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

Joseph Mullins *†

June 19, 2025

Abstract

This paper finds that accounting for the human capital development of children has a quantitatively large effect on the true costs and benefits of providing cash assistance to single mothers in the United States. A dynamic model of work, welfare participation, and parental investment in children introduces a framework for calculating costs and benefits when individuals respond to incentives. The model provides a tractable outcome equation in which a policy's effect on child skills can be understood through its impact on two economic resources in the household – time and money – and the share of each resource, in combination with childcare quality, as factors in the production of skills. These key causal parameters are cleanly identified by policy variation through the 1990s. The model also admits simple and interpretable formulae for optimal nonlinear transfers, with novel features arising from the child skill formation channel. Using a broadly conservative empirical strategy, estimates imply that optimal transfers are up to 50% more generous than the US benchmark, with quite different labor supply incentives.

1 Introduction

A proper accounting of the costs and benefits of cash assistance programs for households with children must include the long-run impacts on children's human capital. This paper precisely articulates and quantitatively validates the argument by estimating a model in which single mothers work, participate in government assistance programs, and shape their children's human capital through time and money investments. Accounting for the role played by time and money as economic resources has implications for the optimal size and shape of transfers to these households, which a formal analysis of optimal taxation in the spirit of Mirrlees (1971) will demonstrate.

A growing empirical literature establishes the motivating facts. First, maternal time and household income have both been shown to play a causal role in shaping child skill outcomes (Duncan, Morris, and Rodrigues, 2011; Dahl and Lochner, 2012; Akee, Copeland, Costello, and Simeonova, 2018; Bernal and Keane, 2010, 2011). Second, an individual's skills shape their life-cycle outcomes across multiple economic and social dimensions (Cunha, Heckman, and Schennach, 2010; Heckman, Stixrud, and Urzua, 2006; Heckman, Pinto, and Savelyev, 2013). Third, interventions that boost skill outcomes in childhood have been shown to reap large long-run returns (Heckman, Hyeok, Pinto, Peter, Moon, Savelyev, and Yavitz, 2010; Bailey, Sun, and Timpe, 2021), including those that provide material economic support to disadvantaged households (Barr, Eggleston, and Smith, 2022; Hoynes, Schanzenbach, and Almond, 2016).

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

†For support and useful feedback, many thanks to Chris Flinn, Gianluca Violante, Meta Brown, Lance Lochner, Tobias Salz, David Cesarini, Timothy Christensen, Simon Mongey, Roy Allen, Job Boerma, Anmol Bhandari, Jeremy Lise, Ellen McGrattan, and Mariacristina De Nardi. All remaining errors are my own.

Put together, existing evidence[1] suggests that children's human capital development can shape the weighing of benefits and costs of social programs. Despite this evidence, Aizer, Hoynes, and Lleras-Muney (2022) point out that the United States Congressional Budget Office does not account for such impacts when evaluating the fiscal effects of proposed legislation; nor have economists provided a coherent framework for doing so. This paper provides such a framework for a specific class of policies (cash assistance that conditions on work, earnings, and participation behavior) for a specific population (single mothers) and uses the framework to show, theoretically and quantitatively, how children's skill formation affects the optimal size and shape of transfers to households with children.

The framework requires four key ingredients: (1) an economic model in which choices are articulated with respect to the class of policies considered; (2) a technology of skill formation that maps these choices to child human capital outcomes; (3) a mapping from skills to economic resources in the long-run; and (4) credible sources of identification for each of these three mechanisms. Describing ingredients (1)-(3) along with their sources of empirical discipline ("identification") will provide a useful overview of this paper's methodological choices and limitations before highlighting some quantitative results.

The paper develops a dynamic model in Section 2 that carefully replicates the complex cash assistance policies available to single mothers in the United States. These fall into three categories:

1. Food stamps: the Supplemental Nutritional Assistance Program (SNAP);
2. Cash welfare: Assistance to Families with Dependent Children (AFDC) before 1996, Temporary Assistance of Needy Families (TANF) thereafter; and
3. Taxes: the Earned Income Tax Credit (EITC) and the Child Tax Credit (CTC) both specifically provide refundable and non-refundable payments to households with children.

In the model, single mothers decide whether to work and whether to participate in assistance programs. Their earnings combine with prevailing tax and benefit formulae to determine net income. They then decide how much of their net income and non-work time to invest in each of their children. These, along with childcare inputs, determine future cognitive and behavioral skills through a human capital production function. In the baseline model, childcare quality is a heterogeneous, exogenous random variable that is latent and permitted to be correlated with mothers' other latent traits. An extension of the baseline model allows quality to be purchased out of net income.

Mothers' decisions are dynamic: they must weigh the benefits and costs of their decisions today against the future human capital of their children. The setup follows a conceptual tradition of thinking about child development as the outcome of a human capital investment problem (Becker and Lewis, 1973; Becker and Tomes, 1976). In the model, social assistance policies shape net income and hours at home through their generosity and incentive structure, and the effect on these economic resources has a spillover to children's development through mothers' investment decisions. A key assumption in the model is that mothers are unobservably different from each other in their preferences and labor market productivity. These differences in "type" turn out to be crucial in determining the distributional impacts of different policy reforms, which the paper's counterfactual exercises will demonstrate.

The Panel Study of Income Dynamics (PSID) and its Child Development Supplement (CDS) provide the necessary data (Section 4) for estimating the model (Section 5). The PSID provides longitudinal data on the work, earnings, and program participation for mothers of children in the CDS, while the CDS itself provides measures on skill outcomes and maternal time investment for children in the household. The mothers in the selected sample worked and raised their children through the 1990s, which was a period of historically significant changes in US safety net policies. Most significantly, the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) in 1996 drastically altered the nature of cash welfare: it introduced federally mandated time limits on welfare participation, imposed work requirements on participants, and gave States legislative freedom to re-allocate funds, design program features, and

[1] Additional evidence on the long-run returns to skill-boosting interventions can be found in García, Heckman, Leaf, and Prados (2020); Kline and Walters (2016); Chetty, Friedman, Hilger, Saez, Schanzenbach, and Yagan (2011); Bastian and Michelmore (2018); Aizer, Eli, Ferrie, and Lleras-Muney (2016); Bailey, Hoynes, Rossin-Slater, and Walker (2020)

change benefit formulae. The 1990s also saw several large expansions in the EITC, which significantly enhanced the financial returns to work for single mothers. A number of papers using a variety of methods have found that these reforms played a substantial role in reducing welfare participation and increasing labor force participation during this time (Hoynes, 1996; Grogger, 2002, 2003; Meyer, 2002; Chetty et al., 2013a; Chan, 2013). The time period therefore provides a useful laboratory in which to study and estimate the responsiveness of single mothers to changes in work incentives. The longitudinal dimension of the data also unveils a substantial degree of heterogeneity in mothers' costs of work and participation, as well as their earnings potential, which will prove to be critical in determining widely varying distributional impacts of these historical changes to the safety net.

As an outcome of the model, child skills can be written as a linear function of money resources (net household income), time resources (the mother's non-market time), and the quality of childcare relative to maternal care. The coefficients of this outcome equation come directly from the technology of skill formation and map any policy's impacts on economic resources to skill impacts. They can be credibly identified in the presence of plausibly exogenous variation in net income and mothers' work behavior, which the policy variation in the data across states and over time provides. The model structure also provides a control function approach to identification of these parameters, which produces broadly similar point estimates with greater precision. The preferred set of results – which come from a quasi-bayesian procedure that imposes weak priors on plausible parameter values – assign modest causal impacts to both time and money in skill formation that nonetheless produce substantial skill impacts from persistent and large reforms.

The long-run economic value of cognitive and behavioral skills is not directly estimable from the PSID-CDS data, with CDS children having reached only young adulthood. In the absence of truly credible numbers, this paper opts for conservative ones instead, focusing on the economic value of skills from labor market earnings. The PSID-CDS data permit estimates of the effect of cognitive and behavioral skills on earnings in young adulthood. A simple quadratic forecasting model of earnings from Current Population Survey (CPS) data provides a means to extrapolate these effects to net present values over the life-cycle.

The model allows one to then imagine – as this paper does in Section 3 – the problem of a government that wishes to maximize the welfare of single mothers subject to a net present value resource constraint. At the margin the government considers the return to offering a dollar to households for a particular choice against the cost, which consists of a mechanical component (that dollar) and a behavioral component (the change in total costs when individuals respond to this change in incentives). Classically, this is where the problem ends. In this model, the government must also consider the direct impact of that dollar on the net present value of future resources through child skills, and the behavioral impacts on the child's future human capital through mothers' responses in work decisions. Adding this mechanism to the problem, Section 3 offers two formulae that provide insight into the determinants of the optimal size and shape of cash transfers. The two expressions indicate that overall generosity is increasing in the factor share of money in child skill formation, while the shape of optimal transfers is determined by the net impact of positive work hours (relative to not working) on skills. Negative net skill impacts – which for example prevail when the quality of maternal care is higher than external care and the financial returns to work are small – lead the planner to optimally dampen labor supply incentives relative to the case without skill formation. The second formula formalizes this result in terms of a wedge between optimal allocations with and without moral hazard. This allows for a quantitative decomposition in Section 6 of the contributions of the various mechanisms to the shape of optimal policies.

To summarize, while aggregate allocations depend only on the importance of money in skill formation, the shape of these allocations depends on the relative factor shares of money and time in skill production, as well as the relative quality of maternal care. Labor supply incentives are – roughly – decreasing in the relative quality of maternal care, decreasing in the factor share of time, and increasing in the factor share of money. The theoretical analysis of Section 3 therefore provides a tight theoretical link between the parameters of the skill production function and optimal policy, underscoring the importance of estimating these parameters as credibly as possible.

Counterfactuals and decompositions using the estimated model in Section 6 suggest that child skill formation weighs heavier than the more traditional fiscal trade-offs. The end result is an optimal non-linear system of transfers that is far more generous and shaped quite differently from the existing system in the US.

When using the sample in the year 2000 as a benchmark, average net income under the optimal policy is just over \$2,000/month, a number that is 50% higher than under existing policies. The implications for shape are also quantitatively significant, with individuals facing much higher participation and marginal tax rates than in the benchmark. A quantitative decomposition of the planner's optimal allocation, using the theory of Section 3, demonstrates that skill formation concerns are the primary determinant of the shape of optimal policies.

