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

Alexander M. Gelber
Matthew C. Weinzierl

Equalizing Outcomes vs. 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
JEL No. H21,I30

ABSTRACT

Empirical research suggests that parents'economic resources affect their children's future earnings
abilities. Optimal tax policy, defined as the policy that maximizes the aggregate present-value dynastic
utility of existing families, therefore will treat 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 1.28% of consumption in our baseline case.

Alexander M. Gelber
The Wharton School
University of Pennsylvania
1403 Steinberg-Dietrich Hall
3620 Locust Walk
Philadelphia, PA 19104
agelber@wharton.upenn.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 (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. Throughout the paper, we define "optimal policy" as the policy that maximizes social welfare, i.e., the aggregate present-value dynastic utility of existing families.

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 intratemporal and intertemporal social-welfare maximizing 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

---

[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 2011, for example).

abilities. Lower marginal tax rates encourage greater parental earnings and disposable income, and because of endogenous 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, lower marginal tax rates on low incomes will, in expectation, differentially benefit the low-skilled members of 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 parents to mimic low-ability parents by earning less and accepting the greater probability of having low-ability children. In doing so, it worsens the distortionary effects of marginal taxes on effort, so this force is likely to lead to higher marginal tax rates at low incomes relative to high incomes. The relative strength of these forces determines how social-welfare maximizing 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, social-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 (consistent with our simulations in Section 4 below). 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 social-welfare maximizing policy in this case borrows from future generations to fund greater after-tax income for parents in the current generation. Together, these intra- and inter-generational transfers 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 social-welfare maximizing 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 two equally-sized ability categories each, we estimate the effect of parents' after-tax income on the probability that low- and high-wage parents have children in each of these ability categories, and we calculate the transition matrices between ability categories across generations. Then, using 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 social-welfare maximizing policy, and we find that it redistributes substantially more toward low-ability parents and earlier generations than does the status quo policy. As a consequence, the mean ability level increases across generations under the social-welfare maximizing policy relative to the status quo, with a smaller share of the population having low ability and a larger share having high ability. We then calculate the increase in aggregate welfare due to the improved evolution of the ability distribution. We find that the gain is equivalent to a permanent increase in disposable income–i.e. an increase in disposable income for all generations–of more than one percent. Gains are substantially larger if we use three ability types rather than two, though our estimates are less precise in that case.

This paper introduces a new element to the active literature in dynamic optimal taxation. Following the original contribution of Mikhail Golosov, Narayana Kocherlakota, and Aleh Tsyvinski (2003), most work in this area has considered the impact of stochastic and exogenous skill processes on the optimal taxation of an individual over his lifetime.[3] Emmanuel

---

[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.

[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).

Farhi and Iván Werning (2010a) extend that approach to characterize optimal taxation across generations, noting in their opening sentence 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 social-welfare maximizing 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, 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. equalizing the distribution of opportunities). For example, 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

---

[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.

such as the design of life-cycle tax and training policies and social insurance programs.

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 social-welfare maximizing policy.[7] This technique may prove useful in other contexts.

The paper proceeds as follows. Section 1 describes the model. Section 2 derives analytical conditions that describe the social-welfare maximizing 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 social-welfare maximizing tax policy in our context. Section 5 concludes. An Appendix contains details of the analytical and empirical results.

1   Model

Individuals are linked in families, with one individual per generation in each family. Generations are indexed by $t \in \{1, 2, ..., T\}$. Each individual has one of a fixed set of possible income-earning abilities, or "types," denoted w and 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 after-tax 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 designed for type i'. Disposable income is in turn determined by both the tax system and individual utility maximization and is equal to earned income less taxes.

1.1 Planner's problem

The planner's problem is to choose an allocation of earned income y and disposable income c for each type i in each generation t. These allocations may differ across generations. The planner's objective is to maximize the present-value utility of the population of families starting from generation t subject to the constraints that disposable income must be funded by output (feasibility) and that individuals respond to the tax system (incentive compatibility).

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 convenient in a variety of dynamic optimal tax contexts such as Kapicka's (2006) analysis of human capital. In the context of this paper, however, there is a more fundamental reason to impose this restriction: realism. 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 strong normative aversion to such history-dependence across generations, so we will impose it on the policy here.

Formally, the planner's problem is as follows:

Problem 1 *Planner's Problem*

$$\max_{\{c_\tau^i, y_\tau^i\}_{\tau=t, i=1}^{\infty, I}} \sum_i p^i U_t^i \tag{1}$$

where U_t^i, the present-value expected utility of a family with parents of type i, is defined recursively as

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

This is maximized subject to feasibility:

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

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:

$$U_\tau^i \geq U_\tau^{i'|i} \quad \text{for all generations } \tau \text{ and types } i, i', \tag{3}$$

and $U_\tau^{i'|i}$ denotes the utility obtained by an individual of type i when claiming to be type i' :

$$U_\tau^{i'|i} = u\left(c_t^{i'}\right) - v\left(\frac{y_t^{i'}}{w_t^i}\right) + \beta \sum_{j=1}^{I} p^j \left(w_t^i, c_t^{i'}\right) U_{t+1}^j. \tag{4}$$

A technical note: expression (4) shows the usefulness of assuming a fixed distribution of types, as the next step of the recursion, U_{t+1}^j, is not directly affected by the deviation.

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 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. 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.

Second, we assume that all forms of parental resources matter equally for a child's ability. In fact, earned and unearned income may be associated with different effects: if parents work more, they will have more resources but may have fewer hours to spend investing in their children, whereas increases in unearned income may more unambiguously increase child ability. Implicitly, this assumption is similar to assuming either that the allocation of parental time has no effect on children's abilities, or that parents who adjust their work hours hold constant the time allocated to developing their children's abilities and offset that adjustment with time spent on other leisure. Again, in making this assumption, 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 the separate effect of these factors on child outcomes. Moreover, Gelber and Michell (2012) find no evidence that parental time with children (as the primary activity) is affected by single mothers' increase in work hours due to EITC expansions.

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 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 social-welfare maximizing policy

Our analysis of the planner's problem in expressions (1) through (4) generates two results. First, we characterize the distortion to an individual's choice of how much to earn, the classic subject of optimal tax analyses since Mirrlees (1971). Second, we derive a necessary condition on social-welfare maximizing 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 $v'\left(y_{t+s}^k/w_{t+s}^k\right)/[w_{t+s}^k u'\left(c_{t+s}^k\right)]$ 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 chooses to satisfy:

$$\frac{v'\left(y_t^i/w_t^i\right)}{w_t^i u'\left(c_t^i\right)} = 1 + \beta \sum_j \frac{\partial p^j\left(w_t^i, c_t^i\right)}{\partial c_t^i} \frac{U_{t+1}^j}{u'\left(c_t^i\right)}. \tag{5}$$

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 (5), 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+s}^i and y_{t+s}^i imply a distortion to a parent's private choice.[8]

Lemma 1 *Intratemporal Distortion: Let λ denote the multiplier on* (2) *and $\mu_{+\tau}^{i'|i}$ denote the*

---

[8]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 (6) is different than the condition (5) characterizing the parent's choice.

