, and $I = 10$

	α_a	α_c^1	α_c^2	α_c^3	α_c^4	α_c^5	α_c^6	α_c^7	α_c^8	α_c^9	α_c^{10}
$I = 10$	0.83	1.00	0.55	0.42	0.29	0.26	0.26	0.19	0.17	0.14	0.00
$I = 5$	0.86	0.67	0.32	0.22	0.14	0.00	n/a	n/a	n/a	n/a	n/a
$I = 4$	0.88	0.57	0.23	0.17	0.00	n/a	n/a	n/a	n/a	n/a	n/a

All cases yield similar values for α_a and a similar pattern of values for the $\left\{\alpha_c^j\right\}_{j=1}^{I}$ parameters. The larger value for α_c^1 in the $I = 10$ case is natural, as the group of parents for whom it applies has a lower ability level than in the case of $I = 4$. All cases yield very small values for α_c^I.

The parameter values provide a bit less information on the robustness of our baseline results in this case than for the other robustness checks, as the increase in the number of types makes the implications of the parameters less directly apparent. As in Figure 2 of the baseline analysis, in Appendix Figure 1 we show the ability distribution under the optimal and calibrated status quo policies (for the fifth, middle, generation).

[11] In the computational calibration of the $I = 10$ case, we restrict the α_c parameters to be less than or equal to unity. If we leave them unrestricted, the simulation chooses implausibly large values ranging from 2.61 to 0.96, with a value of α_i of zero. This binding restriction is not necessary in the other cases, since all of the α_c parameters are less than unity even when the upper bound in the simulation is much larger.



Appendix Figure 1: Ability distribution under calibrated status quo policy (filled-in columns) and social-welfare maximizing policy (empty outlined columns).

As in the baseline case, the optimal policy achieves an improvement in the ability distribution. We can calculate the increase in welfare due to this improvement, just as in the baseline case (see Table 6). In Appendix Table 12 we provide these welfare gains for the overall economy and by type in the $I = 10$, $I = 5$, and the baseline $I = 4$ case.

Appendix Table 12: Welfare gains for $I = 4$, $I = 5$, and $I = 10$

	Overall welfare gain	Type-specific gains 1 (lowest)	2	3	4	5	6	7	8	9	10 (highest)
$I = 10$	3.46%	6.2%	3.8%	3.3%	3.2%	3.1%	3.0%	3.0%	2.9%	2.8%	3.0%
$I = 5$	2.54%	3.4%	2.5%	2.3%	2.2%	2.3%	n/a	n/a	n/a	n/a	n/a
$I = 4$	1.75%	2.2%	1.6%	1.6%	1.6%	n/a	n/a	n/a	n/a	n/a	n/a

These results suggest that the welfare gains from optimal policy are robust, and are likely to increase, as we increase the fineness of the ability distribution by adding additional types.

7 Robustness of results to including redistribution from higher incomes in calibration

Here, we address the possibility that our results depend on the limited income range over which we perform the calibration and optimal policy simulation. The concern this limited range generates is that higher income earners pay substantial net taxes under both the existing U.S. tax system and the optimal tax policy. Those net taxes are used to fund lower average taxes (i.e., greater redistribution) to low earners in each system. In principle, the greater redistribution that this implies under the existing U.S. system could diminish the extent to which the optimal policy can take advantage of the endogeneity of ability. Intuitively, for example, there could be a satiation point of redistribution beyond which the effects on children's abilities are minimal, and if the U.S. system has reached that point, perhaps there is little room for improvement. The data we use for the calibration do not extend sufficiently high up the income scale for us to address this concern directly (i.e., by simply including higher ability types). Nevertheless, there is a simple way to gauge the potential consequences of including higher earners, which we now describe.