Having established the quantitatively crucial role played by skill formation in determining the optimal static policy, Section 6 then turns to a broader class of policies with features that more closely resemble existing welfare programs. Numerical results here indicate that expanding transfers to include time limits, welfare ordeals, or work requirements does not meaningfully improve upon the planner's objective, nor do they present any trade-offs among diffent types in the model.

All of the optimal reforms involve large increases in transfers to all households, which the planner can justify thanks to a large return in future skills that makes the expansion almost revenue neutral in net present value. However the validity of the exercise requires significant extrapolation over changes in resources that are not observed in the data. Section 6 concludes by scaling down the size of the optimal reform. The result is a more modest expansion in transfers that is better than revenue neutral, produces welfare gains for all types, and leads to improvements in the cognitive and behavioral skills of most (but not all) children. For this reform in particular, heterogeneity in the relative quality of care is an important driver of differences in cognitive skill impacts across types.

The quantitative exercises in this paper connect four literatures. The first seeks to discipline theories of optimal taxation with microeconomic evidence, either taking as given a redistributive objective (Saez, 2001, 2002; Blundell and Shephard, 2011) or building in a role for transfers through incomplete insurance markets (Heathcote, Storesletten, and Violante, 2017). To the extent that papers in this literature consider transfers aimed at households with children (Guner, Kaygusuz, and Ventura, 2020; Bruins, 2019; Ho and Pavoni, 2020), they do not consider the human capital spillovers to children and how this affects policy conclusions.[2] Another literature studies the investment behavior of parents and the resultant human capital outcomes of children through a technology of skill formation, either modeling explicitly the demand for inputs as a maximization problem (Del Boca, Flinn, and Wiswall, 2014; Bernal, 2008; Griffen, 2019; Brilli, 2022; Caucutt, Lochner, Mullins, and Park, 2020; Gayle, Golan, and Soytas, 2015), or estimating directly the demand for inputs in reduced form (Cunha, Heckman, and Schennach, 2010; Attanasio, Meghir, and Nix, 2020). A third empirical literature documents direct empirical evidence on the role played by economic resources in shaping short and long-run child outcomes.[3] Finally, by using the model to study the behavior of single mothers under welfare reform, this paper repeats past efforts to understand program participation and labor supply behavior and quantify the welfare effects of policy reforms (Hoynes, 1996; Chan, 2013; Keane and Wolpin, 2010).

2 Model

This section describes a dynamic model of work, program participation, and child investment. The model features the minimal set of ingredients to allow for interpretation of the data and articulation of the tradeoffs faced by a government that can appropriate and redistribute resources. It features a particular combination of Cobb-Douglas production and log preferences that greatly simplifies the interaction between child skills in any period and the rest of the dynamic problem, as in Del Boca, Flinn, and Wiswall (2014). The model extends the framework of Del Boca et al. (2014) in that it adopts a multidimensional skill technology, and embeds the simplification inside a dynamic, complex, policy environment. It also augments the production technology to consider the quality of non-maternal care,

[2] A close counterpart that does not study optimal policy is Caucutt and Lochner (2020). The authors estimate a model of household investment in children with lifecycle borrowing constraints. They use the model to evaluate policies that alleviate borrowing constraints at different life-cycle stages.

[3] In addition to the papers cited above, Aizer et al. (2022) provide a useful review of this literature.

following Chaparro, Sojourner, and Wiswall (2020). Unlike Del Boca et al. (2014), the model considers only one parent households.

2.1 Environment, Demographics, and Initial Conditions

Time is discrete, with one period in the model equal to one year. Mothers are the sole decision-making agents in the economy. Throughout the paper, y indexes calendar years and t enumerates the number of periods since the beginning of an individual's decision problem, with the first period beginning at $t = 0$. Births are determined by an exogenous sequence $\boldsymbol{b} = \{b_f\}_{f=1}^{F}$ of birth years for each mother, where $F = |\boldsymbol{b}|$ is the total number of births and f indexes each child. Let $T(\boldsymbol{b})$ indicate the finite horizon of the problem, which in the baseline model is equal to $b_F + 18$, the period in which the last-born child reaches maturity.[4] The remainder of the paper suppresses dependence of the finite horizon T on the the birth sequence $\boldsymbol{b}$. To introduce notation and lay out the remaining dimensions of the environment, below is the vector s of variables that will characterize an agent's decision problem:

$$s = \left\{ \begin{array}{cr} k, & \text{(a variable that indexes mother's type)} \\ \boldsymbol{b} = \{b_f\}_{f=1}^{F}, & \text{(the period } b_f \text{ in which each child } f \text{ is born)} \\ Age_0, & \text{(the age of the mother in the first period)} \\ \mathbf{Z} = \{Z_t\}_{t=0}^{T(B)}, & \text{(the full sequence of policy environments the mother will face)} \\ \varepsilon, & \text{(the stochastic component of mother's labor market productivity)} \\ \omega, & \text{(cumulative periods of welfare participation)} \\ t & \text{(the number of periods since the beginning of the problem).} \end{array} \right\}$$

The economy in each year y is populated by a measure $\pi_y(s)$ of single mothers in state s. Each mother in the economy belongs to one of K discrete types, indexed by k. An individual's type is known to them but unobservable to the econometrician and to the planner. It determines their preferences, their productivity in the labor market, and their productivity in producing child skills.

In addition to the state vector s, the skills of each child f will also determine individual payoffs. Children are characterized at time t by a dynamic 2-dimensional vector of characteristics

$$\theta_{f,t} = [\theta_{f,t,c},\ \theta_{f,t,b}],$$

representing their stock of *cognitive* (c) and *behavioral* (b) skills. These skills can only evolve after the child is born and before they reach age 18. When a child reaches age 18 their skills are no longer malleable, indicating the end of the investment problem. Let $a(f,t) = t - b_f$ be the age of each child f in period t. Two vectors will be useful when describing the model and its solution:

$$\mathcal{A}(t) = \{f : 0 \leq a(f,t) < 18\}, \qquad \boldsymbol{a} = \{a(f,t) : \forall\ f \in \mathcal{A}(t)\}.$$

The first contains all children in the household who are in their developmental period, the second contains the age of each of those children. The dependence of these variables on the period t and the birth sequence $\boldsymbol{b}$ will remain implicit in the remainder of the paper.

2.1.1 Initial conditions

At $t = 0$, mothers randomly draw a birth sequence $\boldsymbol{b}$, initial age Age_0, and policy sequence $\mathbf{Z}$ from an exogenous joint distribution, $\Pi_{\mathbf{X}}$. She draws her type k from a discrete distribution $\Pi_K(\cdot|\mathbf{X})$ where $\mathbf{X} \supset \{\boldsymbol{b}, Age_0, \mathbf{Z}\}$ is a larger set of exogenous observables. She draws an initial wage shock ε_0 independently from a distribution Π_{ε}^{0}, and she begins with cumulative participation $\omega_0 = 0$.

[4] Concluding the model when all children mature is without loss of generality here, since in this final period, no decisions can be made to influence future pay-off relevant states. This would not be true if, for example, mothers were permitted to save through a risk-free bond or accumulate human capital themselves.

Outside of the vector s, each child's initial skill is set as a deterministic function of type, $\theta_{f,0} = \mu_{\theta,k,-1}$. A technology of skill formation determines the stochastic evolution of each skill once children enter their developmental period at age 0.

2.2 Choices

In every period mothers makes a discrete choice, j, that determines her weekly hours of work ($H_j \in \{0, 15, 30, 45\}$), whether she participates in food stamps ($S_j \in \{0, 1\}$), and whether she participates in cash welfare ($A_j \in \{0, 1\}$).[5] As in Chan (2013), participation in welfare will imply participation in food stamps, but not vice versa. This results in 4 hours choices and 3 overall program participation choices, giving 12 discrete choices in total.

In addition to the discrete choice, each mother decides each period on private consumption (C), leisure (l), the fraction of non-market time invested in each developing child ($\phi_\tau = \{\phi_{\tau,f}\}_{f\in\mathcal{A}}$) and money investment in each developing child ($x = \{x_f\}_{f\in\mathcal{A}}$).[6] An extension of the baseline model also allows for mothers to make an additional choice regarding the quality of childcare arrangement, φ_f, while working (this quality is determined exogenously in the baseline model).

2.3 Technology and Government Policies

This section introduces the determinants of a mapping $\boldsymbol{Y}_j(s)$ from an individual's discrete choice (j), and their current state (s, which includes individual and policy variables), to net income for the period, as well as the subsequent constraints on the allocation of time and money resources to private consumption or investment in children.

2.3.1 Wages and Earnings

Each mother faces a wage $W(s)$ in the labor market that generates earnings $E_j(s) = W(s)H_j$ where H_j is the hours choice associated with choice j. Wages are determined by the equation:[7]

$$\log(W(s)) = \gamma_{W,k,0} + \gamma_{W,k,1}Age_t + \gamma_{W,k,2}Age_t^2 + \sigma_\varepsilon\varepsilon$$

where $Age_t = t - Age_0$ is the age of the mother at time t and ε is a discrete random variable that takes up to K_ε uniformly spaced values on the interval $[-1, 1]$. Letting $\mathcal{E} = \{\varepsilon_1, ..., \varepsilon_{K_\varepsilon}\}$ be the grid space, transition probabilities are symmetric with reflection at the top and bottom of the grid:

$$\Pi_\varepsilon(\varepsilon_{k'}|\varepsilon_k) = \mathbf{1}\{k' = k\}\pi_\varepsilon + \mathbf{1}\{k' = \min\{k+1, K_\varepsilon\}\}(1-\pi_\varepsilon)/2 + \mathbf{1}\{k' = \max\{k-1, 1\}\}(1-\pi_\varepsilon)/2. \tag{1}$$

The heterogeneous vector $\gamma_{W,k}$ determines fixed differences in the life-cycle profile of labor market productivity across types. Since type is unobserved, both sources of heterogeneity in individual wages are unobserved by the econometrician. Since type is known to the individual, the life-cycle profile determined by $\gamma_{W,k}$ is known, while future realizations of ε are not. The parameter σ_ε determines the dispersion of the shocks.

2.3.2 Government Policies

Let Z_t be the vector of government policy rules that apply to the individual at time t. This section offers a general summary of the components of Z_t that determine the *net transfer* from the government, which is given by the sum of food stamps (SNAP), cash welfare (AFDC/TANF), and taxes.

[5] The introduction and the next section identify food stamps in the US as the Supplemental Nutrition Assistance Program (SNAP), and cash welfare as Aid to Families with Dependent Children (AFDC), and later Temporary Assistance for Needy Families (TANF). The model will follow the rules of eligibility and payments for these programs as closely as possible.