multiplier on (3). *The solution to the Planner's Problem satisfies:*

$$\frac{v'\left(y_{t+s}^{j}/w_{t+s}^{j}\right)}{w_{t+s}^{j}u'\left(c_{t+s}^{j}\right)} = \mathbb{A}_{t+s}^{j}\frac{\left(\mathbb{B}_{t+s}^{j}+\mathbb{C}_{t+s}\right)}{\left(\mathbb{B}_{t+s}^{j}+\mathbb{D}_{t+s}\right)}\left(1+\beta\sum_{k}\frac{\partial p^{k}\left(w_{t+s}^{j},c_{t+s}^{j}\right)}{\partial c_{t+s}^{j}}\frac{U_{t+s+1}^{k}}{u'\left(c_{t+s}^{j}\right)}\right) \tag{6}$$

where

$$\mathbb{A}_{t+s}^{j} = \frac{1}{1-\beta\sum_{k}\frac{\partial p^{k}\left(w_{t+s}^{j},c_{t+s}^{j}\right)}{\partial c_{t+s}^{j}}R_{t+s+1}^{k}}, \tag{7}$$

$$\mathbb{B}_{t+s}^{j} = \beta^{s}\pi_{t+s}^{j} + \sum_{\tau=0}^{s-1}\beta^{s-\tau}\sum_{i}\sum_{i'}\mu_{t+\tau}^{i'|i}\left(\pi_{t+s}^{j}\big|_{c_{t+\tau}^{i}} - \pi_{t+s}^{j}\big|_{c_{t+\tau}^{i'}}\right), \tag{8}$$

$$\mathbb{C}_{t+s} = \sum_{j'}\mu_{t+s}^{j'|j} - \sum_{j'}\mu_{t+s}^{j|j'}, \tag{9}$$

$$\mathbb{D}_{t+s} = \sum_{j'}\mu_{t+s}^{j'|j} - \frac{\frac{1}{w_{t+s}^{j'}}v'\left(\frac{y_{t+s}^{j}}{w_{t+s}^{j'}}\right)}{\frac{1}{w_{t+s}^{j}}v'\left(\frac{y_{t+s}^{j}}{w_{t+s}^{j}}\right)}\sum_{j'}\mu_{t+s}^{j|j'}, \tag{10}$$

where $\pi_{t+s}^{j}\big|_{c_{t+\tau}^{i}}$ *is the probability that a generation* $(t+s)$ *descendant of parent type* i *from generation* $t+\tau$ *is of type* j *and* $\sum_{i}\pi_{t+s}^{j}\big|_{c_{t+\tau}^{i}}$ *is denoted by the unconditional probability* π_{t+s}^{j}.

Lemma 1 shows that the product $\mathbb{A}_{t+s}^{j}\frac{\left(\mathbb{B}_{t+s}^{j}+\mathbb{C}_{t+s}\right)}{\left(\mathbb{B}_{t+s}^{j}+\mathbb{D}_{t+s}\right)}$ is the optimal wedge distorting the parent's choice of earned income. This distortion can be divided into two components: a wedge present in a conventional model and a new wedge due to endogenous ability.

In a conventional model, parental resources have no effect on children's abilities, so $\partial p^{k}\left(w_{t+s}^{j},c_{t+s}^{j}\right)/\partial c_{t+s}^{j} = 0$. If this is true, expressions (7) and (8) imply that $\mathbb{A}_{t+s}^{j} = 1$ and $\mathbb{B}_{t+s}^{j} = \beta^{s}\pi_{t+s}^{j}$. In that case, the optimal distortion is driven by binding incentive constraints in the current generation. Then, the ratio $\left(\mathbb{B}_{t+s}^{j}+\mathbb{C}_{t+s}\right)/\left(\mathbb{B}_{t+s}^{j}+\mathbb{D}_{t+s}\right)$ yields the conventional optimal marginal income tax rate at income y_{t}^{i}. Note that $\mathbb{C}_{t+s} < \mathbb{D}_{t+s}$ when higher-skilled types are tempted to mimic lower-skilled types, so this ratio is less than one in a conventional model, which implies a positive marginal tax rate.

In the model with endogenous future ability, $\mathbb{A}_{t+s}^{j}$ and $\mathbb{B}_{t+s}^{j}$ take more complicated values and affect the optimal distortion to income, as shown in expressions (7) and (8). The terms $\mathbb{C}_{t+s}$ and $\mathbb{D}_{t+s}$ are unchanged from the conventional model.

The value of $\mathbb{A}_{t+s}^{j}$ depends only on the present value of the expected net revenue gain

11

from raising the disposable income of parent j. The planner values such revenue gains, while the parent does not. That gain is the weighted sum of net revenues obtained across types over time, with the weight on type k in generation $t + s + 1$ representing the increase in probability that children of parent type j will have type k when c_{t+s}^j is increased slightly. Intuitively, larger $\mathbb{A}_{t+s}^j$ means that the planner generates greater net revenue gain from having the parent obtain a larger disposable income. Note that this effect holds whether or not incentive constraints bind, as $\mathbb{A}_{t+s}^j$ simply multiplies the right-hand side of (6). Therefore, social-welfare maximizing policy entails a smaller distortion to parent j's effort, encouraging the parent to earn more and thus retain greater disposable income.

The value of $\mathbb{B}_{t+s}^j$ measures how an increase in c_{t+s}^j affects the incentive problems of taxing earlier generations who can affect the probability that their generation-$(t + s)$ descendants have the type j. For example, suppose type j is a high skill type, so that if type i is higher than i' in generation t, $\pi_{t+s}^j|_{c_{t+\tau}^i} > \pi_{t+s}^j|_{c_{t+\tau}^{i'}}$, $\mu_{t+\tau}^{i'|i} > 0$, and $\mu_{t+\tau}^{i|i'} = 0$. Then, $\mathbb{B}_{t+s}^j$ is larger for high-skilled types in a model with endogenous ability than in a model without endogenous ability. This force decreases the optimal distortion to the high-skilled individual's private optimum. Intuitively, we should decrease the marginal distortion on type j if doing so reduces earlier generations' incentive problems. A similar logic holds if j is a low-skilled type. Then, $\pi_{t+s}^j|_{c_{t+\tau}^i} < \pi_{t+s}^j|_{c_{t+\tau}^{i'}}$ while $\mu_{t+\tau}^{i'|i} > 0$, and $\mu_{t+\tau}^{i|i'} = 0$, so that $\mathbb{B}_{t+s}^j$ is smaller 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. Thus, larger distortions are required on low-skilled types when these intergenerational connections are introduced.

In the end, the sign of the effect of this form of endogenous ability on optimal distortions 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 chlidren of low-skilled parents, then $\mathbb{A}$ is likely to be high for these parents, reducing the optimal distortion. At the same time, $\mathbb{B}$ is likely to be small because increasing this low-skilled parent's resources makes it harder to incentivize previous generations to work hard, increasing the optimal distortion. 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–the case

we will consider in the numerical simulations below due to the limitations of the available data. In the case of two ability types, only one of the incentive constraints will bind within any given generation, allowing us to write result (6) more concisely for each ability type. We provide those expressions in the Appendix.

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+s}^j can be shown to equal:

$$\frac{\pi_{t+s}^j}{u'\left(c_{t+s}^j\right)}\lambda = \frac{1}{\beta^s}\left(\beta^s\pi_{t+s}^j + \sum_{j'}\mu_{t+s}^{j'|j} - \sum_{j'}\mu_{t+s}^{j|j'}\right).$$

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

$$\sum_j \frac{\pi_{t+s}^j}{u'\left(c_{t+s}^j\right)} = \sum_k \frac{\pi_{t+s+1}^k}{u'\left(c_{t+s+1}^k\right)}. \tag{11}$$

This condition, parallel to the Symmetric Inverse Euler Equation in Weinzierl (2011), shows that the social-welfare maximizing 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 (11) may not hold. Instead, a modified version of it applies, which we state in the following proposition and derive in the Appendix.[9]

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

$$\sum_j \left[\frac{\pi_{t+s}^j}{u'\left(c_{t+s}^j\right)}\frac{1-\beta\sum_k\frac{\partial p^k\left(w_{t+s}^j,c_{t+s}^j\right)}{\partial c_{t+s}^j}R_{t+s+1}^k}{1+\beta\sum_k\frac{\partial p^k\left(w_{t+s}^j,c_{t+s}^j\right)}{\partial c_{t+s}^j}\frac{U_{t+s+1}^k}{u'\left(c_{t+s}^j\right)}}\right] = \sum_k \left[\frac{\pi_{t+s+1}^k}{u'\left(c_{t+s+1}^k\right)}\frac{1-\beta\sum_l\frac{\partial p^l\left(w_{t+s+1}^k,c_{t+s+1}^k\right)}{\partial c_{t+s+1}^k}R_{t+s+2}^l}{1+\beta\sum_l\frac{\partial p^l\left(w_{t+s+1}^k,c_{t+s+1}^k\right)}{\partial c_{t+s+1}^k}\frac{U_{t+s+2}^l}{u'\left(c_{t+s+1}^k\right)}}\right]. \tag{12}$$

---

[9]Note that results (11) and (12) 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).

