NBER WORKING PAPER SERIES

EQUALIZING OUTCOMES AND EQUALIZING OPPORTUNITIES: OPTIMAL TAXATION WHEN CHILDREN'S ABILITIES DEPEND ON PARENTS' RESOURCES

Alexander M. Gelber
Matthew C. Weinzierl

Working Paper 18332
http://www.nber.org/papers/w18332

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2012

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. 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, Monica Singhal, and Aleh Tsyvinski for comments and suggestions. We thank Gordon Dahl, Lance Lochner, Dave Rapson, and Larry Kotlikoff for sharing data. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w18332.ack

Equalizing Outcomes and Equalizing Opportunities: Optimal Taxation when Children's Abilities Depend on Parents' Resources
Alexander M. Gelber and Matthew C. Weinzierl
NBER Working Paper No. 18332
August 2012, Revised August 2013
JEL No. H21,I30

ABSTRACT

Empirical research suggests that parents'economic resources affect their children's future earnings abilities. Optimal tax policy therefore treats future ability distributions as endogenous to current taxes. We model this endogeneity, calibrate the model to match estimates of the intergenerational transmission of earnings ability in the United States, and use the model to simulate such an optimal policy numerically. The optimal policy in this context is more redistributive toward low-income parents than existing U.S. tax policy. It also increases the probability that low-income children move up the economic ladder, generating a present-value welfare gain of more than two and one half percent of consumption in our baseline case.

Alexander M. Gelber
Goldman School of Public Policy
University of California at Berkeley
2607 Hearst Ave
Berkeley, CA 94720
and NBER
agelber@berkeley.edu

Matthew C. Weinzierl
Harvard Business School
277 Morgan
Soldiers Field
Boston, MA 02163
and NBER
mweinzierl@hbs.edu

Introduction

Economists have long recognized that parents' resources and investment in their children may be key determinants of their children's outcomes (Arleen Leibowitz 1974; Gary Becker and Nigel Tomes 1976; Becker 1981; Becker and Tomes 1986). This paper is motivated by recent evidence that increasing the disposable incomes of poor parents raises the performance of their children on tests of cognitive ability. That finding suggests that current tax policy may affect the future distribution of underlying income-earning abilities in the taxpayer population, the key determinant of how difficult a tradeoff between efficiency and equality society will face in the future. The dominant modern model of optimal taxation is unable to take this effect into account, as it assumes that the distribution of ability is entirely exogenous. Our paper is an analytical and numerical exploration of the implications for optimal policy of relaxing this assumption.

First, we generalize a standard dynamic Mirrleesian optimal tax model to include the effect of parental disposable income on children's abilities. In the standard model of James Mirrlees (1971), ability is exogenously given. In our model, a child's ability depends on three components: parental ability, which is exogenous to the parent and child; parental disposable income, which is endogenously chosen by parents given the tax system; and a stochastic shock. These components imply that the process generating children's skills in our model is partly exogenous, partly endogenous, and stochastic. By combining these components, our model introduces a novel element to the recent literature on dynamic optimal taxation that seeks, among other goals, to capture the complex process by which society's ability distribution is determined.[1]

Using this model, we derive analytical conditions that reveal the key effects of endogenous ability on optimal intratemporal and intertemporal policy. On the intratemporal margin, we find contradictory forces at play. First, marginal income tax rates are lower on parents whose economic resources matter more for their children's expected abilities. Lower marginal tax rates encourage greater parental earnings and disposable income, and because of endogenous

[1] We abstract from other aspects of the ability distribution that are also currently being studied, such as the lifecycle path of earnings (see Matthew Weinzierl 2011 or Emmanuel Farhi and Ivan Werning 2013, for example).

ability, these parents thus produce higher-skilled children from whom society will be able to collect more tax revenue. Evidence suggests the impact of parental resources on child skills is largest among parents with low incomes, so this force is likely to lead to lower marginal tax rates on low incomes relative to high incomes. Second, if low-ability parents enjoy a high relative return to earning disposable income (due to its larger effects on their children), incentive problems are relaxed relative to a setting with exogenous ability, so marginal distortions on low earners can be reduced (thus reinforcing the first factor). Third, and working in the opposite direction, lower marginal tax rates on low incomes will in expectation differentially benefit low earners in prior generations because low-income parents are more likely than high-income parents to have low-ability children. This differential benefit increases the temptation for high-ability ancestors to feign low ability by earning less and accepting the greater probability of having low-ability descendants. In doing so, it worsens the distortionary effects of marginal taxes on effort. This factor works against the others, acting as a force for higher marginal tax rates at low incomes. The relative strength of these forces determines how optimal marginal tax rates differ from a conventional policy.

On the intertemporal margin, we derive a condition showing that the allocation of resources should equalize the cost of raising welfare across generations, taking into account not only the marginal utilities of individuals in each generation (as in a conventional model), but also the effects of the current distribution of resources on future generations' tax payments and utility levels. As a result, welfare-maximizing policy takes advantage of its potential to shape the ability distribution of future generations. For example, suppose hypothetically that the ability distribution is stable across generations under an existing tax policy. A conventional optimal policy model would recommend that generations be treated similarly, as each generation resembles the next. Our model may recommend a different approach. Namely, the optimal policy in this case borrows from future generations to fund greater after-tax income for parents in the current generation.

Together, optimizing policy along these intra- and inter-generational margins can generate an upward trajectory for the ability distribution across generations, generating more productive future populations and greater welfare overall.

Second, we calibrate our model to empirical evidence and solve numerically for the optimal policy. The model calibration requires empirical estimates of key statistics describing the transmission of ability across generations under an existing tax policy. To generate these estimates, we study the effect of policy changes in the U.S. Earned Income Tax Credit (EITC) on the ability levels of taxpayers' children. Our empirical approach adapts the strategy of Gordon Dahl and Lance Lochner (forthcoming) in order to generate estimates relevant to the calibration exercise we perform.[2] Specifically, dividing matched parents and children in the National Longitudinal Survey of Youth (NLSY) and Children of the National Longitudinal Survey of Youth (CNLSY) into five equally-sized ability categories each, we estimate the effect of parents' after-tax income on the probability that parents in each wage category have children in each category, and we calculate the transition matrices between ability categories across generations. Then, using a smooth approximation of Laurence Kotlikoff and David Rapson's (2007) estimates of effective marginal tax rates in the United States as the status quo tax policy, we find the values of the model's parameters that yield a model output that best matches the target statistics, when optimizing households take that policy as given.

We use the calibrated model to simulate Utilitarian-optimal policy, and we find that the pattern of optimal average and marginal tax rates is very different than the status quo. The optimal policy redistributes substantially more toward low-ability parents and earlier generations than does the status quo policy. The increase in redistribution toward low earners is driven by the Utilitarian objective assumed in the conventional optimal policy model. Nevertheless, the increase in redistribution generates an upward shift in the mean ability level across generations relative to the status quo, with a smaller share of the population having lower abilities and a larger share having higher abilities. We calculate the increase in aggregate welfare due to *only* the improved evolution of the ability distribution. We find that the gain is equivalent to a 2.54 percent permanent increase in disposable income in our baseline case–i.e. an increase of more than two and one-half percent in disposable income for all generations.

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

[2] Viewed in isolation, we see the empirical work as merely a secondary contribution of the paper, as our work is closely related to the Dahl and Lochner analysis. The estimates it generates are primarily useful as inputs to the calibration and simulation of the model.

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.[3] Emmanuel Farhi and Iván Werning (2010) extend that approach to 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 assumed, however, that children's skills are independent of their parents' abilities and their parents' economic resources, leaving unaddressed a core part of the "family risk" that is their focus.[4] We take up the complementary analysis. That is, 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).[5]

The core conceptual contribution of this paper is to take into account the dynamic interaction between exogenous and endogenous components of skill heterogeneity.[6] We consider how choices by parents affect the abilities taken as given by their children, and how these abilities in turn affect the set of choices available to children. This interaction is a central factor in policy design, in that it is the crux of the tradeoff between redistributing to the poor later (i.e. equalizing the distribution of outcomes) and investing in their skills now (i.e.

[3] Contributions include Kocherlakota (2005), Golosov and Tsyvinski (2006), Albanesi and Sleet (2006), Golosov, Tsyvinski, and Werning (2007), Farhi and Werning (2010b), and Weinzierl (2011).

[4] Farhi and Werning do consider a simple form of parental investment in children's human capital in the two-period version of their analysis, but children do not exert effort in that version.

[5] Other examples include the following. Marek Kapicka (2006a, 2006b) allows a deterministic skill process to be endogenous. 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. 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.

[6] Kapicka (2006a,b) has heterogeneity in natural ability, but each type is fixed for life, and all types share the same human capital production function. 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. Anderberg (2009) has human capital and an exogenous shock interact, but human capital investment decisions are made by agents before their ability heterogeneity is realized.

equalizing the distribution of opportunities). Our findings suggest one way in which society might increase equality of both outcomes and opportunities. That is, if future skill levels among the poor can be increased through current transfers, the net benefit of those transfers to society will be increased. Though its application is most apparent across generations, the interaction between natural ability and human capital investments is also relevant for issues such as the design of life-cycle tax and training policies and social insurance programs.[7]

The paper proceeds as follows. Section 1 describes the model. Section 2 derives analytical conditions that describe the optimal policy both within and across generations. Section 3 calibrates the model to existing U.S. tax policy and new empirical evidence on the transmission of ability across generations. Section 4 uses the calibrated model to simulate and characterize both the structure and welfare implications of optimal tax policy in our context. Section 5 concludes. An Appendix contains details of the analytical and empirical results.

1 Model

The formal model largely follows the standard setup of modern dynamic Mirrleesian analysis. Individuals obtain utility from consumption and disutility from exerting work effort. Each individual has an unobservable ability to earn income, which he or she combines with an unobservable level of work effort to determine pre-tax income, which is observable. The social planner designs a tax system that generates a mapping from pre-tax income to disposable income. Individuals optimally choose work effort knowing this tax system, and thereby produce income, pay taxes, and enjoy the remaining disposable income as consumption. We assume no intergenerational transfers, so disposable income equals consumption for each generation.

