Designing Cash Transfers in the Presence of Children's Human Capital Formation

Joseph Mullins *†

March 23, 2019

Abstract

This paper studies the design of antipoverty programs in the United States and their impact on child development. I introduce a model in which mothers work, participate in cash assistance programs, and invest time and money in their children. A technology of skill formation maps these investments to child development outcomes. The model is estimated by combining exogenous variation in policies, model-based orthogonality restrictions, and a classification routine that sorts mothers into types based on a set of revealed preference inequalities. This approach relaxes typical assumptions regarding the distribution of latent variables in dynamic models. My estimates indicate that fluctuations in income and labor supply have non-negligible impacts on child human capital, and that mothers exhibit significant heterogeneity in their preferences for labor supply, program participation, and investment. A counterfactual study, that fixes the policy environment just prior to welfare reform in 1996, implies that major changes to cash assistance in the United States resulted in a significant redistribution of maternal welfare across latent types, for approximately the same fiscal cost to the government. The optimal policy is a Negative Income Tax, which mothers value on average as equivalent to an additional $82 per week in consumption over the 17 year investment period, and raises skills to the order of 1.3 percentage points in the probability of high school graduation. These gains are unevenly distributed across types.

*Dept. of Economics, University of Minnesota. Email: mullinsj@umn.edu

†Many thanks to Chris Flinn, Gianluca Violante, Meta Brown, Lance Lochner, Tobias Salz, David Cesarini, Timothy Christensen, Simon Mongey, Roy Allen, and seminar participants at the following: Queen's, NYU, Wash U, SoLE 2017 Annual Meetings, York University, Minnesota, Oregon, and the Workshop on Family and Aging in Solstrand (2019). All remaining errors are my own. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B140035 to New York University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

1 Introduction

The main ambition of this paper is to establish quantitative guidelines for choosing the parameters of government antipoverty programs, in settings where the human capital of future generations may be at stake. A growing body of research suggests that relieving economic disadvantage in early childhood leads to developmental gains (Duncan et al., 2011; Dahl and Lochner, 2012; Hoynes et al., 2016; Aizer et al., 2016), while a contrapositive statement, that economic disadvantage presents a developmental adversity for children, also appears to hold (Currie and Almond, 2011).

From a normative perspective, while the problem of designing taxes and transfers to maximize welfare has received considerable theoretical and empirical attention (Mirrlees, 1971; Saez, 2001, 2002; Heathcote et al., 2017), the apparent link between economic resources and child development demands that we return to this classic problem with one eye on a new, quantitatively relevant externality (Cunha and Heckman, 2007; Heckman et al., 2010). In addition, policymakers may hold inherent interest in the distributional consequences of differently designed transfer programs on child skill outcomes. It is to these questions that I devote attention in this paper.

To fix ideas, consider two different transfer programs of equivalent cost. The first program incentivises maternal labor force participation by subsidizing income, and therefore reduces maternal hours at home. The alternative is a negative income tax, consisting of a payment standard that is reduced at a certain rate with family income, imposing an implicit tax on earnings, and therefore discouraging labor supply. Relative to the first case, mothers spend more time at home and have lower net incomes. Ranking these two programs requires us to make a trade-off between the value of income and the value of hours spent at home. Which of these programs leads to better child outcomes? And which is valued more by mothers? Complicating this policy problem is the observation that, even if we could quantitatively decide upon the productive value of time and money, doing so is insufficient for making policy-relevant calculations. In order for child outcomes and maternal welfare to be fully articulated with respect to policy options, we must also learn exactly how mothers respond to the different incentive structures embedded in cash transfer programs.

Thus, our conclusions about policy will be decided by the answer to two fundamental questions. First, how important is family income and maternal hours at home for child development? Second, precisely how do mothers make their labor supply and program participation decisions in response to program incentives? These questions, and their respective answers, are given precise meaning in this paper through the development of a behavioral model, in which a mother makes two sets of decisions. First, she evaluates the relative value of her leisure hours against the returns to labor, deciding how much to work, and whether to participate in government-provided transfer programs. Second, she decides how much of her remaining time and money to invest in her child (or number of children), solving a dynamic investment problem. The latter investment channel provides some formal logic for why family income and hours at home matter for children.

The importance of time and money can then be mapped to structural parameters of the model, in particular to the parameters of the *technology of skill formation*, while the response of mothers to different program incentives is determined by the distribution of preferences in the population. Therefore, to statistically mobilize the insights of the model, I must be able to credibly identify these objects from data. I develop an identification argument that combines cross-sectional variation in policies with a control function condition to learn about the importance of money and time investments for child skill development, and uses repeated observation of mothers' decisions to learn about latent preferences, without imposing strong parametric assumptions on this distribution.

The approach to identification in this paper is constructive in that it proposes a natural estimation procedure, which is consistent when the distribution of latent heterogeneity in the model is described by a discrete number of types. In practice, estimates are produced by imposing that mothers obey a number of revealed preference inequalities for their correct type, and that a set of orthogonality restrictions implied by the model hold.[1] The estimation procedure iteratively sorts mothers in the sample into type-specific bins, and estimates the production parameters, in a routine that closely follows a K-means clustering method. Asymptotics for this style of procedure are sound (Bonhomme and Manresa, 2015). This general approach of estimating structural parameters during a time of exogenous policy changes follows in the tradition of related work (Blundell et al., 1998, 2016; Bernal, 2008; Voena, 2015), however the analytical tractability of my model permits a particularly tight link between key parameters and data.

I estimate the model using data on single mothers and their children from the *Panel Study of Income Dynamics* (PSID) and its *Child Development Supplement* (CDS). The PSID provides panels of hours, earnings, and program participation for mothers of children in the CDS, while the CDS itself provides measures on skill outcomes and maternal time investment. These data are sufficient to implement the estimation procedure described above. I also exploit the longitudinal dimension of the data to address the joint issues of *scaling* and *aggregation* of skills. That is, how do we assess the magnitude of skill impacts, aside from measuring their effect in standard deviations of test scores? Further, when assessing changes in multiple skill dimensions, how should these changes be weighed against each other? I tackle this problem by connecting measured skills to whether each child in the sample graduates from High School. This ties skill outcomes to an important socioeconomic outcome, giving an interpretable scale to survey skill measures, and providing a reasonable (if imperfect) aggregation mechanism.

Three major programs exist in the United States with the express purpose of alleviating poverty: (1) the Supplemental Nutritional Assistance Program (SNAP); (2) Temporary Assistance for Needy Families (TANF); and (3) the Earned Income Tax Credit (EITC). These policies show great variation in their generosity, and the extent to which they reward labor force participation. Accordingly, I build into the model the complex program structure that reflects the realities faced by mothers in

[1] These include the typical orthogonality conditions of linear instrumental variables.

the United States, during the time period that applies to my sample. This time period saw quite dramatic changes to the landscape of government cash assistance, with a concerted policy shift towards programs that encourage self-sufficiency and discourage dependance on government support. First, the 1990s saw a significant expansion in the EITC, particularly in the period between 1993 and 1995. Second, in 1996, the *Personal Responsibility and Work Opportunity Reconciliation Act* (PRWORA) was signed into law. This act affected a transition in 1997 from the former welfare program *Aid to Families with Dependant Children* (AFDC) to a new program: *Temporary Assistance for Needy Families* (TANF). Three aspects of the reform combined to encourage labor force participation among welfare recipients. First, entitlements to TANF are subject to a 60 month time limit for participants[2]. Second, recipients are required to spend 20 hours a week in work or work-related activities. Failure to meet this requirement is punishable by sanctions. Third, many states restructured their benefit calculation formulae to reduce the effective marginal tax rate on earnings. This series of institutional changes was successful in shifting a sizeable number of participants off welfare rolls and into the labor force (Hoynes, 1996; Grogger, 2002, 2003; Meyer, 2002; Chetty et al., 2013).

I use the estimated model to conduct three main counterfactual exercises. In the first, I hold constant the welfare environment in 1996 (thereby undoing welfare reform and later EITC expansions), and examine the implications for welfare and child development. While there are no significant changes in the aggregate, the counterfactual shows that the 1996 reforms significantly redistributed resources (and consequently welfare) to mothers with higher labor force attachment and lower welfare dependance. Next, I move to two normative policy exercises, in which average maternal welfare and child skills are alternatively maximized subject to a revenue constraint. In the first case, the solution to this problem is a negative income tax, which delivers welfare gains equivalent to $82 in additional consumption every week, and gains in child skills associated with an additional 1.3 percentage points in the probability of high school graduation. The skill-maximizing policy, in contrast, consists of an initial earnings subsidy, with a maximum credit of just over $4000 per year. This yields skill gains of 1.9 percentage points in high school graduation, and welfare gains equivalent to $20 a week in additional consumption. In both settings, the aggregate numbers conceal important differences in skill and welfare impacts across types. For example, mothers with low elasticities of labor supply will benefit less from earnings subsidies, while mothers with lower costs of program participation will lose welfare when programs with universal eligibility are introduced. The converse statements also hold, and the same implications hold for the children of these mothers.

This pattern of results points to three key insights for policy. First, that different design choices are likely to have statistically relevant consequences, both aggregate and distributional, for

[2] These time limits apply only to the block grants awarded to the States by Federal Government. States can, if they choose, independantly fund support beyond the time limit.

the human capital formation of children. Second, the weight placed on child outcomes relative to maternal welfare has crucial implications for the shape of optimal tax schedules: emphasising the chosen skill aggregate of this paper leads to negative marginal tax rates (earnings subsidies) which are not otherwise implied. Finally, essential and latent heterogeneity in mothers' capacity to earn, preferences for work, and costs of program participation suggest an additional and quantitatively relevant trade-off for policymakers, and imply that neither of the optimal policies considered here are Pareto improving.

There are, as always, some necessary and important caveats to issue when interpreting results. First and foremost, the model provides a stylized environment, within which the outside option to maternal time is fixed. As such, we cannot discuss the potential impact of policies that, say, jointly subsidize employment and childcare. Second, I consider changes to policies that also apply to populations outside of single mothers. The model says little about how to extrapolate impacts to these populations, or the fiscal consequences of doing so. Finally, for a complete analysis, one would have to consider the issues of selection into and out of the current sample of mothers through endogenous responses in marital and fertility decisions. Nevertheless, I contend that the quantitative lessons of this study are useful, and the methodology outlined provides a way forward for future extensions in the above-mentioned directions.

2 A Model of Child Development

In this section I develop a model that positions mothers as the sole decision-makers in a family unit. The model can be thought of as an extension of the framework introduced in Del Boca, Flinn, and Wiswall (2014), which itself follows in the tradition of Becker and Lewis (1973) and Becker and Tomes (1976). I extend this model in several meaningful ways. First, I adopt a multidimensional skill technology, which broadens the policy implications and importance of the study, since behavioral traits are influential in shaping many life-cycle outcomes (Heckman et al., 2006). Furthermore, theory and evidence suggest that such skills are crucial in the development of future skills (Cunha and Heckman, 2007; Cunha et al., 2010). Second, I embed the model in a detailed policy environment, which potentially adds complexity to the problem, as well as additional dynamic concerns, given the possibility of time limits on welfare participation.

We will see that, despite the added complexity, the model still permits a parsimonious solution. Indeed, it will become apparent that a broad class of dynamic elements can be added without interacting with the child investment problem. In more technical language, we will see that child skills enter additively into the mother's value function, and hence do not factor in complexity concerns when adding other dimensions to the model.

Naturally, some features of this model are quite stylized, yet we will see that the practical advantages are significant, allowing me to include otherwise intractable components. Furthermore, despite a relatively rich dynamic structure, I am able to derive a linear equation for child outcomes, permitting the estimation of production parameters with instrumental variables.

As a simplification, this model does not include fertility and divorce decisions. In this setting, family structure is irrelevant to the production problem, while fertility is taken as exogenous. This is a direction for future extension. Finally, to enable cleaner exposition, I develop here the case for a mother with one child. In Appendix C, I show how this can be easily extended to multiple children when estimating the model.

2.1 Environment, Production and Preferences

Time in the model is discrete and indexed by t at an annual frequency. Let a_t indicate the age of the child in year t, and consider the case in which $t = 0$ is the year of birth. Under this assumption, $a_t = t$ and I drop, for simplicity, the t subscript from age.

Let θ_{it} be a N_θ-dimensional vector[3] that describes the behavioral traits and capabilities of a child at time t. A periods after birth, the child reaches maturity. Maturity is defined as the stage beyond which an individual's skills (θ) no longer develop.[4] In every period $t = 0, ..., A-1$, mothers choose their consumption (c), leisure (l), labor supply (h), and program participation (p). Importantly, they can also choose to make investments in their child in the form of money (x) and time (τ). Period A signifies the end of the mother's decision problem, and at this point they receive a terminal payoff from final child abilities, which can be written as $V_{A,i}(\theta)$. In each period $t < A$, utility is derived from consumption, leisure, hours worked, child abilities, and program participation, which has cardinal value determined by the function, u_i. Utility at time t is written as

$$U_{it} = u_{it}(c_{it}, l_{it}, h_{it}, \theta_{it}, p_{it})$$

Finally, future payoffs are discounted exponentially by a factor, β. I assume that mothers make their decisions to maximize this discounted stream of payoffs.

The economic substance of the problem is introduced by resource constraints. First, mothers cannot consume and invest beyond their budget constraint, which is in turn determined by the hours of labor, h, she supplies to the market, and her participation in the Government's transfer programs. Second, the sum of mothers' leisure, time investment, and labor supply must equal the number of hours in one period, which we can normalize to 1. Thus, mothers face a trade-off between earned income through the supply of labor, and hours that may be spent in child-relevant activities or private leisure. I let $\mathcal{H}$ indicate the set of work hours from which mothers can choose. In the empirical application that follows, I make labor supply a discrete choice with $\mathcal{H} = \{0, 10, 20, 30, 40, 50\}$. Finally, let $p \in \{0, 1, 2\}$ indicate the mother's decision to participate

[3] In my empirical application, I set $N_\theta = 3$

[4] I set $A = 17$ in this paper.

in the Government's transfer programs, the structure of which I formalize below. These choices correspond to no programs (0), food stamps (1) or food stamps as well as welfare (2). In what follows, let S_{it} be a vector that tracks all variables that are relevant for determining the household budget in period t. With this variable, we can write the budget function as $\mathbf{Y}(h, p, S_{it})$. S_{it} will contain variables that are relevant to the determination of wages, non-labor income, as well as the parameters that summarize federal transfer policies.

The evolution of each child's skills $\theta_{i,t}$ over time is determined by a production function. I specify that this technology takes the familiar Cobb-Douglas form:[5]

$$\theta_{it+1,k} = \psi_{it} x^{\delta_{x,k,a}} \tau^{\delta_{\tau,k,a}} \prod_{j=1}^{N_\theta} \theta_{ia,j}^{\delta_{\theta,k,j}} \tag{2.1}$$

Here, $\delta_{x,k,a}$ ($\delta_{\tau,k,a}$) is the Cobb-Douglas share of money (time) investment in the production of skill k at age a. Similarly, $\delta_{\theta,k,j}$ indicates the share of the current stock of skill j in the production of next period's future stock of skill k. The coefficient ψ_{it} represents total factor productivity in production, which I assume is determined as:

$$\log(\psi_{it}) = \mu_{\theta,i} + \eta_{it}, \tag{2.2}$$

where $\mu_{\theta,i}$ is a mother-specific fixed effect[6], and η_{it} is an iid, developmental shock. The presence of $\mu_{\theta,i}$ reflects the idea that some mothers will simply be more productive in shaping child outcomes. Since this may be freely correlated with mothers' ability to generate income, there is significant empirical content to this assumption.

This specification introduces particular dynamics to skill production. For example, the returns to monetary investment at age a are not solely determined by the shares $\delta_{x,a}$, but also by how skills at age $a+1$ shape the formation of skills at age $a+2$, and so on. One should also note the dependance of the development process on a. This is a natural way to specify production, since it is well known that developmental sensitivities change as children age.

2.2 Family Budget and The Policy Environment

To implement the solution described above it is necessary to unpack the components of the budget function, $\mathbf{Y}(h, p, S_{it})$, and specify the members of the state vector, S_{it}. We can decompose the family budget into labor earnings and three sources of government transfers, which are food stamps,

[5]This specification follows Del Boca et al. (2014), with an extension to multidimensional skills.
[6]This is permitted to depend on observables, at no conceptual cost.

welfare, and taxes. We can write

$$
\begin{aligned}
\mathbf{Y}(h,p,S_{it}) =& \underbrace{W_{it}h}_{=E_{it}} \\
&+ \mathbf{1}\{p\geq 1\}T^F(E_{it};Z_{F,it}) \\
&+ \mathbf{1}\{p=2\}T^A(E_{it},L_{it};Z_{A,it}) + T^T(E_{it};Z_{T,it}) \\
=& \mathbf{Y}\Big(h,p,W_{it},L_{it},\underbrace{Z_{F,it},Z_{A,it},Z_{T,it}}_{Z_{it}}\Big)
\end{aligned}
$$

In this expression, $E_{it} = W_{it}h$ is earned income from labor hours supplied to the market, and (T^F, T^A, T^T) are functions that map these variables to food stamp receipt, welfare receipt, and net taxes. Each transfer is indexed by the set of policy parameters $(Z_{F,it}, Z_{A,it}, Z_{T,it})$, which are mapped to each mother according to the relevant calendar year and her reported state of residence. Furthermore, the functions T^F and T^A incorporate conditions defining the eligiblity of each mother given her income E_{it} and accumulated periods of welfare use, L_{it}. The latter variable applies only after the introduction of time limits for welfare participation.

Members of $(Z_{A,it}, Z_{F,it})$ include the parameters that define gross and net income tests for eligibility, as well as gross and net income for payment calculations. Members of $Z_{T,it}$ include the marginal income tax rates at different income brackets, standard deductions for single-headed households, and the parameters defining federal and state EITC payments. Further details on the computation of these tax functions and the exact members of each policy vector can be found in Appendix F.

2.3 The Family Production Problem

Having described the environment and maternal preferences, we can move on to stating the dynamic programming problem faced by the family unit. I proceed by stating the problem, then describing the relevant notation. Recall that the terminal period of the problem occurs at A, the time period at which the child matures. For simplicity, I suppress here the subscript i and ask the reader to bear in mind that this problem is defined uniquely for each mother i in the sample. I state below the problem for every time period $t < A$:

$$V_t(\theta_t, S_t, \psi_t) = \max_{c,l,x,\tau,h\in\mathcal{H},p} \Big\{u(c,l,h,\theta_t,p) + \beta\mathbb{E}[V_{t+1}(\theta_{t+1}, S_{t+1}, \psi_{t+1})\ |S_t]\Big\} \tag{2.3}$$

Subject to the constraints:

$$c + x \leq \mathbf{Y}(h,p,S_t) \tag{2.4}$$

$$l + \tau + h = 1 \tag{2.5}$$

$$\theta_{k,t+1} = \psi_t x^{\delta_{x,k,a}} \tau^{\delta_{\tau,k,a}} \prod_{j=1}^{N_\theta} \theta_{j,t}^{\delta_{\theta,k,j}},\ \text{for } k = 1,2,..,N_\theta \tag{2.6}$$

$$L_{t+1} = L_t + \text{TimeLimit}_t \mathbf{1}\{p=2\}\mathbf{1}\{L_t < \mathcal{L}\} \tag{2.7}$$

In this setup, equation (2.4) is the budget constraint. The function $\mathbf{Y}$ defines the budget set for the family as a function of parental labor supply, h, program participation, p, and relevant economic state variables, S_t, which are permitted to evolve dynamically with parental decision making. Most importantly, S_t includes all variables that summarize the policy environment faced by mothers, including the welfare and tax policy applicable at time t. One particularly important dynamic component of the decision-making problem will involve the introduction of time limits. The variable TimeLimit_t indicates whether time limits are present in the current policy environment. If $\text{TimeLimit}_t = 1$, equation (2.7) indicates that mothers must track their accumulated welfare use, L_t, up to the exogenous limit $\mathcal{L}$. Finally, (2.5) describes the time constraint faced by the mother.