In words, this proposition reflects that the true cost of raising social welfare through allocating disposable income to each generation depends on more than just the marginal utilities of disposable income of that generation's members. In particular, if extra resources granted to the current generation result in increased output and, therefore, tax revenue from future generations, or if they result in increased utilities for future generations, then the cost of raising social welfare is lower.

More formally, comparing equations (11) and (12), there are two terms that affect the latter but not the former. Both terms rely on the effects of an increase in the disposable income of an individual in one generation on the ability distribution of that individual's children. For example, the term $\beta \sum_k \left[\partial p^k \left(w_{t+s}^j, c_{t+s}^j \right) / \partial c_{t+s}^j \right] R_{t+s+1}^k$, which equals $1 - \left(\mathbb{A}_{t+s}^j \right)^{-1}$, is the present value of the net change in future taxes paid by that individual's children when c_{t+s}^j increases, while the term $\beta \sum_k \left[\partial p^k \left(w_{t+s}^j, c_{t+s}^j \right) / \partial c_{t+s}^j \right] \left[U_{t+s+1}^k / u' \left(c_{t+s}^j \right) \right]$ is the present value, in disposable income units, of the net increase in utility enjoyed by those same children. The greater the values of these terms, the less costly it is to achieve any given increase in welfare through increases in the earlier generation's disposable income.

The effects on social-welfare maximizing policy captured by these two terms are theoretically ambiguous; to build intuition for their effects, we consider a specific, empirically relevant scenario in the remainder of Section 2. 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.[10] Conventional policy designed to satisfy the expression (11) would treat generations symmetrically, and those allocations would satisfy equation (12). But that conventional policy fails to take advantage of the endogeneity of the ability distribution.

Consider, instead, a policy that transfers resources from generation $t+s+1$ to generation $t+s$ and, in particular, increases the resources available to the low-ability workers in generation $t+s$. Such a policy would violate the conventional expression (11), as it would lower the marginal utilities of disposable income for generation $t+s$ and raise them for generation $t+s+1$, increasing the left-hand side and decreasing the right-hand side of (11). Intuitively, the conventional perspective implies that the relative cost of raising welfare under such a

---

[10]Formally, suppose $\sum_{k=1}^K \pi_{t+s}^k = \sum_{k=1}^K \pi_{t+s+1}^k$ for all $K \leq I$ and $\left| \partial p^j \left(w_{t+s}^{k+1}, c_{t+s}^{k+1} \right) / \partial c_{t+s}^\kappa \right| \leq \left| \partial p^j \left(w_{t+s}^k, c_{t+s}^k \right) / \partial c_{t+s}^k \right|$ for $\kappa > k$.

policy is higher in the recipient generation $t + s$.

Such a policy is consistent with the true social-welfare maximizing policy condition (12), however, because endogenous ability reduces the true relative cost of raising welfare in the previous generation $t + s$. To see this, note that the policy will (inadvertently) increase the population proportions of higher-ability workers in generation $t + s + 1$. This shift in the distribution of π^k_{t+s+1} 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, increasing the right-hand side of (12). Intuitively, the true relative cost of raising welfare through generation $t+s+1$ is higher than the conventional perspective infers, because that generation is now higher-skilled and less able to generate further ability increases. Therefore, equation (12) may be satisfied with a policy that treats generations asymmetrically and generates greater welfare.

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 (12) 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.

3   Model calibration under existing U.S. tax policy

In this section, we calibrate the model of Section 1 to empirical results on 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 in the first generation on their children's abilities and the transition matrix between the first and second generations of parent and child ability types.

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.[11] Their study examines a specific 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.

---

[11]See Joseph Hotz and Karl Scholz (2003) and 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 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.

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 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.[12]

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

---

[12]We use year 2000 dollars throughout.

payments).[13] Child outcomes are denoted w_{ia}, which are a function of the child's and parents' characteristics and income. $\chi_a^{s_{ia}}(y_{ia})$ denotes EITC income, which is a function of pre-tax income, y_{ia}. Taxes other than the EITC are denoted $T_a^{s_{ia}}(P_{ia})$. The EITC schedules vary within a year based on income and number of children, and the EITC schedules also vary across years. The superscript s_{ia} on the EITC and tax functions denotes which schedule a child's family is on; the tax schedules may vary based upon the number of children in the household and marital status. Family disposable income is $c_{ia} = y_{ia} + \chi_a^{s_{ia}}(y_{ia}) - T_a^{s_{ia}}(y_{ia})$.

We use $\chi_a^{IV}(y_{i,a-1}) \equiv \chi_a^{s_{i,a-1}}(\hat{E}[y_{i.a}|y_{i.a-1}]) - \chi_{a-1}^{s_{i,a-1}}(y_{i,a-1})$ to instrument for the change in family disposable income from age $a-1$ to age a.[14] 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.

---

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

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

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 parent of a given ability level has a child of a given (possibly different) ability level. We adapt the Dahl and Lochner empirical specification by estimating the following model:

$$D_{ia}^{lm} = x_i'\alpha + \Delta c_{ia}\beta + W_{i,a-1}\delta + \Phi(y_{i,a-1}) + \eta_{ia} \qquad (13)$$

using χ_a^{IV} as an instrument for Δc_{ia}. We relate a binary dummy D_{ia}^{lm} equal to one when the child is in ability category l and the parent is in hourly wage category m, 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.[15] We run separate a regression for each parent-child ability pair. In our main specification, for example, there are two parent types and two child types, which implies that we estimate four separate regressions, in each of which the dependent variable is a dummy that equals one when the parent-child ability pair in question occurs. Intuitively, 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, given that the parent was in a given wage category, 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.[16]

---

[15] 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.

[16] 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.

We divide parents into two wage categories (P_1 and P_2) and divide children into two test score categories (C_1 and C_2). Each category comprises one-half of the distribution of wages or test scores, respectively. The subscript 1 indicates the lower half of the distributions, and the subscript 2 indicates the upper half. 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 social-welfare maximizing 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. To illustrate this tradeoff, in the Appendix, we show results with three 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. The results with three types show broadly similar patterns to the results with two types, though the impacts of the policy on the ability distribution and welfare are substantially larger than in the two-type baseline case.

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.[17] We control for the child's initial test score category (i.e. upper or lower half), but the results are very similar when we instead control for linear or higher-order terms in the child's initial test score. We have measured parent wage category using their wages at the beginning of the sample period, so that their wages are not affected by subsequent EITC variation. To calculate the hourly wage, we divide earnings by hours worked for NLSY survey respondents. Over 99% of respondents are mothers.[18]

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

---

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

[18]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.

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).[19] It is important to emphasize that our results are similar when we use several alternative specifications: a linear specification (which is less compatible with our formal model but which allows us to include zero values of parental income); a specification in which we add 1 to income before logging it (which clearly allows us to log income, at the cost of adding an arbitrary value to income before logging it); and a specification in which we simply log income and discard observations in which income is zero (whose sample size is substantially reduced from the sample size we use in our regressions).

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. The probability that a high-wage parent has a high-ability child, and the probability that a low-wage parent has a low-ability child, is 26.9%. The probability that a low-wage parent has a high-ability child, and the probability that a high-wage parent has a low-ability child, is 23.1%.

---

[19] 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.

3.1.2 Empirical Results

We show the main empirical results in Table 1.[20] We show the results of four separate regressions, each one corresponding to a binary dependent variable defined by one of the four possible combinations of parent and child ability categories. 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 (percentage points)

	(1) Child category 1	(2) Child category 2
(A) Parent category 1	-0.036	0.688
	(0.249)	(0.284)**
(B) Parent category 2	-0.646	-0.007
	(0.321)**	(0.206)

Notes: The dependent variable is a dummy variable that equals 1 if the parent is in a given hourly wage category and the child is also in a given ability category (which is potentially different from the parent's ability category). The child's 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). The categories are halves of the distribution (of wages or ability for parents and children, respectively), with category 1 constituting the lower half of the values and category 2 constituting the larger half of the values. Parent wage is measured at the beginning of the sample period, and child test score is 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,

---

[20]As a preliminary step, we can consider a simplified empirical exercise 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 (13) above, including a binary dummy measuring whether the lagged child test score is above or below the median. Increases in parental disosable 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 0.665 percentage points. Evaluating this at the mean of parental income ($30,598.22), 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 2.19 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.

an indicator for positive lagged income, gender, age, number of siblings, and dummies for each child's lagged test score category. Four separate regressions are run corresponding to each of four possible combinations of two child categories and two parent categories used to form the dependent variable. The number of observations in each regression is 6,902, corresponding to 2,108 mothers and 3,714 children. The table shows the coefficient, with the standard error below in parentheses, for each of the four regressions. 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.