[7]Modeling this interaction is challenging, however, and one technical contribution of this paper is a novel formal simplification of the dynamics of the endogenous ability distribution. Rather than having parental resources directly affect the levels of children's abilities, we locate the effects of parental resources on the distribution of children across a fixed set of abilities. In combination with history-independence, the natural assumption that taxes on individuals do not depend on the income of their parents or children, our use of a fixed set of abilities with an endogenous distribution rather than an endogenous set of abilities substantially simplifies the computations of the optimal policy. The alternative approach, in which types vary continuously with parental resources, means that a planner has to specify allocations for all possible deviation paths. This technique may prove useful in other contexts.

The intergenerational focus of this paper requires some additional structure.[8] Individuals are linked in families, with one individual per generation in each family. Generations are indexed by $t \in \{1, 2, ..., T\}$. Each individual's unobservable ability is taken from a fixed set of possible values, or "types," indexed by $i \in \{1, 2, ..., I\}$. The distribution of individuals across types is exogenous in the first generation, but in subsequent generations it is endogenous and 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 the disposable income c of type i'.

1.1 Planner's problem

As in standard optimal tax analysis, we model the social planner as specifying a menu of allocations of earned income y and disposable income c. We describe here the planner's problem in the case in which children's ability may depend on parents' resources (but not on parents' time allocation); we later explore the results in the case in which children's ability also depends on parents' time allocation. By the Revelation Principle we can restrict attention to menus in which the planner intends a specific (y_t^i, c_t^i) bundle for each type i in each generation t. These allocations may differ across generations. The planner's objective is to maximize social welfare. Following the standard approach, we define social welfare as the present-value utility of the population of families starting from the first generation; that is, the objective is Utilitarian. The planner's maximization problem is constrained in two ways: first, feasibility, specifying that disposable income must be funded by output; second, incentive compatibility, specifying that individuals choose optimally among the offered bundles (i.e. maximize their own utility taking the tax system as given).

We also impose the constraint that taxes may depend on only the *current* generation's characteristics and choices. In other words, taxes are restricted to be independent of history and cannot depend on the income of the taxpayer's parents or children. This restriction is standard, as well as convenient, in a variety of dynamic optimal tax contexts such as

[8]The Mirrlees model's emphasis on ability is particularly well-suited to addressing the heart of the issue in this paper–namely, how parental resources affect the opportunities available to children.

Kapicka's (2006) analysis of human capital. Moreover, history independence captures the explicit tax system in a realistic way: no tax system does or, we conjecture, ever will levy taxes on a child that depend in any direct way on that child's parents' characteristics. There seems to be a normative aversion to such history-dependence across generations, so we will impose it on the policy here.

Some aspects of policy, such as subsidies for children's education like Pell Grants or 529 education savings plans, may seem to violate our assumption of history dependence because they condition policy on parents' resources. The bulk of these incentives lie outside of and are small relative to the overall history-independent tax and transfer system. More importantly, these policies do not condition on the income of the child, a necessary component of history-dependent policies. To see this, note that an optimal history-dependent tax policy in this model would condition redistributive transfers to low earners in one generation on whether their parents were low or high earners. In particular, it would reduce transfers to those low earners whose parents were high ability, thereby discouraging those parents from underinvesting in their children.[9] Policies such as Pell grants and 529 savings plans, even if conditioned on parental income levels, are qualitatively different, because they do not depend on the child's earnings. In fact, these policies are similar to the supplements to parental resources that our results recommend. Of course, our assumption of history independence is nevertheless a simplification, and understanding the impact of realistic intergenerational history dependence would be a valuable avenue for future research.

Formally, the planner's problem is as follows:

Problem 1 *Planner's Problem*

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

where U_t^i, the present-value expected utility of a family with generation-t parents of type i, is

[9] Or, it would tax more lightly those high earners whose parents were low earners, to encourage this investment. Note that taxes on bequests or *inter vivos* gifts are potential targets for such history-dependent taxation, though they are usually history-independent in reality.

defined recursively as

$$U_t^i = u\left(c_t^i\right) - v\left(\frac{y_t^i}{w_t^i}\right) + \beta \sum_{j=1}^{I} p^j\left(w_t^i, c_t^i\right) U_{t+1}^j. \tag{2}$$

This is maximized subject to feasibility:

$$\sum_i p^i R_1^i \geq \bar{R}, \tag{3}$$

where $\bar{R}$ is an exogenous revenue requirement, and R_t^i is the expected present value of all current and future tax revenue of a family with parents of type i, defined recursively as follows

$$R_t^i = \left(y_t^i - c_t^i\right) + \beta \sum_j p^j\left(w_t^i, c_t^i\right) R_{t+1}^j;$$

and incentive compatibility for each generation:

$$U_t^i \geq U_t^{i'|i} \text{ for all generations } t \text{ and types } i, i', \tag{4}$$

where $U_t^{i'|i}$ denotes the utility obtained by a generation-t parent of type i when claiming to be type i' :

$$U_t^{i'|i} = u\left(c_t^{i'}\right) - v\left(\frac{y_t^{i'}}{w_t^i}\right) + \beta \sum_{j=1}^{I} p^j\left(w_t^i, c_t^{i'}\right) U_{t+1}^j. \tag{5}$$

1.2 Limitations

Some apparent limitations of the setup deserve clarification.

First, while the setup 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, 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. The empirical evidence we have concerns the effect on a child's ability of transfers to parents through the tax system; 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, throughout the paper to make this aspect of our analysis clear. If data were available that divided parental expenditure into consumption and investment in children's abilities, a more subtle optimal policy could result. If such investment were particularly valuable for children with low-ability parents, for example, optimal policy may include income-dependent subsidies for investment offset by higher tax rates on income. Complicating such a policy would be the incentives it would provide for misclassification of spending across categories.

Second, we assume in the problem above that the allocation of parental time has no effect on children's abilities. We later explore the case in which children's ability depends on both parental income and parents' hours worked. If parents work more, they could 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). We find that, on net, parental time allocation has only a small effect on our baseline results. However, as we discuss, our empirical estimate of the effect of parents' hours worked on children's ability is more suggestive than our estimate of the effect of parent income.

Third, only tax policy is modeled in this paper, but that does not imply that other policies play no role. Our empirical estimates take as given the existing set of non-tax policies and institutions, such as schools, 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, again an assumption we make to match the empirical evidence to which we calibrate the model.

Fourth, in the terminology of Becker and H. Gregg Lewis (1973), we assume that quality of children is valued and affected by parental resources, but we abstract from the effect of resources on the quantity of children. Valuing new lives is beyond the scope of this paper, and empirical work has produced inconsistent evidence on the effect of tax policy on fertility.

Finally, we do not constrain parent and child distributions of ability to be the same, as they might be in some steady state. Again, we are motivated by the data: wage distributions have shown secular time trends in the data across generations (e.g. Claudia Goldin and Lawrence Katz 2007), and test scores have secularly increased over time as in the well-known

"Flynn effect" (e.g. James Flynn 1987).

2 Analysis of optimal policy

Our analysis of the planner's problem in expressions (1) through (5) generates two results. First, we characterize the distortion to an individual's choice of how much to earn. Second, we derive a necessary condition on optimal allocations across generations that modifies the conventional model's recommendation in an intuitive but powerful way.

2.1 Optimal marginal distortion to earned income

The classic object of study in optimal tax models is the marginal tax rate, or the distortion to the individual's marginal choice between disposable income and leisure. Formally, the ratio $v'\left(y_t^i/w_t^i\right)/[w_t^i u'\left(c_t^i\right)]$ equals one if an individual sets the marginal disutility of labor equal to the marginal utility of consuming the income that labor earns. Any factor reducing the marginal utility of earnings (such as a positive marginal tax rate) causes this ratio to be less than one, distorting the individual's choice of labor effort.

In the model above, in the absence of taxes, parent i in generation t would solve her own planning problem. Formally, she 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 $p^j\left(w_t^i, c_t^i\right)$, will also choose optimally for themselves. That parent's optimal private choice will satisfy the following condition:

$$\frac{v'\left(y_t^i/w_t^i\right)}{w_t^i u'\left(c_t^i\right)} = 1 + \beta \sum_j \frac{\partial p^j\left(w_t^i, c_t^i\right)}{\partial c_t^i} \frac{U_{t+1}^j}{u'\left(c_t^i\right)}. \tag{6}$$

To see this result, take the first-order conditions of expression (2), the utility of parent i in generation t, with respect to c_t^i and y_t^i, subject to the no-tax condition that $c_t^i = y_t^i$, and simplify. In words, 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 (6), relative to a model in which they took only their own disposable income into account. Recall that the right-hand side equals one in a conventional model without endogenous ability distributions.

Lemma 1 (proved in the Appendix) establishes that the planner's first-order conditions for c_t^i and y_t^i imply a distortion to a parent's private choice.[10]

Lemma 1 *Intratemporal Distortion: Let $\mu_t^{i'|i}$ denote the multiplier on* (4)*. The solution to the Planner's Problem satisfies, for all $t \in [1, 2, ...T]$ and all $j \in [1, 2, ...I]$,*

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

where

$$\mathbb{A}_t^j = \frac{1}{1 - \beta \sum_k \frac{\partial p^k\left(w_t^j, c_t^j\right)}{\partial c_t^j} R_{t+1}^k}, \tag{8}$$

$$\mathbb{B}_t^j = \beta^t \pi_t^j + \sum_{\tau=1}^{t-1} \beta^{t-\tau} \sum_i \sum_{i'} \mu_\tau^{i'|i} \left(\pi_t^j|_{c_\tau^i} - \pi_t^j|_{c_\tau^{i'}}\right), \tag{9}$$

$$\mathbb{C}_t^j = \sum_{j'} \mu_t^{j'|j} - \frac{1 + \beta \sum_k \frac{\partial p^k\left(w_t^{j'}, c_t^j\right)}{\partial c_t^j} \frac{U_{t+1}^k}{u'\left(c_t^j\right)}}{1 + \beta \sum_k \frac{\partial p^k\left(w_t^j, c_t^j\right)}{\partial c_t^j} \frac{U_{t+1}^k}{u'\left(c_t^j\right)}} \sum_{j'} \mu_t^{j|j'}, \tag{10}$$

$$\mathbb{D}_t^j = \sum_{j'} \mu_t^{j'|j} - \frac{\frac{1}{w_t^{j'}} v'\left(\frac{y_t^j}{w_t^{j'}}\right)}{\frac{1}{w_t^j} v'\left(\frac{y_t^j}{w_t^j}\right)} \sum_{j'} \mu_t^{j|j'}, \tag{11}$$

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 compare this paper's setting to two simpler, hypothetical benchmarks. We identify three effects.