I close the model by specifying the form of utility, u_t, and the terminal value function:

$$u_t(c, l, h, \theta, p) = \alpha_c \log(c) + \alpha_l \log(l) + \alpha_\theta \sum_k \log(\theta_k) - v_t(h, p) \tag{2.8}$$

$$V_A(\theta) = (1-\beta)^{-1} \alpha_\theta \sum_k \log(\theta_k) \tag{2.9}$$

where v_t is the disutility from work and program participation, which is allowed (in principle) to be time varying. Since θ is multidimensional, it is certainly conceivable that mothers will value components of the skill vector differently, applying different weights to cognitive and behavioral skills. Such an admission in this model, in which investments x and τ are unidimensional and largely unobservable, would create a difficult identification problem without a transparent solution. Instead, I will assume that α_θ is a scalar and that, subject to appropriate scaling,[7] all skills are valued equally. Finally, for notational convenience, we can write the production function in the following vector notation:

$$\log(\theta_{t+1}) = \delta_{x,a} \log(x) + \delta_{\tau,a} \log(\tau) + \delta_\theta \log(\theta_t) + \log(\psi_t) \tag{2.10}$$

For reasons made clear in Appendix C, this assumption is pivotal in reducing the computational complexity of the problem. Since θ is a vector, $\delta_{x,a}$ and $\delta_{\tau,a}$ are N_θ dimensional vectors of Cobb-Douglas shares. Assuming log utility, as in (2.8), facilitates the derivation of closed-form expressions for investment policies. This is also demonstrated in the model solution found in Appendix C. Finally, I specify that the terminal value V_A is equal to the discounted present value of utility derived from each child's final abilities over an infinite horizon [8].

Assumptions (2.8) and (2.6) lead to the following simplification of the dynamic program. In Appendix C, I formally derive the following expressions in the general case that allows for multiple children. Let me relegate technical details to that section, presenting here the substantive

[7] I deal with the issue of how to scale test scores in Section 4.1.2.

[8] This specification can be easily made to correspond exactly to the right value for a fully-specified infinite horizon problem.

model implications and some informal discussion. The value function can be written as additively separable in the skill vector θ:

$$V_t(\theta, S_t, \psi_t) = \alpha'_{V,t} \log(\theta_t) + \alpha_{V,t+1} \log(\psi_t) + \nu_t(S_t) \tag{2.11}$$

with the component value function ν given by the recursive formulation:

$$\nu_t(S_t) = \max_{h \in \mathcal{H}, p} \left\{ \tilde{u}_t(\mathbf{Y}(h, p, S_t), h, p) + \beta \mathbb{E}[\nu(S_{t+1}) \mid S_t] \right\} \tag{2.12}$$

and $\tilde{u}_t$ an augmented utility function, where the utility derived from income and hours choices is adjusted by the developmental importance of time and money in period t:

$$\tilde{u}_t(Y, h, p) = \overline{\alpha}_{c,t} \log(\mathbf{Y}(h, p, S_t)) + \overline{\alpha}_{l,t} \log(1 - h) - v_t(h, p)$$

In this formulation, the parameters $\overline{\alpha}_{c,t}$ and $\overline{\alpha}_{l,t}$ are aggregates that represent the total value of income and leisure, respectively. They can be written as:

$$\overline{\alpha}_{c,t} = \alpha_c + \beta \alpha'_{V,t+1} \delta_{x,a} \tag{2.13}$$

$$\overline{\alpha}_{l,t} = \alpha_l + \beta \alpha'_{V,t+1} \delta_{\tau,a} \tag{2.14}$$

We see that the utility derived from income is composed of two terms: α_c is of course the value derived from private consumption, while $\beta \alpha_{V,t+1} \delta_{x,t}$ is the marginal return of investment to each skill, scaled by the value derived from each skill next period, $\alpha_{V,t+1}$, and discounted by β. The value derived from skills in each period, $\alpha_{V,t}$ can itself be expressed recursively:

$$\alpha_{V,t} = \alpha_\theta \mathbf{1}_{N_\theta} + \beta \delta'_\theta \alpha_{V,t+1} \tag{2.15}$$

where $\mathbf{1}_{N_\theta}$ is a N_θ-sized column vector of ones. The value from skills is the sum of utility derived from each skill today (α_θ) and the return of each skill to the production of future skills, scaled by the value of these in the next period ($\beta \alpha_{V,t+1} \delta_\theta$). Since terminal utility V_A also takes this log-linear form, we can see how this recursion holds in each period. Finally, preservation of the log-additivity of the value function leads to the following proportional investment rules:

$$x_t = \underbrace{\frac{\beta \alpha'_{V,t+1} \delta_{x,a}}{\overline{\alpha}_{c,t}}}_{=\phi_{x,t}} \mathbf{Y}(h, p, S_t) \tag{2.16}$$

$$\tau_t = \underbrace{\frac{\beta \alpha'_{V,t+1} \delta_{\tau,a}}{\overline{\alpha}_{l,t}}}_{=\phi_{\tau,t}} (1 - h) \tag{2.17}$$

This formulation of the problem greatly simplifies computation, since the value function in terms of the skill vector θ can be solved in closed form. The key to this result is that, subject to a log transformation, the current realization of θ does not effect the productivity of investments. Thus, the value of current θ can be expressed as a linear combination of period utility and the discounted return to future skills, as defined by the share of θ in production, δ_θ. This is formalized in expression (2.15). Since we have log-utility, these coefficients describe in (2.16) the relative share of income spent on the child and in (2.17) the relative share of non-labor hours spent in time investment. The proportional investment rules, when substituted into the dynamic program, allow us to simplify the problem to one of labor supply and program participation, as shown in (2.11) and (2.12). The coefficients $(\overline{\alpha}_{c,t}, \overline{\alpha}_{\tau,t})$ define the labor supply problem, and adjust each period to reflect the relative importance of time and money in the production of child skills.

2.4 Parametric Assumptions on Utility and Heterogeneity

Thus far I have made no restrictions on the distribution of structural primitives across mothers. To cement this before bringing the model to data, I assume that each mother i is one of K specific types, $k(i) \in \{1, 2, ..., K\}$. Thus the preference parameters $(\alpha_{l,i}, \alpha_{\theta,i})$ take a type-specific value $(\alpha_{l,k(i)}, \alpha_{\theta,k(i)})$. Furthermore, I assume that the function v_{it} takes the form:

$$v_{it} = (\alpha_{h,0,k(i)} + \alpha_{h,1,k(i)} AGE_{it})\mathbf{1}\{h > 0\} + \alpha_{F,k(i)}\mathbf{1}\{p \geq 1\} + \alpha_{A,k(i)}\mathbf{1}\{p = 2\}$$

where AGE_{it} is mother i's age at time t. In the next section, I will introduce an approach to identification that is not restricted to a particular number of types, however the choice of K is ultimately subject to data limitations. One aspect of preferences that I have not allowed to be heterogeneous is the discounting parameter, β, which I externally set to $\beta = 0.95$.

2.5 Specifying Information and the Wage Process

Since the vector θ is continuously distributed, being able to solve for this component of the value function in closed form is highly advantageous. However, we are still left with ν, the value function from the augmented labor supply problem described above. While S_{it} could in principle contain a wide range of endogenous, dynamic variables, in this paper we consider just one: the mother's number of accumulated periods of welfare use, L_{it}. In the pre-reform era, this variable is irrelevant, however once time limits are introduced, mothers must begin to make trade-offs between welfare use today and welfare use in the future.

Notice that the exogenous components of S_{it} include, among other things, the wage, W_{it}, and the policy vectors $Z_{F,it}, Z_{A,it}, Z_{T,it}$. This is a high-dimensional, continuous vector, which introduces another potential intractability to the problem. To negotiate this, I will assume that mothers have perfect information about the future path of policy variables, and that wages follow a process:

$$\log(W_{it}) = \gamma_{w,0,k(i)} + \gamma_{w,a,k(i)} AGE_{it} + \epsilon_{w,it}$$

where $\epsilon_{w,it}$ follows as an AR(1) process:

$$\epsilon_{w,it+1} = \rho_{w,k(i)}\epsilon_{w,it} + \sigma_{w,k(i)}\xi_{it}, \xi_{it} \sim \mathcal{N}(0, 1).$$

I follow Kopecky and Suen (2010) and use the method of Rouwenhorst (1995) to approximate this process with a discrete markov transition matrix.

3 Identification

Having derived the necessary properties of the model's solution for a single mother, in this section I return to the problem of specifying how this model relates to available data, and negotiating the identification issues that arise from a realistic specification of the model. First, assume that we have a panel dataset of observations $\mathcal{X}_i$ for each mother i, and $\mathcal{X}_i$ takes the form:[9]

$$\mathcal{X}_i = \{(W_{it}, N_{it}, P_{it}, H_{it}, \theta_{it}, Z_{it})_{t=0}^{T}, \tau_{is}\}_{i=1}^{N}.$$

To clarify notation, notice that while I used h and p to denote the mother's choice variables in the previous section, H_{it} and P_{it} signify their respective random variables in the data. Second, it is assumed that in some period, s, the mother's time investment is observed. While not crucial for identification, this assumption corresponds to the data I use in estimation, and can be put to good use. Finally, it must be clarified that the wage, W_{it}, is only observable when mother i supplies positive hours to the market in period t, $H_{it} > 0$.

When considering identification, it will help to review the parameters of the specific version of the model I will take to the data. Each of the K types share the parameter vectors defining Cobb-Douglas shares:

$$\delta = (\delta_\theta, \delta_x, \delta_\tau)$$

while each $k \in \{1, 2, ..., K\}$ will have type-specific parameters:

$$\zeta_k = (\alpha_k, \mu_{\theta,k}, \gamma_{w,k}, \sigma_{w,k})$$

which define, respectively, preferences, productivity in the production of skills, and the parameters defining the wage process. The challenge to credible identification arrives when deciding on the joint distribution of these latent characteristics. While specification of the model already relies on particular functional form restrictions, the credibility of my empirical analysis will be greatly improved if the model can be estimated without imposing *a priori* distributional assumptions on this latent vector, and without asserting a particular correlation structure with important observables. For example, a key determinant of inference regarding the effect that different welfare programs have on child outcomes is the absolute and relative magnitude of δ_x and δ_τ, the vectors of Cobb-Douglas shares of money and time in the technology of skill formation. If there is a positive relationship between mother's TFP, $\mu_{\theta,i}$, and wages (say through $\gamma_{w,i}$), then this may lead to upward bias in estimates of money's share in production. Similarly, if there is a negative relationship between $\mu_{\theta,i}$

[9] This is not precisely the same as the dataset that I will use, but a close enough approximation to facilitate exposition. In reality, the panel may be unbalanced, for example, nor will we see development outcomes in every period.

and the preference for leisure $\alpha_{l,i}$ or the fixed cost of work, $\alpha_{h,0,i}$, then this may result in downward bias in the estimates of time's share in production. In the next section, I derive a linear estimation equation, such that the challenge of credibly estimating key causal parameters can be explored in the familiar setting of linear regression.

Typically, identification problems like the one described above can be negotiated either by exploiting exogenous cross-sectional variation, or by using any available panel dimension in the dataset. My identification strategy in this paper exploits both, making use of exogenous policy variation as well as repeated observation of maternal decisions and outcomes, from which the distribution of latent variables may be inferred. Loosely speaking, the crucial assumptions that I am making in order to achieve semiparametric identification of the model, are as follows:

1. Government policies, Z_{it}, sufficiently vary (where sufficiency is defined by a rank assumption) mothers' labor supply decisions and total household income.
2. There is sufficient variation in each mother's budget set (where sufficiency is defined by a support condition) over time such that their labor supply and program participation decisions asymptotically reveal their preferences.

In principle, an identification argument could be developed without relying on this latter assumption, and therefore without relying on an increasing panel dimension, however the key advantage to this approach is that it does not require variation in the budget set to be *independent* of latent variables. To identify the model using purely cross-sectional information would require much stronger assumptions on the support of budget variation induced by government policies, which I view as being overly demanding in this setting.

Instead, my approach rests on the observation that, when assisted by sufficient variation in families' budget sets, particular choice comparisons reveal mothers' preferences as $T \to \infty$. Specifically, I introduce a score criterion that, for a particular vector of preferences, counts the number of times a choice is dominated by an alternative option in the choice set. Under particular conditions, and for a careful definition of choice comparisons, this criterion is uniquely minimized in the limit by each mother's true preferences. This result implies that, in theory, each mother's latent preference type could be estimated consistently from their respective panels. While this would allow my assumptions on preferences to be completely unrestricted, it introduces problematic asymptotic properties known widely as the "incidental parameter problem" (Arellano and Bonhomme, 2011). Adopting the grouped heterogeneity assumption, as I have here, allows for the standard $\sqrt{N}$-asymptotics to hold, as argued by Bonhomme and Manresa (2015) . In accordance with this assumption, in Section 4 I outline a classification based estimator that relies on the score criterion to consistently group mothers into preference types, prior to a step in which I estimate the full set of model parameters.

Following work by Bonhomme and Manresa (2015) and Bonhomme et al. (2017), classification estimators of this style have grown increasingly popular. While the approach proposed by these authors involves selection of moments that will converge to unique limits for a particular type, k, the nature of my problem demands a different strategy. Since two mothers of the same type k who live in different states face different policy environments, it is virtually impossible to choose moments that will share a unique limit across states. In my approach, instead of using panel moments to distinguish between types, I use their ordinal rankings of choices in their choice set, and introduce a score criterion that will be uniquely minimized by the correct preference type as the panel dimension grows.

In the remainder of this section I will first discuss the identification of production parameters, δ, and then discuss support conditions that will permit identification, at the panel level, of maternal preferences without restrictions on their space or distributional structure.

3.1 Identification of Production Parameters

First, in a critical step, a convenient estimating equation can be derived by taking the outcome equation (2.6) and substituting in the investment policies (2.16) and (2.17). We obtain the expression:

$$\log(\theta_{it+1}) = \delta_{x,t}\log(Y_{it}) + \delta_{\tau,t}\log(1 - H_{it}) + \delta_{\theta,t}\log(\theta_{it}) + \varepsilon_{it} + \mu_{\theta,i} + \eta_{it} \tag{3.1}$$

$$\varepsilon_{it} = \delta_{x,t}\log(\phi_{x,it}) + \delta_{\tau}\log(\phi_{\tau,it}) \tag{3.2}$$

Recall that $\phi_{x,it}$ and $\phi_{\tau,it}$ are the marginal propensities to invest money and time in the child, and that these are functions (see for example expression (C.13)) of each mother's preferences $(\alpha_{l,i}, \alpha_{\theta,i})$ and the Cobb-Douglas shares $(\delta_x, \delta_\tau, \delta_\theta)$. A useful advantage of this model is that it allows us to write outcomes as a function of the two key endogenous variables that are affected by transfer policies: total income, and maternal labor supply. In the absence of the model, we would not have a coherent sense of how to use family income and maternal hours as meaningul proxies for largely unobserved investments. The model presented here shows us one particular path forward.

For two different reasons, regressing skills in $t+1$ on income, hours at home, and skills at t will produce biased estimates. First, mothers' productivity $\mu_{\theta,i}$ is quite believably positively correlated with variables that determine net income, Y_{it} (in particular, wages and non-labor income). Second, even without this feature, the *missing investment* problem persists: the propensity to invest money and time is correlated with income and hours through their joint dependence on maternal preferences. Both of these issues can be navigated under the presence of instruments that sufficiently shift hours and income, where "sufficiency" here is defined by the typical rank assumptions of linear instrumental variables.

So, if family budget sets are sufficiently exogenously varied, we have estimates of a key set of parameters: those that define the way time and money enter in the production of childrens' human capital. While a method to credibly identify the importance of time and money is a useful step forward, these alone do not allow us to forecast the impacts of counterfactual policies, which are mediated through the labor supply and program participation decisions of mothers. These, in turn, are determined in the model by the budget function $\mathbf{Y}$, a function of wages, policies, and mothers' preferences, α. In the next section, I sketch a panel data approach to identification of this latent distribution, which motivates an iterative estimation procedure using classification by minimum score.

3.2 Identification of Preferences

Assume for the purposes of exposition that we have identified, through rank assumptions on our instrumental variables, the Cobb-Douglas shares, δ. In this section I discuss the conditions under which the latent preference vector, $\alpha = [\alpha_l, \alpha_\theta, \alpha_h, \alpha_F, \alpha_A]'$, is revealed by a mother's decisions as the length of the panel, T, increases.

The argument for identification proceeds in the following four steps:

1. Given the model solution for investment policies, the ratio α_θ/α_l can be inverted from a single observation of time investment, τ.
2. Showing that for any distinct pair α,α' there is a state S such that these this pair disagrees on the ordinal ranking of particular choices.
3. Showing that a score criterion that utilizes these choice comparisons is minimized, in the limit, only by a preference vector that preserves these ordinal rankings.

Combination of points (2) and (3) implies that only the true preference vector will, in the limit as $T \to \infty$, minimize the score criterion when there is sufficient variation in wages and government policies.

First, it will simplify this discussion to note that the recursive definition of $\alpha_{V,t}$, as given by (2.15), can be written as a linear transformation of α_θ:

$$\alpha_{V,t} = \Gamma_t(\delta)\alpha_\theta,$$

where δ is the full vector of production parameters. Under this notation, the utility coefficients can be written[10]:

$$\overline{\alpha}_{c,t} = 1 + \beta\delta_{x,t}\Gamma_{t+1}(\theta)\alpha_\theta \tag{3.3}$$

$$\overline{\alpha}_{l,t} = \alpha_l + \beta\delta_{\tau,t}\Gamma_{t+1}(\theta)\alpha_\theta \tag{3.4}$$

$$\frac{\tau}{1-h} = \frac{\beta\delta_{\tau,t}\Gamma_{t+1}(\theta)\alpha_\theta}{\alpha_{l,t}} \tag{3.5}$$

Fixing estimates of the production parameters, δ, we see that a single observation of $\tau/(1-h)$ is sufficient to back out the ratio α_θ/α_l. Thus, the remaining task is to use preference orderings to

[10]Adopting once more the simplifying assumption of one child

identify the remaining parameters: $\alpha_l, \alpha_F, \alpha_A, \alpha_h$. In Appendix D I show how support conditions on the following three terms, for all choices of h and p, are sufficient to distinguish any two distinct preference parameters

$$\log\left(\frac{\mathbf{Y}(h+\Delta,p,S)}{\mathbf{Y}(h,p,S)}\right), \log\left(\frac{\mathbf{Y}(p,h,S)}{\mathbf{Y}(p,0,S)}\right), \log\left(\frac{\mathbf{Y}(2,h,S)}{\mathbf{Y}(0,h,S)}\right).$$

where Δ is any fixed deviation in hours of work. In words, these three terms represent the financial return to (1) working an additional Δ hours, (2) joining the labor force, and (3) participating in welfare. They capture the incentives for choice at the intensive margin of labor supply, the extensive margin of labor supply, and the extensive margin of program participation. There are two candidate sources of variation in these three terms. First, an increase in wages will increase the returns to work at both margins, while diminishing the returns to program participation. Second, changes in policies over time will also provide crucial variation. For instance, changes in the implicit marginal tax rate (through changes in benefit formulae or the tax code) will have the same effect as wages, while changes in the generosity of welfare benefits will move each term in the opposite direction. If the space of preferences is unrestricted, then each of these terms must have full support on $\mathbb{R}^+$ for identification to hold. While this is a burdensome condition, it is only required under this very relaxed assumption regarding the support of preferences. When implementing this strategy under the grouped heterogeneity assumption, it appears that there is sufficient variation to separate sample members into well-defined types (see Section 4 for more details).

Identification by Minimum Score

If, under appropriate budget variation, only the true vector of maternal preferences preserves the true ranking of choices across the support of S, the task remains to leverage this insight into an estimation criterion. To do this, I define a score criterion which counts, for a given preference vector, the number of times the deterministic component of utility is dominated by an alternative in the choice set. We define:

$$Q_{T,i}(\alpha, S_i) = \sum_{t=1}^{T} \sum_{d\in\mathcal{H}\times\{0,1,2\}} q(D_{it}, d, S_{it})\mathbf{1}\{\Delta\tilde{u}_t(D_{it}, d, S_{it}, \alpha, \delta) < 0\}$$

where q is a function that selects particular comparisons in a data-dependent fashion. In addition, to make use of observed time use information introduced in part (1) of my identification argument, I augment this score criterion with a quadratic term:

$$\overline{Q}_{T,i}(\alpha, S_i) = Q_{T,i}(\alpha, S_i) + \sum_{t\in\{1997,2002\}} \left(\frac{\tau_{it}}{1-H_{it}} - \phi_{it}(\alpha,\delta)\right)^2$$

where τ_{it} is measured time investment across all children, and $\phi_{it}(\alpha, \delta)$ is the fraction of time invested in all children implied by the model parameters (α, δ). Notice that if utility is strictly deterministic then this function is trivially minimized by α_i, since by definition of optimality

$Q_{T,i}(\alpha_i, S_i) = 0$. In the data, we will undoubtedly see cases in which this restriction is violated, which can be easily rationalized by assuming small, iid, shocks ($\epsilon_{h,d}$) to the utility derived from each discrete choice pair (h, d). Allowing for this case requires more care in selecting the alternatives through the function, q.