Two of the four coefficient estimates in Table 2 are significantly different from zero at the 5% level: parental income has a negative (as expected), substantial, and statistically significant impact on the probability that a high-ability parent has a low-ability child, and parental income has a positive (as expected), substantial, and statistically significant impact on the probability that a low-ability parent has a high-ability child. The point estimates are moderate-sized and reasonable. They show that a 1% increase in parental income causes a 0.688 percentage point increase in the probability that a parent in the low ability category has a child in the high ability category, and that a 1% increase in parental income causes a 0.646 percentage point decrease in the probability that a parent in the high ability category has a child in the low ability category. Evaluating these point estimates at the mean of parental income, they imply that a $1,000 increase in parental income causes a 2.25 percentage point increase in the probability that a parent in the low ability category has a child in the high ability category, and that a $1,000 increase in parental income causes a 2.11 percentage point decrease in the probability that a parent in the high ability category has a child in the low ability category.

The estimates in Table 1 provide four of the statistics targeted by our calibration. The remaining targets are the elements of the empirical ability transition matrix across generations. Using the same dataset and definition of types as in the analysis just described, we can readily generate that 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.

	(1) Child category 1	(2) Child category 2
(A) Parent category 1	0.538	0.462
(B) Parent category 2	0.462	0.538

Table 2: Empirical ability transition matrix

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.

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:[21] the expected ability of the child of a parent of type j with disposable income $c_1^{j'}$ is

$$E\left[\ln w_2 | w_1^j, c_1^{j'}\right] = \phi + \alpha_i \left(\rho \ln w_1^j + (1-\rho)\ln \bar{w}\right) + \alpha_c^j \ln c_1^{j'}. \tag{14}$$

Expression (14) shows that the child's expected ability is a function of a constant, 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_1^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 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, 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

---

[21] In (14) 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.

function.[22]

We translate the expected ability in expression (14) into an ability distribution for the population of children of parents of type j with disposable income $c_1^{j'}$ by assuming that ability is distributed lognormally with variance σ^2 :

$$\ln w_2 \sim N\left(E\left[\ln w_2 | w_1^j, c_1^{j'}\right], \sigma^2\right). \tag{15}$$

The variance σ^2 represents an exogenous, stochastic shock to child ability common across parent types.

Expressions (14) and (15) indicate that the model calibration will search over values of the following six parameters: $\left\{\phi, \rho, \alpha_i, \{\alpha_c^j\}_{j=1}^I, \sigma\right\}$. 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 Feldman, Otto, and Christiansen 2000.[23] We investigate the robustness of our results to this choice in the Appendix. This leaves five parameters to be chosen.

Given this setup, we can generate the model's values of the targeted statistics as follows. Transition probabilities $\pi_{t+1}^k|_{c_t^j}$ are calculated by finding the positions of the (fixed) children's ability types within the parent-specific distributions (15) for each parent type. The marginal effects of parental disposable income are calculated 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.[24]

Finally, before proceeding with the calibration, we specify the tax system facing individuals, the utility function those individuals maximize, and the ability types they may take. For the status quo tax system, we assume that the Kotlikoff and Rapson (2007) calculations

---

[22] We do not estimate this production function directly using our empirical approach because our empirical approach relies on a fixed effects specification, which would difference out parent ability. Our regression specification estimates a coefficient on parental income that is comparable to the coefficient on parental income in (14).

[23] 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.

[24] Formally, we calculate the difference between the cumulative density of a given child type when the mean of the child's type is $E[\ln w] + 0.01\alpha_c^j$ and when it is $E[\ln w]$.

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.[25] 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. Their estimates are shown in Figure 1 (from Kotlikoff and Rapson 2007):
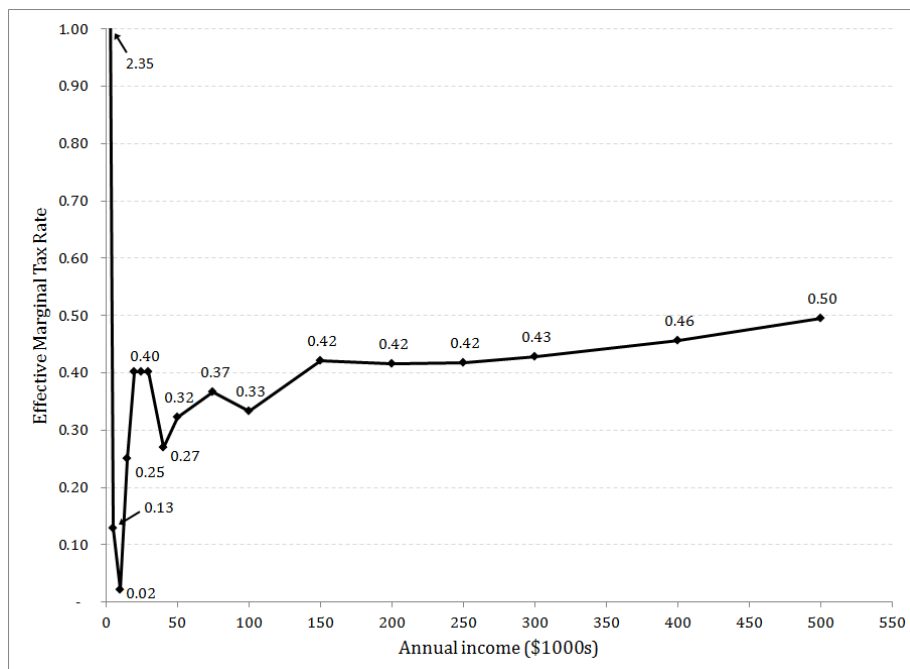
Figure 1: Effective marginal tax rate in the U.S.

Within each generation, the government also makes a grant to all individuals, and these grants may differ across generations due to variation in aggregate earnings. As in the feasibility constraint on the planner, expression (2), the government's budget is balanced in

---

[25] In the simulations, we suppress the first marginal tax rate shown in Figure 1 (by setting it equal to the second marginal tax rate), as the value of 235% can cause problems for the numerical optimization by generating negative disposable income values from positive earned incomes. Negative disposable income values generate a non-monotonic marginal utility of disposable income, with that marginal utility undefined at zero. The results of our simulations are unaffected by this change, as this marginal tax rate is inframarginal for all workers in the simulation.

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.[26]

Finally, guided by the empirical analysis discussed above, we assume ability comes in two fixed types (roughly interpretable as the hourly wage):[27] $w_t^i \in \{6.94,\ 12.95\}$ for all $t = \{1, 2, ..., T\}$. The probability distribution across those types is uniform in the first generation but is endogenously determined in the model for subsequent generations.

3.3 Calibration Results

Table 3 shows the parameter values chosen by the simulation.

Table 3: Parameters estimated by simulation					
Parameter	ϕ	α_i	α_c^1	α_c^2	σ
Value under status quo policy	0.07	0.63	0.63	0.46	0.72

---

[26]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: a simulation with $\theta = 2.5$ yields lower labor effort as a share of total time but nearly the identical α_c^j parameter values and welfare gain from the optimal policy.

[27]These values are the mean reported wages for below-median and above-median wages in the NLSY sample we use.

Recall that α_i and α_c^j are the weights on the two channels, ability and economic resources, through which parents affect their child's ability. To aid in building intuition for these values, rewrite expression (14) as follows:

$$\exp\left(E\left[\ln w_2 | w_1^j, c_1^j\right]\right) = 1.07 \left(w_1^j\right)^{\alpha_i \rho} (\bar{w})^{\alpha_i(1-\rho)} \left(c_1^j\right)^{\alpha_c^j}, \tag{16}$$

where $1.07 = \exp(\phi)$. 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 values of α_c^j in Table 3 suggest that parental resources play a greater role among low-ability parents, consistent with the empirical evidence (since $\alpha_c^1 > \alpha_c^2$).[28] 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.