First, suppose the planner had full information on individuals' (endogenous) abilities. In that case, expressions (9) through (11) would imply that $\mathbb{B}_t^j = \beta^t \pi_t^j$ 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

[10] It is important to be clear about our terminology. By "distorting" the parent's choice, we mean simply that the condition characterizing the planner's first-order condition (7) is different than the condition (6) characterizing the parent's choice.

expected net revenue gain from raising the disposable income of parent j. In particular, the gain is 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. The planner values that revenue gain, while the parent does not. Intuitively, larger $\mathbb{A}_t^j$ means that the planner generates greater net revenue gain from having the parent obtain a larger disposable income, implying that optimal policy entails a smaller downward distortion (or larger upward distortion) to parent j's effort. We call this factor the "revenue effect." 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[11] at low incomes will be optimal.

Second, suppose the planner cannot observe ability, but parental resources have no effect on children's abilities–in other words, the conventional Mirrlees model. In that case, $\partial p^k\left(w_t^j, c_t^j\right)/\partial c_t^j = 0$, so $\mathbb{A}_t^j = 1$. Then, the wedge would be simply $\left(\mathbb{B}_t^j + \mathbb{C}_t^j\right)/\left(\mathbb{B}_t^j + \mathbb{D}_t^j\right)$. In that conventional setting, $\mathbb{B}_t^j = \beta^t \pi_t^j$, 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. Note that $\mathbb{C}_t^j < \mathbb{D}_t^j$ when higher-skilled types are tempted to mimic lower-skilled types, so this ratio is less than one, which implies a positive marginal tax rate in the conventional setting.

Introducing endogenous, unobservable ability has two effects on the optimal intratemporal wedge in addition to the revenue effect identified above. In particular, both $\mathbb{B}_t^j$ and $\mathbb{C}_t^j$ take on more complicated values than in either of these simpler benchmark cases.[12] We call these two effects the "relative return effect" and the "ancestor incentive effect" to aid in intuition.

The "relative return effect" relates to the value of $\mathbb{C}_t^j$, which captures how the differential value of extra disposable income to parents of different abilities interacts with the distortionary tax system. In particular, while $\mathbb{C}_t^j < \mathbb{D}_t^j$ in a conventional model when higher-skilled types are tempted to mimic lower-skilled types, the size of $\mathbb{C}_t^j$ will increase once endogenous

[11] Throughout this section, we use the terminology of distortions and marginal taxes interchangeably. This assumes an implementation of the distortions through an explicit tax system, which is straightforward in our setting without capital.

[12] The term $\mathbb{D}_{t+s}^j$ is unchanged from the conventional model, unless parental work effort enters the child's ability production function. We explore that case below.

ability is included if the marginal effect of additional parental disposable income on children's abilities is larger for lower-ability parents.[13] To see this, suppose that $w^{j'} > w^j$, so that type j' is the higher type. If $\frac{\partial^2 p^k\left(w_t^j, c_t^j\right)}{\partial c_t^j \partial w_t^j} = 0$, implying that the marginal effect of parental resources is independent of parental ability, then $\mathbb{C}_t^j$ is unchanged from in the standard model, while if $\frac{\partial^2 p^k\left(w_t^j, c_t^j\right)}{\partial c_t^j \partial w_t^j} < 0$, implying that the marginal effect of parental resources is decreasing in parental ability, then $\mathbb{C}_t^j$ increases. Meanwhile, $\mathbb{D}_t^j$ is unchanged. When $\mathbb{C}_t^j$and $\mathbb{D}_t^j$ have more similar values, the optimal marginal taxes on low-skilled parents are smaller. Intuitively, if the extra resources granted to low-income families are of less value to high-ability (and thus high-income) parents, high-income parents will be less tempted in this model to claim low ability than in a standard model, so smaller distortions at low incomes will be required.

The "ancestor incentive effect" relates to the value of $\mathbb{B}_t^j$, which measures how an increase in c_t^j affects the incentive problems of taxing earlier generations who can affect the probability that their descendants have the type j. For example, suppose $w^{i'} > w^i$, so that type i' is the higher type in generation τ, and $w^{j'} > w^j$, so that type j' is the higher type in generation t. In this case, $\pi_t^j|_{c_\tau^i} > \pi_t^j|_{c_\tau^{i'}}$, $\mu_\tau^{i'|i} = 0$, and $\mu_\tau^{i|i'} > 0$, so that $\mathbb{B}_t^j$ is smaller, and the optimal marginal distortion is larger, for low-skilled types in a model with endogenous ability than in a model without. Intuitively, a smaller distortion on a low-skilled type raises the temptation for previous generations to work less and produce low-skilled descendants. The same logic holds in reverse: if j is a high skill type, $\pi_t^j|_{c_\tau^i} < \pi_t^j|_{c_\tau^{i'}}$, $\mu_\tau^{i'|i} = 0$, and $\mu_\tau^{i|i'} > 0$. Then, $\mathbb{B}_t^j$ is larger and the marginal distortion is smaller for high-skilled types. Intuitively, we should decrease the marginal distortion on type j if doing so reduces earlier generations' incentive problems. In this way, the ancestor incentive effect pushes against the revenue and relative return effects, serving to increase marginal taxes at low incomes.

In the end, as this discussion suggests, the effect on optimal distortions of introducing endogenous ability is ambiguous. To get a sense for this ambiguity, consider the case of a low-ability parent. If parental resources have greater marginal effects on the children of low-skilled parents, then $\mathbb{A}$ and $\mathbb{C}$ are likely to be larger for these parents, reducing the optimal distortion due to the revenue and relative return effects. At the same time, $\mathbb{B}$ is

[13] Technically, the size will increase if the weighted sum of these effects, weighted by eventual child type, is larger.

likely to be smaller because increasing this low-skilled parent's resources makes it harder to incentivize previous generations to exert effort, increasing the optimal distortion due to the ancestor incentive effect. On net, the optimal distortion could be smaller or larger than in the conventional model.

Further intuition can be obtained by examining the case of only two ability types. In the case of two ability types, only one of the incentive constraints will bind within any given generation, allowing us to write result (7) more concisely for each ability type. We provide those expressions in the Appendix.

One of the lessons of Lemma 1 is that a two-period version of the model in this paper would obscure key aspects of the optimal policy problem. To see this, consider the two novel terms in the lemma, $\mathbb{A}_t^j$ and $\mathbb{B}_t^j$. $\mathbb{A}_t^j$ depends on how a marginal increase in current disposable income affects the tax revenue raised from all future generations. $\mathbb{B}_t^j$ depends on the incentive constraint multipliers for all previous generations. Any two-period model will neglect one of these two channels.

2.2 Allocations across generations

We now turn to analyzing intertemporal allocations. In a conventional model, the planner's first-order condition for c_t^j can be shown to equal:

$$\frac{\pi_t^j}{u'\left(c_t^j\right)}\lambda = \frac{1}{\beta^t}\left(\beta^t \pi_t^j + \sum_{j'} \mu_t^{j'|j} - \sum_{j'} \mu_t^{j|j'}\right).$$

Summing across types and combining with the same condition for generation $t+1$ immediately yields a condition on allocations across generations.

$$\sum_j \frac{\pi_t^j}{u'\left(c_t^j\right)} = \sum_k \frac{\pi_{t+1}^k}{u'\left(c_{t+1}^k\right)}. \tag{12}$$

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, such as in N. Gregory Mankiw and Weinzierl (2010).

With endogenous ability, expression (12) may not hold. Instead, a modified version of it applies, which we state in the following proposition and derive in the Appendix.[14]

Proposition 1 *The solution to the Planner's Problem satisfies*

$$\frac{1}{\Lambda_t}\sum_j \frac{\pi_t^j}{u'\left(c_t^j\right)} \frac{1-\beta\sum_k \frac{\partial p^k\left(w_t^j,c_t^j\right)}{\partial c_t^j} R_{t+1}^k}{1+\beta\sum_k \frac{\partial p^k\left(w_t^j,c_t^j\right)}{\partial c_t^j}\frac{U_{t+1}^k}{u'\left(c_t^j\right)}} = \frac{1}{\Lambda_{t+1}}\sum_k \frac{\pi_{t+1}^k}{u'\left(c_{t+1}^k\right)} \frac{1-\beta\sum_l \frac{\partial p^l\left(w_{t+1}^k,c_{t+1}^k\right)}{\partial c_{t+1}^k} R_{t+2}^l}{1+\beta\sum_l \frac{\partial p^l\left(w_{t+1}^k,c_{t+1}^k\right)}{\partial c_{t+1}^k}\frac{U_{t+2}^l}{u'\left(c_{t+1}^k\right)}} \tag{13}$$

where

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

To understand Proposition 1 intuitively, recall the meaning of the Symmetric Inverse Euler Equation in expression (12), namely 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 economy. In a conventional model, the average inverse marginal utility of disposable income in a generation determines the cost of raising welfare through transfers to a generation. In our model, that cost also depends on three novel factors that we now discuss.