In the estimation procedure, I consider the following. Let $D \in \mathcal{D} = \mathcal{H} \times \{0, 1, 2\}$ be the discrete choice, consisting of an hours choice H and a program choice P. Let H^+ and H^- be the hours choices above and below H in the set $\mathcal{H}$. Our definition of q is as follows:

$$\begin{aligned} q(D_{it}, d) = \mathbf{1}\{p = P_{it}\} \Big(& \mathbf{1}\{H_{it} < \overline{h}\}\mathbf{1}\{h = H^+_{it}\} + \mathbf{1}\{h > 0\}\mathbf{1}\{h = H^-_{it}\} \\ & + \sum_{h' \in \mathcal{H}} \mathbf{1}\{H_{it} = 0\}\mathbf{1}\{h' = h\} + \mathbf{1}\{H_{it} > 0\}\mathbf{1}\{h = 0\}\Big) \\ & + \mathbf{1}\{h = H_{it}\}\mathbf{1}\{\text{TimeLimits}_{it} = 0\}\big(\mathbf{1}\{P_{it} = 2\}\mathbf{1}\{p = 0\} + \mathbf{1}\{P_{it} \leq 1\}\mathbf{1}\{p = 2\}\big) \end{aligned} \tag{3.6}$$

In words, q (1) fixes the participation choice, and compares hours choices above and below the observed hours choice as well as to not working; (2) when $H_{it} = 0$ is observed, compares this choice to each $h > 0$; and (3) fixes the hours choice and compares non-partication to participation, or vice-versa. A formal treatment of score criterions of this style is given in Fox (2007). It suffices for us to note that, in this definition of q, $q(d, d') = 1$ if and only if $q(d', d) = 1$. Consider then that in the probability limit, for a given state S_{it}, $Q_{T,i}$ is comprised of individual components of the form:

$$P(d|S_{it})\mathbf{1}\{\tilde{u}_t(d, \alpha, S_{it}) < \tilde{u}_t(d', \alpha, S_{it})\} + P(d'|S_{it})\{\tilde{u}_t(d, \alpha, S_{it}) > \tilde{u}_t(d', \alpha, S_{it})\}$$

where $P(\cdot|S_{it})$ are the choice probabilities. In expectation, this component of $Q_{T,i}$ is minimized by a choice of α such that the ranking of choice probabilities, $P(\cdot|S_{it})$ is matched by the ranking of utilities, $\tilde{u}_t(\cdot, \alpha)$. The true preference vector, α_i, will satisfy this condition whenever the shocks to utility, $\epsilon_{h,p}$, are exchangeable (Fox, 2007).

This reasoning is sufficient to conclude that α_i must be in the set of parameters that minimize $\overline{Q}_{T,i}$ in the limit, but does not guarantee that such a set is a singleton. This latter implication is given, however, by our assumption that there is sufficient variation in the budget set for each mother. As was discussed already, when the space of preferences is unknown, identification is only guaranteed by unbounded variation in the returns to work and program participation, but these support conditions become less burdensome when the space of preferences is limited to discrete types, as I do when I practically implement these insights in estimation. In this section I have provided a constructive, albeit informal, analysis of identification, which provides clear suggestions for how estimation of the model should proceed. In this sense it is also transparent, in terms of clearly specifying the information from the data that pins down key parameters.

4 Estimation

Estimation of the model's parameters is motivated by the identification analysis of the previous section, in particular by the result that a well-defined score criterion can reveal the preferences (up to scale) of each mother in the sample. This observation suggests that we can use the information revealed by mothers' choices to classify them into preference types. While in principle one could use the score to uniquely estimate the preferences of each mother in the sample, this approach introduces a bias term that depends on the relative rate at which N and T grow in the sample (Arellano and Bonhomme, 2011). As a solution to this incidental parameter bias, Bonhomme and Manresa (2015) suggest adoption of a *grouped heterogeneity* framework. With this assumption, instead of estimating a latent vector for each type, mothers can be classified into one of K latent types.[11] This procedure admits regular $\sqrt{N}$-asymptotics, and therefore standard inference.[12] This assumption can be written as:

$$\zeta_i = \zeta_{k(i)}, \;\; k(i) \in \{1, 2, ..., K\}$$

such that each mother corresponds to one of a set of K types.

Estimation now proceeds in two stages. In the first stage, I conduct a procedure that iteratively classifies mothers into preference type, and estimates the production parameters $\delta = [\delta_x, \delta_\tau, \delta_\theta]$ using moment conditions implied by the linear estimation equation, (3.1). In the second stage, fixing type classification and the production parameters, I estimate the remaining parameters by matching empirical moments separately for each type. I describe each stage in more detail below.

Stage 1

Let $\mathcal{K} = \{k(i)\}_{i=1}^N$ denote a particular assignment of each mother i to a type $k(i) \in \{1, 2, ..., K\}$, and let $\alpha = \{\alpha_k\}_{k=1}^K$ be the preferences of each type (up to scale). Fixing δ, the classification estimator proceeds by solving:

$$\hat{\mathcal{K}}, \hat{\alpha} = \arg\min_{\mathcal{K},\alpha} \mathcal{Q}(\mathcal{K}, \alpha) = \arg\min_{\mathcal{K},\alpha} \sum_{i=1}^{N} \overline{Q}_{T,i}(\alpha_{k(i)}).$$

Because of the computationally complex task of simultaneously minimizing over α while classifying by type, I follow Bonhomme and Manresa (2015) and take a K-Means clustering approach, which uses the following iterative loop:

1. Fixing $\hat{\mathcal{K}}$, estimate $\hat{\alpha} = \arg\min \mathcal{Q}(\hat{K}, \alpha)$
2. Fixing $\hat{\alpha}$, obtain a new classification $\mathcal{K}'$ by setting $k(i) = \arg\min \overline{Q}_{T,i}(\alpha_k)$.

[11] While this assumption may seem restrictive, it is almost omnipresent among papers that structurally estimate dynamic decision models with latent heterogeneity.

[12] Of course, if the discrete heterogeneity assumption is violated in the data, bias reappears. Bonhomme et al. (2017) introduces asymptotic bounds on this bias, as a function of the dimension of the true distribution of latent variables.

3. If $\hat{\mathcal{K}} = \mathcal{K}'$, terminate the loop. Otherwise, return to step (1).

Now, fixing the classification $\mathcal{K}$, I turn to the estimation of δ. Recall first that identification was established by invoking rank assumptions on the policy instruments, Z_i with respect to H_i and Y_i. This insight is mobilized by the moment condition:

$$g_1(Z_{it}, \delta) = \mathbb{E}[(\varepsilon_{it} + \mu_{\theta,i} + \eta_{it}) \otimes Z_{it}] = 0,$$

which is equivalent to linear instrumental variables. However, to add precision to the estimates, I also make use of the fact that the classification $\mathcal{K}$ can be used as a control in this regression, since the error terms ε_{it} and $\mu_{\theta,i}$ are functions of type. This is equivalent to making use of the moment condition:

$$\mathbb{E}[\varepsilon_{it} + \mu_{\theta,i} + \eta_{it} | H_{it}, Y_{it}, k(i)] = \mathbb{E}[\varepsilon_{it} + \mu_{\theta,i} + \eta_{it} | k(i)].$$

Thus in addition to g_1 we include:

$$g_2(H_{it}, Y_{it}, \delta, \mathcal{K}) = \mathbb{E}[\eta_{it} \otimes [D'_{k(i)}, \log(1 - H_{it}), \log(Y_{it})]'] = 0,$$

where $D_{k(i)}$ is a vector of dummy variables that indicates i's type classification. This moment condition makes use of the assumption that, if we knew each mother's latent type, we could control for this information, offering an alternative to the instrumental variables strategy.

Stage 1 of the procedure iterates on the two steps outlined in this section:

1. Fix δ, and estimate $\mathcal{K}$ using the clustering routine described above.
2. Fix $\mathcal{K}$, and estimate δ using g_1 and g_2 in a nonlinear GMM procedure.
3. If $\delta' = \delta$, terminate. Otherwise, set $\delta = \delta'$ and return to Step (1).

Appendix E gives further details on computation, and the specifics of implementing the nonlinear GMM procedure.

Stage 2

Stage 1 produces consistent estimates of $\mathcal{K}$, the production parameters δ, and estimates of preferences α up to scale. However, the procedure does not make full use of the available information in the data, and does not permit estimation of the parameters governing wages. I therefore conduct a second stage that estimates the full set of parameters for each type using indirect inference (II). Grouping the data by type assignment ($\hat{\mathcal{K}}$) from the previous section, I calculate a vector of statistics for each group: $\mathcal{M}_k$. For each k, an identical vector of statistics can be simulated from the model, $\mathcal{M}_k(\alpha_k, \delta)$. Fixing a weighting matrix on these, the II estimator is:

$$\hat{\alpha}_k = \arg\min \left(\mathcal{M}_k - \mathcal{M}_k(\alpha, \hat{\delta})\right)' W \left(\mathcal{M}_k - \mathcal{M}_k(\alpha, \hat{\delta})\right),$$

where $\hat{\delta}$ is the estimate of production parameters from the first stage. In Appendix E I provide further details regarding selection of the moments, choice of estimator, and implementation of this stage.

4.1 Data

To answer the empirical and policy questions outlined in Section 1, I use data from the *Panel Study of Income Dynamics* (PSID) and its *Child Development Supplement*. The PSID is a dynastic, longitudinal survey taken annually from 1968 to 1997, and biennially since 1997. It collects important information on a range of economic and demographic indicators. The CDS consists of three waves, collected in 1997, 2002 and 2007. Any child in a PSID family between the ages of 0 and 12 at the time of the 1997 survey was considered eligible. These surveys contain a broad array of developmental scores in cognitive and socioemotional outcomes as well as information on the home environment of the child. One important feature of the survey is the availability of time use data, which is collected by the participants' completion of time diaries. I provide further details below.

4.1.1 Description of Variables and Sample Selection

From the PSID survey I collect data on mothers' labor supply, labor income, total family income, welfare receipt and some demographic variables. The CDS is comprised of several questionnaires. I use two in particular: the child interview and the *primary caregiver* (PCG) interview. From the child interviews, I use measures of cognitive ability as reflected by a battery of test scores. To measure child attributes, I utilize two measures of cognitive ability and two measures of behavioral traits. For cognitive ability, I use the Letter-Word (LW) and Applied Problems (AP) modules of the Woodcock-Johnson Aptitude test. For socioemotional or "non-cognitive" abilities, I use constructed scales that measure *externalizing* (BPE) and *internalizing* (BPN) behavioral problems. This gives, in total, four measures of child attributes that we use to track human capital outcomes. In the next section, I explore how each measure is related to adult outcomes, as proxied by high school graduation.

Finally, the CDS asks participant children to fill out a "time diary". This portion of the survey requires participants to record a detailed, minute by minute timeline of their activities for two days of the week: one random weekday and one random day of the weekend. Activities were subsequently coded at a fine level of detail. When necessary, children are assisted in completion of the time diary by the PCG. These diaries provide a unique snapshot into the daily life of the child. From this data I construct a measure of maternal time investment by taking a weighted sum[13] of the total hours of time use in which the mother is recorded as actively participating in each diary activity.

Since much of the focus of the antipoverty programs considered here is on single mothers, I make sample restrictions in order look more closely at this subpopulation. Since family structure in the data is quite dynamic and tends to fluctuate, a stand must be taken on how to restrict

[13] $\frac{5}{7}$ for the weekday, and $\frac{2}{7}$ for the weekend

the sample. Following Fang and Silverman (2004), I restrict attention to mothers of CDS children who never marry. This leaves me with a sample of 959 mothers, and 1,405 children. Table A.1 reports descriptive statistics from the sample. We see that mothers in this sample have low levels of education (63% have a High School diploma, while 23% have less) with modest wages and modest earnings. Additionally, we see that this sample is heavily reliant on welfare, with 50% having reported welfare use at least one, and 17% having used welfare for at least 5 years. As further evidence of the relative disadvantage in my sample, notice that the rate at which CDS children from this sample graduate from high school is 74%, relative to the national average of 83%.[14]

Before describing the anchoring procedure, it is worth using the data to address some of the modelling concerns laid out in the Introduction and in Section 2. In particular, the omission of savings and childcare from the model may cause issues when interpreting estimates and analyzing counterfactuals. To assess the severity of this issue, I use the Wealth Module of the PSID in 1999 to compute total cash assets for the households in my sample, and I use the 1997 time diary from the CDS to collect data on time spent in formal care. Table A.1 reports particular statistics of interest using these data. We can observe that the majority of households report no cash assets, while even the 75th percentile amounts to modest savings. Similarly, it appears that 87% of children report spending no time in a formal care arrangement in 1997 (83% for children under the age of 6). From this we can infer that the use of formal care is rare in our sample. Of course, while this may help with interpretation of our estimates (one should think about the exercise as identifying the value of maternal time, holding outside options fixed), there is ultimately no guarantee that these choice variables are invariant to the policies we consider here.

4.1.2 Anchoring Test Scores

In order to give some weight to the scale of cognitive and behavioral scores, I next inspect how each score is associated with the probability of high school graduation. To do this, I estimate a series of linear probability models to give each skill outcome an interpretable scale. That is, for each skill k, if $\tilde{\theta}_{i,k}$ is the raw score of skill k for child i, I estimate the model:

$$HSG_i = \mathbf{1}\{\text{const}_k + \tilde{\theta}_i' \beta_{HSG} - U_i \geq 0\}, \qquad U_i \sim \text{unif}[0, 1] \tag{4.1}$$

Thus, I re-weight each raw score to get $\log(\theta_{i,k}) = \beta_{HSG,k}\tilde{\theta}_{i,k}$, so each skill has been "anchored" in the language of Cunha and Heckman (2008), to the probability of high school graduation. Estimates are presented in Table A.2. In accordance with the joint outcome equation (specification 5), which finds no significant role for internalizing behaviors in determining high school graduation, I proceed to estimate the model on just the other three skills: the two measures of cognitive skills (LW and AP) and a single measure of behavioral skills (BPE).

[14] Using 2014 data from the National Center for Education Statistics: https://nces.ed.gov/programs/coe/indicator_coi.asp

4.2 Estimates

4.2.1 Stage 1: Clustering and Production

Tables A.3 and A.4 show estimates of the parameters of the technology of skill formation, from the iterative GMM procedure, setting the number of types, K, to 7.[15] Estimates are scaled in percentage points of the probability of graduating from high school. Table A.3 reports estimates of the parameters defining the Cobb-Douglas shares of x_t, τ_t, and θ_t in the production of next period's skills, θ_{t+1}. The age-dependance of $\delta_{x,a}$ and $\delta_{\tau,a}$ is parameterized linearly, as

$$\delta_{x,a} = \gamma_{\delta,x,0} + \gamma_{\delta,x,1}a \tag{4.2}$$

$$\delta_{\tau,a} = \gamma_{\delta,\tau,0} + \gamma_{\delta,\tau,1}a. \tag{4.3}$$

These numbers are more readily interpreted graphically, and so in Figure B.1 I plot the Cobb-Douglas shares of money and time as a function of age, with 95% confidence intervals. Several common patterns are immediately clear. First, while both inputs play a significant role in the production of cognitive skills (as measured by the Letter-Word and Applied Problems Scale), it appears that Externalizing Behaviors are not sensitive, at 95% significance, to either input. Second, there is a general downward trend in the shares of each investment with age, with the exception of Applied Problems, which appears to exhibit an upward trend. Turning to estimates of δ_θ in Table A.3, we can see that dynamic complementarities exist, with each skill displaying high self-productivity over time. However, interdependance in skill formation appears limited here, admitting only one significant off-diagonal term: a 1% point increase in Letter-Word scores at age a contributes to a .02% point increase in Applied Problems at age $a + 1$.

To assess the magnitudes of these estimates, I highlight two results from the literature that should be taken into consideration. First, Dahl and Lochner (2012) estimate that a $1,000 increase in annual household income, induced by expansions in the EITC, leads to increases in test scores of between 6% and 9% of a standard deviation. Second, Bernal and Keane (2011) estimate that the effect of a full year of employment among single mothers leads to a reduction in test scores of about 11% of a standard deviation. It should be noted that neither of these estimates correspond to (nor can they be used to infer) any structural parameter in this model. It will however be useful to benchmark my production shares against these numbers by interpreting them as *ceterus paribus* changes in skills, and computing equivalents from my estimates, using the approximate age of the children used in each exercise.

For example, normalizing by standard deviation, my estimates predict that a log unit increase in income at age 8 leads to a 7.8% standard deviation increase in Letter-Word score and a 13.2%

[15]As is typical, K is chosen with consideration for both statisical accuracy and pragmatism. Clustering with type numbers larger than 7 produces clusters with too few members to produce meaningful parameter estimates, and with observable behavior that is indistinguishable from the other clusters.

standard deviation increase in the Applied Problems score. Since an extra $1,000 per year will, in most cases, amount to significantly less than a log unit increase in income,[16] we can infer that my estimates may predict somewhat smaller cognitive impacts than those found in Dahl and Lochner (2012), but only if one assumes that the estimated impact is derived only from increases in income, and not confounded by any changes in maternal labor supply.[17] In addition, I find that when a mother joins the labor force and works 40 hours, the loss in output from time inputs for a child of age 3 will be 12.9% of a standard deviation in Letter-Word and 8.8% of a standard deviation in Applied Problems scores.[18]. Based on these calculations, I conclude therefore that these estimates produce skill changes that are quite comparable in magnitude to those found in Bernal and Keane (2011).

As a final comment on magnitudes, we can use the estimates of δ_θ to aggregate the effect of a log unit increase in investment in every period from birth to 17. Excluding the impact on Behavioral Problems, for which I estimated non-significant production shares, these amount to 3.9 percentage points for money investments and 4.3 percentage points for time investments. These numbers correspond, respectively, to 43% and 48% of the gap between the high school graduation rate in this sample and the national average.

Table A.4 reports estimates of the parameters that are introduced as control variables in our estimation procedure. Jointly, $\gamma_{K,k,0}$ and $\gamma_{K,k,1}$ absorb the terms $\mu_{k(i)}$ and $\varepsilon_{k(i),a}$ in equation (3.1)[19], and I additionally include dummy variables for scores measured in years 1997 and 2002. Figure B.2 shows some interesting differences in the age-dependance of fixed effects across types, without demonstrating any strong differences across levels at any particular age. Since the relationship between $\mu_{k(i)}$ and $\varepsilon_{k(i),a}$ in the model is unrestricted, no *a priori* restrictions are imposed on these estimates across types.

The model provides an important observation when interpreting the parameter vectors δ_x and δ_τ as effect sizes: neither constitute an immediately policy relevant parameter. The model dictates that money and time are inextricably linked by labor supply, and so when we talk about the effect of a $1 increase in income, the true effect is ambiguous unless we know from where this $1 comes. If it comes from a lump sump transfer, the model dictates a reduction in hours, leading to an increase in time investment and an additional impact on outcomes. Conversely, if it arrives from an increase in labor supply, then we must consider the off-setting impact of a reduction in time

[16]When net income is $10,000 a year, this would be approximately 0.1 log units, or 0.2 if we factor in the two-year design of Dahl and Lochner's (2012) study.

[17]One might imagine, for example, that women on the "plateau" region of the EITC schedule could reduce their labor supply and increase time investment in response to expansion of the maximum credit.

[18]To compute this number, I normalize by the standard deviation of scores at age 5, which corresponds to the outcome variable used in the study.

[19]In order to properly capture $\varepsilon_{k(i),a}$, a flexible polynomial in a should be introduced, however I found little sensitivity of the results to the inclusion of extra terms beyond a linear approximation.

investment. In order to conduct sensible policy analysis with these estimates, it is clearly necessary to uncover from the data the distribution of preferences that dictate labor supply decisions. This is made possible by the classification algorithm and second stage of the estimation procedure, which I shall shortly discuss.