Table 4 shows how the output of the model simulations under the status quo tax policy compares to the empirical estimates of the statistics shown in Tables 1 and 2.

---

[28]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.

	Data			Status quo policy		
		Child			Child	
	Parent	*1*	*2*	Parent	*1*	*2*
Transition	*1*	0.538	0.462	*1*	0.538	0.462
matrix	*2*	0.462	0.538	*2*	0.462	0.538
	Parent	*1*	*2*	Parent	*1*	*2*
Marginal effect of c	*1*	-0.0004	0.0069	*1*	-0.0035	0.0035
on child ability	*2*	-0.0065	-0.0001	*2*	-0.0026	0.0026

Table 4. Transition matrices and marginal effects

The simulation does well in matching the empirical targets. The model's transition matrix is especially well-matched to the data. The calibration is able to match the sign of the marginal effect in three of the four cases and, most important, in both of the statistically-significant (from Table 1) off-diagonal elements.[29]

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, i.e., social-welfare maximizing, policy by comparing it to the status quo policy used in that calibration. To illustrate the forces affecting the social-welfare maximizing policy, we use ten generations, and we show robustness to this choice in the Appendix.

First, Table 5 shows average and marginal tax rates for each type under the social-welfare maximizing ("optimal") and status quo policies.[30] Average tax rates are calculated as the ratio $(y - c) / y$. For marginal tax rates, we compare the marginal rates imposed by the status quo policy to the marginal tax rates that would implement the social-welfare maximizing

---

[29]Note that the assumption of two types forces the marginal effects of parental resources to be equal and of opposite sign for each parent across the two types of children.

[30]These hold for all but the initial and final generations, which differ slightly given their endpoint status.

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 (7), (8), (9), and (10). Table 5 shows the results.

	Table 5. Marginal and average tax rates			
	Marginal tax rate		Average tax rate	
Type	Optimal	Status Quo	Optimal	Status Quo
Low	11%	25%	-34%	-11%
High	-17%	40%	13%	6%

Table 5 shows how the social-welfare maximizing policy differs from the status quo. It is substantially more redistributive, generating large transfers to low-skilled parents. Intuitively, the social-welfare maximizing policy wants these parents to have higher after-tax incomes because those resources translate into higher ability for their children, while leaving high-ability parents with less after-tax income has only weak effects on their children's abilities. Making these large transfers, however, tempts the high-ability parents (of both the current and prior generations) to work less, so the social-welfare maximizing policy imposes negative marginal distortions at high incomes to encourage high-skilled workers to earn high incomes. These distortions make the allocation for lower types less attractive to those with high ability (who expect to have children with higher ability on average).

The social-welfare maximizing policy also adjusts intertemporal allocations to capitalize on the endogeneity of ability. Table 6 reports the difference between the planner's "budget balance" as a share of aggregate income in each generation under the social-welfare maximizing 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 6. Intertemporal allocations										
	Economy's budget balance									
	(as percent of output)									
Generation	1	2	3	4	5	6	7	8	9	10
Optimal - Status Quo	-0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9

Table 6 shows that the social-welfare maximizing policy borrows from future generations to fund greater investment in the skills of the current generation relative to the status quo, which we assume treats all generations similarly. 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 6 is that society can benefit by having later generations contribute, through higher taxes, to improving the ability distribution generated by earlier generations.[31]

These differences in tax policy affect the evolution of the ability distribution. We report the transition matrices for types across generations under the social-welfare maximizing ("optimal") and status quo policies. Table 7 repeats the transition matrix from Table 2 for the status quo model between the first and second generations. It also shows the comparable transition matrix for the social-welfare maximizing policy.

Table 7. Transition matrices					
Optimal policy			Status quo policy		
	Child			Child	
Parent	*1*	*2*	Parent	*1*	*2*
1	0.496	0.504	1	0.538	0.462
2	0.434	0.566	2	0.462	0.538

The social-welfare maximizing policy enables a greater share of the children of low-skilled parents to move up the skill ladder than does the status quo policy. The social-welfare

---

[31]The United States is running substantial yearly budget deficits as of 2012 and did in 2005 when the Kotlikoff and Rapson tax rates are calculated. Our "status quo policy" abstracts from this issue.

maximizing policy also allows a slightly greater share of the children of high-skilled parents to obtain high skills as well, so the overall mean of ability shifts up. Intuitively, Tables 6 and 7 show that the social-welfare maximizing policy takes resources from high-skilled parents and later generations to support low-skilled parents and earlier generations. This moves resources from those for whom the effect of resources on a child's ability is low to those for whom they have substantial value (that is, in the context of Table 4's estimates, $\alpha_c^2 < \alpha_c^1$), and it supports the maintenance of a higher ability distribution over time.

As these transition matrices imply, the evolution of the ability distribution is different under the social-welfare maximizing and the status quo policies. Figure 2 shows the ability distribution over time under the two policies. This figure shows the substantial shift toward a higher ability distribution under the social-welfare maximizing policy that results from the greater progressivity of the social-welfare maximizing policy; the social-welfare maximizing policy leads to 3.5 percent fewer individuals of the low type and 3.5 percent more individuals of the high type. Figure 2 also shows that the shift in the ability distribution is accomplished almost entirely between the first and second generations and is persistent in future periods.
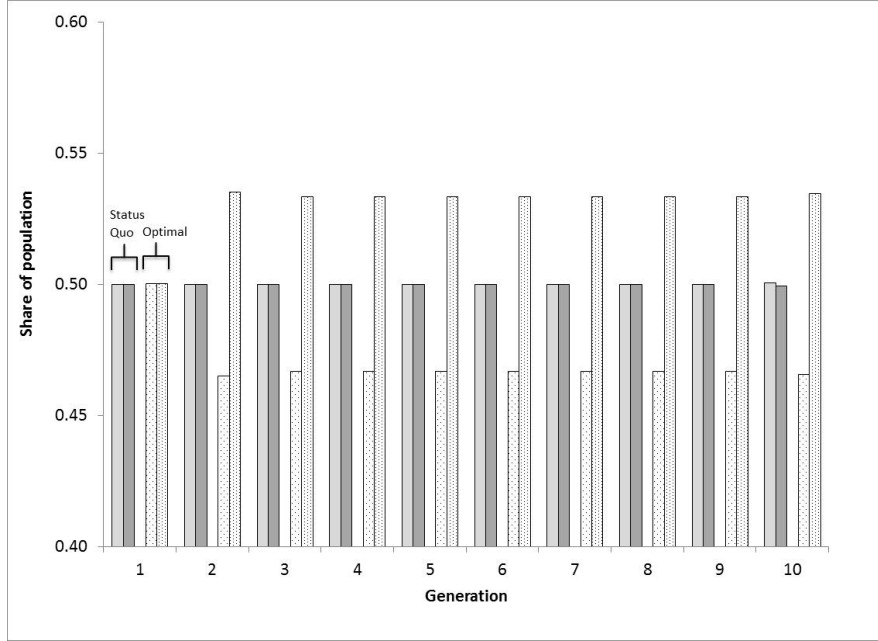
Figure 2: Ability under status quo policy (solid-filled columns) and
social-welfare maximizing policy (dot-filled columns).

Welfare is much higher under the social-welfare maximizing policy, and it is more eq-
uitably distributed. In fact, the welfare gain of moving from the status quo policy to the
social-welfare maximizing policy is large: it is equivalent to a 6.4% permanent increase in
disposable income. But this gain is explained by more than simply the effect of policy on the
ability distribution. In particular, the social-welfare maximizing policy's Utilitarian founda-
tion tends to value equality, so the greater redistribution to low-skilled parents under the
social-welfare maximizing policy than under the status quo policy generates a large 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 social-welfare maximizing model for all generations; we call this the "adjusted status
quo." Suppose further that we hold fixed the 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

33

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 social-welfare maximizing 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 8 shows the results for the baseline case of ten generations.