The key to our approach is to recognize that the primary way in which higher earners are relevant to the topic of this paper is as a source of net tax revenue. Of course, there is some movement of children from the families in our sample to higher-income status as adults, and vice-versa, but the calibrated monotonically-declining pattern of the $\left\{\alpha_c^i\right\}_i$ parameters—which reach zero at a parental wage of approximately \$20—strongly suggests that the potential for tax policy to affect these transitions is negligible. However, we recognize that one limitation of these results is that they simplify the optimal tax problem by assuming that higher earners are relevant to the issues we study only insofar as they represent a source of net tax revenue.

Once higher earners are relevant only for their potential to supply funds for redistribution to the households in our sample income range, we can proxy for their inclusion by simply relaxing the planner's feasibility constraint. In particular, we can assign to the planner a revenue requirement $\bar{R}$ that is negative, whereas our assumption in the baseline model was $\bar{R} = 0$.

The essence of our approach is to relax the feasibility constraint on policy when it operates on only a portion of the income distribution so that it treats households similarly to how they are treated by the policy when it can operate on the full income distribuiton. The specifics are as follows. The Congressional Budget Office provides data on wages and average tax rates for the four lower quintiles and the next 10, 5, 4, and 1 percentiles of the U.S. income distribution. Those data show the third quintile of households in 2006 earned \$46,000 on average, similar to what the highest type in the $I = 5$ sample described above (who has a wage of \$21.38) would earn at a full time job. In other words, our sample of households represents (approximately) the lower 60 percent of the U.S. income distribution. We modify our calibrated Status Quo model from the robustness check above with $I = 5$ types to allow for a negative revenue requirement. We then search for the value of that negative revenue requirement that causes the calibrated Status Quo policy to yield average tax rates toward this lower 60 percent of households that approximate the actual rates observed in the CBO data. This value captures the extent to which the feasibility constraint on policy toward the lower 60 percent of households ought to be relaxed to account for the resources available from higher-income households. We perform a similar exercise for the optimal policy. In particular, we simulate a conventional (exogenous ability)

Mirrleesian optimal tax policy, using the same functional forms and parameters as in the baseline case, for the CBO's full income distribution. Then, we simulate a conventional Mirrleesian optimal tax policy toward only the lower 60 percent of households, but we allow for a negative revenue requirement. We search for the negative revenue requirement that generates (conventionally) optimal average tax rates toward the lower 60 percent of households (our sample) that approximate the optimal rates chosen by the planner facing the full distribution. As with the Status Quo policy, this negative revenue requirement captures the extent to which having a wider income distribution relaxes the feasibility constraint on optimal policy and, thus, the extent to which the feasibility constraint on the optimal policy toward the lower 60 percent of households ought to be relaxed. Appendix Table 13 shows the results of these simulations. It gives the average tax rates in the Status Quo and conventional optimal policies toward both the CBO's full income distribution and our sample of households (in the latter case, including the relevant negative revenue requirements). For clarity, we show the rates from the full-distribution case only on the first three quintiles of the distribution, that is, the quintiles that correspond to our sample households.

Appendix Table 13. Average tax rates

	Full distribution			Sample distribution (with negative revenue requirement)	
Quintile	Optimal policy (exogenous ability)	Status Quo (CBO data)	Type	Optimal policy (exogenous ability)	Status Quo (CBO data)
1	-1,538%	-298%	1 (lowest)	-1,519%	-253%
			2	-606%	-154%
2	-158%	-22%	3	-369%	-114%
			4	-235%	-80%
3	-29%	10%	5 (highest)	-45%	-17%

Next, we apply the negative revenue requirement calculated in this way for the conventional (exogenous ability) optimal tax problem and apply it to the optimal policy simulation exercise as described in the main analysis. In other words, we relax the planner's problem by the amount of resources that the conventional planner required to treat the households in our restricted sample similarly to how they would be treated as part of policy toward an unrestricted income distribution. The calibrated parameter values are shown in Appendix Table 14, with the results from the $I = 5$ analysis above also shown for reference:

Appendix Table 14: Parameter values chosen in calibration

	α_a	α_c^1	α_c^2	α_c^3	α_c^4	α_c^5
Value under status quo policy	0.86	0.40	0.23	0.17	0.12	0.00
Baseline calibration value	0.86	0.67	0.32	0.22	0.14	0.00

As would be expected, the $\left\{\alpha_c^i\right\}_i$ values decrease once a much larger level of disposable income is available to parents in the calibrated status quo. Intuitively, to match the empirical patterns of ability across generations, the effects of parental resources must be smaller in magnitude when those resources are substantially greater.