4.2.2 Observable heterogeneity across clusters

To finish dicussion of the first stage estimates, I analyze the extent to which the clustering routine, based on revealed preference information in the panel dimension, unveils heterogeneity in mother's preferences for work, program participation, and investment. The first row of Table A.5 indicates that the classification routine produces distinct types of roughly equal sizes in the population. In Figure B.3, I plot the group-year level means of hours worked and rates of program participation for each estimated classification group. This graph plainly shows large differences across groups in propensities to work and participate in welfare, as well as differences in exposure to the welfare and tax reforms of the 1990s. These differences will crucially determine, in upcoming policy experiments, the extent to which each group values particular transfer programs. In particular, taste for work and responsiveness to earnings subsidies will determine each mother's tendency to benefit from programs that increase the incentive for labor supply. In Figure B.4, I compute group level means in hours, wages, welfare, and my measurement of ϕ_τ (taken from two observations of time investment in 1997 and 2002). This graph further identifies important variation in the propensity of mothers to invest (given by the relative weight of skills, α_θ, and leisure, α_l, in utility), which will contribute to final inequality in skill outcomes. Finally, in Figure B.5 I plot the means of actual skill outcomes by type, as well as group-level rates of high school graduation.

Examining these three figures reveals mothers in group (1) to be the clear outlier across types. While types (1) and (5) show the lowest propensity to work, and receive the lowest wages, types (1) and (3) exhibit the highest degrees of welfare dependance. Type (1) also appears, in Figure B.4, to be the only type with a significantly different marginal propensity to invest (ϕ_τ).

Turning back to Figure B.5 and examining first the high school graduation panel, we see that types (1) and (5) display much lower rates of graduation, while the remaining groups appear to graduate at rates that are comparable to the 2014 national average of 83%.[20] Interestingly, turning to the other panels, only group (1) displays a disadvantage in cognitive skills, while groups (1) and (5) exhibit similar levels of disadvantage in behavioral skills. This suggests that, at least for the chosen life cycle outcome of high school graduation, observed inequality may be explained by different skill deficiencies for these two groups. I estimate that behavioral skills are not significantly elastic with respect to investments of time and money, suggesting that, for children in type (5) households, an important portion of the observed skill gap cannot be closed by policies aiming to increase either income or hours resources.

[20] Taken from the NCES, see footnote 14

Finally, Table A.8 presents cluster-level means in other important observable characteristics. The clearest observable pattern offered by this table is that type (1) mothers are younger, have more children, commence fertility at younger ages, score lower on the CDS-administered aptitude test, and have fewer years of education. While type (5) mothers also appear to have higher total births, which may partly explain their similarly low labor force attachment, few other significant empirical differences emerge from this tabulation.

While this section considers differences across groups that are observable, we will return to group differences across latent dimensions after introducing and discussing the model estimates in the next section.

4.2.3 Stage 2: Full Procedure

Figures B.6 through B.9 show the model's fit, for each type, of the annual participation rate in welfare, mean of weekly work hours, mean labor force participation, and mean food stamp participation. In this regard, the model performs reasonably well given the relatively small number of parameters used to fit the time series. Readers may observe that the model tends to "smooth out" otherwise abrupt changes in the time series, but does capture responses to large and salient policy changes, such as the introduction of time limits in 1996.

Tables A.5 and A.6 offers exact estimates of preference, wage, and non-labor income parameters for each type from this second stage. While the estimates confirm the existence of essential heterogeneity across all dimensions, this was in some sense guaranteed by the first stage, which documented such heterogeneity in the raw empirical patterns provided by revealed preference information. I finally note the first row of Table A.5, which reports the classified proportions of each type from the first stage clustering procedure.

Heterogeneity in elasticities across clusters Finally, I use the estimated structural parameters to compute Marshallian elasticities for each classified group. This is achieved by increasing wages in every state by 10% relative to the baseline simulation, and calculating the mean response in work at both the extensive and intensive margin. I find that all groups except (2) and (7) display extensive marginal elasticities that dominate the intensive margin in magnitude, which is consistent with the prior literature. Most strikingly, the structural estimates point to quite significant heterogeneity in elasticities across groups.

Groups (1) and (5) produce the largest elasticities, and Figure B.3 will remind readers that these groups displayed a strong hours response to changes in the policy environment through the 1990s. This is also true of groups (3) and (4), which are predicted to have the next highest elasticities (in terms of point estimates). Types (2) and (7), as Figure B.3 shows, have always had high labor force attachment and high rates of full time employment, despite changes in earnings and policies. Accordingly, estimates predict elasticities close to zero for these groups. Overall, the model produces a weighted average of extensive marginal elasticities of 0.44, with lots of variation around this mean. These results appear to replicate the findings of Attanasio et al. (2018), who find large variation in labor supply elasticities for women.

5 Counterfactual Analysis

With estimates of the technology of skill formation, and estimates of the distribution of maternal preferences, I now have the requisite architecture to explore the impact of counterfactual policy environments. While there is a large literature examining the labor supply and welfare impacts of different antipoverty policy frameworks,[21] we have the unique opportunity in this setting to jointly consider child outcomes in addition to maternal welfare. Furthermore, I will undertake in this section the problem of calculating optimal cash transfer schedules for this population, in the spirit of other studies such as Saez (2002), Blundell and Shephard (2011), and Heathcote et al. (2017). This section is built around two counterfactuals, each designed to answer the following questions. First, what was the impact of the 1996 PRWORA Act and its concomitant changes in federal tax policy on maternal welfare and child skill outcomes? Second, given estimates of skill production and maternal preferences, what is the shape of an optimal transfer policy?

To answer the first question, I design a counterfactual that simulates maternal work, participation, and investment behavior under the assumption that the tax and welfare environment "freezes" in 1996, just before the introduction of reform. To answer the second question, I solve for an optimal "one kink" transfer function, in which a planner seeks to maximize maternal welfare subject to cost neutrality. Further details on each simulation are provided in their respective sections.

When it comes to the effect of counterfactual policies, my attention in this paper is devoted to two outcomes in particular. The first is maternal welfare, the effect on which I report by computing a consumption equivalence measure: the additional dollars in consumption at baseline that would deliver an equivalent change in welfare. I refer to this measure in the remaining sections as a "willingness to pay" for the counterfactual. The second is the final skill outcomes of children at maturity, which I report in the anchored scale of their joint association with the probability of High School graduation. This scaling approach in turn suggests a skill aggregation mechanism, in which one supposes that the probability of High School graduation is a linear function of the two cognitive and one behavioral skill measured in this paper. Accordingly, in discussions below, I refer to the "effect" on High School graduation as the sum of each scaled skill change. It must be emphasised that while the individual skill effects are interpretable as causal, the same cannot be said of this number, which is merely a particular lens through which skill outcomes can be aggregated.

[21] See Chan (2013) for an excellent, modern example and review.

5.1 No Reform Counterfactual

To evaluate the impact of the large, aggregate changes to the cash assistance environment that occurred in the years following welfare reform, I must choose from a menu of imperfect counterfactual hypotheticals. Here I simulate outcomes under the assumption that the welfare and tax environment remained fixed from 1996 until the last child in my sample matures. An important limitation to bear in mind is that I do not consider here the potential general equilibrium effects of counterfactuals policies. One should bear in mind that large policy changes such as this may induce equilibrium adjustments in the labor and marriage markets, and to endogenous fertility choices. My decision to exclude these features from the model was influenced by the fact that allowing for mothers to make choices in the marriage market, and make fertility decisions, would greatly increase the complexity of this analysis and remove much of the empirical transparency. The reader should consult Brown et al. (2015) and Gayle et al. (2015) for a more thorough treatment of these model features, and in this paper be mindful of the biases that such an omission could create. It is worth noting, on the other hand, that some recent studies have found only modest effects, if any, of policy changes on marriage and fertility, and that these results appear quite sensitive to specification (Grogger and Bronars, 2001; Bitler et al., 2004; Kearney, 2004). Low et al. (2018) offer complementary evidence suggesting that welfare reform did have a signiciant impact on marital dissolution decisions. Understanding these issues is indeed a very fruitful direction for future research.

Table A.9 reports the results from this study. We see that the counterfactual costs \$3.88 *less* per mother, per week, than the real set of policies, and results in a statistically insignificant loss in mean welfare. The aggregate result conceals a reasonable degree of heterogeneity however. The only significant beneficiaries under the counterfactual are type (1) mothers, who were the principal recipients of welfare prior to reform. Offsetting their gains are the statistically significant welfare losses of mothers in groups (2), (4), (6), and (7). These groups are characterized either by strong labor force attachment, or a strong response in labor supply to the reform period. Once can more easily understand the difference in welfare effects by simply expecting the changes in cost per mother, per week, shown in column 7. Undoing reform redirects government transfers back toward type (1) mothers, or conversely, we can think of the reform as directing transfers away from this type to others. The welfare changes merely reflect this change in resources. This analysis echoes that of Moffitt (2015) and Hoynes and Schanzenbach (2018), who both study the redistribution of government transfers across observable dimensions in the data. This analysis provides a structural interpretation of this policy change.

Turning to skill outcomes, it appears that there is no significant mean impact on child skills. However, just as in the case of welfare, inspecting impacts by type unveils heterogeneity in skill impacts. Namely, while the children of type (5) mothers appear to enjoy skill improvements under the counterfactual, type (2), (6), and (7) children experience skill losses that are statistically significant. It can be noted that, while signifcant, the magnitudes of these changes are quite modest, with all effect sizes less than a fifth of a percentage point of high school graduation in magnitude.

5.2 Optimal Policy Counterfactual

While the above counterfactual provides some revealing lessons about how mothers have valued recent changes to the welfare landscape, and the potential impacts on skill outcomes, these lessons are incomplete from a normative perspective. Accordingly, I now dedicate my attention to the numerical solution of an optimal policy problem, in which a planner seeks to maximize an equally weighted average of each mother's *ex ante* welfare, measured at the time of their child's birth. Since the model treats fertility as exogenous, I take the simplest permissable fertility process, considering the outcomes of one child only, from birth until maturity.

The planner may choose any transfer policy from the space of continuous, "one kink" policies, which can be defined in terms of four parameters:

$$\mathbf{T}(E) = \mathcal{T}_0 - \tau_0 \min\{E, \mathcal{K}\} - \tau_1 \max\{E - \mathcal{K}, 0\}$$

where $\mathcal{T}_0$ is the payment given to those with no source of income, τ_0 the initial rate of taxation, $\mathcal{K}$ is the "kink point", the point at which the rate of taxation changes, and τ_1 is the second rate of taxation. Thus, they maximize the welfare function with choices of $(\mathcal{T}_0, \tau_0, \mathcal{K}, \tau_1)$ subject to a revenue constraint:

$$\max_{\mathcal{T}_0,\tau_0,\mathcal{K},\tau_1} \sum_k \pi_k \mathbb{E}_0 \sum_{t=0}^{T} \beta^t u_k(c_{kt}, l_{kt}, h_{kt}, \theta_{kt})$$
$$\text{s.t.} \sum_k \pi_k \sum_t \mathbb{E}_0 \mathbf{T}(E_{kt}) = R$$

where $\mathbb{E}_0$ denotes the expectation operator at $t = 0$.

It is worth reviewing the fundamental determinants of the shape of the optimal transfer scheme. Saez (2002) argues that if the extensive marginal elasticity of labor supply is large relative to the intensive margin, then the optimal schedule may initially feature negative marginal tax rates ($\tau_0 < 0$), such as those provided by the EITC. Chon and Laroque (2010) show that heterogeneity in preferences can lead to a similar conclusion. Since both channels are active in my model, it is possible that negative marginal tax rates may be optimal. An additional channel, related to the importance of time and money in the production of child skills, may also be active in this setting. As δ_x becomes large relative to δ_τ, mothers value money increasingly over time, which will ultimately determine how they value work incentives over unconditional transfers.

For lack of a most appropriate choice in revenue constraint, R, for this sample, I construct one by simulating the average fiscal cost per mother, per week, under California's welfare and tax code in the year 2010. As a baseline, I calculate maternal decisions from birth to maturity of a single child, integrating over potential wage paths. I then average over types using the classified proportions,

π_k, from Table A.5. Given my estimates, this amounts to an average expenditure of \$52.72 per household, per week. Accordingly, Table A.10 reports the aggregate and type-level impacts of the optimal policy relative to this baseline. In order to give changes in welfare a measure in units, for each type k I compute a consumption-based equivalent, equal to the additional consumption awarded in each period and state of the world that would deliver the same change in value to the mother as the counterfactual policy. These are reported in the final column of Table A.10.

The solution to the planner's problem is depicted in Figure B.11, along with the implied transfer schedules for a mother of one child in California in 2010, with and without welfare participation. In contrast with the baseline set of policies, there is no self-selection into the transfer schedule, which necessitates that it be less generous than any comparison baseline. The policy dictates a weekly payment of \$97 per week, which is phased out with other income at a rate of 55%. At \$176 in weekly earned income, the tax rate changes to 40%. Readers may note that this rate is high, and is a function of the new program's universal nature, which expands the government's liabilities to a broader population than typical welfare participants.

Turning to Table A.10, we see that the welfare gains from the policy are equivalent, on average, to \$80 in additional consumption in every period. The size of the welfare gain is in part due to the creation of an income support program that can be administered without self-selection and participation costs: I assume here that the new program can be delivered in the same manner as the EITC, through the extant income tax apparatus. To partially account for this issue, the welfare calculations I report here do not count the heterogeneous utility cost borne by welfare participants in the baseline case.[22] The broad pattern of results from this counterfactual mirrors those from the previous exercise. Faced with a lesser incentive to work, mothers reduce their labor supply, which leads to a reduction in net income and an increase in hours at home. The net effect on child outcomes is positive and significant, with a point estimate of a 1.29 percentage point increase in the rate of high school graduation.

Rows 2 through 8 of Table A.10 reveal that the aggregate story masks a good deal of heterogeneity in the response to this change in incentives. While all types reduce, on average, their labor supply as a result of the new policy, notice that net income increases for type (5), which in part explains their strong consumption equivalent valuation of the policy change. Accordingly, children of these households enjoy the greatest gains in skills, to the order of 5.6 percentage points. Revisiting Figure B.5, which shows average skill outcomes for children across each group, a gain of this magnitude would reduce the graduation gap between Group 5 children and the rest of the sample by about 25%.

It is essential to note that this policy change is not Pareto improving: types(2) and (7) endure

[22]One should note that this does not completely account for this issue, however, and that more work is needed to assess the source of these utility costs, which are a common (and empirically crucial) modeling device for welfare participation. See Moffitt (1983), Hoynes (1996), and Chan (2013) for three other examples in the literature.

statistically significant losses in welfare, while types (4) and (5) enjoy the greatest gains in welfare from the policy. In the baseline, these types exhibited low welfare dependance, and estimates indicated that in part this was due to participation costs. As a result, the universal program redirects public funds toward these groups (on average about \$34 and \$88 more per week), which drives my welfare findings. In particular, group (5) experiences both an increase in net income as well as a reduction in work hours, which drives the large positive impacts on children in these households.

5.3 Other Counterfactuals

5.3.1 Maximizing Skill Outcomes

While the traditional objective of a planner in normative policy analysis is to maximize some weighted average of welfare, one might wonder what the implications for policy are if the planner is tasked with maximizing skill outcomes. Using the aggregate of High School Graduation (HSG), I use the estimated coefficients in the anchoring equation (4.1) and assume that the new objective of the planner is:

$$\max_{\mathcal{T}_0,\tau_0,\mathcal{K},\tau_1} \sum_k \pi_k \mathbb{E}[HSG_k]$$

subject to the same revenue constraint as the original problem. While the tradeoffs for this maximization problem are similar to before, the planner now disregards how mothers themselves value time and money, and instead focus on the relative importance of these resources for the final skill outcomes of each mother's child.

Figure B.12 shows the optimal schedule for skill outcomes, while Table A.11 shows the average and type-level impacts of the policy. In constrast to the welfare-maximizing policy, the planner now favors an initial negative marginal tax rate, effectively providing a work subsidy to poor families. The universal payment is \$43 a week, which increases at a rate of 60%, until earnings reach \$64/week. This amounts to a maximum weekly credit of \$81, or just over \$4000 per year. From this point, the credit is clawed back at a rate of 20% and workers' earnings are eventually taxed at the same rate. This result suggests that even though time has (at most ages) a larger share in the production of skill outcomes, a negative income tax leads mothers to work too little relative to the planner's new objective.

While there is no statistically significant average impact on hours or net income, the mean welfare gain is equivalent to \$21.54 in additional consumption every week. Types (2) and (5) experience significant increases in weekly net income, while types (1), (4), and (5) undergo significant changes in labor supply. As before, groups (4) and (5) are the main beneficiaries of the new policy (in terms of welfare), but the gains to children are more evenly distributed. While in the welfare maximizing policy, only children in group (5) experienced a significant increase in skill outcomes, in the skill-maximizing policy all children except those in type (1) households exhibit statistically significant skill gains. My point estimate for the average increase in high school graduation is 1.87 percentage points.

5.3.2 Optimal Policy by Type

To further unpack the implications of preference heterogeneity, I solve for the optimal policy for each type individually, under the same budget constraint. The schedules of these individually optimal policies are presented in Figure B.13, and they show large variation in the optimal shape of the transfer schedule. While most types individually appear to prefer negative marginal tax rates, the optimal schedule for type (7) dictates a modest but negative initial tax credit. Across types, optimal schedules appear different because of a combination of differences in preferences for leisure, preferences for child outcomes, costs of labor force participation, and underlying wage processes. While traditional analyses of optimal tax problems assume homogeneous preferences, my revealed preference analysis of mothers in this particular panel of data suggested fundamental differences in behavior, which delivers direct implications for optimal policy, as well as trade-offs for a planner that must design programs that cater to different tastes. Since the space of policies I consider here is relatively primitive, there appears to be little scope for policymakers to escape this trade-off through a more sophisticated mechanism, however this is a promising direction for future research.

5.3.3 Optimal Policy without Children

Finally, Figure B.14 compares the original welfare-maximizing scheme to one in which mothers have no children. In order to maintain a reasonable comparison to the baseline case, I simulate this by setting $\alpha_\theta = 0$ for each mother in the baseline calculation. It appears that the optimal schedule when child development is ignored is not drastically different from the baseline case. The fact that the transfer schedule in this case is less generous than the baseline is a function of the fact that mothers' reduction in labor supply is more dramatic without children than with one. It is unclear whether such a comparison is consistent across a richer distribution of fertility histories, so I simply use this exercise to note that the addition of children to the problem has discernible implications for labor supply behavior and optimal policy.

6 Conclusion

In this paper I develop a model in which mothers work, participate in welfare, and raise children by solving a dynamic investment problem. Investments of time and money produce cognitive and behavioral skills, and so cash transfers influence child outcomes by shaping maternal incentives to work and the household budget. The model I propose is analytically tractable, allowing me to establish a tight link with data. I use this to estimate the model by deriving orthogonality restrictions from exogenous policies, and a control function condition. Simultaneously, I classify mothers into latent types using the revealed preference information contained in their labor supply and program participation decisions. The estimates imply that the design of antipoverty programs has important implications for both maternal welfare and child outcomes. These are borne out by two counterfactuals, the substance of which suggests that a shift in policies towards earnings subsidies has had no significant impact on aggregate welfare, but significantly redistributed welfare (as measured by consumption equivalence) away from mothers with high welfare dependance and low labor force attachment. In evaluating these results, it is important to consider two tradeoffs presented. First, the welfare-maximizing program and the program that maximizes skill are quite different. The problem of deciding on how to balance these goals (perhaps by placing a dollar valuation on skill output) is a difficult one, but worthy of attention in future work. Second, the model unveils rich heterogeneity, even among a small number of types, in preferences. These are influential in determining who wins and who loses in the optimal policy relative to the baseline, and also in determining the shape of optimal transfer schedules. Policies that address this underlying heterogeneity would seem most promising. One important future exercise would be to determine a weighting that delivers a pareto-improving optimal policy, if one can be found, in the spirit of Blundell and Shephard (2011).

Finally, some important caveats must be applied to this work, and these in turn suggest fruitful directions for future research. In particular, many of the results hinge on the importance of maternal time for developmental outcomes, and does not permit the use of substitutes for this investment, such as childcare. Of equal potential importance are considerations regarding the marital and fertility decisions of single mothers. A modeling approach that featured a broader population of household types, as well as the self-selection between demographic categories, would add important weight to the counterfactual analysis. As it stands, we can consider only the results as they apply to the estimated population, without considering the potential for adopting a truly universal policy reform.