[6] Appendix C demonstrates that assuming either fully private or fully public investments in children entail an identical outcome equation for child skills and is therefore without loss of generality.

[7] Minimum wages are not imposed for this analysis.

Time Limits The introduction of this paper identified time limits as an important component of the changes to welfare policies introduced by PRWORA. This is modeled as a lifetime limit on cash transfers from AFDC/TANF, denoted by Ω.

Let $Z_{TL,t} \in \{0,1\}$ indicate whether time limits are imposed at time t. When $Z_{TL,t} = 1$, the endogenous state variable ω tracks accumulated periods of welfare usage. It evolves according to:

$$\omega_{t+1} = \omega_t + AZ_{TL} \tag{2}$$

where A is the mother's welfare participation choice in period t.

Work Requirements Another prominent feature of welfare reform was the introduction of work requirements, which mandated that all non-exempt welfare participants either meet a minimum weekly hours requirement of 30 hours, or participate in state administered job search or employment training activities. Given the inexact nature of how work requirements are enforced, this paper models work requirements as a parameterized policy that simultaneously makes work less costly for individuals who are participating in welfare, and while also imposing additional nonpecuniary costs on welfare participants who do not work. This assumption is based on the lack of evidence that employment or training programs had any effect on the wages of participants, but does appear to have an effect on the number of individuals who work while participating in welfare (Mullins, 2025). $Z_{R,t}$ is a random binary variable that indicates whether policy environment at time t involves work requirements. $Z_{R,t}$ will feature in mothers' utility (described below) but not explicitly in determining welfare payments.

Transfers The net transfer from the government to a household with food stamps choice S, welfare choice A, and earnings E is given by the function:

$$\mathcal{T}(s, S, A, E) = ST^S(Z_{S,t}, E, \tilde{F}_t) + A\mathbf{1}\{\omega_t < \Omega\}T^A(Z_{A,t}, E, \tilde{F}_t) + T^T(Z_{T,t}, E, \tilde{F}_t)$$

where $\tilde{F}_t$ is the total number of dependent children, and the triple $(Z_{S,t}, Z_{A,t}, Z_{T,t})$ contains all the parameters that determine net transfers from food stamps (S), welfare (A), and income taxes (T). These vectors include variables determining gross and net income eligibility requirements, benefit standards that determine overall generosity of programs based on family size, earnings disregards for welfare payments, marginal tax rates, and child-specific tax credits such as the Earned Income Tax Credit (EITC) and Child Tax Credit (CTC). Appendix G gives a thorough description of how these functions are calculated. Eligibility standards are imposed in the model by setting transfers to zero when a mother is not eligible for a program.

It is convenient to simplify the expression for net household income to the relationship $\boldsymbol{Y}_j(s)$ where

$$\boldsymbol{Y}_j(s) = E_j(s) + \mathcal{T}(s, S_j, A_j, E_j(s)).$$

Figure 1 provides a sense of the variation in incentives created by differences in policy over time and across states. It plots the monthly transfer for a single mother of two in two states – California and Florida – as a function of monthly earnings in the years 1990 and 2000. To capture incentives at the extensive margin, Figure 1 also plots the corresponding participation tax rates (PTR), showing substantial variation in work incentives across states and over time. Of particular note is the large reduction in participation rax rates over time: a consequence of welfare reform and expansions in the EITC.

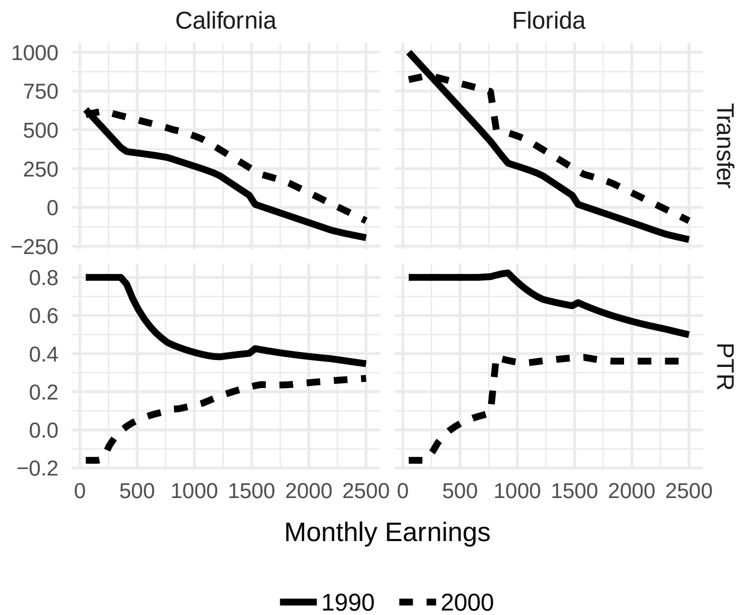
2.3.3 Resource Constraints

A budget constraint and time constraint each apply to the mother's decisions:

$$C + \sum_{f\in\mathcal{A}} x_f \leq Y_j(s) \tag{3}$$

$$l + \sum_{f\in\mathcal{A}} \phi_{\tau,f} \leq 1. \tag{4}$$

Figure 1: Transfer Policies and Tax Rates: Example



This figure shows a comparison of transfers to single mothers with two children in Florida and California in the years 1990 and 2000. The bottom panel shows the participation tax rate (PTR) for each combination of state and year, while the top panel shows the monthly transfer in year 2000 USD.

Equation (4) is a constraint that dictates how *non-market* time is allocated as a fraction of the total, and applies independently of the hours choice.

Although standard, these resource constraints highlight the role for government transfer policies to shape child development outcomes through the effect they have on time and money constraints in the household.

2.3.4 Child Development

Technology of Skill Formation Mothers can contribute to their developing children's skills in each period t by purchasing market inputs (x_f) and by spending a fraction $\phi_{\tau,f}$ of their non-market time in enriching activities with the child. While $\phi_{\tau,f}$ determines the quality of the child's environment when the mother is present, they face a different quality when the mother is working, $\varphi_{\tau,f}(s)$, which is determined exogenously.[8] Overall time input quality, τ, is aggregated according to a Cobb-Douglas function, with endogenous factor shares equal to the fraction of time the child spends in each arrangement, $h_j = H_j/112$:

$$\tau_f = \phi_{\tau,f}^{1-h_j} \varphi_{\tau,f}(s)^{h_j}.$$

Naturally, under this specification the quality of external care (φ) is only relevant when work hours (and hence h) are positive.

Indexing inputs and skills by t, evolution of each child skill $i \in \{c,b\}$ is also determined by a Cobb-Douglas production function, which is specified for a child of age $a = a(f,t)$ as:

$$\theta_{f,t+1,i} = \exp(\mu_{\theta,k,a,i} + \eta_{f,t,i})\tau_{f,t}^{\delta_{\tau,a,i}} x_{f,t}^{\delta_{x,a,i}} \theta_{f,t,c}^{\delta_{\theta,c,i}} \theta_{f,t,b}^{\delta_{\theta,b,i}}, \qquad i \in \{c,b\} \tag{5}$$

where $\delta_{\tau,a,i}$ is the Cobb-Douglas share of time investment at age a for the production of skill i, $\delta_{x,a,i}$ is the equivalent for money investment, and $\delta_{\theta,i',i}$ is the share of skill i' in the production of skill i. Each mother may differ in their

[8]This quality, though exogenously determined, is permitted to be correlated with preferences and technology through dependence on type.

innate ability for raising children, represented by the time invariant productivity parameter $\mu_{\theta,k,a} = [\mu_{\theta,k,a,c},\ \mu_{\theta,k,a,b}]$, while the vector $\eta_{f,t} = [\eta_{f,t,c},\ \eta_{f,t,b}]$ is a mean zero[9] shock to total factor productivity that is independent over time and across individuals. Importantly, no such restriction is placed on $\mu_{\theta,k,a}$ which may be freely correlated with preferences and productivities through its dependence on type.

2.4 Preferences

Mothers derive utility each period from their private consumption (C), the fraction of non-market time spent in leisure (l), the skills of all children currently born (θ), and the discrete choice (j). Outcomes are ranked by the expected discounted present value of utility and a terminal payoff that depends on the final skills of all children:

$$V_t = \mathbb{E}_t \left\{ \sum_{r=t}^{T} \beta^{r-t} U_j(C_r, l_r, s_r\theta_r, \epsilon_r) + \beta^{T+1-t} V_{T+1}(s_{T+1}, \theta_{T+1}) \right\} \tag{6}$$

where $\mathbb{E}_t$ is the expectation operator conditional on the mother's information set at the beginning of time t, before the realization of preference shocks ϵ and development shocks η. In order to write the functional form of utility, let $h_j = H_j/112$ be the fraction of weekly hours spent in the labor market. Utility is given by

$$U_j(C, l, s, \theta, \epsilon) = \log(C) + \alpha_l[(1-h_j)\log(l) + h_j \log(\ell_k)] + \alpha_{\theta,k} \left(\sum_{f:a(f,t)\geq 0} \sum_{i=c,b} \alpha_{\theta,i} \log(\theta_{f,t,i}) \right) + \alpha_j(s) + \epsilon_j$$

and the terminal payoff from child skills is:

$$V_{T+1}(s, \theta) = (1-\beta)^{-1} \alpha_{\theta,k} \sum_{f=1}^{F} \sum_{i=c,b} \alpha_{\theta,i} \log(\theta_{f,i}).$$

Notice that the leisure value of time at home (l) is aggregated with the leisure value of time in the market (ℓ_k), with factor shares equal to the fractions of time spent in each place. The leisure value of market time, ℓ_k, is taken as fixed and parameterizes the cost of working additional hours. For example, in the case without children when $l = 1$, if $\ell_k < 1$, this specification gives a linear disutility in additional work hours with heterogeneous coefficients across types.

The innate utility derived from each discrete choice j consists of two stochastic components: a state-dependent component $\alpha_j(s)$ and an idiosyncratic component ϵ_j. Furthermore, the vector of innate utilities from discrete choices, $\{\alpha_j(s)\}_{j=0}^{J-1}$, is restricted in the model to take the form:

$$\alpha_j(s) = -\alpha_{A,k}A_j - \alpha_{S,k}S_j - \alpha_{H,k}\mathbf{1}\{H_j > 0\} - Z_{R,t}A_j\mathbf{1}\{H_j < 30\}\alpha_{R,1} + Z_{R,t}A_j\mathbf{1}\{H_j > 0\}\alpha_{R,2}$$

where $\alpha_{A,k}$ and $\alpha_{S,k}$ are heteregeneous costs of participating in AFDC/TANF and SNAP, $\alpha_{H,k}$ is a heterogeneous cost of participating in the labor market, $\alpha_{R,1}$ parameterizes how work requirements may increase the cost of not working while participating in welfare, and $\alpha_{R,2}$ parameterizes how work requirements may decrease the cost of work while participating in welfare.