First, if the transfer raises individual $j's$ investment in her children's abilities, resulting in increased tax revenue from future generations, these revenue gains offset the costs of the transfer. Formally, this factor is captured in the expression $\beta\sum_k\left[\partial p^k\left(w_t^j,c_t^j\right)/\partial c_t^j\right]R_{t+1}^k$, and it 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.

[14] Note that results (12) and (13) will hold for the special case in which consumption for each type and the ability distribution across types are exactly constant, regardless of whether the corresponding allocation is optimal from a welfare standpoint. Thus, these results are necessary but not sufficient qualities of the optimal policies (without and with endogenous ability, respectively).

Second, if the transfer raises individual $j's$ 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 matter for intertemporal allocations as well, i.e., through their effects on Λ_t. Note that Λ_t is equal to one 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[15] that were preventing low-income parents from having the disposable income to make relatively high-return investments 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. Intuitively, the more that binding incentive constraints are preventing investments in children in either the current or preceding generations, the more the planner wants to use transfers to relax those incentive constraints.

As one might expect, the overall implications of these three novel factors for optimal policy are theoretically ambiguous; to build intuition for their effects, consider a specific, empirically plausible scenario. Namely, suppose that mean ability is stable over time and the effects of parental resources on a child's ability are largest at lower skill levels.[16] Conventional policy designed to satisfy the expression (12) would treat generations symmetrically, and those allocations would satisfy equation (13). However, that conventional policy fails to take advantage of the endogeneity of the ability distribution.

[15] That in the usual way, i.e., high-ability parents are tempted to claim lower ability to obtain a more generous tax treatment.

[16] Formally, suppose $\sum_{k=1}^K \pi_t^k = \sum_{k=1}^K \pi_{t+1}^k$ 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 $\kappa > k$.

Consider, instead, a policy that transfers resources from generation $t+1$ to generation t and, in particular, increases the resources available to the low-ability workers in generation t. Such a policy would violate the conventional expression (12), as it would lower the marginal utilities of disposable income for generation t and raise them for generation $t+1$, increasing the left-hand side and decreasing the right-hand side of (12). Intuitively, the conventional perspective implies that the relative cost of raising welfare under such a policy is too high in the recipient generation t; the resources ought to stay with the future generation.

Such a policy is consistent with the true optimal policy condition (13), however, because of endogenous ability. To see why, note that the policy will increase the population proportion of higher-ability workers in generation $t+1$. This shift in the distribution of π_{t+1}^{k} will put greater weights 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, transfers from generation $t+1$ to generation t increase the cost of raising social welfare in generation $t+1$. This offsets what seemed to be a problem, namely that those transfers increased the cost of raising social welfare in generation t by directly lowering marginal utilities of income. Mathematically, then, this policy will increase both the left-hand and right-hand sides of (13). Therefore, equation (13) may be satisfied with a policy that treats generations asymmetrically and generates greater welfare. In other words, transfers from future to earlier generations generate gains for all generations: early generations gain from having higher disposable incomes, and future generations gain from having improved ability distributions.

This example implies that an optimal policy making use of the endogeneity of the ability distribution may differ from the conventionally-optimal policy. While result (13) does not prove that such a superior policy equilibrium exists, the simulations of Section 4 show that the scenario described above fits the empirical evidence from the United States, and that the potential welfare gains from such a policy are substantial.

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. If children's ability depends on their parents' hours worked, 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*

$$\max_{\left\{c_t^i, y_t^i\right\}_{t=1,i=1}^{T,I}} \sum_i p^i U_1^i \tag{14}$$

where

$$U_t^i = u\left(c_t^i\right) - v\left(\frac{y_t^i}{w_t^i}\right) + \beta \sum_{j=1}^{I} p^j \left(w_t^i, c_t^i, \frac{y_t^i}{w_t^i}\right) U_{t+1}^j \tag{15}$$

This is maximized subject to feasibility:

$$\sum_i p^i R_1^i \geq \bar{R}, \tag{16}$$

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

$$\sum_i p^i \left[\left(y_t^i - c_t^i\right) + \beta \sum_j p^j \left(w_t^i, c_t^i, \frac{y_t^i}{w_t^i}\right) R_{t+1}^j\right] \geq \bar{R}, \tag{17}$$

and incentive compatibility for each generation:

$$U_t^i \geq U_t^{i'|i} \quad \textit{for all generations } t \textit{ and types } i, i', \tag{18}$$

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

$$U_t^{i'|i} = u\left(c_t^{i'}\right) - v\left(\frac{y_t^{i'}}{w_t^i}\right) + \beta \sum_{j=1}^{I} p^j \left(w_t^i, c_t^{i'}, \frac{y_t^{i'}}{w_t^i}\right) U_{t+1}^j. \tag{19}$$

Simplifying as in the model without labor effort in the ability production function, we obtain the following Lemma parallel to the first.

Lemma 2 *Intratemporal Distortion with Parental Labor as an Input to Child Ability: Let $\mu_t^{i'|i}$ denote the multiplier on* (18) *. 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 $\frac{\frac{1}{w_t^{j\prime}}\partial p^k\left(w_t^{j\prime},c_t^j,\frac{y_t^j}{w_t^{j\prime}}\right)}{\partial y_t^j}<0$. First, the term $\mathbb{A}_t^j$ now captures that extra parental effort will lower future revenues, so the optimal downward distortion to 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 on lower types. Intuitively, incentive constraints are looser 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. In the Appendix, we show that including parental time as modeled here has only small effects on the results from our baseline quantitative analysis, which we describe in the next Section.

3 Model calibration under existing U.S. tax policy

In this section, we calibrate the model of Section 1 to estimates of the effect of parental resources on children's ability under existing U.S. tax policy. We focus our calibration on matching empirical estimates of statistics related to the transmission of ability across generations under the status quo tax policy. In particular, we minimize the distance (i.e. sum of squared deviations) between the model's output and the empirical estimates of the marginal effects of parental resources on their children's abilities, the transition matrix between generations, and the expected log wage within generations.

3.1 Generating 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 (forthcoming) study the effect of expansions of the EITC in the 1990s on children's test score outcomes.[17] Their study examines a specific

[17] See Joseph Hotz and Karl Scholz (2003) or Nada Eissa and Hilary Hoynes (2005) for detailed descriptions of the EITC program and a summary of related research. The marginal effect of parental resources on child ability is difficult to estimate for at least two reasons. First, it is difficult to find plausibly exogenous variation in parents' disposable income levels. Second, it is difficult to find data on parents' disposable income and wage levels linked to measures of their children's outcomes. Several papers have estimated the effect of parents' income on their children's achievement levels (e.g. Dahl and Lochner forthcoming; Kevin Milligan and Mark Stabile forthcoming; Christine Paxson and Norbert Schady 2007; Randall Akee, William Copeland, Gordon Keeler, Adrian Angold, and Jane Costello 2010; Katrine Løken, Magne Mogstad, and

context, and we must generalize outside of the specific features of this context with caution. While recognizing this caveat, we choose to examine this context because we believe that it represents one of the best available opportunities to study the effect of tax policy toward parents on children's outcomes in the United States. We refer readers to their paper for a full description of their empirical strategy and its motivation, but we briefly describe their empirical strategy here, often borrowing from their description of it.

The size of the EITC, which is a refundable tax credit primarily benefitting low- and middle-income families, depends on earned income and the number of qualifying children. The EITC tax schedule has three regions. Over the "phase-in" range, a percentage of earnings is transferred to individuals. Over the "plateau" region, an individual receives the maximum credit, after which the credit is reduced (eventually to zero) in the "phase-out" region. Near the period studied in this paper, the EITC was expanded substantially in the tax acts of 1986, 1990, and 1993. The largest expansion of the EITC was in 1993. This reform increased the additional maximum benefit for taxpayers with two or more children, which reached $1400 in 1996. The phase-in rate for the lowest-income recipients increased from 18.5% to 34% for families with one child and from 19.5% to 40% for families with two or more children.

Dahl and Lochner ask how the EITC and other tax and transfer programs affect the cognitive achievement of disadvantaged children through their effects on parental income. 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's analysis uses the Children of the National Longitudinal Survey of

Matthew Wiswall 2012; Karen Macours, Schady, and Renos Vakis 2012). However, they have not estimated the effect of parents' disposable income on children's wage rates in large part because linking the income of children's parents when the children were young to children's wage outcomes when they have grown into adults requires a long panel of data in which all of these variables are linked. This coincidence of data is unlikely in circumstances with suitable exogenous variation in parents' disposable income. In fact, our paper suggests a new empirical object of interest that should be studied in future work: the effect of parents' disposable income on children's wages.

Youth, 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 five repeated measures of cognitive test scores per child taken every other year. The data are longitudinal, implying that it is possible to first-difference the data to remove child fixed effects. They use measures of child ability based on standardized scores on the Peabody Individual Achievement Tests (PIAT), which measures oral reading ability, mathematics ability, word recognition ability, and reading comprehension. 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. 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.[18]

3.1.1 Regression Specification

We estimate a model similar to Dahl and Lochner's, using the same basic sample of data they use (described more fully in their paper and below), but we use it to obtain a slightly different empirical object. Motivated by our model above and 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. 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).[19] 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

[18] We use year 2000 dollars throughout.

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

family disposable income from age $a-1$ to age a.[20] 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 possibly different) ability level. We adapt the Dahl and Lochner empirical specification by estimating the following model:

$$D_{ia}^{l} = x_i'\alpha + \Delta c_{ia}\beta + W_{i,a-1}\delta + \Phi(y_{i,a-1}) + \eta_{ia} \tag{20}$$

using χ_a^{IV} as an instrument for Δc_{ia}. We relate a binary dummy D_{ia}^{l} equal to one when the

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

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.[21] 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.[22]

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

[21] 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.

[22] 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.

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.[23]

In our main specification, we divide parents into five wage (ability) categories $\{P_i\}_{i=1}^5$ and divide children into five test score categories $\{C_i\}_{i=1}^5$. Each category comprises one quintile of the sample distribution of wages or test scores, respectively, with subscript i indicating the quintile of the distribution, where $i = 1$ is the lowest quintile. Because there are five parent types, we estimate five separate regressions, in each of which the dependent variable is a dummy that equals one when the child has ability in the ith 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 considerations: more categories will 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. In the Appendix, we show results with 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 analysis with ten types yield similar results as does our benchmark five-type model.