A Tables

Table A.1: Descriptive Statistics

	Value	Std. Dev
Mother		
Earnings (25th percentile)	0.00	-
Earnings (50th percentile)	520.55	-
Mean Earnings	897.91	1168.37
Wage (25th percentile)	5.28	-
Wage (50th) percentile)	8.13	-
Mean Wage	9.63	7.17
Mean Net Income	2438.19	1757.90
Cash Assets (50th percentile)	0.00	-
Cash Assets (75th percentile)	785.36	-
Cash Assets (90th percentile)	2890.89	-
Used Welfare Once	0.49	-
Used Welfare $\geq$ 5 years	0.17	-
$<$ High School	0.26	-
High School	0.63	-
College	0.10	-
Mean No. children	2.85	1.34
Mean Panel length	14.30	4.70
N	959	-
Child		
Graduate High School	0.74	
No Formal Care	0.87	
No Formal Care (Age $\leq$ 6)	0.83	
N	1405	

All dollar amounts are deflated to 1997 values, with income variables reported in monthly averages. When applicable, standard deviations of variables are reported in the right hand column. Cash Assets are collected at the household level from the 1999 Wealth Module, while Childcare time is measured using the 1997 time diary.

Table A.2: Estimates of anchoring equations

	Specification (1)	(2)	(3)	(4)	(5)
LW	0.0129 [0.011, 0.015]				0.0124 [0.008, 0.0017]
AP		0.0122 [0.010, 0.014]			0.0044 [0.0022 0.017]
BPE			-0.0232 [-0.027, -0.019]		-0.0215 [-0.029, -0.014]
BPN				-0.0193 [-0.024, -0.014]	0.0014 [-0.008, 0.011]

Parameters are estimated using a linear probability model, with 95% confidence intervals shown. These estimates are calculated the full sample of PSID-CDS children with available data on scores and high school completion.

Table A.3: Production Parameter Estimates - Cobb Dougls Shares

	LW	AP	BE
$\gamma_{\delta,x,0}$ (Intercept)	0.69	0.08	-0.36
	[0.20 ,1.18]	[-0.06 ,0.22]	[-1.00 ,0.28]
$\gamma_{\delta,x,1}$ (Slope)	-0.03	0.00	0.05
	[0.03 ,-0.10]	[-0.00 ,0.01]	[-0.05 ,0.15]
$\gamma_{\delta,\tau,0}$ (Intercept)	1.61	0.57	0.06
	[0.52 ,2.70]	[0.23 ,0.92]	[-1.00 ,1.11]
$\gamma_{\delta,\tau,1}$ (Slope)	-0.10	-0.06	-0.04
	[-0.20 ,-0.00]	[-0.10 ,-0.03]	[-0.15 ,0.08]
$\delta_{\theta,LW}$	0.88	0.02	0.00
	[0.86 ,0.90]	[0.01 ,0.02]	[-0.03 ,0.03]
$\delta_{\theta,AP}$	0.05	0.87	0.01
	[-0.04 ,0.13]	[0.84 ,0.89]	[-0.10 ,0.11]
$\delta_{\theta,BE}$	0.00	0.00	0.90
	[-0.01 ,0.02]	[-0.00 ,0.01]	[0.88 ,0.92]

This table shows estimates of the parameters of the technology of skill formation for Letter-Word score (LW), Applied Problems (AP) and Externalizing Behaviors (BE), using the iterative GMM procedure described in the text. The parameters $\gamma_{s,0}$ and $\gamma_{s,1}$ report the estimated slope and intercept, respectively, for $s \in \{x, \tau\}$ as in equations (4.2) and (4.3). 95% confidence intervals are shown, calculated using the standard asymptotic formula.

Table A.4: Production Parameter Estimates - Controls

	LW	AP	BE
$\gamma_{K,1,0}$ (Intercept)	4.01	-1.25	0.43
	[-13.11 ,21.12]	[-6.88 ,4.37]	[-11.91 ,12.78]
$\gamma_{K,1,1}$ (Slope)	0.54	0.63	-0.42
	[-0.96 ,2.04]	[0.12 ,1.15]	[-2.13 ,1.29]
$\gamma_{K,2,0}$ (Intercept)	-0.16	2.51	1.06
	[-16.30 ,15.99]	[-2.91 ,7.92]	[-12.24 ,14.35]
$\gamma_{K,2,1}$ (Slope)	0.96	0.32	-0.39
	[-0.47 ,2.39]	[-0.17 ,0.82]	[-2.12 ,1.33]
$\gamma_{K,3,0}$ (Intercept)	3.36	-0.02	-1.36
	[-13.69 ,20.40]	[-5.92 ,5.87]	[-13.68 ,10.97]
$\gamma_{K,3,1}$ (Slope)	0.82	0.54	-0.17
	[-0.69 ,2.32]	[0.01 ,1.07]	[-1.84 ,1.51]
$\gamma_{K,4,0}$ (Intercept)	-1.10	-0.32	-1.30
	[-16.67 ,14.47]	[-5.62 ,4.97]	[-12.99 ,10.39]
$\gamma_{K,4,1}$ (Slope)	1.02	0.56	-0.20
	[-0.40 ,2.45]	[0.07 ,1.05]	[-1.85 ,1.44]
$\gamma_{K,5,0}$ (Intercept)	1.82	-0.40	-0.67
	[-15.35 ,18.98]	[-6.25 ,5.45]	[-15.09 ,13.74]
$\gamma_{K,5,1}$ (Slope)	0.85	0.60	-0.35
	[-0.69 ,2.39]	[0.07 ,1.14]	[-2.16 ,1.46]
$\gamma_{K,6,0}$ (Intercept)	0.97	-0.54	-0.50
	[-14.92 ,16.86]	[-5.98 ,4.91]	[-12.40 ,11.39]
$\gamma_{K,6,1}$ (Slope)	0.87	0.57	-0.20
	[-0.57 ,2.31]	[0.06 ,1.07]	[-1.87 ,1.46]
$\gamma_{K,7,0}$ (Intercept)	2.05	-0.93	-0.68
	[-13.57 ,17.68]	[-6.38 ,4.52]	[-12.76 ,11.40]
$\gamma_{K,7,1}$ (Slope)	0.86	0.63	-0.13
	[-0.56 ,2.28]	[0.12 ,1.13]	[-1.80 ,1.53]
Dummy - 1997	-32.60	-7.89	-3.64
	[-34.98 ,-30.22]	[-8.64 ,-7.14]	[-6.51 ,-0.76]
Dummy - 2002	3.83	0.98	0.14
	[3.07 ,4.60]	[0.72 ,1.23]	[-0.90 ,1.19]

This table shows estimates of the parameters of the technology of skill formation for Letter-Word score (LW), Applied Problems (AP) and Externalizing Behaviors (BE), using the iterative GMM procedure described in the text. $\gamma_{K,k,0}$ and $\gamma_{K,k,1}$ represent the intercept and age coefficient for children of mothers of type k. 95% confidence intervals are shown, calculated using the standard asymptotic formula.

Table A.5: Second Stage Estimates: Preferences

Type	1	2	3	4	5	6	7
				Proportions			
π_k	0.10	0.30	0.17	0.18	0.10	0.07	0.08
				Preferences			
α_l	4.09	0.66	18.29	0.83	9.97	11.25	0.83
	[2.67 ,8.10]	[0.56 ,0.74]	[15.27 ,37.41]	[0.63 ,1.03]	[8.70 ,20.66]	[10.38 ,21.95]	[0.77 ,1.05]
α_θ	10.71	2.79	48.15	3.58	40.58	32.90	2.25
	[4.50 ,13.18]	[1.41 ,3.77]	[47.84 ,49.97]	[1.76 ,4.66]	[32.65 ,48.38]	[28.16 ,40.87]	[1.38 ,3.54]
α_F	1.28	0.32	1.02	0.18	1.36	25.51	0.57
	[0.89 ,2.23]	[0.21 ,0.47]	[0.36 ,1.38]	[0.13 ,0.25]	[0.60 ,1.62]	[20.79 ,40.52]	[0.41 ,0.89]
α_A	0.53	0.20	1.17	0.15	0.73	0.90	0.25
	[0.27 ,0.70]	[0.14 ,0.26]	[0.79 ,1.69]	[0.10 ,0.18]	[0.02 ,0.90]	[0.63 ,1.40]	[0.15 ,0.37]
$\alpha_{h,0}$	3.54	4.02	8.04	3.03	7.62	9.74	8.18
	[3.27 ,3.92]	[3.20 ,4.40]	[5.69 ,9.35]	[2.74 ,3.56]	[6.28 ,9.47]	[7.01 ,11.66]	[7.05 ,8.81]
$\alpha_{h,1}$	-0.31	-0.59	-0.61	-0.37	-0.72	-0.87	-0.77
	[-0.37 ,-0.26]	[-0.67 ,-0.50]	[-0.77 ,-0.55]	[-0.45 ,-0.33]	[-0.84 ,-0.55]	[-0.93 ,-0.79]	[-0.83 ,-0.71]

This table reports estimates from the second stage estimation procedure. Parameters are as described in the text.

Table A.6: Second Stage Estimates: Wages

Type	1	2	3	4	5	6	7
				Proportions			
π_k	0.10	0.30	0.17	0.18	0.10	0.07	0.08
				Wages			
Constant ($\gamma_{w,0,k}$)	0.49	1.60	0.50	0.98	1.46	1.06	1.09
	[0.19 ,0.85]	[1.47 ,1.78]	[0.16 ,0.86]	[0.85 ,1.24]	[1.16 ,2.33]	[0.49 ,1.37]	[0.82 ,1.51]
Age ($\gamma_{w,a,k}$)	0.04	0.02	0.04	0.04	0.04	0.06	0.03
	[0.03 ,0.06]	[0.01 ,0.03]	[0.03 ,0.05]	[0.03 ,0.05]	[0.04 ,0.07]	[0.04 ,0.07]	[0.02 ,0.04]
$\sigma_{w,k}$	0.84	0.61	0.67	0.65	0.64	0.86	0.80
	[0.73 ,0.94]	[0.57 ,0.64]	[0.62 ,0.81]	[0.60 ,0.70]	[0.57 ,0.75]	[0.69 ,1.01]	[0.67 ,0.92]
$\rho_{w,k}$	0.48	0.58	0.84	0.62	0.50	0.46	0.50
	[0.36 ,0.54]	[0.52 ,0.67]	[0.44 ,0.87]	[0.50 ,0.76]	[0.36 ,0.70]	[0.26 ,0.60]	[0.34 ,0.61]

This table reports estimates from the second stage estimation procedure. Parameters are as described in the text. Brackets show 95% confidence intervals, computed by bootstrap with 100 resamples.

Table A.7: Second Stage Model Fit

Moment	Source	Type 1	2	3	4	5	6	7
$\phi_{\tau,1997}$	Data	0.38	0.44	0.39	0.41	0.37	0.40	0.38
	Model	0.35	0.46	0.47	0.39	0.30	0.41	0.40
$\phi_{\tau,2002}$	Data	0.25	0.33	0.32	0.31	0.33	0.32	0.33
	Model	0.28	0.29	0.32	0.32	0.24	0.32	0.26
$\hat{\gamma}_{w,k,0}$	Data	1.44	1.49	1.08	1.53	1.40	1.39	1.54
	Model	1.47	1.51	1.02	1.55	1.41	1.43	1.53
$\hat{\gamma}_{w,k,a}$	Data	0.01	0.02	0.03	0.02	0.01	0.02	0.01
	Model	0.02	0.02	0.02	0.00	-0.02	0.02	0.01
$\mathbb{C}(\hat{\epsilon}_{W,t+1}, \hat{\epsilon}_{W,t})$	Data	0.10	0.16	0.29	0.20	0.13	0.19	0.14
	Model	0.13	0.19	0.32	0.20	0.05	0.22	0.12
$\mathbb{V}(\hat{\epsilon}_{W,t})$	Data	0.62	0.39	0.58	0.52	0.67	0.40	0.33
	Model	0.32	0.38	0.42	0.23	0.12	0.23	0.34

This table reports the model's fit of the statistics used in the Second Stage Estimation procedure, in addition to those shown in Figures B.6 - B.9. The variables $\hat{\gamma}$ and $\hat{\epsilon}$ refer to estimated coefficients and residuals from the regression: $\log(W_{it}) = \hat{\gamma}_{W,k,0} + \hat{\gamma}_{W,k,a}\text{Age}_{it} + \hat{\epsilon}_{it}$. Due to selection bias on wages, these coefficients are not themselves consistent estimates of the corresponding structural wage parameters.

Table A.8: Cluster Characteristics

	Age '97	Age 1st Birth	PC	Tot. Births	Years Educ.
$k = 1$	28.10	18.88	25.70	3.47	13.00
	(0.72)	(0.33)	(0.64)	(0.16)	(1.38)
$k = 2$	34.08	20.58	28.93	2.98	16.95
	(0.77)	(0.49)	(0.76)	(0.16)	(2.51)
$k = 3$	30.10	19.69	28.46	2.73	15.80
	(0.88)	(0.47)	(0.61)	(0.19)	(2.13)
$k = 4$	28.16	19.96	28.50	2.69	16.40
	(0.44)	(0.21)	(0.41)	(0.07)	(1.21)
$k = 5$	28.24	19.63	28.59	3.36	15.36
	(0.56)	(0.30)	(0.50)	(0.13)	(1.67)
$k = 6$	30.29	20.44	29.31	2.74	17.32
	(0.50)	(0.33)	(0.44)	(0.10)	(1.54)
$k = 7$	32.39	21.10	28.36	2.56	16.63
	(0.72)	(0.51)	(0.66)	(0.12)	(2.13)

This table presents summary satistics for each cluster group. Columns report: (1) Age of mother in 1997; (2) Age of mother in year of first birth; (3) Mother's raw score from the Passage Comprehension module of the Woodcock-Johnson aptitude test; (4) Mother's total number of births recorded in the PSID; and (5) Mother's total years of education in 1997. Standard errors for each statistic are shown in parentheses. Brackets show 95% confidence intervals, computed by bootstrap with 100 resamples.

Table A.9: Impacts of the No Reform Counterfactual

	ΔHSG	ΔLW	ΔAP	ΔBPE	ΔY	ΔH	Cost	Cons. Value
Overall	-0.02	0.02	-0.00	-0.03	4.94	0.21	-3.88	-0.28
	[-0.09 , 0.04]	[-0.02 , 0.06]	[-0.02 , 0.01]	[-0.07 , 0.00]	[2.15 , 9.00]	[-0.52 , 1.30]	[-6.69 , -1.90]	[-0.92 , 0.45]
$k = 1$	0.07	0.15	0.01	-0.08	-8.80	-5.24	18.98	15.93
	[-0.37 , 0.45]	[-0.12 , 0.31]	[-0.05 , 0.06]	[-0.28 , 0.06]	[-18.86 , 11.56]	[-7.56 , -2.18]	[14.25 , 25.09]	[11.14 , 19.68]
$k = 2$	-0.06	-0.01	-0.01	-0.03	18.98	2.47	-14.70	-3.42
	[-0.10 , -0.01]	[-0.04 , 0.03]	[-0.02 , -0.00]	[-0.06 , -0.01]	[12.57 , 36.40]	[1.71 , 3.42]	[-25.76 , -8.70]	[-3.88 , -2.43]
$k = 3$	0.09	0.08	0.02	-0.01	9.85	0.13	-3.54	-0.20
	[-0.02 , 0.20]	[0.01 , 0.13]	[-0.00 , 0.03]	[-0.05 , 0.03]	[4.49 , 13.24]	[-0.89 , 1.45]	[-7.16 , 1.54]	[-1.39 , 1.63]
$k = 4$	-0.11	-0.04	-0.02	-0.05	-1.23	0.52	-5.82	-2.37
	[-0.18 , -0.03]	[-0.08 , 0.02]	[-0.03 , -0.00]	[-0.09 , -0.01]	[-4.00 , 2.09]	[-0.37 , 1.39]	[-7.26 , -3.22]	[-2.58 , -1.64]
$k = 5$	0.17	0.10	0.03	0.04	4.71	-0.07	4.98	0.06
	[0.02 , 0.43]	[0.02 , 0.21]	[0.00 , 0.07]	[-0.01 , 0.15]	[1.13 , 8.13]	[-0.78 , 0.33]	[1.01 , 9.26]	[-0.29 , 0.62]
$k = 6$	-0.08	-0.02	-0.01	-0.04	11.43	1.39	-12.86	-3.66
	[-0.15 , -0.02]	[-0.07 , 0.01]	[-0.03 , -0.00]	[-0.09 , -0.01]	[-0.06 , 22.84]	[-0.81 , 2.90]	[-20.02 , -5.74]	[-4.39 , -2.17]
$k = 7$	-0.08	-0.03	-0.01	-0.03	12.08	1.93	-11.10	-3.57
	[-0.11 , -0.03]	[-0.05 , 0.00]	[-0.02 , -0.01]	[-0.06 , -0.01]	[6.01 , -0.03]	[1.45 , 3.09]	[-19.58 , -6.13]	[-4.02 , -2.49]

This table shows the aggregate impact of "undoing" welfare reform, as described in the text, as well as the average impact by latent type. Letter-Word (LW), Applied Problems (AP), and Externalizing Behaviors (BPE) are all scaled by their association with the probability of graduating high school (in percentage points), while the column HSG reports the total effect. ΔY and ΔH display changes in net household income and hours worked per week. "Cost" reports the change in average weekly cost per mother, and the column "Cons. Value" reports the change in welfare, as measured by equivalent additional consumption in every period of the simulation. Brackets show 95% confidence intervals, computed by bootstrap with 100 resamples.

Table A.10: Impacts: Optimal Policy

	ΔHSG	ΔLW	ΔAP	ΔBPE	ΔY	ΔH	Cost	Cons. Value
Overall	1.29	0.47	0.21	0.62	-46.80	-12.05	0.00	82.02
	[0.31 , 2.29]	[-0.24 , 1.40]	[0.02 , 0.43]	[-0.09 , 1.33]	[-58.40 , -38.19]	[-14.55 , -9.39]	[-26.77 , 17.06]	[53.44 , 122.16]
$k = 1$	0.08	-0.24	-0.03	0.36	-57.31	-7.02	-5.18	-1.51
	[-1.09 , 0.99]	[-0.96 , 0.50]	[-0.26 , 0.17]	[-0.21 , 0.88]	[-78.88 , -37.81]	[-10.10 , -4.60]	[-20.07 , 3.33]	[-8.94 , 2.63]
$k = 2$	0.80	-0.08	0.11	0.77	-85.75	-5.00	-90.07	-11.69
	[-0.71 , 2.19]	[-1.04 , 0.91]	[-0.18 , 0.39]	[-0.24 , 1.71]	[-93.03 , -71.44]	[-11.66 , 1.75]	[-147.38 , -45.90]	[-18.64 , -6.43]
$k = 3$	0.86	0.23	0.11	0.52	-51.90	-14.41	-7.06	44.14
	[-0.43 , 2.41]	[-0.88 , 1.51]	[-0.19 , 0.42]	[-0.15 , 1.20]	[-74.33 , -17.58]	[-19.07 , -10.59]	[-35.98 , 26.69]	[3.19 , 92.21]
$k = 4$	0.47	0.23	0.03	0.21	-57.22	-19.56	34.57	152.28
	[-0.71 , 2.17]	[-0.71 , 1.62]	[-0.17 , 0.36]	[-0.21 , 0.64]	[-74.10 , -36.09]	[-22.05 , -15.63]	[14.43 , 50.56]	[65.82 , 272.70]
$k = 5$	5.65	3.10	1.10	1.46	66.72	-9.52	88.04	254.66
	[3.59 , 8.58]	[1.73 , 4.34]	[0.66 , 1.63]	[0.51 , 2.67]	[47.61 , 78.44]	[-13.73 , -5.33]	[79.79 , 91.00]	[148.90 , 300.00]
$k = 6$	0.59	-0.15	0.06	0.68	-79.10	-9.85	-46.50	-9.18
	[-1.49 , 2.32]	[-1.33 , 1.21]	[-0.31 , 0.41]	[-0.44 , 1.63]	[-121.25 , -66.17]	[-18.82 , -4.00]	[-102.16 , -5.35]	[-17.63 , 13.73]
$k = 7$	0.70	-0.09	0.09	0.70	-75.18	-6.21	-66.07	-11.11
	[-0.68 , 2.37]	[-0.94 , 1.14]	[-0.17 , 0.40]	[-0.25 , 1.64]	[-90.95 , -66.04]	[-13.78 , 0.11]	[-123.95 , -26.04]	[-18.99 , -5.28]

This table shows the aggregate impact of the welfare maximizing policy, as described in the text, relative to the California 2010 policy environment. Average impacts by latent type are also shown. Letter-Word (LW), Applied Problems (AP), and Externalizing Behaviors (BPE) are all scaled by their association with the probability of graduating high school (in percentage points), while the column HSG reports the total effect. ΔY and ΔH display changes in net household income and hours worked per week. "Cost" reports the change in average weekly cost per mother, and the column "Cons. Value" reports the change in welfare, as measured by equivalent additional consumption in every period of the simulation. Brackets show 95% confidence intervals, computed by bootstrap with 100 resamples.