Mothers differ in terms of how they value the skills of their children and each discrete choice. In the latter case, the utility derived from each choice j consists of three time invariant components $(\alpha_{F,k}, \alpha_{A,k}, \alpha_{H,k})$ that are fixed for each mother over time, and a component ϵ that is identically and independently distributed over time and across mothers. The vector ϵ is distributed as a nested logit with three layers. The outermost nest contains each of the program participation choices, the middle layer contains the binary decision of whether or not to work, while the innermost layer contains the hours choice. Let $\sigma = (\sigma_P, \sigma_{H,0}, \sigma_{H,1})$ be a three-dimensional vector containing the dispersion parameters for each layer.

The heterogeneous weight on child skills ($\alpha_{\theta,k}$) is a scalar variable that scales a homogenous aggregate given by the pair $(\alpha_{\theta,c}, \alpha_{\theta,b})$. Later sections of the paper discuss identification of the model and will demonstrate that this latter pair of weighting parameters need not be identified for the purposes of policy and prediction.

[9] This is a normalization without loss of generality.

2.5 Information

Uncertainty about future wage realizations and utility shocks are the two key sources of risk for agents in the model. Formally, while the pair $(\epsilon_t, \varepsilon_t)$ is observable at time t before decisions are made, future realizations of these variables are unknown. The same information structure is imposed for realizations of the vector of TFP shocks in child development, η_t. Section 2.8 demonstrates that these shocks enter additively in mothers' dynamic values, having no affect on decision-making. In this sense, assumptions on whether these shocks are known or forecastable have no effect on the model's solution properties.

As is implied by their membership of the state vector s, mothers have perfect information about the path of fertility outcomes (the number and timing of all births, $\boldsymbol{b}$) as well as the path of policy variables, $\mathbf{Z}$. Appendix C.2 discusses the computational complexity of the model and how this assumption swaps the very high dimensional space spanned by Z_t and $\boldsymbol{b}$ for the (much smaller) size of the sample, M. This simplification is necessary for model tractability, though it does have some empirical content which Appendix **?** also discusses.

2.6 Adult Outcomes

The life-cycle earnings of children is a policy-invariant function of their skills once they reach age 18. In order to weigh the short-run costs of transfers to households with children against the long-run benefits, assume a linear relationship:

$$\mathcal{Y}(\theta) = \tilde{\Gamma}_{18,c} \log(\theta_c) + \tilde{\Gamma}_{18,b} \log(\theta_b) \tag{7}$$

where $\mathcal{Y}(\theta)$ is the net present value of resources (in the form of labor market earnings) contributed to the economy by a child who reaches maturity with skills θ.

2.7 Recursive Formulation

Mothers choose contingent plans for consumption, leisure, time and money investment, and discrete choices, to maximize the objective written in (6) subject to intratemporal constraints on time and money and intertemporal dynamics defined by time limits, wage risk, and the technology of skill formation. This section formulates the problem recursively, so that the next section can reduce it to a discrete choice dynamic program for estimation.

Integrating out taste shocks (ϵ) and TFP shocks (η) which are both independently distributed over time, the full decision problem has a recursive formulation in terms of the state vector (s, θ):

$$V_t(s,\theta) = \int \max_{j,c,l,x,\phi_\tau} \left\{ U_j(c,l,s,\theta,\epsilon) + \beta \mathbb{E}[V_{t+1}(s',\theta')|s,\theta,j,x,\phi_\tau] \right\} dF_\epsilon(\epsilon) dF_\eta(\eta)$$

subject to the resource constraints (3)-(4), the transition rules governed by (2) and (1), and the technology of skill formation governed by (5). Given the finite horizon of the problem, the recursion can be initialized at the terminal period:

$$V_T(s,\theta) = \int \max_{j,c,l,x,\tau} \{U_j(c,l,s,\theta,\epsilon) + \beta V_{T+1}(\theta)\} \, dF_\epsilon(\epsilon).$$

Even though the time period t belongs to the state vector s, it helps exposition to make explicit the dependence of the value function V on t.

2.8 Model Solution and Empirical Content

The two most important features of the model's solution are (1) the derivation of an indirect utility function, with dynamic coefficients that evolve according to the developmental importance of money and time investments of children in the household; and (2) an outcome equation that describes skill evolution as a function of net household income and mothers' labor supply decisions, with an endogenous residual that depends on mothers' unobserved preferences and abilities. These two features of the model spell out the model's key mechanisms in terms of a small number of parameters, and provide clarity regarding challenges and solutions for identification of these parameters.

The key to the following analytic derivations is that $\log(\theta)$ is, thanks to the Cobb-Douglas specification, additively separable in all lagged inputs, and that mothers value $\log(\theta)$ in a linear fashion. Thus, the discounted present value of skills today (and hence investment) has a recursive representation:

$$\Gamma_{a,i} = \begin{cases} (1-\beta)^{-1}\alpha_{\theta,i} & \text{for } a \geq 18 \\ \alpha_{\theta,i} + \beta\left[\delta_{\theta,i,c}\Gamma_{a+1,c} + \delta_{\theta,i,b}\Gamma_{a+1,b}\right] & \text{for } 0 \leq a < 18 \\ \beta\Gamma_{a+1,i} & \text{for } a < 0 \end{cases} \tag{8}$$

Appendix C shows in more detail that this additive separability results in an additively separable value function:

$$V_t(s,\theta) = \nu_t(s) + \alpha_{\theta,k}\sum_{f:a(f,t)\geq 0}\left(\Gamma_{a(f,t),c}\log(\theta_c) + \Gamma_{a(f,t),b}\log(\theta_b)\right)$$

where

$$\nu_t(s) = \int \max_j \left\{u_{k,j}(\mathbf{Y}_j(s),\boldsymbol{a}) + \alpha_j(s) + \epsilon_j + \beta\mathbb{E}[\nu_{t+1}(s')|s,j]\right\} dF_\epsilon(\epsilon) \tag{9}$$

subject to the transition rules that govern accumulated welfare use (2) and wages (1). Since F_ϵ is specified as a nested logit, the integral in (9) has an analytical solution, as does the probability of any choice j given the state s. The function u is an indirect utility function evaluated at optimal input choices, which the next section derives.

2.8.1 Indirect Utility and Optimal Inputs

Indirect utility derives from the solution to the maximization problem:

$$u_{k,j}(y,\boldsymbol{a}) = \max_{C,l,x,\phi}\Bigg\{\log(C) + \alpha_l[(1-h_j)\log(l) + h_j\log(\ell_k)] \\ + \alpha_{\theta,k}\left(\sum_{f\in\mathcal{A}}\Gamma_{x,a(f,t)}\log(x_f) + \Gamma_{\tau,a(f,t)}[(1-h_j)\log(\phi_{\tau,f}) + h_j\log(\varphi_{\tau,f})\right)\Bigg\}$$

subject to the resource constraints:

$$C + \sum_f x_f \leq y, \qquad l + \sum \phi_{\tau,f} \leq 1. \tag{10}$$

The coefficients $\Gamma_{x,a}$ and $\Gamma_{\tau,a}$ are the return to a log-point increase in time and money inputs:

$$\Gamma_{x,a} = \beta(\delta_{x,a,c}\Gamma_{a+1,c} + \delta_{x,a,b}\Gamma_{a+1,b}), \quad \Gamma_{\tau,a} = \beta(\delta_{\tau,a,c}\Gamma_{a+1,c} + \delta_{\tau,a,b}\Gamma_{a+1,b}).$$

This results in optimal input choices:

$$\phi_{x,f} = x_f/y = \frac{\alpha_{\theta,k}\Gamma_{x,a(f,t)}}{1+\sum_{a'\in\boldsymbol{a}}\alpha_{\theta,k}\Gamma_{x,a'}} \tag{11}$$

$$\phi_{\tau,f} = \frac{\alpha_{\theta,k}\Gamma_{\tau,a(f,t)}}{\alpha_l+\sum_{a'\in\boldsymbol{a}}\alpha_{\theta,k}\Gamma_{\tau,a'}}. \tag{12}$$

To simplify identification and estimation of the free parameter $\varphi_{\tau,f}(s)$, the model assumes that:

$$\varphi_{\tau,f}(s) = \overline{\varphi}_k\frac{\alpha_{\theta,k}\Gamma_{\tau,a(f,t)}}{\alpha_l+\sum_{a'\in\boldsymbol{a}}\alpha_{\theta,k}\Gamma_{\tau,a'}}$$

where the parameter $\overline{\varphi}_k$ parameterizes the relative quality of non-maternal care for type k. This results in an indirect utility function:

$$u_{k,j}(y,\boldsymbol{a}) = (1+\alpha_{\theta,k}\Gamma_x(\boldsymbol{a}))\log(y) + h_j\left(\alpha_l\log(\ell_k) + \alpha_{\theta,k}\Gamma_\tau(\boldsymbol{a})\log(\overline{\varphi}_k)\right) + g_k(\boldsymbol{a}) \tag{13}$$

where

$$\Gamma_x(\boldsymbol{a}) = \sum_{a\in\boldsymbol{a}}\Gamma_{x,a}, \qquad \Gamma_\tau(\boldsymbol{a}) = \sum_{a\in\boldsymbol{a}}\Gamma_{\tau,a}$$

and $g_k(\boldsymbol{a})$ is an additive constant that depends on preferences and the ages of all children, but is not affected by any past or future decisions (and so can be ignored henceforth). Notice that this expression for indirect utility encodes *dynamic* incentives related to the importance of money and time inputs for each child in the household. At the same time, if discrete choices do not affect any other pay-off relevant state variables, the dynamic model reduces to a static discrete choice model, with coefficients on net income and market time that are functions of household demographics. In the baseline model, the only other source of these dynamics appears when individuals are subject to time limits in participation on welfare ($Z_{TL,t} = 1$).