	Table 8. Welfare		
	Welfare (in utils)		Welfare Gain
Type	Status quo policy	Adjusted status quo	(Percent of disposable income)
Low	6.813	6.845	1.30%
High	6.926	6.957	1.27%
Overall	6.869	6.901	1.28%

As these results show, the social-welfare maximizing policy has the potential to generate a welfare gain equivalent of more than one percent of aggregate disposable income simply by shifting the ability distribution over time.[32] The gains are slightly larger among low-skilled parents, as would be expected, but high-skilled parents gain nearly as much from the more efficient tax system, 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, the number of generations, the assumed persistence of type across generations, and the number of types. Though the results change somewhat when we vary these factors, the main qualitative and quantitative lessons of the baseline analysis persist. In particular, social-welfare maximizing policy that takes advantage of endogenous ability is more redistributive than the status quo, generates an upward shift in the ability distribution, and generates a sizeable welfare gain equivalent to approximately one percent of aggregate income.

---

[32]The Appendix shows substantially larger welfare gains if we expand the simulation to three wages types. We use two wage types as the baseline simulation because we have more confidence in the empirical results for two types.

We do find a more dramatic result however, when we allow for three types of agents. While we consider our results from the two-type case to be our most solid, as the statistical significance of the empirical estimates is greatly diminished in the cases with three (or more) types, with three types of parents and children the welfare gains from reform are much larger than in the baseline: i.e., around six percent of total income. Intuitively, the high wage earners in the three-type case provide a large source of funds for redistribution to the lower- and medium-skilled workers' families. These results suggest that increasing the number of types beyond three could increase the welfare gains still further. From all of our results, then–both in the two-and three-type cases, and across the other robustness checks we perform–we conclude that the available data suggest that the welfare gains are at least approximately 1% of consumption.

5   Conclusion

In this paper, we explore one aspect of the policy problem relating to equalizing the distribution of individuals' economic outcomes or equalizing the distribution of opportunities: 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, defined as the policy that maximizes the aggregate present-value dynastic utility of existing families. An optimal tax policy 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 social-welfare maximizing tax policy when children's abilities depend on both inherited characteristics and parental (financial) resources. If the ability distribution is stable under an existing policy and the effect of parental resources is largest for the poor, these results imply that a social-welfare maximizing policy may be more redistributive toward low-ability parents and earlier generations than would be a standard optimal tax policy that ignores the effect of a parent's disposable income on

a child's ability.

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 social-welfare maximizing policy in this calibrated model and compare it to an estimated version of the existing U.S. tax code. The social-welfare maximizing 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 one 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.

This paper also makes a methodological contribution that may be useful for the study of topics in which endogenous ability plays an important role, such as life-cycle taxation, social insurance, and early childhood education. Rather than allowing the set of abilities to be endogenous, we fix a set of abilities and allow the distribution across them to be endogenous. This choice substantially reduces the scale of the optimization problem, particularly if history-independence of policy is imposed, as is natural in many contexts.

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. Some dimensions, such as time spent with children, may bear on the social-welfare maximizing tax policy directly. Others, such as spending on specific inputs to child ability such as education, will have implications for a broader class of policies. This paper demonstrates that such analyses may yield further insights into how current tax policy can use its potential to affect, not 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] 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).

[9] 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.

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

[11] Farhi, Emmanuel, and Ivan Werning. 2010b. "Insurance and Taxation over the Life Cycle." National Bureau of Economic Research Working Paper 16749.

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

[13] 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.

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

[15] 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.

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

[17] 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.

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

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

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

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

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

[23] 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.

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

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

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

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

[32] 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.

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

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

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

[36] 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+s}^k and y_{t+s}^k:

$$
\begin{bmatrix}
u'\left(c_{t+s}^k\right)\left(1 + \beta \sum_j \frac{\partial p^j\left(w_{t+s}^k, c_{t+s}^k\right)}{\partial c_{t+s}^k} \frac{U_{t+s+1}^j}{u'\left(c_{t+s}^k\right)}\right) \\
\left(\beta^s \pi_{t+s}^k + \sum_{\tau=0}^{s-1} \beta^{s-\tau} \sum_i \sum_{i'} \mu_{t+\tau}^{i'|i} \left(\pi_{t+s}^k\big|_{c_{t+\tau}^i} - \pi_{t+s}^k\big|_{c_{t+\tau}^{i'}}\right) + \sum_{k'} \mu_{t+s}^{k'|k} - \sum_{k'} \mu_{t+s}^{k|k'}\right)
\end{bmatrix}
$$

$$
= \lambda \beta^s \pi_{t+s}^k \left(1 - \beta \sum_j \frac{\partial p^j\left(w_{t+s}^k, c_{t+s}^k\right)}{\partial c_{t+s}^k} R_{t+s+1}^j\right),
$$

$$
\frac{1}{w_{t+s}^k} v'\left(\frac{y_{t+s}^k}{w_{t+s}^k}\right)
\begin{pmatrix}
\beta^s \pi_{t+s}^k + \sum_{\tau=0}^{s-1} \beta^{s-\tau} \sum_i \sum_{i'} \mu_{t+\tau}^{i'|i} \left(\pi_{t+s}^k\big|_{c_{t+\tau}^i} - \pi_{t+s}^k\big|_{c_{t+\tau}^{i'}}\right) \\
+ \sum_{k'} \mu_{t+s}^{k'|k} - \frac{\frac{1}{w_{t+s}^{k'}} v'\left(\frac{y_{t+s}^k}{w_{t+s}^{k'}}\right)}{\frac{1}{w_{t+s}^k} v'\left(\frac{y_{t+s}^k}{w_{t+s}^k}\right)} \sum_{k'} \mu_{t+s}^{k|k'}
\end{pmatrix}
= \lambda \beta^s \pi_{t+s}^k
$$

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 (6), for each ability type in generation $t+s$, is as follows:

$$
\frac{v'\left(y_{t+s}^i/w_{t+s}^i\right)}{w_{t+s}^i u'\left(c_{t+s}^i\right)} =
\begin{bmatrix}
\left(\frac{1 + \beta \sum_k \frac{\partial p^k\left(w_{t+s}^j, c_{t+s}^j\right)}{\partial c_{t+s}^j} \frac{U_{t+s+1}^k}{u'\left(c_{t+s}^j\right)}}{1 - \beta \sum_k \frac{\partial p^k\left(w_{t+s}^i, c_{t+s}^i\right)}{\partial c_{t+s}^i} R_{t+s+1}^k}\right) \\
\left(\frac{\beta^s \pi_{t+s}^i + \sum_{\tau=0}^{s-1} \beta^{s-\tau} \mu_{t+\tau}^{i|j} \left(\pi_{t+s}^i\big|_{c_{t+\tau}^j} - \pi_{t+s}^i\big|_{c_{t+\tau}^i}\right) - \mu_{t+s}^{i|j}}{\beta^s \pi_{t+s}^i + \sum_{\tau=0}^{s-1} \beta^{s-\tau} \mu_{t+\tau}^{i|j} \left(\pi_{t+s}^i\big|_{c_{t+\tau}^j} - \pi_{t+s}^i\big|_{c_{t+\tau}^i}\right) - \frac{\frac{1}{w_{t+s}^j} v'\left(\frac{y_{t+s}^i}{w_{t+s}^j}\right)}{\frac{1}{w_{t+s}^i} v'\left(\frac{y_{t+s}^i}{w_{t+s}^i}\right)} \mu_{t+s}^{i|j}}\right)
\end{bmatrix}
\tag{17}
$$

$$
\frac{v'\left(y_{t+s}^j/w_{t+s}^j\right)}{w_{t+s}^j u'\left(c_{t+s}^j\right)} = \frac{1}{1 - \beta \sum_k \frac{\partial p^k\left(w_{t+s}^j, c_{t+s}^j\right)}{\partial c_{t+s}^j} R_{t+s+1}^k} \left(1 + \beta \sum_k \frac{\partial p^k\left(w_{t+s}^j, c_{t+s}^j\right)}{\partial c_{t+s}^j} \frac{U_{t+s+1}^k}{u'\left(c_{t+s}^j\right)}\right) \tag{18}
$$