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.[24] We control for the child's initial test score category (i.e. by quintile), 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%

[23] 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.

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

of respondents are mothers.[25]

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.

As we discuss later, the formal model whose moments we will match to the data will be specified in 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).[26] 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

[25]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.

[26]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.

use in our regressions).

Appendix Table 1 shows summary statistics. Children's mean age is 11.31 years old. Nearly half of the children are male. Respondents work a mean of 1,692.68 hours per year. The mean calculated hourly wage is \$8.16/hour.

3.1.2 Empirical 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 (20) 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.

We show the main empirical results in Table 1. For each regression, we show the estimated effect β and its standard error in parentheses.

Table 1: Empirical marginal effects of parental resources on child ability distribution, in percentage points

	Child type						
Parent type	1	2	3	4	5	N	n
1 (lowest)	-0.06	-0.01	0.24	0.89	-1.06	1,192	421
	(1.00)	(1.07)	(1.02)	(1.31)	(1.45)		
2	-0.54	0.17	0.16	-0.06	0.27	1,531	422
	(0.34)	(0.34)	(0.33)	(0.25)	(0.28)		
3	-0.15	-0.53	0.32	0.17	0.19	1,461	421
	(0.29)	(0.41)	(0.41)	(0.34)	(0.29)		
4	-0.63	0.28	0.40	0.13	-0.18	1,397	422
	(1.38)	(1.47)	(1.40)	(1.21)	(1.00)		
5 (highest)	0.24	0.09	-0.59	-0.30	0.56	1,321	422
	(0.38)	(0.48)	(0.57)	(0.41)	(0.56)		

Notes: The dependent variable is a dummy variable that equals 1 if the child is in a given ability quintile. 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 quintile 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 five regressions estimated in a parent quintile. N differs across regressions because the number of missing observations differs across parents; n differs slightly across regressions because the total number of parents is not a multiple of five (and the results are not sensitive to the this allocation across quintiles). 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.

The point estimates of the coefficients in Table 1 sum to zero across child types (within a parent type). The signs of these point estimates are generally what one would expect: higher parental income usually increases the probability that a child is high-ability and decreases the probability that a child is low-ability. In no case, however, are the point estimates statistically significantly different from zero. The regressions in Appendix Table 4 additionally control for the first-difference of parent hours worked and again show a very small impact of parent hours on child ability, as we discuss further in the Appendix.

The estimates in Table 1 provide some the statistics targeted by our calibration. The additional targets are the elements of the empirical ability transition matrix between generations and the expected log wage within a generation. Using the same dataset and definition of types as in the analysis just described, we can readily generate the transition matrix by calculating the fraction of the sample from each parent wage category who began the sample period with the child test score in each category. The results are in Table 2.[27]

Table 2: Empirical ability transition matrix

	Child type						
Parent type	1	2	3	4	5	N	n
1	0.27	0.21	0.19	0.16	0.17	1,192	421
2	0.23	0.20	0.20	0.19	0.18	1,531	422
3	0.19	0.21	0.21	0.20	0.19	1,461	421
4	0.18	0.19	0.19	0.23	0.21	1,397	422
5	0.14	0.19	0.20	0.22	0.25	1,321	422

See notes to Table 1.

The expected log wage is also readily calculated, as the average of the log of the five ability levels shown above, to be 2.06.

3.2 Model specification

We next describe how the model produces quantities corresponding to these target statistics, and we specify some components of the model required for simulation.

[27] We do not show the standard errors on these estimates, as they are nearly uniform at approximately 0.01.

The quantities corresponding to the targeted statistics are generated by the model as follows. In the planner's problem, the production function for a child's ability was left unspecified. Here, we impose a particular, tractable form:[28] the expected ability of the child of a parent of type j with disposable income $c_1^{j'}$ is

$$E\left[\ln w_{t+1} | w_t^j, c_t^{j'}\right] = \alpha_i \left(\rho \ln w_t^j + (1-\rho)\ln \bar{w}\right) + \alpha_c^j \ln c_t^{j'}. \tag{21}$$

Expression (21) shows that the child's expected ability is a function of the parent's ability, a fixed "mean" ability, 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 the empirical evidence on income, e.g. Steven Haider and Gary Solon 2006).

This log-linear functional form concisely captures the basic forces at work in this model determining the transmission of ability across generations. Namely, it allows us to adjust the role of parental ability in determining a child's ability through the parameter ρ. It also allows us to vary the relative importance of this channel and a second channel, parental resources, by adjusting the parameters α_i and α_c^j. Note that the dependence of α_c^j, the parameter controlling the importance of parental disposable income, on j, the parental ability type, establishes a direct connection between the exogenous and endogenous components of the ability production function.

The specification in expression (21) imposes no restrictions on whether the marginal value of parental resources is increasing or decreasing in the child's innate ability. In particular, our assumed production function for child ability allows α_c^j to vary with parental type j in expression (21). Depending on how α_c^j varies with j—a relationship we will estimate in our simulations—the marginal value of parental resources may increase, be constant, decrease, or exhibit complex nonlinearities as innate ability increases.[29]

[28] In (21) and elsewhere, in the model specification we make several assumptions for the sake of tractability. This fact notwithstanding, we view our model specification and calibration as demonstration that taking into account the effect of parents' resources on children's abilities can have important implications for optimal tax policy.

[29] 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

We translate the expected ability in expression (21) 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 :

$$\ln w_{t+1} \sim N\left(E\left[\ln w_{t+1}|w_t^j, c_t^{j'}\right], \sigma^2\right). \tag{22}$$

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 expressions (21) and (22) 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 expression (22) into transition probabilities among types across generations.[30]

Applying this procedure, we can generate the transition probabilities $\pi_{t+1}^k|_{c_t^j}$ for all parent and child types. This structure also enables us to calculate the marginal effects of parental disposable income 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. Formally, to compute the marginal effect of c^j, we calculate the semi-elasticity of the probability of each child type with respect to parental disposable income. That is, we calculate the change in the probability of each child type associated with an incremental increase in the log of parental resources c^j.

Expressions (21) and (22) indicate that the model calibration will search over values of

income in (21).

[30] 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 (21). 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.

the following parameters: $\left\{\rho, \alpha_i, \{\alpha_c^j\}_{j=1}^I, \sigma\right\}$, such that with $I = 5$ there are eight values to estimate. As a baseline case, we will impose $\rho = 0.5$ for the parameter controlling the transmission of ability across generations. This assumption is based on the voluminous evidence surveyed in Marcus Feldman, Sarah Otto, and Freddy Christiansen 2000.[31] We show the robustness of our results to this choice in the Appendix. We also impose the value of $\sigma = 0.76$ from the NLSY sample. This leaves six parameters to be chosen by this calibration.

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. 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. These authors' detailed calculations go well beyond statutory personal income tax schedules and include a wide array of transfer programs (such as Social Security, Medicare, Medicaid, Food Stamps, and low-income benefit programs such as the Earned Income Tax Credit) as well as corporate income taxes, payroll taxes, and state and local income and sales taxes. Our computational procedure requires a smooth tax function, so we take a fifth-order polynomial approximation of the Kotlikoff-Rapson schedule over annual incomes up to $75,000, a level well above the point at which the EITC is fully phased out. (Other polynomial approximations, including a third-order polynomial approximation, yield very similar results). This approximation and the corresponding original Kotlikoff-Rapson estimates are shown in Figure 1.

[31] Feldman et al. find a range of heritability estimates from 0.28 to 0.38 (their h^2) and a "cultural transmission" estimate (their b^2) of 0.27 (see their Table 4.3). The mapping between these channels and our "ability" channel is imperfect. The two channels together could explain nearly two-thirds of the variance in a characteristic. But while all of the former channel is contained with our notion of "ability," it is not clear that all of the latter is so contained. We use 0.50 as a reasonable middle ground.

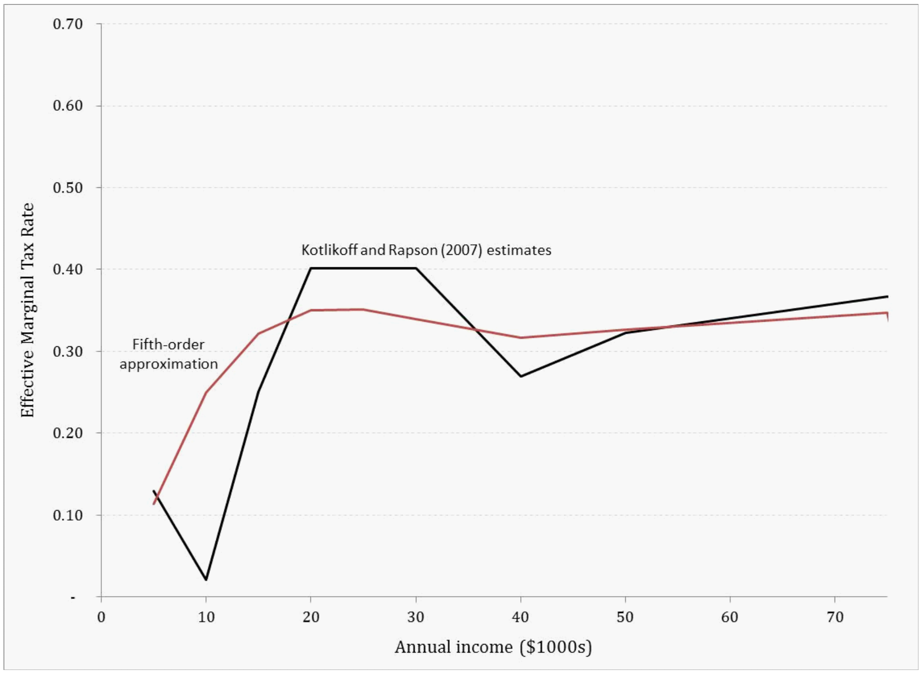


Figure 1: Effective marginal tax rates in the U.S.