Choice probabilities Let $v_j(s)$ be the choice-specific value of option j in state s:

$$v_j(s) = u_{k,j}(\boldsymbol{Y}_j(s), \boldsymbol{a}) + \beta \mathbb{E}[\nu_{t+1}(s')|s, j].$$

Given the vector $\{v_j(s)\}_{j=0}^{J-1}$, the optimal choice probabilities $P_j(s)$ have the standard analytical expression following a nested logit formula.

2.8.2 Outcome Equation

Substituting the optimal input rules (11)-(12) into the production function yields a linear vector autoregressive system with one lag period:

$$\log(\theta_{f,t+1}) = \mu_{\theta,k,a} + \delta_{x,a}\log(Y_t) + \delta_{\tau,a}\log(\overline{\varphi}_k)h_t + \delta_\theta \log(\theta_{f,t}) + e_{f,t} + \eta_{f,t} \tag{14}$$

where the error term $e_{f,t}$ is defined as:

$$\delta_{\tau,a}\log(\phi_{\tau,f}) + \delta_{x,a}\log(\phi_{x,f}).$$

Each of $(\theta_{f,t}, \mu_{\theta,k}, \delta_{x,a}, \delta_{\tau,a})$ are two-dimensional vectors with one component for each skill outcome. The matrix δ_θ defines the persistence of this system through self-productivity of skills:

$$\delta_\theta = \begin{bmatrix} \delta_{\theta,c,c} & \delta_{\theta,b,c} \\ \delta_{\theta,c,b} & \delta_{\theta,b,b} \end{bmatrix}.$$

Since the error term $e_{f,t}$ is a function of preferences only, and invariant to policy, equation (14) demonstrates that the effects of any welfare policy reform on child outcomes can be calculated by forecasting its effects on labor supply and net household income. This result suggests a clear route to identification of production parameters (which Section 5.2 will discuss) and illustrates how changes to taxes and transfers affect child outcomes through their impact on net income and labor supply. While the effect of net income on skills depends only on child age, the effect of additional hours is heterogeneous and depends on the relative quality $\overline{\varphi}_k$ of external care.

2.9 Model Extensions

In order to evaluate the robustness of its main policy conclusions, this paper also considers two extensions to the baseline model. The first extension allows for mothers themselves to accumulate human capital through labor market participation. This is achieved by allowing for the transition matrix of the wage shock, Π_ε, to depend on whether or not mothers are working at least 30 hours a week. Since this dynamic decision affects payoffs even after children mature, this version of the model extends the finite horizon until mothers reach age 65, allowing them to continue to make labor supply and food stamp participation decisions.

The second extension allows for the quality of care φ to be purchased at a uniform price. The main substantive implication of this framework appears in the child's outcome equation, with the addition of an interaction term between net income and hours of work.

In both cases, estimation yields parameter estimates that undermine the mechanism of interest for these model extensions, and they do not feature in the policy analysis of the paper. Appendix D provides more details. Appendix D also presents the results of a regression analysis that tests one of the model's stronger implications: that investment choices do not depend on the child's current stock of skills. The optimal time investment rules (12) implies a test of this implication using a specification with mother by year fixed effects and finds no strong evidence to reject this property of the model.

3 Planner's Problem

3.1 The Planner's Objective

This section studies the design of taxes and transfers by introducing a planner that would like to maximize the welfare of single mothers subject to a linear opportunity cost of resources. Formally, the planner ranks policies according to the objective:

$$\boldsymbol{W}_y = \int (\boldsymbol{\mu} V_t(s,\theta) + \boldsymbol{\lambda} R_t(s,\theta)) d\pi_y(s,\theta) \tag{15}$$

where $R_t(s,\theta)$ is the expected net present value of resources (net of costs) derived from policies for a family in state s with child skills θ. The parameter $\boldsymbol{\mu}$ is the weight placed on mother's welfare, V_t, which is defined under optimal behavior given the policies decided on by the planner. The parameter $\boldsymbol{\lambda}$ is the Lagrange multiplier on the planner's net present value resource constraint, the other components of which are not modeled. The measure π_y is defined for year y over states s (including the period t) and child skills θ.[10] The logic of this setup is that transfers to single mothers are one of many programs administered by the planner. The parameter $\boldsymbol{\lambda}$ captures the opportunity cost of spending an extra dollar on transfers to this group, or equivalently the value of transferring a marginal dollar to one of these other programs. The ratio $\boldsymbol{\mu}/\boldsymbol{\lambda}$ indexes the planner's overall taste for distribution to single mothers and will be sufficient for the analysis below.

Section 2 showed that the value function V_t is additively separable in log-skills. Key to this representation was a pair of recursive coefficients $(\Gamma_{a,c}, \Gamma_{a,b})$ that summarized the present discounted value of cognitive (c) and behavioral (b) skills for a child at age a. The Planner's problem is also characterized by a pair of coefficients $(\tilde{\Gamma}_{a,c}, \tilde{\Gamma}_{a,b})$, that have an almost identical derivation. Appendix F demonstrates the resource function R_t, just like V_t, has a recursive, additive, representation:

$$R_t(s,\theta) = r_t(s) + \sum_{f:a(f,t)\leq 18} (\tilde{\Gamma}_{a(f,t),c} \log(\theta_{f,c}) + \tilde{\Gamma}_{a(f,t),b} \log(\theta_{f,b}))$$

where:

$$r_t(s) = \sum_j P_j(s) \left[E_j(S) - \boldsymbol{Y}_j(s) + \tilde{\Gamma}_x(s) \log(\boldsymbol{Y}_j(s)) + \tilde{\Gamma}_\tau(s) h_j \log(\overline{\varphi}_k) + \beta \mathbb{E}[r_{t+1}(s')|s,j] \right]$$

The planner values the skills of children because when they mature at age 18, they contribute a total:

$$\mathcal{Y}(\theta) = \tilde{\Gamma}_{18,c} \log(\theta_c) + \tilde{\Gamma}_{18,b} \log(\theta_b)$$

in net present value earnings to the planner's resource constraint.[11] The net present value to the planner of log-skills at younger ages can then be expressed recursively as:

$$\tilde{\Gamma}_{a,i} = \beta \left(\delta_{\theta,i,c} \tilde{\Gamma}_{a+1,c} + \delta_{\theta,i,b} \tilde{\Gamma}_{a+1,b} \right), \qquad i \in \{c,b\}.$$

Substituting in the outcome equation (14) then delivers a formula for how the planner values the skill return to log income $(\log(\boldsymbol{Y}_j(s)))$ and hours of work $(h_j \log(\overline{\varphi}_k))$ in the above expression:

$$\tilde{\Gamma}_x(s) = \beta \sum_{a \in \boldsymbol{a}} \left(\delta_{x,a,c} \tilde{\Gamma}_{a+1,c} + \delta_{x,a,b} \tilde{\Gamma}_{a+1,b} \right) \tag{16}$$

$$\tilde{\Gamma}_\tau(s) = \beta \sum_{a \in \boldsymbol{a}} \left(\delta_{\tau,a,c} \tilde{\Gamma}_{a+1,c} + \delta_{\tau,a,b} \tilde{\Gamma}_{a+1,b} \right) \tag{17}$$

[10] Note that this objective places no weight on future cohorts of mothers and children. This assumption is without loss of generality when assessing transfer programs that are static, which are the main policies of interest in this paper. In dynamic settings, maintaining this assumption provides a tractable benchmark with which to rank policies without making additional assumptions on the evolution of childbirth patterns in the future.

[11] Counting total earnings in the resource constraint is consistent with the treatment of earnings in much of public finance (Kline and Walters, 2016; Hendren and Sprung-Keyser, 2020). It will also obtain if an altruistic planner can equate the marginal utility of children's consumption with the marginal cost of resources $\boldsymbol{\lambda}$.

The coefficients $(\tilde{\Gamma}_x(s), \tilde{\Gamma}_\tau(s))$ sum the return to money and time inputs over the ages of each child in the household s.

The recursive representation of the planner's problem facilitates numerical computation of optimal policies in Section 6 when features of the policy – such as time limits – create dynamic interactions between choices and payoffs. The next section examines optimal policies for a key benchmark of interest in the paper: a static function $\boldsymbol{y}$ that maps earnings to net income. In this case, the dynamic objective can reduce to a static[12] problem that resembles classic public finance analyses (Mirrlees, 1971; Diamond, 1980; Saez, 2001).

3.2 Optimal Static Transfers

This section considers the optimal design of a policy function $\boldsymbol{y}$ that assigns each individual a resource allocation taking only earned income as an input, i.e. $\boldsymbol{Y}_j(s) = \boldsymbol{y}(E_j(s))$.

This case simplifies both the decision problem for the planner and for individuals in the economy. Note that this policy removes any participation choice from the choice set, reducing the number of discrete choices to the four choices of hours, $H_j \in \{0, 15, 30, 45\}$. It also removes dynamics from the agent's discrete choice, allowing them to solve the static problem:

$$\max_j u_{k,j}(\boldsymbol{y}(E_j(s)), \boldsymbol{a}) + \epsilon_j.$$

Similarly, the planner's choice of policy $\boldsymbol{y}$ in year y does not affect the distribution of payoff relevant states at $y+1$. Thus, if the planner is free to re-optimize at $y+1$, they can solve a static version of the problem:[13]

$$\max_{\boldsymbol{y}} \int \left[\boldsymbol{\mu}\mathbb{E}[\max_j u_{k,j}(\boldsymbol{y}(E_j(s)), \boldsymbol{a}) + \epsilon_j] + \boldsymbol{\lambda}\mathbb{E}\left[E_j(s) - \boldsymbol{y}(E_j(s)) + \tilde{\Gamma}_x(s)\log(\boldsymbol{y}(E_j(s))) + \tilde{\Gamma}_\tau(s)h_j\log(\overline{\varphi}_k)\right]\right] d\pi_y(s) \tag{18}$$

Appendix F rearranges first order conditions for $\boldsymbol{y}(0)$ and each $\boldsymbol{y}(e)$ to derive two formula that the optimal allocation $\boldsymbol{y}$ must satisfy. The first is an expression for optimal generosity:

$$\mathbb{E}[\boldsymbol{y}(E_j(s))] = \mathbb{E}\left[\frac{\boldsymbol{\mu}(1+\alpha_{\theta,k}\Gamma_x(s))}{\boldsymbol{\lambda}} + \tilde{\Gamma}_x(s)\right] \tag{19}$$

where the expectation is taken over states (s, including t) and choices j. Define

$$\boldsymbol{w}(s) = \frac{\boldsymbol{\mu}(1+\alpha_{\theta,k}\Gamma_x(s))}{\boldsymbol{\lambda}} + \tilde{\Gamma}_x(s)$$

as the "first best" allocation for a household in state s. In the absence of any behavioral responses, the planner would allocate resources to each household s by equating the marginal benefit – the marginal utility of income plus the marginal resource return – with the marginal cost ($\boldsymbol{\lambda}$), resulting in an allocation $\boldsymbol{w}(s)$. Equation (19) provides a useful insight: the planner's target for net income at the optimum depends only on marginal utilities and the factor share of money. As the returns to money inputs grow, the planner seeks to capitalize on the long-run returns by increasing overall generosity. Quantitatively, this formula provides the key to understanding why all of the optimal policies in Section 6 involve large increases in tranfers to single mothers.