6.2 Proof of Proposition 1

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

$$
\left(\beta^s \pi_{t+s}^k + \sum_{\tau=0}^{s-1} \beta^{s-\tau} \sum_i \sum_{i'} \mu_{t+\tau}^{i'|i} \left(\pi_{t+s}^k|_{c_{t+\tau}^i} - \pi_{t+s}^k|_{c_{t+\tau}^{i'}} \right) + \sum_{k'} \mu_{t+s}^{k'|k} - \sum_{k'} \mu_{t+s}^{k|k'} \right)
$$

$$
= \quad \lambda \beta^s \frac{\pi_{t+s}^k}{u'\left(c_{t+s}^k\right)} \frac{1 - \beta \sum_j \frac{\partial p^j \left(w_{t+s}^k, c_{t+s}^k\right)}{\partial c_{t+s}^k} R_{t+s+1}^j}{1 + \beta \sum_j \frac{\partial p^j \left(w_{t+s}^k, c_{t+s}^k\right)}{\partial c_{t+s}^k} \frac{U_{t+s+1}^j}{u'\left(c_{t+s}^k\right)}}.
$$

Then, sum each side over k:

$$
\sum_k \left(\beta^s \pi_{t+s}^k + \sum_{\tau=0}^{s-1} \beta^{s-\tau} \sum_i \sum_{i'} \mu_{t+\tau}^{i'|i} \left(\pi_{t+s}^k|_{c_{t+\tau}^i} - \pi_{t+s}^k|_{c_{t+\tau}^{i'}} \right) + \sum_{k'} \mu_{t+s}^{k'|k} - \sum_{k'} \mu_{t+s}^{k|k'} \right)
$$

$$
= \quad \sum_k \lambda \beta^s \frac{\pi_{t+s}^k}{u'\left(c_{t+s}^k\right)} \frac{1 - \beta \sum_j \frac{\partial p^j \left(w_{t+s}^k, c_{t+s}^k\right)}{\partial c_{t+s}^k} R_{t+s+1}^j}{1 + \beta \sum_j \frac{\partial p^j \left(w_{t+s}^k, c_{t+s}^k\right)}{\partial c_{t+s}^k} \frac{U_{t+s+1}^j}{u'\left(c_{t+s}^k\right)}}.
$$

The left-hand side simplifies dramatically, so that we obtain:

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

This holds for all s, yielding the Proposition.

6.3 Appendix Table 1

Appendix Table 1: Summary Statistics		
(1) Variable	(2) Mean	(3) Standard deviation
Household Income	30,598	22,859.84
Hours worked of respondent	1,692.68	785.42
Hourly wage	9.57	18.24
Child age	11.31	2.02
Child male (dummy)	0.497	0.500
2x2 Parent category 1-child category 1 (dummy)	0.269	0.443
2x2 Parent category 1-child category 2 (dummy)	0.231	0.421
2x2 Parent category 2-child category 1 (dummy)	0.231	0.421
2x2 Parent category 2-child category 2 (dummy)	0.269	0.443
3x3 Parent category 1-child category 1 (dummy)	0.125	0.331
3x3 Parent category 2-child category 1 (dummy)	0.119	0.323
3x3 Parent category 3-child category 1 (dummy)	0.089	0.285
3x3 Parent category 1-child category 2 (dummy)	0.106	0.308
3x3 Parent category 2-child category 2 (dummy)	0.115	0.319
3x3 Parent category 3-child category 2 (dummy)	0.112	0.316
3x3 Parent category 1-child category 3 (dummy)	0.102	0.303
3x3 Parent category 2-child category 3 (dummy)	0.100	0.300
3x3 Parent category 3-child category 3 (dummy)	0.132	0.338

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 next four rows show the means of the binary dependent variables used in the regressions in Table 1. These binary dependent variables take the value of 1 when the child is in the ability category indicated and the parent is in the ability category indicted, and 0 otherwise (as described in the text). The final nine rows show the means of the binary dependent variables used in the regressions in Appendix Table 7 below. The number of observations is 6,902, corresponding to 2,108 mothers and 3,714 children. Income is measured in year 2000 dollars.

6.4 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	
(1) Child in high ability category	
Income	.665 (.261)***
R-squared	.204

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 observations is 6,902, corresponding to 2,108 mothers and 3,714 children. 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.

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

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

6.5.1 Value of ρ

First, ρ indicates the role of parental ability, relative to a mean ability level, in determining a child's ability. As shown in expression (14), 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 six other values of ρ, namely 0.25, 0.33, 0.40, 0.60, 0.67, 0.75. Appendix Table 3 shows the parameter values chosen by these simulation, as well as the baseline case of $\rho = 0.50$ in bold type for reference.

Appendix Table 3: Parameters with alternative ρ values					
	ϕ	α_i	α_c^1	α_c^2	σ
$\rho = 0.25$	-0.15	0.68	0.68	0.57	0.79
$\rho = 0.33$	-0.18	0.63	0.88	0.66	1.13
$\rho = 0.40$	0.07	0.62	0.64	0.51	0.88
$\boldsymbol{\rho = 0.50}$	**0.07**	**0.63**	**0.63**	**0.46**	**0.71**
$\rho = 0.60$	0.19	0.56	0.65	0.52	1.16
$\rho = 0.67$	-0.17	0.70	0.75	0.51	0.82
$\rho = 0.75$	-0.13	0.64	0.78	0.59	1.44

Appendix Table 4 shows the transition matrices across generations for these alternative values of ρ.

Appendix Table 4. Transition matrices under social-welfare maximizing policy with alternative ρ values

$\rho = 0.25$	Child		$\rho = 0.33$	Child		$\rho = 0.40$	Child	
Parent	1	2	Parent	1	2	Parent	1	2
1	0.497	0.504	1	0.511	0.489	1	0.505	0.495
2	0.431	0.570	2	0.442	0.558	2	0.438	0.562

$\boldsymbol{\rho = 0.50}$	**Child**	
Parent	**1**	**2**
1	**0.496**	**0.504**
2	**0.434**	**0.566**

$\rho = 0.60$	Child		$\rho = 0.67$	Child		$\rho = 0.75$	Child	
Parent	1	2	Parent	1	2	Parent	1	2
1	0.513	0.487	1	0.495	0.505	1	0.514	0.486
2	0.445	0.555	2	0.436	0.565	2	0.447	0.553

Appendix Table 5 shows the overall welfare gains from the improved ability distributions generated by the social-welfare maximizing policies, for each value of ρ.

Appendix Table 5. Welfare with alternative ρ values

	Welfare (in utils)		Welfare Gain
	Status quo policy	Adjusted status quo	(Percent of disposable income)
$\rho = 0.25$	6.871	6.904	1.34%
$\rho = 0.33$	6.864	6.887	0.89%
$\rho = 0.40$	6.867	6.893	1.01%
$\boldsymbol{\rho = 0.50}$	**6.869**	**6.901**	**1.28%**
$\rho = 0.60$	6.863	6.883	0.79%
$\rho = 0.67$	6.869	6.900	1.26%
$\rho = 0.75$	6.862	6.880	0.73%

These tables yield two lessons. First, the variation in results, and particularly in the estimated welfare gains, are relatively small in magnitude, so that the main lessons from the baseline analysis are robust to variations in ρ. Second, the welfare gain estimates vary non-monotonically with ρ. To build intuition for this result, start by noting that the cases with larger estimated welfare gains have lower chosen values for σ, the standard deviation of the shock to ability. In fact, the correlation between these series is -0.925. When abilities are subject to greater randomness, the power of optimal policy to reliably target transfers to children of low ability and the ability of it to provide incentives to parents through differential treatment of high- and low-ability children are both diminished. This pattern begs the question of why σ varies non-monotonically with ρ. Recall that the calibration exercise matches an empirical transition matrix. In the extreme case where $\rho = 0$, the model's only option for matching the data is to have the effects of parental resources on child ability be relatively uniform (i.e., the relative size of α_c^1 compared to α_c^2 is small), so that

the children of high-income parents have an advantage, as in the data. As ρ increases, the natural heritability of ability will overshoot the empirical transition matrix. Then, the calibration will have two options for reducing heritability to match the data: either a larger shock value σ or effects of parental resources that are particularly concentrated at low incomes. While the former option will imply only limited potential for policy gains, in the latter case transfers to low-income parents are especially potent and the gains from social-welfare maximizing policy are larger. The model's vacillation between these two options for matching the empirical transition matrices is apparent in the tables above.