Note that our data do not allow us to extend our calibration directly to higher incomes, a limitation that could, in principle, affect our results because both the existing and optimal tax policies would redistribute substantial resources from higher earners. We show an extension of our analysis in the Appendix that suggests our results are robust to this potential concern.

The government's tax system also includes a grant to all individuals, which is constant across generations, as are tax rates. As in the feasibility constraint on the planner, expression (3), the government's budget is balanced in present value, where we set $\beta = 1.00$, reflecting no discounting of utility across generations. In the Appendix, we show that our results are robust to a modest degree of discounting, but note that there is no growth in this economy, so any discounting reflects solely a preference for the utility of earlier generations.

The individual utility function takes a separable, isoelastic form

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

where γ controls the concavity of utility from disposable income, σ controls the elasticity of labor supply, and θ is a taste parameter affecting the level of labor effort. Again, we choose this functional form for the sake of tractability and because it helps in illustrating the key features of the model in a straightforward way. We set $\gamma = 2$ and $\sigma = 3$ to be consistent with mainstream estimates of these parameters (which implies that the Frisch elasticity of labor supply is $\frac{1}{2}$). We choose $\theta = 2.5$ so that hours worked in the simulation approximately match the average labor supply in the population.[32]

Finally, guided by the empirical analysis discussed above, we assume ability comes in $I = 5$ fixed types (roughly interpretable as the hourly wage):[33] $w_t^i \in \{3.01, 5.55, 7.74, 10.58, 21.38\}$ 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.[34]

3.3 Calibration Results

To calibrate the model, we minimize a weighted sum of squared errors, where the targets are the marginal effects and transition matrix shown in Tables 1 and 2 as well as the mean log wage. We weight the squared errors by the inverse of the targets' standard errors, which has the effect of putting much greater weight on the more-precisely-estimated transition matrix elements and the mean log wage. We use ten generations ($T = 10$) in the simulations, allowing for several generations surrounding the middle generation that we use as the target for the calibration exercise. We show robustness to this choice in the Appendix.

Table 3 shows the parameter values chosen by the simulation.

Table 3: Parameter values chosen in calibration

	α_i	α_c^1	α_c^2	α_c^3	α_c^4	α_c^5
Value under status quo policy	0.86	0.67	0.32	0.22	0.14	0.00

Recall that α_i and α_c^j are the weights on the two channels, ability and economic resources,

[32] That is, effort comprises approximately 40 percent of available time in the simulation. If the maximum hours sustainably available for work are approximately 80 per week, this yields total hours of work around 1600 hours per year. The value of θ is unimportant for the results of interest in our analysis.

[33] These values are the mean reported wages by quintile for the NLSY sample we use.

[34] Note that we also implicitly assume that all adults' consumption has an impact on children. In the U.S. population in 2000, 85 percent of adults over 45 have had a biological child (Health and Human Services 2002), with even larger percentages at older ages.

through which parents affect their child's ability. The product of ρ and α_i 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 3 suggest that parental resources play a greater role among lower-ability parents, consistent with the empirical evidence.[35] Key moments determining the estimates of the α_c^j are the coefficients on parent income in determining child ability from Table 1. Key moments determining both the estimates of the α_c^j and the estimate of α_i are the elements of the transition matrix of parent ability to child ability in Table 2, as these determine the combined role that parent ability and parent resources play in determining child ability.

The simulation does well in matching the empirical targets for which the data is most informative, namely the transition matrix and mean log wage. The simulation yields marginal effects of parental resources that differ substantially from the data, as Table 4 shows for the fifth (middle) generation of the simulation. The calibrated status quo marginal effects exhibit a pattern much closer to what intuition would suggest—negative for lower child types and positive for higher child types—than do the estimated effects in the data. This is not surprising, however, given the statistical insignificance of the empirical estimates and their often-unexpected signs.

[35] Dahl and Lochner (forthcoming), Milligan and Stabile (2008), Paxson and Schady (2007), Akee et al. (2010), and Løken, Mogstad, and Wiswall (2012) find a larger effect of parental income on child achievement among lower-income families than among higher-income families. Consistent with these findings, we find that within each parent ability level the effect of parental income on child achievement is concave.

Table 4. Marginal effects of parental resources

	Data						Calibrated status quo policy				
	Child type						Child type				
Parent type	1	2	3	4	5	Parent type	1	2	3	4	5
1	-0.06	-0.01	0.24	0.89	-1.06	1	-0.29	-0.07	0.03	0.13	0.20
2	-0.54	0.17	0.16	-0.06	0.27	2	-0.13	-0.04	0.01	0.05	0.10
3	-0.15	-0.53	0.32	0.17	0.19	3	-0.08	-0.03	0.00	0.03	0.08
4	-0.63	0.28	0.40	0.13	-0.18	4	-0.05	-0.02	0.00	0.02	0.05
5	0.24	0.09	-0.59	-0.30	0.56	5	0.00	0.00	0.00	0.00	0.00

Table 5 shows that the simulation closely matches the data for the transition matrix between generations (we show the transition between the fifth and sixth generations of the simulation as an illustrative example).

Table 5. Transition matrix

	Data						Calibrated status quo policy				
	Child type						Child type				
Parent type	1	2	3	4	5	Parent type	1	2	3	4	5
1	0.27	0.21	0.19	0.16	0.17	1	0.26	0.23	0.17	0.20	0.14
2	0.23	0.20	0.20	0.19	0.18	2	0.23	0.22	0.17	0.22	0.17
3	0.19	0.21	0.21	0.20	0.19	3	0.20	0.21	0.17	0.23	0.19
4	0.18	0.19	0.19	0.23	0.21	4	0.18	0.20	0.17	0.24	0.21
5	0.14	0.19	0.20	0.22	0.25	5	0.15	0.19	0.16	0.25	0.26

Finally, the simulation matches the mean log wage, 2.06, to two significant digits.

As we show in the Appendix, these results are robust to varying time discounting β, the number of generations T, the assumed persistence of type across generations ρ, and the number of types I.

4 Optimal Policy

In this section, we simulate a many-period version of the planner's problem using the calibration from the previous subsection. We characterize optimal policy by comparing it to the status quo policy used in that calibration.

Table 6 shows average and marginal tax rates for each type under the optimal and status quo policies.[36] 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 optimal allocation. The latter are the wedges that distort individuals' choices of labor effort. In the discussion of Lemma 1, we showed that the wedge for parent of type i in generation t, which we denote as τ_t^i, can be written as $\tau_t^i = 1 - \mathbb{A}_t^k\left(\mathbb{B}_t^k + \mathbb{C}_t\right) / \left(\mathbb{B}_t^k + \mathbb{D}_t\right)$, where $\mathbb{A}_t^k$, $\mathbb{B}_t^k$, $\mathbb{C}_t$, and $\mathbb{D}_t$ are defined above in expressions (8), (9), (10), and (11).

Table 6. Marginal and average tax rates

	Marginal tax rate		Average tax rate	
Type	Optimal	Status Quo	Optimal	Status Quo
1 (lowest)	32%	19%	-235%	-49%
2	42%	29%	-78%	-19%
3	52%	34%	-38%	-7%
4	50%	35%	-4%	3%
5 (highest)	0%	32%	40%	17%

Table 6 shows that the optimal policy has very different average and marginal tax rates than the status quo. The optimal policy is substantially more redistributive, generating large transfers to low-skilled parents. This result is due entirely to the redistributive preferences of the social planner.[37] Nevertheless, these redistributive transfers generate an improved ability

[36] These rates–or close approximations of them–hold for all but the initial and final generations, which differ slightly given their endpoint status.

[37] 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

distribution by capitalizing on the gap between the impact of increased disposable income on the children of low-ability parents and high-ability parents. Optimal policy imposes large and decreasing (with income) marginal distortions to make the allocations for lower types less attractive to those with higher ability, who expect to have children with higher ability on average.

The optimal policy also adjusts intertemporal allocations to capitalize on the endogeneity of ability, as was suggested in the discussion of Proposition 1. Table 7 reports the difference between the planner's "budget balance" as a share of aggregate income in each generation under the optimal policy and under the status quo policy. In other words, it is the additional average tax rate assessed on each generation by the planner, relative to a balanced budget as assumed to hold in the status quo.

Table 7. Intertemporal allocations

	Difference in government budget balance (as percent of output in each generation)									
	1	2	3	4	5	6	7	8	9	10
Optimal policy - Status quo policy	-7.7%	-0.3	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.2	14.1

Table 7 shows that the optimal policy borrows from future generations to fund greater investment 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 7 is that society can benefit by having later generations contribute, through higher taxes, to improving the ability distribution generated by earlier generations.[38]

These differences in tax policy affect the evolution of the ability distribution. We report the transition matrices for types across generations under the optimal and status quo policies.

be optimal to have these workers exert more effort, earn more income, and therefore enjoy more disposable income than in a static model.

[38]The United States is running substantial yearly budget deficits as of 2013 and did in 2005 when the Kotlikoff and Rapson tax rates are calculated. Our "status quo policy" abstracts from this aspect of reality, the causes of which are myriad.

Table 8 repeats the transition matrix from Table 5 for the calibrated status quo model and compares it to the transition matrix under the optimal policy.

Table 8. Transition matrix

	Optimal policy						Status quo policy				
	Child type						Child type				
Parent type	1	2	3	4	5	Parent type	1	2	3	4	5
1	0.18	0.20	0.17	0.24	0.22	1	0.26	0.23	0.17	0.20	0.14
2	0.21	0.22	0.17	0.22	0.18	2	0.23	0.22	0.17	0.22	0.17
3	0.20	0.21	0.17	0.23	0.19	3	0.20	0.21	0.17	0.23	0.19
4	0.18	0.21	0.17	0.23	0.21	4	0.18	0.20	0.17	0.24	0.21
5	0.14	0.18	0.16	0.25	0.26	5	0.15	0.19	0.16	0.25	0.26

The optimal 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. The optimal policy has much smaller effects on the prospects of the children of higher-skilled parents. Intuitively, Tables 6 and 7 show that the optimal policy takes resources from higher-skilled parents and later generations to support lower-skilled parents and earlier generations. This moves resources from those for whom the effect of resources on a child's ability is lower to those for whom they are higher (that is, from smaller to greater values of α_c^i).