The second expression decomposes the optimal allocation into:

$$\boldsymbol{y}(e) = \mathbb{E}[\boldsymbol{w}(s)|E_j(s)=e] + \mathbb{E}\left[\sum_i \eta_{ij}(s)\mathcal{R}_i(s)\middle|E_j(s)=e\right] \tag{20}$$

where the expectation $\mathbb{E}$ is again taken over states (s) and choices j and $\eta_{ij}(s)$ is the semi-elasticity of the probability of choice i ($P_i(s)$) with respect to a change in the utility from choice j. $\mathcal{R}_j(s)$ is the net resource benefit of making

[12]The problem is static in the sense that choices do not affect the state variables in s. It is still dynamic in the sense that the coefficients Γ and $\tilde{\Gamma}$ account for future payoffs in skills.

[13]Alternatively, one could similarly solve for $\boldsymbol{y}$ as a permanent choice and arrive at similarly tractable formulae, with a weighted average of static utilities that accounts for future distributions over states. This would result in different effective weights on each state s if π_y is non-stationary, without changing any of the key insights here.

choice j relative to choice 0 (not working). This is the sum of net earnings and allocations, and the net effect of choice j on skills:

$$\mathcal{R}_j(s) = \underbrace{E_j(s) - \boldsymbol{y}(E_j(s)) + \boldsymbol{y}(0)}_{\text{short-run effect}} + \underbrace{\tilde{\Gamma}_x(s)\log\left(\frac{\boldsymbol{y}(E_j(s))}{\boldsymbol{y}(0)}\right) + \tilde{\Gamma}_\tau(s)\log(\overline{\varphi}_k)}_{\text{long-run effect}=\mathcal{D}_j(s)} \tag{21}$$

Equation (20) offers useful insights into what drives the shape of optimal policy (and therefore labor supply incentives) by offering a formula for the wedge between first-best allocations with and without moral hazard. For example, if the returns to work are small and the care available to mothers is of lower relative equality, this will result in negative values of $\mathcal{D}_j(s)$, leading the planner to optimally reduce $\boldsymbol{y}$ for higher hours choices and dampen labor supply incentives. Section 6 will use (20) to decompose the wedge in a number of ways to assess the relative contributions of short and long-run costs to the overal shape of optimal allocations.

Overall, equations (19) and (20) highlight the central role that the technology of skill formation plays in shaping optimal allocations and underscores the importance of identifying them, along with the behavioral parameters that determine semi-elasticities η, as carefully as possible.

Appendix F.2 extends these results to cases where $\boldsymbol{y}$ is restricted to belong to some parametric class, as will be the case in practice for the quantitative section below. In particular, (19) remains true for all parametric classes in which $\boldsymbol{y}(e)$ can be scaled proportionally. The implications in (20) extend to the case where $\boldsymbol{y}(e)$ is a stepwise function.

4 Data

The model outlines a data generating process for a set of endogenous variables given a set of exogenous variables, where neither set is fully observable. To establish notation, let $\mathbf{X}$ be the vector of observable exogenous variables excluding those that define the policy environment. It contains the mother's initial age (at $t = 0$), the sequence of birth periods ($\boldsymbol{b}$), and years of education[14]. Let Z_t summarize the policy environment in period t, as described in Section 2.3.2, and let $\mathbf{Z} = \{Z_t\}_{t=1}^T$. Finally, let $\mathbf{Y}_t$ be the set of observable endogenous variables in period t, with $\mathbf{Y} = \{\mathbf{Y}_t\}_{t=1}^T$. This includes the discrete choice (j_t), observed hourly wages (if working), observed time investment ($\phi_{f,t}$), and child skill outcomes. Wages, time investment, and child skills will all be observed with additive measurement error. Individual type k, and the wage shock ε are exogenous variables, but are not observable in data.

Before turning to quantification of the mechanisms outlined in Sections 2 and 3, this section briefly describes the data inputs used to form a sample of observations $(\mathbf{X}_m, \mathbf{Y}_m, \mathbf{Z}_m)_{m=1}^M$ from the model. From this point, let $m = 1, 2, ...M$ index mothers in any such sample.

4.1 Data Source and Background

The *Panel Study of Income Dynamics* (PSID) and its *Child Development Supplement* (CDS) is the main source of data for analysis in this paper. The PSID is a dynastic, longitudinal survey taken annually from 1968 to 1997, and biennially since 1997. The main interview provides measures of household members' earnings and program participation, as well as relationships between household members. This allows construction of a complete marriage and fertility history, in addition to an incomplete history of work and participation decisions. Measures of other demographics such as race, education, and state of residence, are also available.

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. The next section provides further details.

[14] Years of education plays no direct role in the model but is allowed to affect the conditional probability of each type.

4.2 Description of Variables and Sample Selection

The CDS is comprised of several questionnaires. This paper uses two in particular: the child interview and the *primary caregiver* (PCG) interview. The Letter-Word (LW) and Applied Problems (AP) modules of the Woodcock-Johnson Aptitude test provide two measures of cognitive ability for children aged 3 and older. Two scales that measure *externalizing* (BPE) and *internalizing* (BPN) behavioral problems are used to measure the behavioral skills of children. This gives, in total, four noisy measures of child attributes that track human capital outcomes (two for each skill).

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. Following Del Boca et al. (2014), maternal time investment is measured in this paper by taking a weighted sum[15] of the total hours of time in which the mother is recorded as actively participating in activities with the child. Following Caucutt et al. (2020), only activity categories that can reasonably be defined as "investment" are included in this sum.[16] To derive a measure of the fraction of non-market time, denoted $\phi_{\tau,t,f}$ in the model, this total is divided by total number of hours that the mother is present in any way with the child.

Historically, single mothers have been the overwhelming majority of participants in child-related transfers, and are considered here as the population of interest. Since family structure is dynamic in reality, there is no perfect way to define this population in the panel dimension. This paper defines the sub-sample of single mothers to be all mothers of CDS children who were unmarried at the time of their first birth. While this model does not consider the later marriage and cohabitation decisions of single mothers, experimental studies find no evidence to suggest that these decisions are responsive to changes in welfare policies (Gennetian and Knox, 2003). At the very least, the magnitudes of the response are likely too small to be meaningfully statistically detected.

Applying this sample restriction results in a sample of 954 mothers and 1,397 children. Table 1 provides some summary statistics from this sample. Unsurprisingly, the selected sample is quite disadvantaged: mothers in this sample have low levels of education (74% have at least a High School diploma or equivalent, while 26% have less, with only 6% having obtained a Bachelor's degree) with modest earnings among those working. Additionally, we see that this sample is heavily reliant on welfare, with 44% having reported welfare use at least once, and 14% having used welfare for at least 5 years during the observed panel. The extent of this economic disadvantage extends to the children of these households. Table 1 reports means for each child's most recently measured Letter-Word and Applied Problems test scores. The results indicate that children in this sample are performing 20-25% of a standard deviation below the national average in these cognitive scores.

Combining the main interview of the PSID and the CDS supplement provides a panel of mothers' work and program participation decisions, earnings, fertility outcomes, time investment in children, and human capital outcomes for those children. This is sufficient to estimate the model, as Section 5 will demonstrate. However, in order to anchor child skill outcomes to a net present value in economic resources, one must link these skills to adult outcomes. The *Transition into Adulthood Supplement* (TAS) provides follow-up survey data on CDS children once they reach adulthood. Appendix E.6 uses data from this supplement on earnings.

The omission of savings from the model may cause concern when interpreting estimates and analyzing counterfactuals. To assess the severity of this issue, Table 1 reports total cash assets for households in the sample using PSID's Wealth Module in 1999, along with time spent in formal care according to the 1997 time diaries. The majority of sample households report no cash assets, and even individuals at the 75th percentile have modest savings.

Finally, to asses the potential role played by informal and informal child care, Table 1 uses the time diaries to calculate the fraction of children who spend some part of their time in formal care. About 91% of children report

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

[16] These categories include a range of learning, play-based, and social activities with the child and excluses television watching and household chores.

Table 1: Descriptive Statistics

	Value
Mother	
Annual Earnings (25th percentile)	8,334
Annual Earnings (50th percentile)	16,639
Mean Annual Earnings	19,227
Cash Assets (50th percentile)	0
Cash Assets (75th percentile)	796
Cash Assets (90th percentile)	2,988
Used Welfare Once (%)	44.2
Used Welfare $\geq$ 5 years (%)	14.0%
$<$ High School (%)	26.0
$\geq$ High School (%)	74.0
Bachelor's (%)	6.1
Mean No. Children	3.06
Mean Panel Length	15.3
M	954
Child	
Mean Letter Word Score	-0.25
Mean Applied Problems Score	-0.22
No Formal Care (%)	90.8
No Formal Care - Age $\leq$ 6 (%)	84.9
Frac. of time w/ mother in investment activities	0.32
Frac. of time w/ other carers in investment activities	0.16
N	1,397

All dollar amounts are deflated to 2000 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.

spending no time in a formal care arrangement in 1997 (85% for children under the age of 5). The scarce use of formal care helps to justify the baseline assumption that the quality of care may be correlated with other individual latent variables but not malleable to formal market interventions. To get a sense of the potential quality of care, Table 1 compares the average fraction of time spent in investment activities (using again the definition of Caucutt et al. (2020)) when children are care for by the mother to the same fraction when in the care of others. Estimates here suggest that maternal care may indeed be of higher quality relative to other options available.

5 Identification and Estimation

Section 3 established that optimal policies depend crucially on production parameters and the responsiveness of labor supply to incentives (summarized by a collection of semi-elasticities). It is especially important then to carefully and credibly establish identification of each set of parameters. Here the properties of the model's solution lend themselves to a clean two-stage procedure. First, the derivation of indirect utility in (13) allows for identification and estimation of behavioral parameters, exploiting both the data's panel dimension and exogenous policy variation. Second, the outcome equation (14) suggests that the same sources of variation will identify production parameters using linear instrumental variables. In this latter case, it is possible to then compare results across specifications that use different sources of variation.