6.5.2 Value of T

We also describe the robustness of our results to variation in T, the number of generations simulated. To do so, we take the baseline simulation's chosen parameter values and simulate different horizons. We find that the number of generations has little effect on the results, aside from the fact that a longer horizon typically allows for greater gains because the years after the ability distribution shifts become relatively more numerous. To show this, we display the results of the welfare gain calculation (as in Appendix Table 6, for example), for a variety of horizons.

Appendix Table 6. Welfare with alternative horizons	
	Welfare Gain
	(Percent of disposable income)
$T = 6$	1.21%
$T = 8$	1.25%
$T = 10$	1.28%
$T = 12$	1.30%
$T = 14$	1.31%

6.5.3 Value of β

Next, we vary the value of β (and thus R, which is equal to $1/\beta$). 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, the estimated welfare gain goes from 1.28 percent of total income to 1.01 percent of total income. The parameter values chosen are extremely similar to those in the baseline simulation. The diminution in welfare gain appears to be parallel to the finding that the welfare gains increase with the time horizon of the simulation. For example, the same sized welfare gain occurs when we use the baseline parameter values with only four generations (namely, 1.01 percent of total income).

6.5.4   Value of I

Finally, we allow for $I = 3$ parent wage and child ability categories, rather than the two categories in the baseline results. In these regressions, the dependent variable is a dummy variable that equals one when a given parent-child combination occurs, and we control for initial child test score tercile in the regressions. Appendix Table 7 shows that the probability that children are in the lowest tercile of ability tends to be lowered by higher parental disposable income, whereas the probability that children are in the highest tercile of ability tends to be raised by higher parental disposable income. The point estimates are usually moderate-sized and reasonable; for example, they show that a 1% increase in parental income causes a 0.216 percentage point decrease in the probability that a parent in the highest ability category has a child in the lowest ability category. Nonetheless, these estimates are not apt to detect an effect of parental disposable income on the outcomes, in the sense that the dependent variable only takes on a value of 1 rarely (on average 11.1% of the time). It therefore should be unsurprising that the results are not more significant. However, the results in Dahl and Lochner and our results from a design with fewer categories (in Table 1 and Appendix Table 2) show that these results rely on useful variation in the data. Moreover, the point estimates in Appendix Table 7 show the expected sign in all six cases where the child's type differs from the parent's (i.e., the off-diagonal elements).

Appendix Table 7: Empirical marginal effects of parental resources on child ability distribution[33]			
	(1) Child category 1	(2) Child category 2	(3) Child category 3
(A) Parent category 1	-.025	.302	.159
	(.174)	(.188)	(.150)
(B) Parent category 2	-.032	-.208	.193
	(.192)	(.180)	(.164)
(C) Parent category 3	-.216	-.145	-.028
	(.196)	(.167)	(.147)

[33]Table notes: The dependent variable is a dummy variable that equals 1 if the parent is in a given hourly wage category and the child is also in a given ability category (which is potentially different from the parent's ability category). The child's 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). The categories are terciles of the distribution (of wages or ability for parents and children, respectively), with category 1 constituting the lowest third of the values, category 2 constituting the middle third of the values, and category 3 constituting the highest third of the values. Parent wage is measured at the beginning of the sample period, and child test score is 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

Appendix Table 8 shows the empirical transition matrices for the three-type case, as in Table 1 for the baseline case.

Appendix Table 8: Empirical ability transition matrix			
	(1) Child category 1	(2) Child category 2	(3) Child category 3
(A) Parent category 1	0.377	0.331	0.291
(B) Parent category 2	0.350	0.332	0.318
(C) Parent category 3	0.272	0.337	0.391

Appendix Table 9 shows how the output of the model simulations under the status quo tax policy compares to the empirical estimates of the statistics shown in Appendix Tables 7 and 8.

Appendix Table 9. Transition matrices and marginal effects								
		Data				Status quo policy		
		Child				Child		
		1	2	3		1	2	3
Transition	1	0.377	0.331	0.291	1	0.376	0.333	0.291
matrix	Parent 2	0.350	0.332	0.318	2	0.348	0.335	0.317
	3	0.272	0.337	0.391	3	0.276	0.331	0.393
		1	2	3		1	2	3
Marginal effect	1	-0.0003	0.0030	0.0016	1	-0.0030	0.0003	0.0027
of parent's c	Parent 2	-0.0003	-0.0021	0.0019	2	-0.0022	0.0001	0.0021
on child ability	3	-0.0022	0.0015	-0.0003	3	-0.0012	-0.0002	-0.0003

The simulation does well matching the empirical targets. The model's transition matrix is especially well-matched to the data. The marginal effects are imprecisely estimated in the data (see Appendix Table 7), but the calibration is able to match the sign of the effect in seven of the nine cases and, in particular, in five of the six off-diagonal elements.

Appendix Table 10 shows the parameter values chosen by the simulation to yield the results of Appendix Table 9.

Appendix Table 10: Parameters estimated by simulation						
Parameter	ϕ	α_i	α_c^1	α_c^2	α_c^3	σ
Value under status quo policy	-0.18	0.63	0.88	0.66	0.42	1.13

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 child's lagged test score category. Nine separate regressions are run corresponding to each of nine possible combinations of three child categories and three parent categories used to form the dependent variable. The number of observations in each regression is 6,902, corresponding to 2,108 mothers and 3,714 children. The table shows the coefficient, with the standard error below in parentheses, for each of the nine regressions. 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.

Next we simulate social-welfare maximizing policy. Appendix Table 11 shows the results for the first generation.

Appendix Table 11. Marginal and average tax rates				
	Marginal tax rate		Average tax rate	
Type	Optimal	Status Quo	Optimal	Status Quo
Low	11%	2%	-412%	-62%
Middle	39%	40%	-132%	-18%
High	-19%	37%	33%	17%

Appendix Table 12 reports the difference between the planner's "budget balance" as a share of aggregate income in each generation under the social-welfare maximizing and status quo policies.

Appendix Table 12. Intertemporal allocations										
	Economy's budget balance									
	(as percent of output)									
Generation	1	2	3	4	5	6	7	8	9	10
Optimal - Status Quo	-1.6	-0.2	-0.1	-0.1	-0.1	-0.1	-0.1	-0.0	-0.4	2.4

These differences in tax policy affect the evolution of the ability distribution. Appendix Table 13 repeats the transition matrix from Appendix Table 9 for the status quo model between the first and second generations. It also shows the comparable transition matrix for the social-welfare maximizing policy.

Appendix Table 13. Transition matrices								
		Optimal policy				Status quo policy		
		Child				Child		
		1	2	3		1	2	3
	1	0.237	0.323	0.440	1	0.376	0.333	0.291
Parent	2	0.288	0.333	0.380	2	0.348	0.335	0.317
	3	0.279	0.332	0.390	3	0.276	0.331	0.393

The evolution of the ability distribution is similar to the results under the two-type model, with approximately the same share of the population having the middle skill level under the social-welfare maximizing and status quo policies, and the social-welfare maximizing policy yielding a six percent greater share with the highest skill and a six percent smaller share with the lowest skill level.

Welfare is much higher under the social-welfare maximizing policy, and it is more equitably distribute. This is shown in Appendix Table 14, which performs the same thought experiment as in the main text, isolating the gain from only the improved ability distribution due to the social-welfare maximizing policy.

Appendix Table 14. Welfare			
	Welfare (in utils)		Welfare Gain
Type	Status quo policy	Adjusted status quo	(Percent of disposable income)
Low	8.126	8.211	6.64%
Middle	8.200	8.282	5.88%
High	8.303	8.380	5.35%
Overall	8.202	8.283	5.98%

These results indicate that the power of social-welfare maximizing policy to adjust the ability distribution is substantially greater–equal to six percent of total output–when the range of wage types considered is expanded.