As these transition matrices imply, the evolution of the ability distribution is different under the optimal and the status quo policies. Figure 2 shows the ability distribution in the fifth generation under the two policies, which closely resembles the distribution in all generations after the initial one. This figure shows the substantial shift toward a higher ability distribution under the optimal policy that results from the greater progressivity of the optimal policy; the optimal policy leads to 2.1 percent fewer individuals of the lowest type and 1.9 percent more individuals of the highest type, with smaller differences for intermediate types.

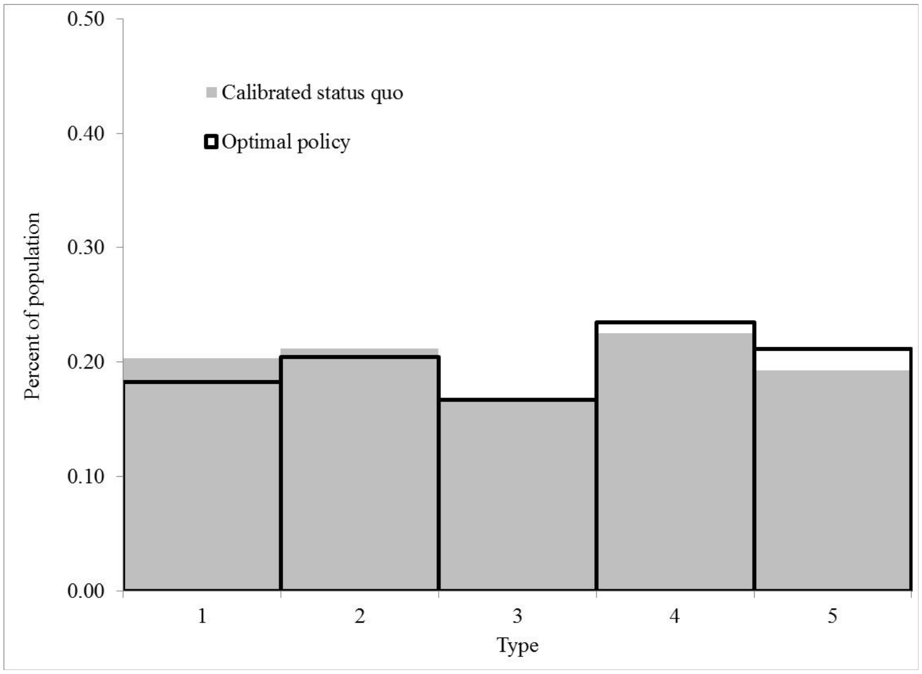


Figure 2: Ability distribution under calibrated status quo policy (filled-in columns) and social-welfare maximizing policy (empty outlined columns).

Welfare is much higher under the optimal policy, and it is more equitably distributed. In fact, the welfare gain of moving from the status quo policy to the optimal policy is enormous: it is equivalent to a 22.7% permanent increase in disposable income. But this very large gain is predominately driven by something other than the effect of policy on the ability distribution. In particular, the optimal policy's Utilitarian foundation places a high value on income equality, so the greater redistribution to low-skilled parents under the optimal policy than under the status quo policy generates most of this large estimated increase in welfare. Because we may be interested in the importance of the endogenous ability channel alone in generating welfare gains, we consider the following thought experiment.

Suppose that the status quo model were granted the distribution of abilities generated by the optimal model for all generations; we call this the "adjusted status quo." Suppose further that we hold fixed the within-period utility levels of all individuals in the status quo model, but we calculate the total welfare for the economy given the adjusted status quo ability distributions. This will generate a greater level of welfare. Now, returning to the status quo tax policy's ability distributions, we calculate the factor by which disposable income would have to rise in the status quo model to reach the welfare of the adjusted status quo. This factor is a measure of the welfare gain due solely to the optimal policy's effects on the ability distribution over time. Similar factors can be calculated for each type of first-generation parent, as well, indicating how the welfare gains through this channel are shared. Table 9 shows the results for the baseline case of ten generations.

Table 8. Welfare

	Welfare level (in utils)		Welfare gain
	Status quo policy	Adjusted status quo	(Percent of disposable income)
Overall	6.39	6.46	2.54%
Type-specific			
1 (Lowest)	6.17	6.27	3.4%
2	6.32	6.39	2.5%
3	6.40	6.46	2.3%
4	6.46	6.52	2.2%
5 (Highest)	6.58	6.64	2.3%

As these results show, the optimal policy has the potential to generate a welfare gain equivalent of around two and one-half 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 efficiency gains and greater equality accruing to future generations raise the current generation's present-value welfare. Gains for future generations follow the same patterns.

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. The qualitative and quantitative lessons of the baseline analysis prove to be robust. In particular, optimal policy that takes advantage of endogenous ability is more redistributive than the status quo, shifts resources from future to earlier generations, generates an upward shift in the ability distribution, and yields a sizeable welfare gain.

5 Conclusion

In this paper, we explore the possibility that equalizing individuals' economic outcomes may help to equalize their children's opportunities: that is, when poor parents have more disposable income, their children's performance improves and they have greater opportunity to succeed. We study the effect that this intergenerational connection has on optimal tax policy, which will take advantage of this relationship to shape the ability distribution over time. But exactly how it will do so depends on complex interactions between natural ability and the returns to investment in human capital. Ours is the first paper we know of to model this complexity and derive policy implications.

We characterize conditions describing optimal tax policy when children's abilities depend on both inherited characteristics and parental (financial) resources. On the intratemporal margin, we highlight competing effects of this endogeneity. If parental resources have greater marginal effects on the children of low-skilled parents, then optimal distortions may be smaller at low incomes because of their positive effects on overall tax revenues and the incentives of high-skilled parents. On the other hand, larger distortions at low incomes have a benefit in encouraging preceding generations to invest in their children's ability pushes in the other direction. In the end, the implications for optimal marginal distortions are ambiguous. On the intertemporal margin, we show that optimality requires a more sophisticated understanding of the cost of raising social welfare through transfers across generations, in particular including the effects of one generation's resources on future generations' tax payments and utilities.

We calibrate our model to microeconometric evidence on the transmission of skills and new estimates of the effects of increases in disposable income on a child's ability, which we obtain by analyzing panel data from the NLSY in the United States. We then simulate optimal policy in this calibrated model and compare it to an estimated version of the existing

U.S. tax code. The schedules of optimal average and marginal tax rates are very different from those in existing policy, as the optimal policy is substantially more redistributive and shifts the ability distribution up over time. This shift in the ability distribution generates a welfare gain equivalent to more than 2.5 percent of total disposable income in perpetuity, with larger gains for the poor. Even higher-skilled members of the current generation gain substantially, however, as the gains in efficiency and equality in future generations raise the current generation's present-value welfare.

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. Incorporating other dimensions of parental influence is another natural next step. We have shown (in the Appendix) 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.

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

[5] 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.

[6] 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.

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

[8] 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.

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

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

[11] 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.

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

[13] Dahl, Gordon, and Lance Lochner. "The Impact of Family Income on Child Achievement: Evidence from Changes in the Earned Income Tax Credit." *American Economic Review* (forthcoming).

[14] 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.

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

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

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

[18] 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.

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

[20] 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.

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

[22] 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.

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

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

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

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

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

[28] 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.

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

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

[31] 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.

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

[33] 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.

[34] 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.

[35] 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.

[36] 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.

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

[38] 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.

[39] 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.

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

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

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

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

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

6 Appendix

6.1 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.

6.1.1 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 (7), for each ability type in generation t, is as follows:

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

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

$$*\left(\frac{\beta^t\pi_t^i+\sum_{\tau=1}^{t-1}\beta^{t-\tau}\mu_\tau^{i|j}\left(\pi_t^i|_{c_\tau^j}-\pi_t^i|_{c_\tau^i}\right)-\mu_t^{i|j}\frac{1+\beta\sum_k\frac{\partial p^k\left(w_t^j,c_t^i\right)}{\partial c_t^i}\frac{U_{t+1}^k}{u'\left(c_t^i\right)}}{1+\beta\sum_k\frac{\partial p^k\left(w_t^i,c_t^i\right)}{\partial c_t^i}\frac{U_{t+1}^k}{u'\left(c_t^i\right)}}}{\beta^t\pi_t^i+\sum_{\tau=1}^{t-1}\beta^{t-\tau}\mu_\tau^{i|j}\left(\pi_t^i|_{c_\tau^j}-\pi_t^i|_{c_\tau^i}\right)-\mu_t^{i|j}\frac{\frac{1}{w_t^j}v'\left(\frac{y_t^i}{w_t^j}\right)}{\frac{1}{w_t^i}v'\left(\frac{y_t^i}{w_t^i}\right)}}\right) \tag{23}$$

$$\frac{v'\left(y_t^j/w_t^j\right)}{w_t^j u'\left(c_t^j\right)} = \frac{1}{1-\beta\sum_k \frac{\partial p^k\left(w_t^j,c_t^j\right)}{\partial c_t^j} R_{t+1}^k}\left(1+\beta\sum_k \frac{\partial p^k\left(w_t^j,c_t^j\right)}{\partial c_t^j}\frac{U_{t+1}^k}{u'\left(c_t^j\right)}\right) \tag{24}$$

6.2 Proof of Proposition 1

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

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

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

Then, sum each side over k:

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

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

We can rewrite this as:

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

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

Rearranging, we obtain

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

This holds for all t, yielding the Proposition.

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

where

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

6.3 Proof of Lemma 2

The first order condition (FOC) for consumption is identical to the baseline case other than the form of $p(\cdot)$:

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

The FOC for income is:

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

Simplifying as in the proof of Lemma 1 yields the result.