The following two sections describe each procedure and the resulting set of estimates in more detail.

5.1 Behavioral Parameters

For ease of exposition, here again is the derived form for indirect utility:

$$u_{k,j}(y,\boldsymbol{a}) = (1+\alpha_{\theta,k}\Gamma_x(\boldsymbol{a}))\log(y) + h_j\left(\alpha_l\log(\ell_k) + \alpha_{\theta,k}\Gamma_\tau(\boldsymbol{a})\log(\overline{\varphi}_k)\right).$$

The full set of parameters that define investment, work, and program participation behavior is:

$$\Theta = (\gamma_W, \sigma_\varepsilon, \pi_\varepsilon, \ell, \alpha_l, \alpha_A, \alpha_S, \alpha_H, \alpha_R, \alpha_\theta, \sigma, \varphi, \Gamma_x, \Gamma_\tau, \sigma_W, \sigma_\tau)$$

where $(\gamma_W, \sigma_\varepsilon, \pi_\varepsilon)$ defines the wage process, the set $(\alpha_l, \ell, \alpha_A, \alpha_S, \alpha_R)$ define the payoffs from each discrete choice, $(\varphi, \Gamma_x, \Gamma_\tau)$ determine how payoffs change with the age and number of children in the household, as well as the fractions of money and time that are invested in the child, and σ sets the dispersion of taste shocks in each layer of the nested logit. The final pair of parameters (σ_W, σ_τ) define the dispersion of measurement error in wage and time inputs, which affect the likelihood of the data but not the model's solution. Notice that the pair of coefficient vectors $(\Gamma_{x,a}, \Gamma_{\tau,a})$ – while being functions of underlying production parameters – are free to directly match how labor supply, participation, and investment behavior changes with the age and number of children in the household. Estimation in the first stage does not impose the cross-equation restrictions between these coefficients and production parameters, allowing (1) relaxation of the assumption that parents correctly perceive the technology of skill formation; and (2) direct estimation of production parameters using skill outcomes only.

The presence of latent heterogeneity (indexed by k) in costs of work and program participation $(\ell_k, \alpha_{A,k}, \alpha_{S,k})$, labor market productivity $(\gamma_{W,0,k}, \gamma_{W,1,k}, \gamma_{W,2,k})$ and investment behavior $(\alpha_{\theta,k})$, prohibits the direct use of cross-sectional variation to estimate key parameters. The data offer three separate solutions to this challenge: (1) a sufficiently long panel dimension with sufficient variation in state variables (Kasahara and Shimotsu, 2009); (2) a sufficient number of observables that "measure" latent type (Bonhomme et al., 2016); and (3) exogenous variation in policies. Appendix E.2 provides a formal discussion of how multiple identification results from the literature apply here, guaranteeing identification of the distribution of latent variables and reduced form choice probabilities.

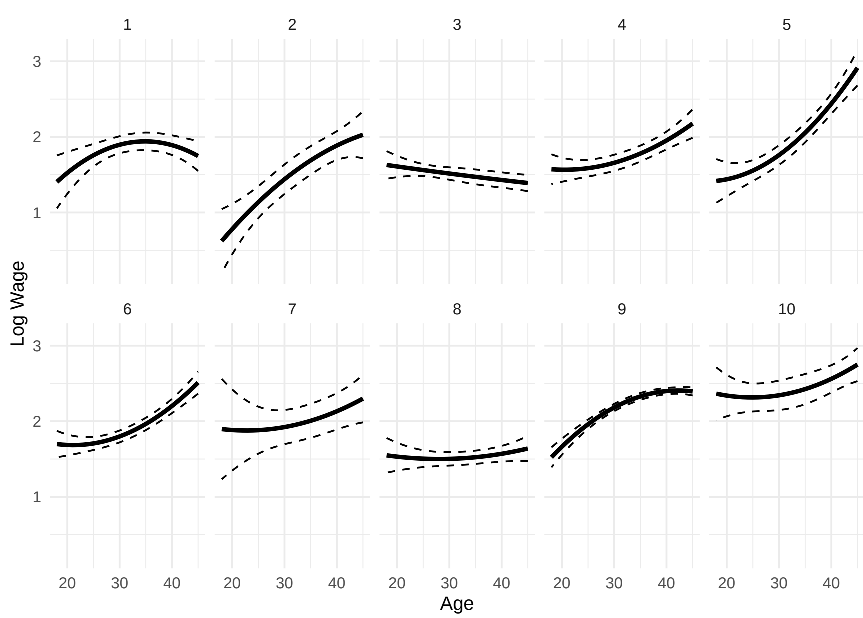
The model imposes lower dimensional parametric restrictions on choice probabilities with straightforward relationships that can now be conditioned on type. Appendix E.3 provides a more rigorous investigation of these relationships.

A useful rubric is that most of the above parameters are determined by average outcomes, while a smaller number are linked to the responsivenes of behavior. For example, the parameters $(\alpha_{S,k}, \alpha_{A,k}, \alpha_{H,k}, \ell_k, \alpha_{\theta,k})$ determine average levels of food stamp enrollment, welfare enrollment, employment, hours, and time investment for each type k. The parameters (Γ_x, Γ_τ) must adjust to match how these levels vary with the age and number of children in the household, and α_R must match how joint work and participation probabilities evolve after the introduction of time limits. Most importantly, the vector $\sigma = (\sigma_P, \sigma_{H,1}, \sigma_{H,2})$ dictate the dispersion of taste shocks for each layer of the nested logit decision problem. Accordingly, σ_P determines the responsiveness of participation behavior to financial incentives, $\sigma_{H,1}$ determines the responsiveness of labor supply at the extensive margin, and $\sigma_{H,2}$ determines the responsiveness of hours, conditional on working. Appendix F.2.1 shows how $\sigma_{H,1}$ and $\sigma_{H,2}$ in particular determine the semi-elasticities that Section 3 identified as key for the shape of optimal policies. Naturally, identification follows from the inverse of these arguments. Finally, since the model exhibits two-period finite dependence once time limits have been introduced, the discount factor β is also identified (Arcidiacono and Miller, 2020) however in practice estimation of this parameter proved difficult. Accordingly, assume throughout that $\beta = 0.98$.

In practice, a maximum likelihood routine delivers estimates of the model's behavioral parameters under the additional assumptions that (1) a multinomial logit defines the distribution of types conditional on other exogenous variables; (2) that observed time inputs and log-wages contain additive measurement errors; (3) the grid space for the wage shocks $\mathcal{E}$ contains $K_{\mathcal{E}} = 5$ points; and (4) that the economy contains $K = 10$ types. Analysis of Bayesian and Akaike Information Criteria support the selection of at least 10 types, with 10 being the highest number feasible due to computational constraints. Appendix E.3 outlines these assumptions in more detail, writes the full likelihood, and provides additional computational details.

Table 2 and Figure 2 present the estimates from this procedure. They document a substantial degree of heterogeneity across types in all dimensions: the weight on child skills (α_θ), costs of work (α_H, ℓ), costs of welfare participation (α_A), and productivity in the labor market (Figure 2). Of particular note is the fact that there is no monotonic relationship between any pairs of these latent parameters, highlighting the importance of allowing for unspecified structure in latent heterogeneity.

Figure 2: Maximum Likelihood Estimates of Wage Profiles by Type



This figure depicts estimated wage profiles for each type from the first stage maxmium likelihood procedure. Dotted lines indicate 95% confidence intervals.

Table 2: Maximum Likelihood Estimates of Behavioral Parameters

	Type-Specific Parameters				
	α_S	α_A	α_H	α_θ	$\log(\ell_k)$
Type 1	4.72	2.13	4.72	0.84	-5.65
	(0.34)	(0.58)	(0.40)	(0.14)	(0.79)
Type 2	1.40	2.52	6.85	0.48	-6.12
	(0.09)	(0.21)	(0.72)	(0.09)	(1.95)
Type 3	1.43	-0.22	3.46	0.97	-5.98
	(0.10)	(0.09)	(0.31)	(0.12)	(0.89)
Type 4	0.88	-0.27	0.17	1.91	0.11
	(0.08)	(0.10)	(0.21)	(0.18)	(0.44)
Type 5	1.18	1.81	0.73	0.88	0.49
	(0.08)	(0.17)	(0.24)	(0.09)	(0.45)
Type 6	3.19	0.59	5.96	1.21	1.50
	(0.22)	(0.18)	(0.53)	(0.17)	(1.13)
Type 7	1.26	2.44	4.57	3.16	4.83
	(0.14)	(0.26)	(0.52)	(0.45)	(1.09)
Type 8	3.61	0.90	1.05	1.94	-2.82
	(0.32)	(0.26)	(0.33)	(0.24)	(0.57)
Type 9	3.59	1.50	2.05	1.65	7.49
	(0.33)	(0.46)	(0.40)	(0.17)	(1.03)
Type 10	2.16	0.57	1.37	2.38	8.92
	(0.18)	(0.13)	(0.39)	(0.24)	(1.16)
	Preferences				
	σ_P	$\sigma_{H,1}$	$\sigma_{H,2}$	$\alpha_{R,1}$	$\alpha_{R,2}$
	1.05	2.42	0.79	0.39	1.07
	(0.10)	(0.16)	(0.06)	(0.09)	(0.16)
	Wage Process		Measurement Error		
	π_ε	σ_ε	σ_τ	σ_W	
	0.73	0.71	0.35	0.53	
	(0.03)	(0.01)	(0.01)	(0.00)	