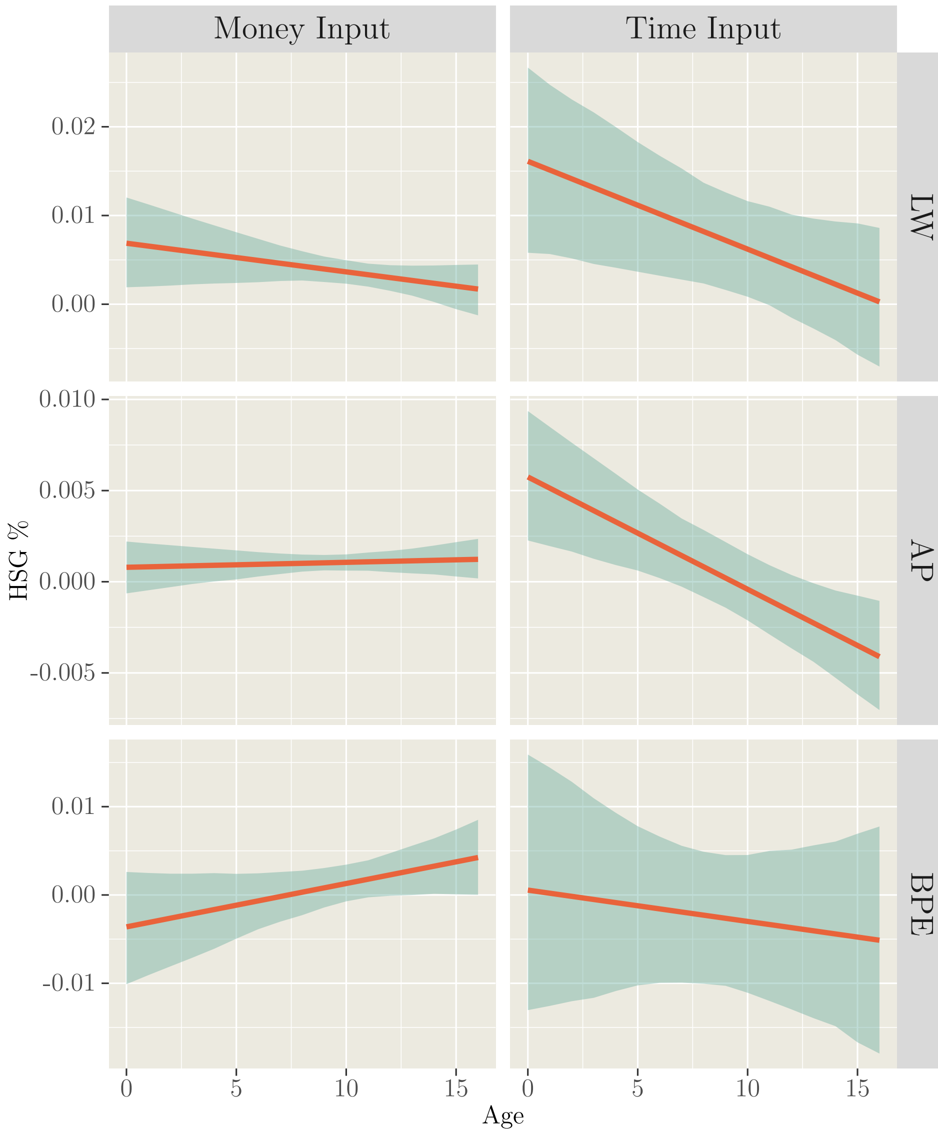
Table A.11: Impacts: Skill Maximizing Policy

	ΔHSG	ΔLW	ΔAP	ΔBPE	ΔY	ΔH	Cost	Cons. Value
Overall	1.87	0.52	0.32	1.04	7.12	-1.66	0.00	21.54
	[0.48 , 3.43]	[-0.07 , 1.38]	[0.07 , 0.56]	[0.01 , 2.04]	[-11.89 , 21.69]	[-5.65 , 1.12]	[-6.50 , 6.37]	[15.78 , 33.58]
$k = 1$	1.09	-0.06	0.17	0.98	-4.25	4.60	-27.62	-11.80
	[-0.27 , 2.99]	[-0.60 , 1.03]	[-0.11 , 0.43]	[-0.10 , 1.97]	[-25.40 , 22.71]	[1.69 , 9.97]	[-33.09 , -22.70]	[-15.46 , -8.83]
$k = 2$	2.06	0.47	0.35	1.23	17.38	0.40	-12.32	-1.76
	[0.36 , 3.68]	[-0.39 , 1.46]	[0.03 , 0.65]	[0.01 , 2.43]	[0.24 , 59.70]	[-2.21 , 4.97]	[-25.13 , -2.25]	[-9.63 , 5.82]
$k = 3$	1.59	0.29	0.26	1.05	3.27	-0.37	-6.41	10.85
	[0.04 , 3.22]	[-0.38 , 1.47]	[-0.02 , 0.54]	[-0.00 , 2.02]	[-11.98 , 15.03]	[-5.69 , 3.17]	[-16.17 , 19.64]	[-3.70 , 41.86]
$k = 4$	1.35	0.29	0.21	0.85	-9.96	-5.15	7.37	36.35
	[0.04 , 2.81]	[-0.45 , 1.43]	[-0.12 , 0.51]	[-0.12 , 1.66]	[-45.37 , 0.80]	[-13.00 , -0.46]	[-6.28 , 19.84]	[18.61 , 66.25]
$k = 5$	3.28	1.69	0.62	0.96	22.54	-6.87	37.65	78.37
	[1.82 , 4.94]	[1.02 , 2.47]	[0.34 , 0.94]	[0.26 , 1.88]	[5.94 , 32.39]	[-12.10 , -3.59]	[31.48 , 41.07]	[54.96 , 110.68]
$k = 6$	2.03	0.44	0.34	1.25	22.27	1.91	-12.22	-1.67
	[0.54 , 4.08]	[-0.33 , 1.53]	[0.05 , 0.68]	[-0.01 , 2.48]	[-11.06 , 47.56]	[-4.15 , 4.72]	[-26.46 , 3.32]	[-9.75 , 11.07]
$k = 7$	2.01	0.48	0.34	1.19	12.97	-0.22	-8.43	2.10
	[0.50 , 4.01]	[-0.28 , 1.61]	[0.04 , 0.66]	[0.03 , 2.42]	[-9.90 , 45.94]	[-4.81 , 2.90]	[-22.68 , 4.18]	[-6.11 , 12.32]

This table shows the aggregate impact of the skill maximizing policy, as described in the text, relative to the California 2010 policy environment. Average impacts by latent type are also shown. Letter-Word (LW), Applied Problems (AP), and Externalizing Behaviors (BPE) are all scaled by their association with the probability of graduating high school (in percentage points), while the column HSG reports the total effect. ΔY and ΔH display changes in net household income and hours worked per week. "Cost" reports the change in average weekly cost per mother, and the column "Cons. Value" reports the change in welfare, as measured by equivalent additional consumption in every period of the simulation.

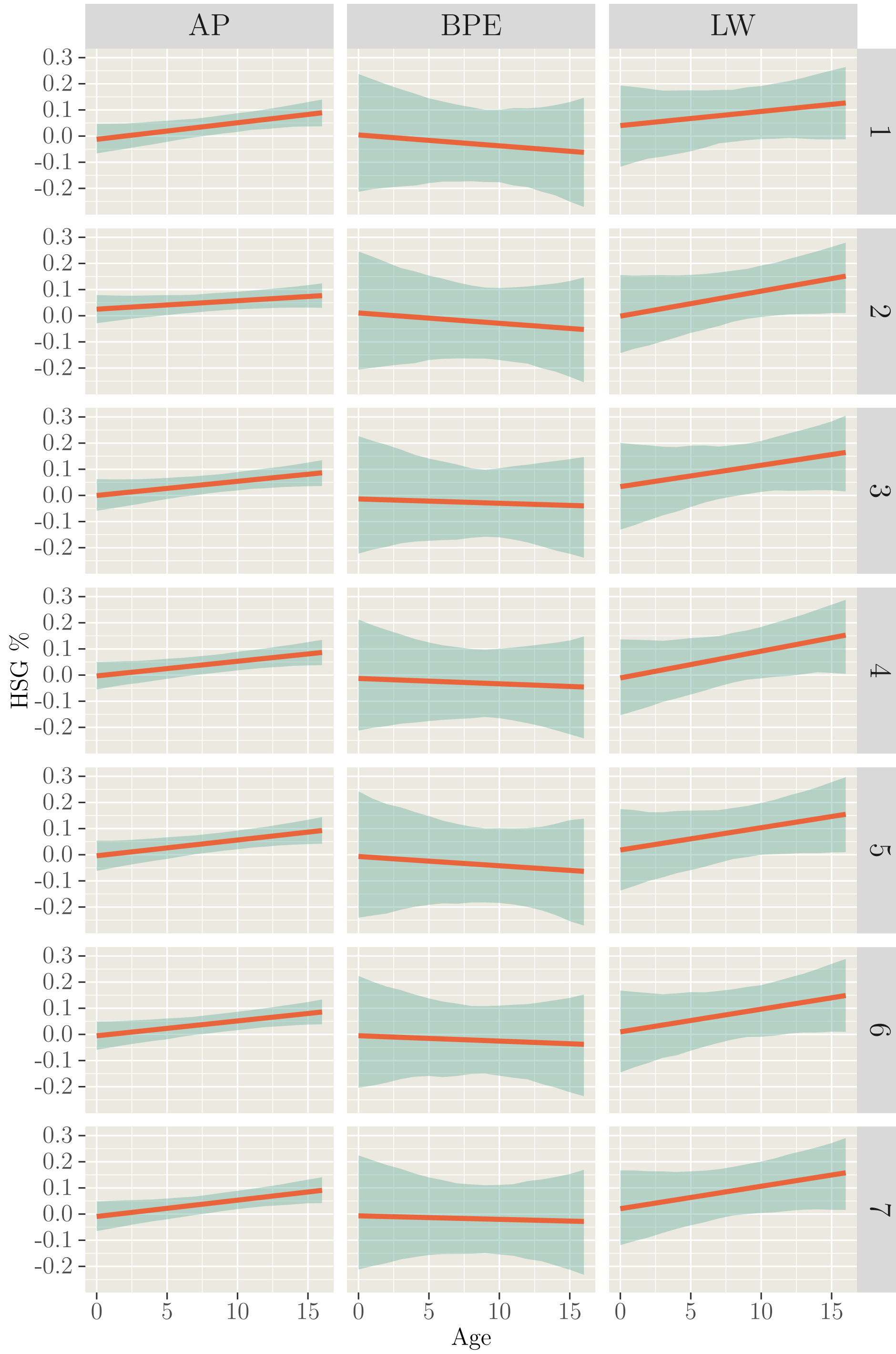
B Figures

Figure B.1: Estimates of Cobb-Douglas Shares



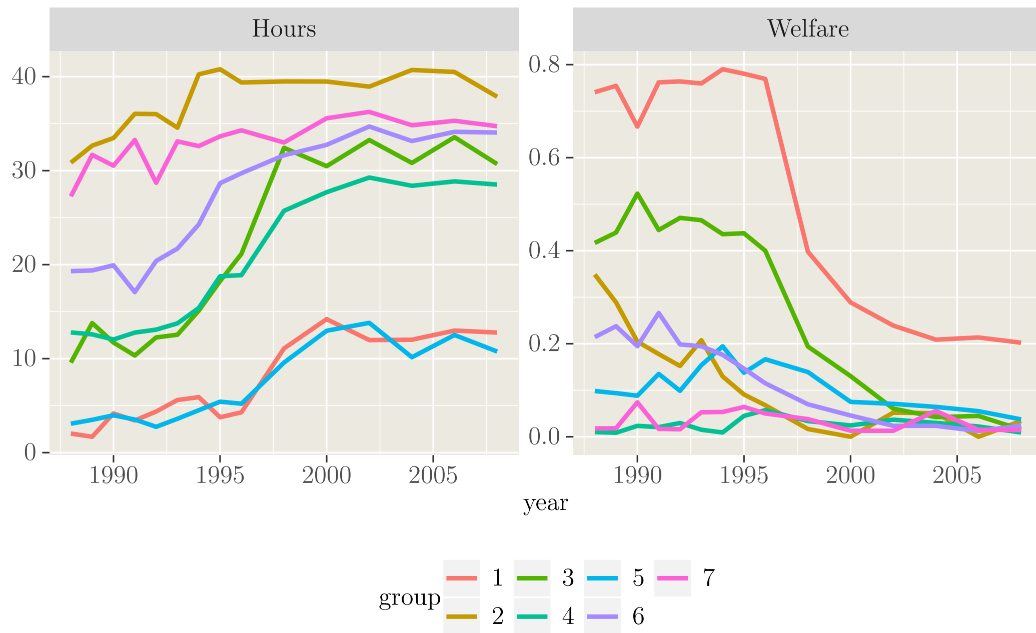
This figure shows estimates of δ_x and δ_τ for the GMM procedure. 95% confidence error bars shown.

Figure B.2: Estimates of Type Fixed Effects



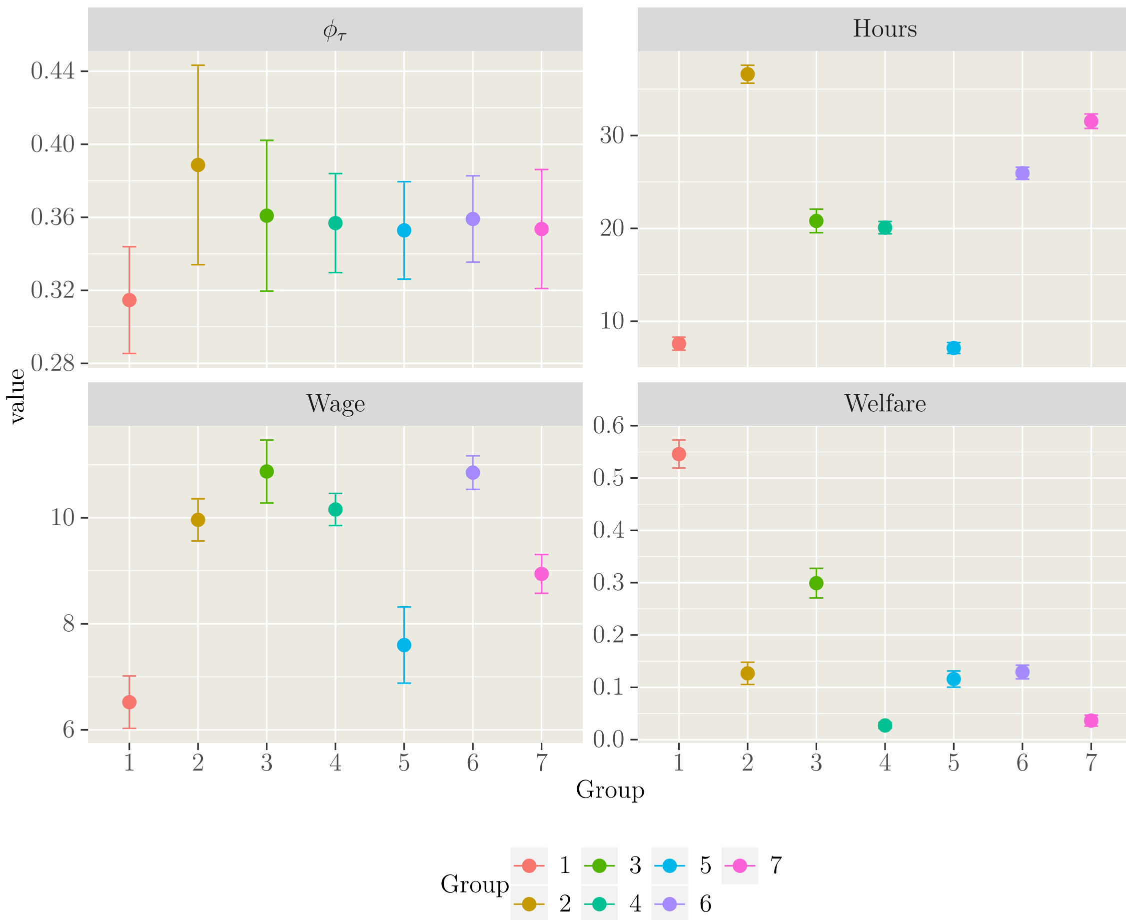
This figure shows estimates of the type and age-specific group effects in production for each age a and type k across each skill. 95% confidence error bars shown.

Figure B.3: Annual Means by Classified Group



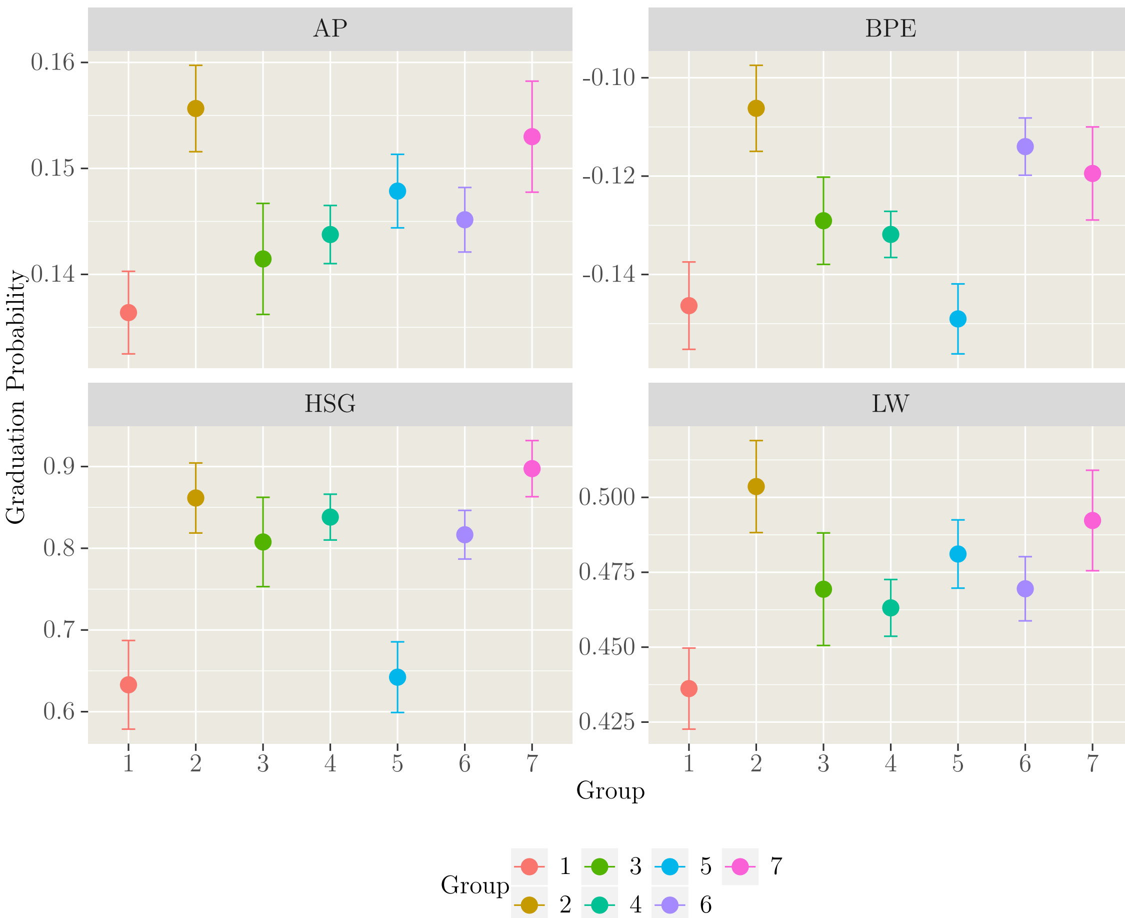
This figure shows the mean level of hours worked and welfare participation for each classified group, where group classification is taken from the the first stage of the estimation procedure.