6.4 Appendix Table 1

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.

6.5 Appendix Table 2

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.

6.6 Appendix Table 3

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.

6.7 Effects of Parental Time Allocation

In this section, we incorporate parental time allocation into the quantitative analysis. First, Appendix Table 4 presents the marginal effects of both parental disposable income and parental work effort on child ability.

Appendix Table 4: Empirical marginal effects of parental resources on child ability distribution, in percentage points

	Parent type				
Child type	1	2	3	4	5
1 (lowest)	-0.13, 0.000020	-0.66, 0.000080	-0.13, -0.0000077	-0.51, 0.000030	0.1558629, -0.000031
	(0.70, 0.000053)	(0.47, 0.000062)	(0.30, 0.000034)	(1.41, 0.00014)	(0.35, 0.000026)
2	0.16, -.0000074	0.21, -0.000045	-0.56, .000084	0.030, 0.000015	0.16, 0.000020
	(0.77, 0.000057)	(0.42, 0.000054)	(.41, 0.000048)	(1.627523, 0.00016)	(0.49, 0.00003)
3	-0.15, 0.0000060	0.15, -0.000025	0.26, -0.000028	0.18, 0.000017	-0.51, 0.0000232
	(0.66, 0.000051)	(0.41, 0.000054)	(0.42, 0.000046)	(1.55, 0.00016)	(0.51, 0.000035)
4	0.56, -0.000032	.037, -0.000014	0.16, -0.0000088	0.63, -0.000080	-0.31, 0.000037
	(0.71, 0.000053)	(0.32, 0.000043)	(0.36, 0.00004)	(1.65, 0.00017)	(0.36, 0.000030)
5 (highest)	-0.43, 0.000013	0.27, 0.0000029	0.26, -0.000039	-0.32, 0.000018	0.50, -0.000050
	(0.57, 0.000043)	(0.36, 0.000046)	(0.31, 0.000035)	(1.31, 0.00013)	(0.43, 0.000031)
N	1,114	1,473	1,411	1,352	1,252
n	403	414	416	413	411

Notes: The dependent variable is a dummy variable that equals 1 if the child is in a given ability quintile. 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 quintile of the parent wage distribution. 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. 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 five regressions estimated in a parent quintile. 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. As in Table 1, parent income tends to have a positive impact on the probability that a child is in low ability categories and a negative impact on the probability that a child is in high ability categories.

We use these estimated marginal effects as additional targets in a calibration exercise similar to that described in the paper. In particular, we target the fifty marginal effects from Appendix Table 4 as well as the transition matrix and mean log wage as in Section 3, calibrating the following reduced-form equation analogous to expression (21) in the paper

$$E\left[\ln w_2 | w_1^j, c_1^{j'}\right] = \alpha_i \left(\rho \ln w_1^j + (1-\rho)\ln \bar{w}\right) + \alpha_c^j \ln c_1^{j'} + \alpha_l^j l_1^{j'}, \tag{25}$$

where $l_1^{j'}$ is the labor effort of the parent of type j'. The calibration therefore searches for values of the parameters $\left\{\rho, \alpha_i, \{\alpha_c^j\}_{j=1}^{I}, \left\{\alpha_l^j\right\}_{j=1}^{I}, \sigma\right\}$ to match the empirical targets, such that with $I = 5$ there are thirteen values to estimate. As in the paper, we assume baseline values for $\rho = 0.50$ and $\sigma = 0.76$.

The results of the calibration exercise are as follows. First, the chosen parameter values are shown in Appendix Table 5.

Appendix Table 5: Parameter values chosen in calibration

	α_i	α_c^1	α_c^2	α_c^3	α_c^4	α_c^5
With effect of labor	0.89	0.63	0.27	0.17	0.10	0.00
Baseline	0.86	0.67	0.32	0.22	0.14	0.00
		α_l^1	α_l^2	α_l^3	α_l^4	α_l^5
With effect of labor		-1.3E-5	-2.0E-4	-8.0E-5	-1.6E-4	-0.21
Baseline		n/a	n/a	n/a	n/a	n/a

As these results indicate, the values for α_i 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 larger negative value for the highest ability type suggests that incentivizing more effort among the higher earners in this group may have some costs in terms of child ability.

As in the baseline case, the simulation does well in matching the empirical targets for which the data is most informative, namely 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 in the paper.

The optimal policy given the calibrated model also strongly resembles that in the paper (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 1.6% more of the population is of the highest type and 1.9% 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 2.13% (compared to 2.54% 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.

6.8 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 $\{\alpha_c^i\}_i$ parameters—which reach zero at a parental wage of approximately \$21—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 $\bar{R}$ from (3) 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 our sample (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 main paper 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 6 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 6. 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 6, with the baseline results from the main paper shown below them for reference:

Appendix Table 6: Parameter values chosen in calibration

	α_i	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 $\{\alpha_c^i\}_i$ 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_i is unchanged. These parameters yield marginal effects of parental resources and transition matrices similar to those in the baseline analysis. They also yield the following marginal and average tax rates:

Appendix Table 7. 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 main 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 main analysis, as shown in Appendix Table 8.

Appendix Table 8: Welfare gains						
	Overall	Type-specific gains				
	welfare gain	1 (lowest)	2	3	4	5 (highest)
With negative revenue requirement	2.50%	2.9%	2.5%	2.4%	2.3%	2.3%
Baseline analysis	2.54%	3.4%	2.5%	2.3%	2.2%	2.3%

6.9 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.

6.9.1 Value of ρ

ρ indicates the role of parental ability, relative to a mean ability level, in determining a child's ability. As shown in expression (21), 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 two other values of ρ, namely 0.40 and 0.60. Appendix Table 9 shows the parameter values chosen by the simulations (as in Table 3). 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 (as in Table 8). The results for the baseline case of $\rho = 0.50$ are given for reference.

Appendix Table 9: Parameter values with alternative ρ values

	α_i	α_c^1	α_c^2	α_c^3	α_c^4	α_c^5
$\rho = 0.40$	0.90	0.50	0.23	0.16	0.11	0.00
$\rho = 0.50$	0.86	0.67	0.32	0.22	0.14	0.00
$\rho = 0.60$	0.83	0.82	0.40	0.27	0.17	0.00

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

	Ability type					Welfare gain
	1	2	3	4	5	α_c^5
$\rho = 0.40$	0.19	0.21	0.17	0.23	0.21	1.7%
$\rho = 0.50$	0.18	0.20	0.17	0.23	0.21	2.5%
$\rho = 0.60$	0.18	0.20	0.17	0.24	0.22	3.2%

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 Appendix Table 6 shows 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 $\{\alpha_c^j\}_{j=1}^I$ values make optimal policy more powerful.

6.9.2 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 3 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 2.52 percent of total income if $T = 8$ compared to 2.54 percent in the baseline case of $T = 10$ and 2.59 percent in the case of $T = 12$.

6.9.3 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 3 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., 2.44 percent of total consumption compared to 2.54 percent).

6.9.4 Value of I

Finally, we allow for $I = 10$ parent wage and child ability categories, rather than the five categories in the baseline results. The wage levels assigned to each category (calculated as in the benchmark case) 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), as in the $I = 5$ baseline case.[39] We omit these targets and the corresponding calibration results for brevity,

[39] 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

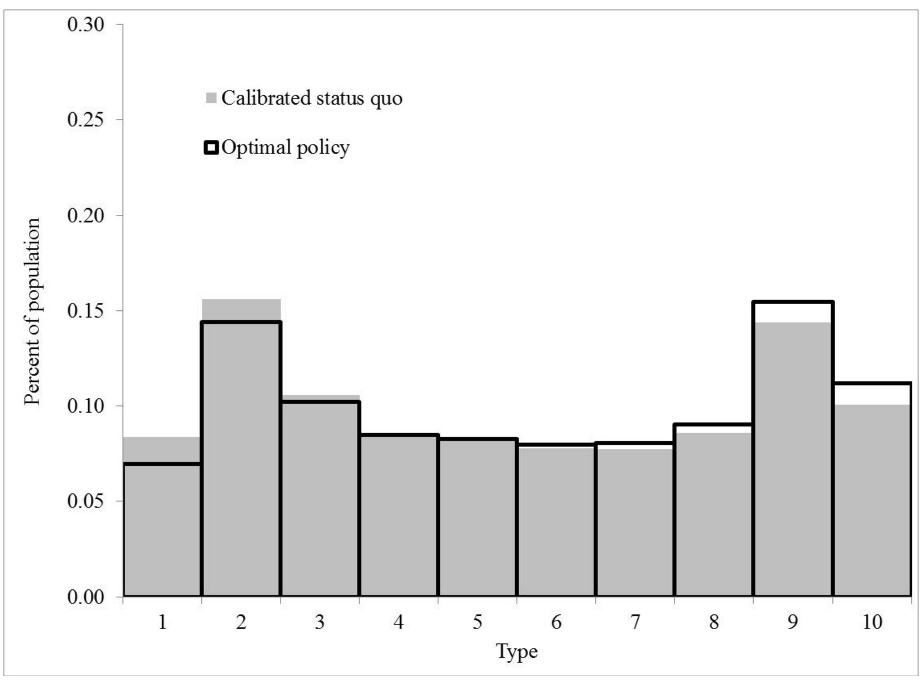
but they resemble those of the baseline case. For reference, we repeat the chosen parameter values from that baseline case.

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

	α_i	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

Both cases yield similar values for α_i and a similar pattern of values for the $\{\alpha_c^j\}_{j=1}^I$ 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 $I = 5$. Both 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).



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

2.61 to 0.96, with a value of α_i of zero. This binding restriction is not necessary in the $I = 5$ case, since all of the α_c parameters are less than unity even when the upper bound in the simulation is much larger, for example 4.0.

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 8). In Appendix Table 12 we provide these welfare gains for the overall economy and by type in both this $I = 10$ and the baseline $I = 5$ case.

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

	Overall	Type-specific gains									
	welfare gain	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

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.