Importantly, the pattern of these parameter values is unchanged by this extension, and the calibrated value of α_a is unchanged. These parameters yield marginal effects of parental resources and transition matrices similar to those in the $I = 5$ analysis. They also yield the following marginal and average tax rates:

Appendix Table 15. Marginal and average tax rates

	Marginal tax rate		Average tax rate	
Type	Optimal	Status Quo	Optimal	Status Quo
1 (lowest)	18%	-35%	-1,537%	-253%
2	22%	3%	-620%	-154%
3	28%	22%	-390%	-114%
4	28%	37%	-237%	-80%
5 (highest)	0%	34%	-48%	-17%

Consistent with the results in the $I = 5$ (and baseline $I = 4$) analysis, the optimal policy remains substantially more redistributive than this calibrated Status Quo policy. This yields a rising ability distribution: the optimal policy leads to 3.5 percent fewer individuals of the lowest type and 3.4 percent more individuals of the highest type, with smaller differences for intermediate types as in the main analysis. The welfare implications are also very similar to those found in the $I = 5$ (and baseline $I = 4$) analysis, as shown in Appendix Table 16.

Appendix Table 16: Welfare gains

	Overall welfare gain	Type-specific gains				
		1 (lowest)	2	3	4	5 (highest)
With negative revenue requirement	2.50%	2.9%	2.5%	2.4%	2.3%	2.3%
Baseline $I = 5$ analysis	2.54%	3.4%	2.5%	2.3%	2.2%	2.3%

8 References

Bernal, Raquel, and Michael P. Keane, 2011. "Child care choices and children's cognitive achievement: The case of single mothers." Journal of Labor Economics 29(3): 459–512.

Best, Michael and Henrik Kleven, 2013. "Optimal Taxation with Career Effects of Work Effort." Working Paper.

Blau, David M. 1999, "The effect of child care characteristics on child development." Journal of Human Resources 34(4): 786–822.

Blau, David M. and Janet Currie, 2004. "Preschool, Day Care, and Afterschool Care: Who's Minding the Kids." In Hanushek, Eric and Finis Welch (eds.), Handbook of the Economics of Education. Elsevier.

Blau, Francine and Adam Grossberg, 1992. "Maternal Labor Supply and Children's Cognitive Development." Review of Economics and Statistics 74(3): 474–481.

Feenberg, Daniel, and Elisabeth Coutts, 1993. "An introduction to the TAXSIM model." Journal of Policy Analysis and Management 12(1): 189–194.

Gelber, Alexander, 2011. "How do 401(k)s Affect Saving? Evidence from Changes in 401(k) Eligibility." American Economic Journal: Economic Policy 3(4):103–122.

Gruber, Jonathan, and Emmanuel Saez, 2002. "The Elasticity of Taxable Income: Evidence and Implications." Journal of Public Economics 84(1): 1–33.

Pence, Karen, 2006. "The Role of Wealth Transformations: An Application to Estimating the Effect of Tax Incentives on Saving." B. E. Journal of Economic Analysis and Policy 5(1): 1–24.

Rothschild, Casey and Florian Scheuer, 2011. "Optimal Taxation with Rent-Seeking." NBER Working Paper 17035.

Ruhm, Christopher J., 2004. "Parental employment and child cognitive development." Journal of Human Resources 39(1): 155–192.