Figure B.4: Means by Classified Group



This figure depicts means of ϕ_τ (the fraction of leisure hours spent actively with the child), wages, welfare participation and hours worked, by classified group. Group classification is taken from the first stage of the estimation procedure. Error bars depict 95% confidence intervals for each estimate.

Figure B.5: Skill Outcomes



This figure shows the mean level of skill outcomes for children in each classified group/type. Skills are weighted by their association with High School Graduation. One standard deviation error bars are shown.

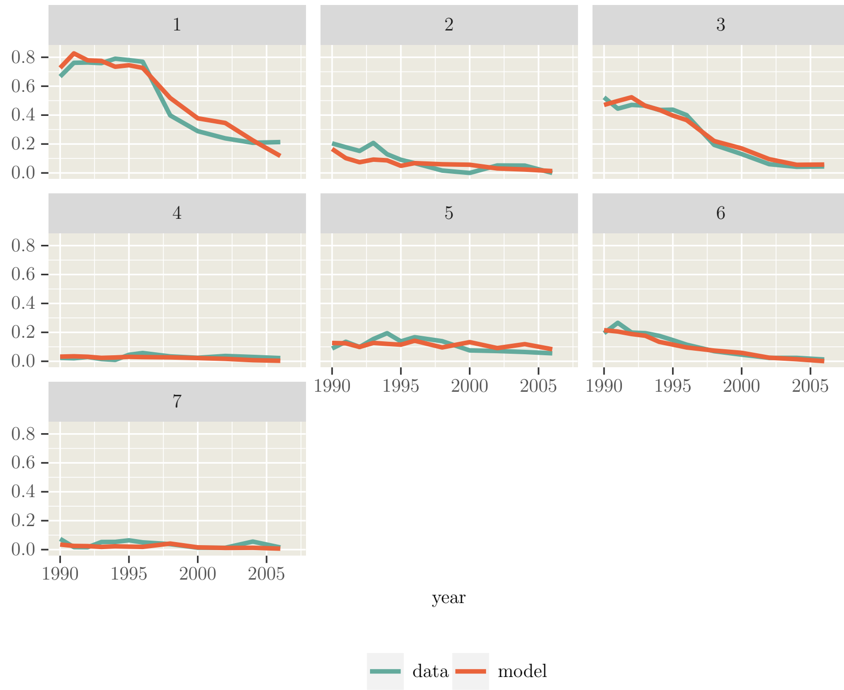


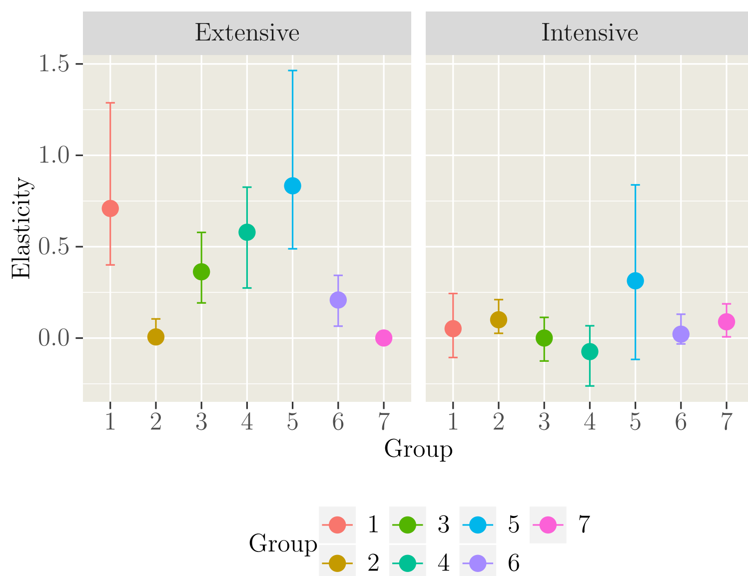
Figure B.6: Model Fit - AFDC/TANF Participation

Figure B.7: Model Fit - Weekly Hours Worked

Figure B.8: Model Fit - Labor Force Participation

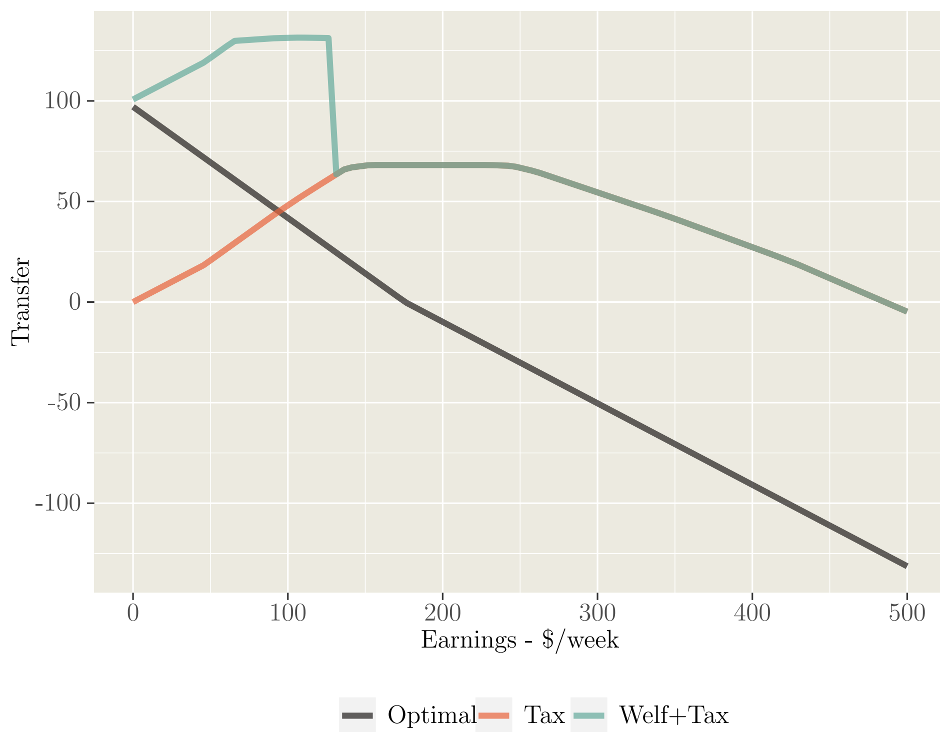
Figure B.9: Model Fit - Food Stamp Participation

Figure B.10: Marshallian Elasticities for Classified Groups



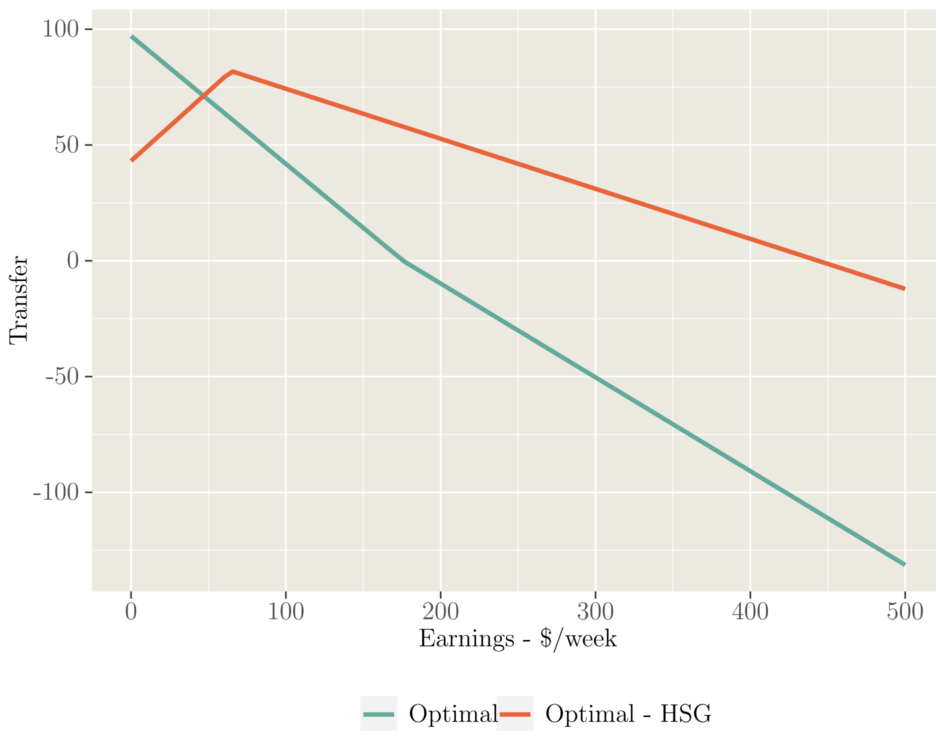
This figure shows Marshallian labor supply elasticities for each classified group, computed using the average response to a 10% wage increase in every period of the simulation. Error bars show 95% confidence intervals, computed using 100 bootstrap resamples.

Figure B.11: Comparison - Optimal Policy vs California



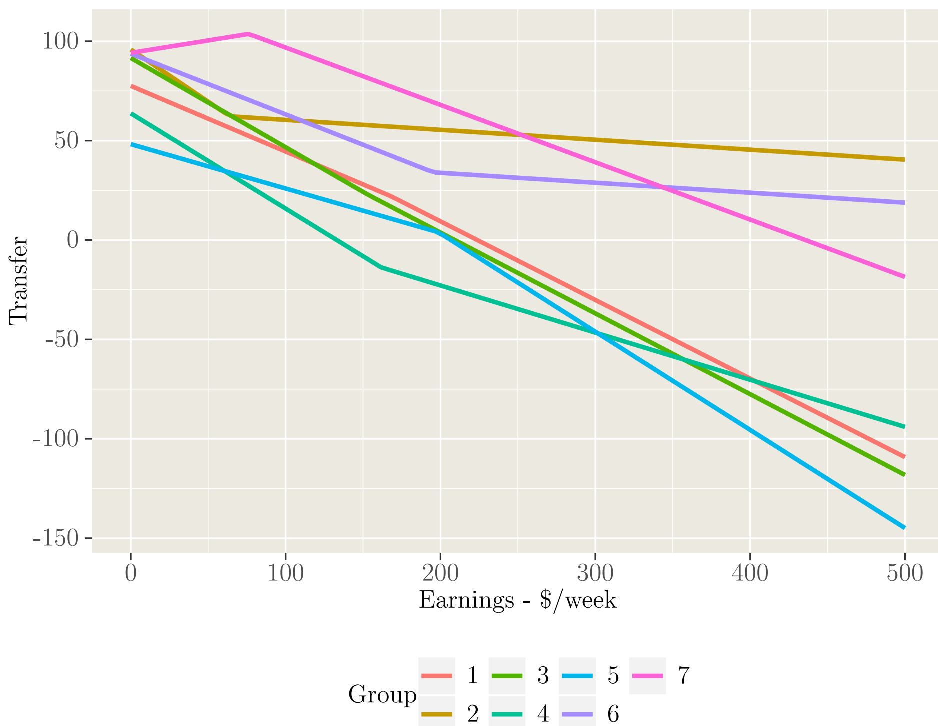
This figure compares the transfer schedule for the optimal policy (welfare maximizing) against the schedule for a mother of one child in California, in 2010. Separate schedules are drawn for the cases of welfare participation and non-participation.

Figure B.12: Comparison - Welfare vs Skill Maximizing Policies



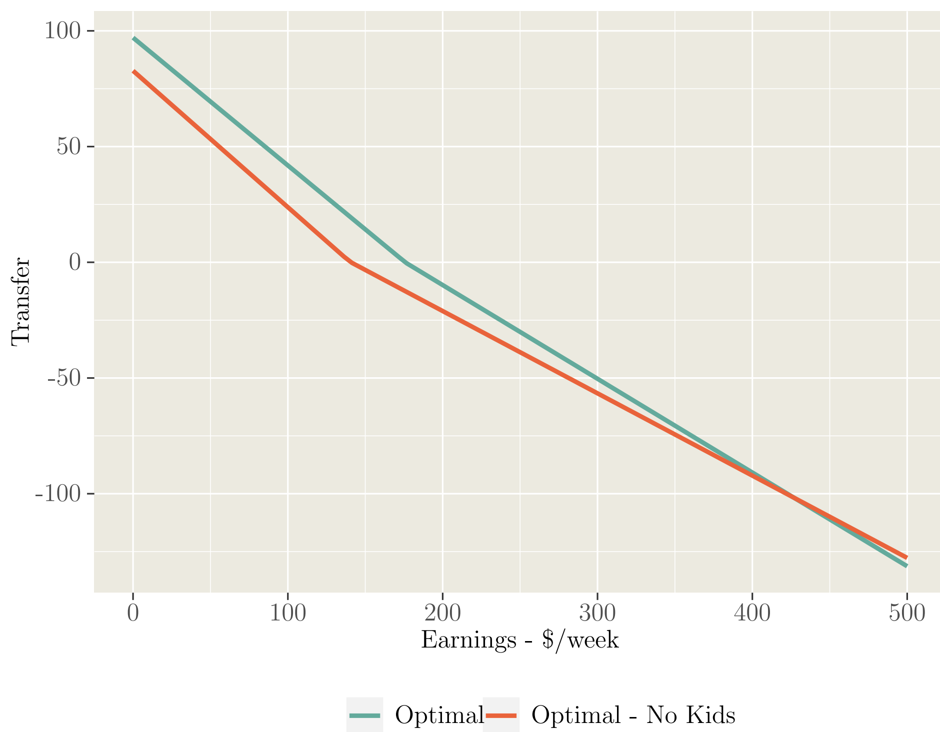
This figure compares the welfare maximizing transfer schedule ("Optimal") against the schedule that maximizes skill outcomes ("Optimal - HSG") as measured by their joint association with the probability of high school graduation.

Figure B.13: Optimal Policy By Type



This figure shows the optimal transfer schedule when welfare is maximized for each type individually, under the same revenue constraint as original problem. "Optimal - k" shows the optimal schedule when calculated for type k.

Figure B.14: Optimal Policy without Children



This figure compares the solution to the welfare maximizing problem in the baseline counterfactual, to the case in which it is applied to women without children. This is achieved by setting the preference weight on child skill outcomes, α_θ, to zero and re-solving the maximization problem.

C General Model and Solution

In this section, I derive the model solution for an arbitrary number of skills, N_θ, and an arbitrary number of children N_K. Let $k = 1, 2, 3, .., N_K$ index each child in ascending order of birth. In this section I assume that birth years are exogenously given and known to the mother. However, as we move through the solution it should be clear that extending the model to relax this assumption need not forbid the key simplification of additive separability in $\log(\theta)$, the skill vector. In fact, this is shown in greater detail in Brown, Flinn, and Mullins (2015), which includes endogenous fertility and marriage decisions. Let $\mathcal{B} = \{b_1, b_2, ..., b_{N_K}\}$ indicate the year in which each child is born. As before, every child matures at age A. I let the problem begin, as before, with the birth of the first child. Therefore we can set $b_1 = 0$, and set the terminal period of the problem at $T_M = b_{N_K} + A$. Now, θ refers to the full vector of skills for each child, and is therefore of dimension $N_\theta \times N_K$. We let $\theta_{k,j}$ refer to the jth skill of the kth child in the family and, similarly, θ_k is the full skill vector for child k. Finally, let $a_k(t) = t - b_k$ indicate the age of child k. Noting that age is collinear in time, I suppress the dependance of a_k on t for notational simplicity.

Preferences and Technology

To solve the problem, we must extend preferences to include multiple children, which we do simply as:

$$u(c, l, \theta) = \alpha_c \log(c) + \alpha_l \log(l) + \sum_{k:\ a_k \geq 0} \alpha_\theta \log(\theta_k) \tag{C.1}$$

$$V_{T_M}(\theta) = (1 - \beta)^{-1} \sum_k \alpha_\theta \log(\theta_k) \tag{C.2}$$

Next, I have to take a stand on the rivalrous nature of time and monetary investment. Ideally, one could assume the existence of public investment categories as well as categories that each child would benefit from privately. In the case of time investment, this would be empirically plausible since categories of time use are observable[23]. Further, this setup can handle an arbitrary number of investment categories: we will see that each would be determined by a proportional investment rule. Yet this exact property removes this assumption of any empirical content in our context, since child outcomes are driven through changes in the log of total income (Y_t) and log of total leisure hours ($1 - H_t$). When investment rules are proportional to total income and total leisure time, we can derive labor supply, program participation, and child outcomes in terms of aggregates that depend only on the total Cobb-Douglas shares of each category. This logic extends to the case of multiple children. If time use is rivalrous, it is true that a child with siblings receives a lesser share of spare leisure hours than an only child, *ceterus paribus*, however this proportion bears no impact once we take logs (in effect, looking at percentage changes in investment). Thus, I assume

[23]Less so, for expenditure categories.

in this paper that there is only one monetary investment and one time investment category, and that each is non-rivalrous across siblings.

Model Solution

Given this set of assumptions, we can write the dynamic program in the following fashion:

$$V_t(\theta_t, S_t, \eta_t) = \max_{c,l,x,\tau,h\in\mathcal{H},p} \Big\{ u(c, l, \theta_t, p) + \beta\mathbb{E}[V_{t+1}(\theta_{t+1}, S_{t+1}, \eta_{t+1}) \; |S_t, h, p] \Big\} \tag{C.3}$$

Subject to the constraints:

$$c + x \leq \mathbf{Y}(s_t, h, p) \tag{C.4}$$

$$\tau + l + h = 1 \tag{C.5}$$

$$\theta_{k,t+1} = \delta_{x,a_k}\log(x) + \delta_{\tau,a_k}\log(\tau) + \delta_{\theta,a_k}\log(\theta) + \eta_{k,t}, \qquad \forall k: \; a_k \geq 0 \tag{C.6}$$

Note that in this general formulation, the vector S_{t+1} is permitted to evolve according to the current state S_t in addition to the labor supply (h) and program participation (p) decisions of the mother. Some of this model's convenient representation could quite conceivably break if the evolution of S_t was further allowed to depend on investments (x, τ). I first propose the following simplification of the model and show that it holds. As before, the key is that the value function is additively separable in $\log(\theta)$:

$$V_t(\theta, S_t, \eta_t) = \sum_{k=1}^{N_k} \alpha_{V,a_k}\log(\theta_{k,t}) + \sum_{k:a_k\geq 0} \alpha_{V,a_k+1}\log(\eta_{k,t}) + \nu(S_t) \tag{C.7}$$

$$\nu(S_t) = \max_{h\in\mathcal{H},p} \Big\{ \overline{\alpha}_{c,t}\log(\mathbf{Y}(S_t, h, p)) + \overline{\alpha}_{\tau,t}\log(1-h) + \beta\mathbb{E}[\nu(s_{t+1}) \mid S_t, h, p] \Big\} \tag{C.8}$$

$$\overline{\alpha}_{c,t} = \alpha_c + \beta \sum_{k:\; 0\leq a_k<A} \alpha_{V,a_k+1}\delta_{x,a_k} \tag{C.9}$$

$$\overline{\alpha}_{\tau,t} = \alpha_c + \beta \sum_{k:\; 0\leq a_k<A} \alpha_{V,a_k+1}\delta_{\tau,a_k} \tag{C.10}$$

$$\alpha_{V,a} = \alpha_\theta + \mathbf{1}\{0 \leq a_k < A\} \cdot \beta\alpha_{V,a+1}\delta_{\theta,a} \tag{C.11}$$

$$x_t = \frac{\beta\sum_{k:\; 0\leq a_k<A} \alpha_{V,a_k+1}\delta_{x,a_k}}{\overline{\alpha}_{c,t}}\mathbf{Y}(S_t, h, p) \tag{C.12}$$

$$\tau_t = \frac{\beta\sum_{k:\; 0\leq a_k<A} \alpha_{V,a_k+1}\delta_{\tau,a_k}}{\overline{\alpha}_{l,t}}(1-h) \tag{C.13}$$

We prove this by first showing that the recursion holds. That is, assume this form holds for V_{t+1} and show that it is preserved at time t. The problem can be stated as:

$$\begin{aligned} V_t(\theta, S_t, \eta_t) = \max_{c,l,x,\tau,h,p} & \Bigg\{ \alpha_c\log(c) + \alpha_l\log(l) + \sum_{k:\; a_k\geq 0} \alpha_\theta\log(\theta) \\ & + \beta\mathbb{E}\left[\nu(S_{t+1}) + \sum_k \alpha_{V,a_k+1}\log(\theta_{k,t+1}) + \sum_{0\leq a_k+1<A} \alpha_{V,a_k+2}\log(\eta_{k,t+1}) \mid S_t, h, p \right] \Bigg\} \end{aligned} \tag{C.14}$$

subject to the constraints given above. First, we can substitute in the production function to get:

$$\begin{aligned} V_t(\theta, S_t, \eta_t) = \max_{c,l,x,\tau,h,p} \Bigg\{ & \alpha_c \log(c) + \alpha_l \log(l) + \sum_{k:\ a_k \geq 0} \alpha_\theta \log(\theta_{k,t}) \\ & + \underbrace{\sum_k (\alpha_\theta + \mathbf{1}\{0 \leq a_k < A\} \cdot \alpha_{V,a_k+1}\delta_{\theta,a_k})}_{=\alpha_{V,a_k}} \log(\theta_{k,t}) \\ & + \beta \sum_{k:\ 0 \leq a_k < A} \alpha_{V,a_k+1}[\delta_{x,a_k} \log(x) + \delta_{\tau,a_k} \log(\tau) + \log(\eta_{k,t})] \\ & + \beta \mathbb{E}\left[\nu(S_{t+1}) + \sum_{0 \leq a_k+1 < A} \alpha_{V,a_k+2} \log(\eta_{k,t+1}) \mid S_t, h, p \right] \Bigg\} \end{aligned} \quad (C.15)$$

As is indicated in the second line of this equation, this step is sufficient to define the recursion for $\alpha_{V,a}$, the value derived from $\log(\theta)$ for a child at age a. Next, we inspect the first order conditions for τ and x, subject to choices of h and p. These yield:

$$\frac{\alpha_c}{c} = \frac{\beta \sum_{k:\ 0 \leq a_k < A} \alpha_{V,a_k} \delta_{x,a_k}}{x} \quad (C.16)$$

$$\frac{\alpha_l}{l} = \frac{\beta \sum_{k:\ 0 \leq a_k < A} \alpha_{V,a_k} \delta_{\tau,a_k}}{\tau} \quad (C.17)$$

$$(C.18)$$

Rearranging these equations gives the proportional investment rules shown in (C.12) and (C.13). Substituting those rules into the value function (C.15) and collecting terms gives the final expression of the value function in (C.7) and (C.8). We can complete this exposition for the solution by noting that the terminal period value function $V_{T_M}(\theta)$ also keeps this additive form, and hence we have the necessary conditions to initiate the recursion.

D Disinguishing Preference Orderings

Let us begin by fixing p, and the state, S, and consider the ranking of two hours choices, h and $h + \Delta$, with $h > 0$. Intuitively, if S can shift the returns to extra work (through policies or wages), then there is some return to Δ extra hours such that one type will prefer to work more while the other does not. For α, h will be ranked above $h + \Delta$ iff:

$$\log(\mathbf{Y}(p, h, S)) + \alpha_l \log(1 - h) > \log(\mathbf{Y}(p, h + \Delta, S) + \alpha_l \log(1 - h - \Delta)$$

This amounts to:

$$\alpha_l > \frac{\log(\mathbf{Y}(h + \Delta, p, S) - \log(\mathbf{Y}(h, p, S))}{\log(1 - h) - \log(1 - h - \Delta)}$$

Assuming, without loss of generality, that $\alpha_l > \alpha_l'$, then α and α' can be "separated" by any S such that:

$$\alpha_l > \frac{\log(\mathbf{Y}(h + \Delta, p, S) - \log(\mathbf{Y}(h, p, S))}{\log(1 - h) - \log(1 - h - \Delta)} > \alpha_l'$$

Now consider the comparison of labor force non-participation ($h = 0$) with work at some level of hours, h. The former will be preferred if:

$$\log(\mathbf{Y}(p,0,S)) + \alpha_l \log(1) > \log(\mathbf{Y}(p,h,S)) + \alpha_l \log(1-h) - \alpha_h$$

which we can rearrange to:

$$\alpha_h - \alpha_l \log(1-h) > \log\left(\frac{\mathbf{Y}(p,h,S)}{\mathbf{Y}(p,0,S)}\right).$$

Just as above, assuming wlog that for some h, the pair α_h, α_l can be distinguished from the pair α_h', α_l' as long as S sufficiently varies the potential returns to joining the labor force, say through changes in non-labor income, wages, or policies.

Finally, to isolate the disutility of welfare participation, α_A, we make a similar comparison holding the hours choice, h, fixed. Participation is preferred to non-participation if:

$$\log\left(\frac{\mathbf{Y}(2,h,S)}{\mathbf{Y}(0,h,S)}\right) > \alpha_A$$

In this case, the returns to participation will increase with the generosity of the program (policy variation), and decrease with non-labor income and wages. While I do not consider distinguishing different values of α_F through any similar choice comparison, it is clear that an analagous comparison applies. When defining the score criterion in the main section of the paper, I do not include such a comparison in the score criterion, and instead exploit the fact that each discrete type can be, by assumption, distinguished by the parameters considered here.

E Estimation Procedure

E.1 First Stage

While the linear estimating equation (3.1) text assumes annual data on skills, in reality the CDS only provides skill measurements every 5 years. To estimate the model, then, I iterate the linear specification forward by four years. This gives an outcome equation:

$$\log(\theta_{i,t+5}) = \sum_{s=0}^{4} \delta_\theta^{4-s}\left(\delta_x \log(Y_{i,t+s}) + \delta_\tau \log(1-H_{i,t+s}) + \mu_{\theta,k(i)} + \varepsilon_{k(i),t} + \eta_{it}\right) + \delta^4 \log(\theta_{i,t}) \quad \text{(E.1)}$$

Stacking the observations of Y and H into vectors, we can simplify this outcome equation as:

$$\log(\theta_{i,t+5}) = \Lambda_x(\delta)\log(\mathbf{Y}) + \Lambda_\tau(\delta)\log(1-\mathbf{H}) + \Lambda_\theta(\delta)\theta_{i,t} + \Lambda_{k(i)} + \tilde{\eta}_{it}$$

Excluding the type fixed effect $\Lambda_{k(i)}$, we get:

$$\log(\theta_{i,t+5}) = \Lambda_x(\delta)\log(\mathbf{Y}) + \Lambda_\tau(\delta)\log(1-\mathbf{H}) + \Lambda_\theta(\delta)\theta_{i,t} + \tilde{\epsilon}_{it}$$

I add to this outcome equation a dummy for the year in which skills were measured and a linear age term, to control for cohort and age effects. For a particular choice of the production parameters

δ and type fixed effects, residuals can be constructed using the above equations, computing the moments:

$$g_1(Z_{it}, \delta) = \mathbb{E}[\tilde{\epsilon}_{it} \otimes f(Z_{it})] = 0, \tag{E.2}$$

$$g_2(H_{it}, Y_{it}, \delta, \mathcal{K}) = \mathbb{E}[\tilde{\eta}_{it} \otimes [D'_{k(i)}, \log(1 - H_{it}), \log(Y_{it})]'] = 0. \tag{E.3}$$

Since the policy vectors Z_{it} consist of many variables, I aggregate them into 4 instruments by making net income calculations using Z_{it}. In a slight abuse of notation, let $\mathbf{Y}(E, p, Z_{it})$ denote the implied net income when annual earnings are E, and program participation p. I compute:

$$f(Z_{it}) = [\mathbf{Y}(10000, 0, Z_{it}), \mathbf{Y}(20000, 0, Z_{it}), \mathbf{Y}(10000, 2, Z_{it}), \mathbf{Y}(20000, 2, Z_{it})]'.$$

To compute the weighting matrix, W, I first estimate the reduced form coefficients of equation (E.1) by OLS, and calculate the variance of the moments using the imputed residuals, using the inverse of this matrix as W. Standard errors are computed with the typical asymptotic variance formula. I run a nonlinear GMM routine using the CDS children of mothers in my sample, setting $t = 1997, 2002$ and $t + 5 = 2002, 2007$. Since the PSID transitioned to a biennial survey from 1997 onward, I must use income and hours measurements only from 1997,1999,2001,2003, and 2005. Missing years are proxied using income and hours measurements from the following year.

E.2 Second Stage

Recall the empirical specification for wages:

$$\log(W_{it}) = \gamma_{w,0,k(i)} + \gamma_{w,a,k(i)} AGE_{it} + \epsilon_{W,it}$$

I model the shocks to the AR(1) process $\epsilon_{W,it}$ as being normally distributed with variance $\sigma_{w,k(i)}$ and autocorrelation $\rho_{w,k(i)}$.

Estimation of the remaining parameters is performed separately by type, using 10 simulations per mother observed in the sample. Statistics included in $\mathcal{M}_k$ are:

- Mean hours worked, by year;
- Mean labor force participation, by year;
- Mean rate of food stamp participation, by year;
- Mean rate of welfare participation, by year;
- Mean of $\phi_{i,\tau}$, in 1997 and 2002;
- The coefficients from regressing log wages on age;
- The correlation and variance of the log-wage residuals in the above regression.

F Description of the Transfer Functions

In this section I describe the computation of the transfer functions (T^F, T^A, T^T). The details of this section very closely follow Chan (2013), which should be consulted for further details.

Welfare

The transfer function T^A includes a benefit computation, and an eligibility test:

$$T_{it}^A = \text{El}_{it}^A \times \text{Ben}_{it}^A \tag{F.1}$$

Where

$$\text{El}_{it}^A = \mathbf{1}\{L_{it} \leq \mathcal{L}_{it}\} \times \mathbf{1}\left\{E_{it} + N_{it} < r_{Agit}e_{Ait}\right\} \times \mathbf{1}\left\{(E_{it} - D_{Aeit})(1 - R_{Aeit}) + N_{it} < r_{Anit}e_{Ait}\right\}. \tag{F.2}$$

Eligibility above is defined as the combination of a time limit, a net income test, and a gross income test. Both tests compare income with a need standard, e_{Ait} which is inflated by some rate (r_{Agit}, r_{Aeit}). Second, the computation of net income involves a fixed disregard on earnings, D_{Aeit} and a percentage disregard. Benefit computation follows similarly:

$$\text{Ben}_{it}^A = \max\{G_{Ait} - (E_{it} - D_{Abit})(1 - R_{Abit}) - N_{it}, 0\} \tag{F.3}$$

The payment standard G_{Ait} sets the generosity of the program when no other sources of income are reported, while the dollar and percentage disregards (D_{Abit}, R_{Abit}) combine to determine net income. Importantly, these policy parameters are a function of the mother's state of residence, the number of dependant children and the year. In this model, these variables are all a function of the mother-year index, it.

Food Stamps

Similarly to welfare, the food stamp transfer function T^F can be written as:

$$T_{it}^A = \text{El}_{it}^F \times \text{Ben}_{it}^F \tag{F.4}$$

Where

$$\text{El}_{it}^F = \mathbf{1}\left\{E_{it} + N_{it} < 1.3e_{Fit}\right\} \times \mathbf{1}\{\underbrace{0.8E_{it} + N_{it} + \text{Ben}_{it}^A - 134}_{=\text{Net}_{it}^F} < e_{Fit}\}. \tag{F.5}$$

In the above expression, e_{Fit} is referred to as the poverty guideline, and the net income includes a standard 20% disregard and \$134 deduction. While the true food stamp benefit formula technically allows for further deductions for child care expenses, child support payments, and shelter expenses, I have insufficient data to calculate these deductions. Finally, given a maximum benefit G_{Fit}, the benefit calculation is:

$$\text{Ben}_{it}^F = \max\left\{G_{Fit} - 0.3\text{Net}_{it}^F, 0\right\}. \tag{F.6}$$

Data Sources for Program Rules

To summarize, the parameter vector Z_{it}^A can be written as

$$Z_{it} = \{r_{Agit}, r_{Aeit}, e_{Ait}, D_{Aeit}, R_{Aeit}, G_{Ait}, D_{Abit}, R_{Abit}, \mathcal{L}_{it}\},$$

while Z_{it}^F can be summarized as

$$Z_{it}^F = \{e_{Fit}, G_{Fit}\}.$$

Parameters on welfare that comprise Z_{it}^A and Z_{it}^F were collected from the Urban Institute's TRIM3 simulation database[24] for years 1985-2011. In addition, since rules on net income calculations were much more simple prior to 1993, I use a 30% disregard across all states [25]. Mothers were merged with program rules based on their state of residence, the year, and the number of children in their household of age 17 or younger.

Taxes

Taxes consist of a federal and a state computation. When earned income is sufficiently low, $\mathcal{T}^T$ will arrive in the form of a net payment (when income tax obligations are exceeded by the EITC). In theory, the relevant parameters to compute taxes include those that define the federal and state EITC programs, state and federal deductions and exemptions, and the marginal income tax rate with their corresponding brackets for state and federal income tax. In practice, I use the TAXSIM model of Feenberg and Coutts (1993), to approximate the tax function. Given the relevant year, state, and family size (in our model these are all exogenous functions of the index, it), TAXSIM computes $\mathcal{T}_{it}^T(e)$ for any given earnings level. Thus, for each it in my sample, I compute $\mathcal{T}_{it}^T(e)$ for earnings levels e on a grid, using increments of \$100, between \$0 and \$100,000[26]. Using this grid, the tax function is approximated using linear interpolation between these grid points.

References

Aizer, A., S. Eli, J. Ferrie, and A. Lleras-Muney (2016): "The Long-Run Impact of Cash Transfers to Poor Families," *American Economic Review*, 106, 935–71.

Arellano, M. and S. Bonhomme (2011): "Nonlinear Panel Data Analysis," *Annual Review of Economics*, 3, 395–424.

Attanasio, O., P. Levell, H. Low, and V. Snchez-Marcos (2018): "Aggregating Elasticities: Intensive and Extensive Margins of Women's Labor Supply," *Econometrica*, 86, 2049–2082.

[24] Source: http://trim3.urban.org/

[25] This approach is taken also in Chan (2013)

[26] This suits as a reasonable upper bound in my sample

Becker, G. and H. G. Lewis (1973): "On the Interaction between the Quantity and Quality of Children," *Journal of Political Economy*, 81.

Becker, G. and N. Tomes (1976): "Child Endowments and the Quantity and Quality of Children," *Journal of Political Economy*, 84.

Bernal, R. (2008): "The Effect of Maternal Employment and Child Care on Children's Cognitive Development," *International Economic Review*, 49, 1173–1209.

Bernal, R. and M. P. Keane (2011): "Child Care Choices and Childrens Cognitive Achievement : The Case of Single Mothers," *Journal of Labor Economics*, 29, 459–512.

Bitler, M. P., J. B. Gelbach, H. W. Hoynes, and M. Zavodny (2004): "The impact of welfare reform on marriage and divorce," *Demography*, 41, 213–236.

Blundell, R., M. Costa Dias, C. Meghir, and J. Shaw (2016): "Female Labor Supply, Human Capital, and Welfare Reform," *Econometrica*, 84, 1705–1753.

Blundell, R., A. Duncan, and C. Meghir (1998): "Estimating Labor Supply Responses Using Tax Reforms," *Econometrica*, 66, 827–861.

Blundell, R. and A. Shephard (2011): "Employment, Hours of Work and the Optimal Taxation of Low-Income Families," *The Review of Economic Studies*, 79, 481–510.

Bonhomme, S., T. Lamadon, and E. Manresa (2017): "Discretizing Unobserved Heterogeneity," .

Bonhomme, S. and E. Manresa (2015): "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 83.

Brown, M., C. Flinn, and J. Mullins (2015): "Family Law Effects on Divorce, Fertility and Child Investment," .

Chan, M. K. (2013): "A Dynamic Model of Welfare Reform," *Econometrica*, 81, 941–1001.

Chetty, R., J. N. Friedman, and E. Saez (2013): "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings," *American ECconomic Review*, 103.

Chon, P. and G. Laroque (2010): "Negative Marginal Tax Rates and Heterogeneity," *American Economic Review*, 100, 2532–47.

Cunha, F. and J. Heckman (2007): "The Technology of Skill Formation," *American Economic Review*, 97.

Cunha, F., J. Heckman, and S. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica*, 78, 883–931.

Cunha, F. and J. J. Heckman (2008): "Formulating , Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43.

Currie, J. and D. Almond (2011): "Chapter 15 - Human capital development before age five," Elsevier, vol. 4, Part B of *Handbook of Labor Economics*, 1315 – 1486.

Dahl, G. B. and L. Lochner (2012): "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit," *American Economic Review*, 102, 1927–1956.

Del Boca, D., C. Flinn, and M. Wiswall (2014): "Household Choices and Child Development," *Review of Economic Studies*, 81.

Duncan, G. J., P. A. Morris, and C. Rodrigues (2011): "Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments." *Developmental psychology*, 47, 1263–79.

Fang, H. and D. Silverman (2004): "On the compassion of time-limited welfare programs," *Journal of Public Economics*, 88, 1445–1470.

Feenberg, D. R. and E. Coutts (1993): "An Introduction to the TAXSIM Model," *Journal of Policy Analysis and Management*, 12, 189–194.

Fox, J. T. (2007): "Semiparametric estimation of multinomial discrete-choice models using a subset of choices," *RAND Journal of Economics*, 93.

Gayle, G.-L., L. Golan, and M. Soytas (2015): "What Accounts for the Racial Gap in Time Allocation and Intergenerational Transmission of Human Capital?" .

Grogger, J. (2002): "The Behavioral Effects of Welfare Time Limits," *American Economic Review*, 92.

——— (2003): "The Effects of Time Limits, the EITC, and Other Policy Changes on Welfare Use, Work, and Income among Female-Headed Families," *Review of Economics and Statistics*, 85, 394–408.

Grogger, J. and S. G. Bronars (2001): "The Effect of Welfare Payments on the Marriage and Fertility Behavior of Unwed Mothers: Results from a Twins Experiment," *Journal of Political Economy*, 109, 529–545.

Heathcote, J., K. Storesletten, and G. Violante (2017): "Optimal Tax Progressivity: An Analytical Framework," *Quarterly Journal of Economics*, 132.

Heckman, J. J., S. Hyeok, R. Pinto, A. Peter, S. H. Moon, P. a. Savelyev, and A. Yavitz (2010): "The rate of return to the HighScope Perry Preschool Program," *Journal of Public Economics*, 94, 114–128.

Heckman, J. J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*.

Hoynes, H. (1996): "Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation under AFDC-UP," *Econometrica*, 64.

Hoynes, H. and D. W. Schanzenbach (2018): "Safety Net Investments in Children," *Brookings Papers on Economic Activity*.

Hoynes, H., D. W. Schanzenbach, and D. Almond (2016): "Long-Run Impacts of Childhood Access to the Safety Net," *American Economic Review*, 106, 903–34.

Kearney, M. S. (2004): "Is There an Effect of Incremental Welfare Benefits on Fertility Behavior? A Look at the Family Cap," *The Journal of Human Resources*, 39, 295–325.

Kopecky, K. and R. Suen (2010): "Finite State Markov-Chain Approximations to Highly Persistent Processes," *Review of Economy Dynamics*, 13, 701–714.

Low, H., C. Meghir, L. Pistaferri, and A. Voena (2018): "Marriage, Labor Supply and the Dynamics of the Social Safety Net," Working Paper 24356, National Bureau of Economic Research.

Meyer, B. D. (2002): "Labor Supply at the Extensive and Intensive Margins : The EITC , Welfare , and Hours Worked," *American economic review*, 92.

Mirrlees, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38, 175–208.

Moffitt, R. (1983): "An Economic Model of Welfare Stigma," *The American Economic Review*, 73, 1023–1035.

Moffitt, R. A. (2015): "The Deserving Poor, the Family, and the U.S. Welfare System," *Demography*, 52, 729–749.

Rouwenhorst, K. (1995): "Asset Pricing Implications of Equilibrium Business Cycle Models," in *Frontiers of Business Cycle Research*, ed. by T. Cooley.

Saez, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68, 205–29.

——— (2002): "Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses," *Quarterly Journal of Economics*, 117, 1039–1073.

Voena, A. (2015): "Yours, Mine and Ours: Do Divorce Laws Affect the Intertemporal Behavior of Married Couples?" *American Economic Review*.