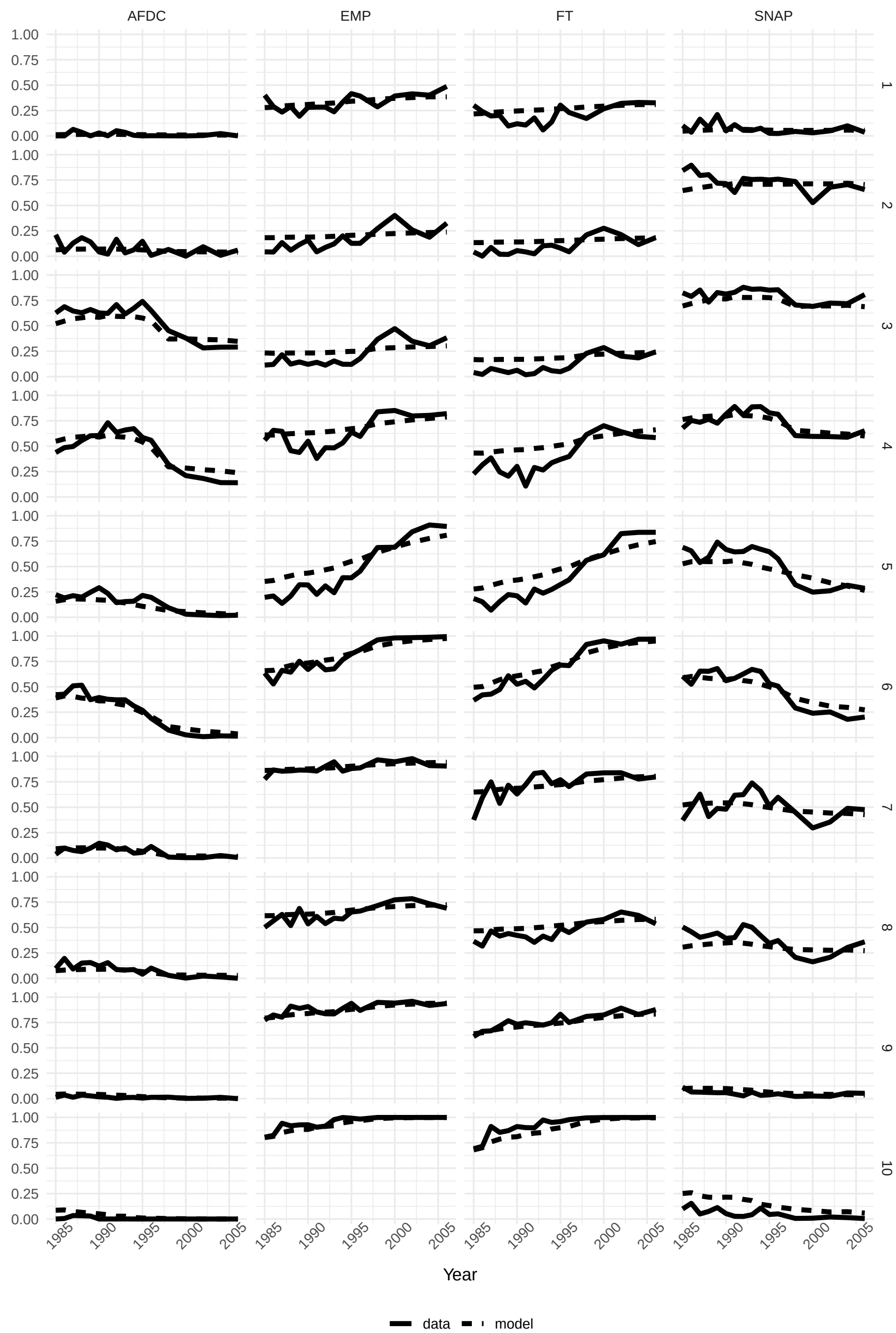
This table reports maximum likelihood estimates of the first stage model parameters using the EM algorithm. Standard errors are reported in parentheses and are calculated using the typical MLE formula (the inverse of the covariance of the score equations).

To assess fit of the model, Figure 3 compares the mean rates of employment (EMP), program participation (AFDC), food stamp participation (SNAP), and full-time employment (FT) for each latent type[17]. The model exhibits excellent fit of the time series in behavior for each type. As discussed in the introduction, the years between 1990 and 2002 saw quite dramatic changes to the policy landscape and labor market conditions for single mothers, which provides important variation to identify the behavioral parameters of the model. A complementary observation is that the panel

[17] Conditional means by type can be calculated using the posterior weights for each mother-year observation that are produced in the expectation step of the EM algorithm.

dimension of the data clearly unveils rich degrees of multidimensional heterogeneity and close inspection of Figure 3 leads to many interesting comparisons across types.

Figure 3: Model Fit by Latent Type



This figure shows the model's fit of mean rates of labor force participation (LFP) and welfare participation (AFDC) by latent type. Means for each type from the data are calculated using posterior probabilities generated by the model's estimates.

Appendix E.3 further examines the estimated model and documents that it produces reasonable estimates of labor supply elasticities. Additional type-specific means show how latent characteristics covary with fertility and education.

5.2 Production Parameters

The linear outcome equation (14) for skills allows for transparent identification and estimation of production parameters. For ease of exposition, outcomes are again given by:

$$\log(\theta_{f,t+1}) = \mu_{\theta,k} + \delta_{x,a}\log(Y_t) + \delta_{\tau,a}\log(\overline{\varphi}_k)h_t + \delta_\theta\log(\theta_{f,t}) + e_{f,t} + \eta_{f,t}$$

where the error term $e_{f,t}$ depends on latent type k, as well as the ages of all children in the household.

The derivation above is convenient because it presents the issue of identification in terms of the familiar linear model: one must find instruments for income and hours worked that are uncorrelated with the residual terms in this outcome equation. There are two sources of endogeneity that prohibit the use of observed net income and hours as instruments. First, the residual $e_{f,t}$ appears due to the fact that investments are unobserved, and is correlated mechanically with work decisions and income through the coefficients $\alpha_{\theta,k}\Gamma_x$ and $\alpha_{\theta,k}\Gamma_\tau$ which determine both work and investment decisions. Second, no assumption has been made to guarantee that mother's TFP, $\mu_{k,\theta}$, is not correlated with other heterogeneous parameters that determine behavior. It is plausible for example that mothers who find employment less costly or are more productive in the labor market may also have innately higher parenting ability.

This paper pursues two solutions. First, if each mother's true unobserved type, $k(m)$, were known, then one could instead use a flexible function of k and $(a, \boldsymbol{a})$ to control for the endogenous residual. In such a case, the variables $Y_{m,t}$ and $h_{m,t}$ are themselves valid instruments. Future sections of the paper will refer to these as the "model-based instruments". Although type is latent, Bonhomme and Manresa (2015) show that classification based on the likelihood from the estimation's first stage is sufficient for regular asymptotics of this second-stage estimator to hold. Given the relatively long panels of the PSID, classification error in this setting is likely to be quite small.

The second approach imposes the weaker condition that, conditional on type, unobserved determinants of skill outcomes are exogenous with respect to only policy variation.[18] The remainder of the paper refers to these as the "strict" or "policy" instruments. In either case, identification is guaranteed by rank conditions on the instruments formed from the set of exogenous variables.

Appendix E.5 augments the linear outcome equation in two ways. First, to allow for the fact that measured skills contain additive error and appear in the data at five year intervals.[19] Second, to include a control function that employs the exogenous variation provided by each set of instruments. The final result is a quasi-likelihood that lends itself to estimation by least squares, maximum likelihood, or Monte Carlo methods (as below).

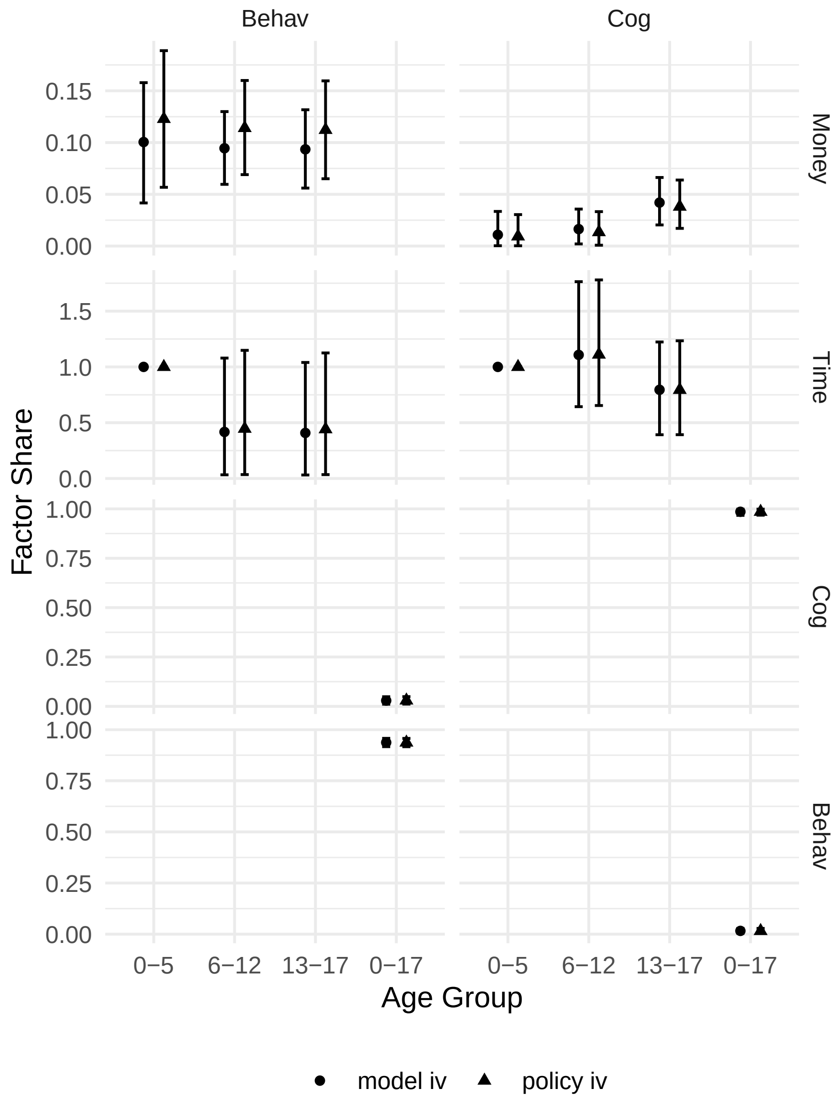
The tables and figures below present the results of a Markov Chain Monte Carlo (MCMC) routine that samples from a posterior distribution given the quasi-likelihood and given some weak priors on key production parameters. The inclusion of these priors ensures theoretical bound constraints and helps with precision in the presence of (potentially) weak instruments. Appendix E.5 provides further details. Figure 4 plots estimates (posterior means) of the factor shares $(\delta_x, \delta_\tau, \delta_\theta)$ along with 95% credibility intervals from the sampled posterior distributions. Estimates of δ_θ suggest very high persistence of both behavioral skills with relatively modest (though still positive) cross-skill factor shares. Estimates of the factor share of money in the production of behavioral and cognitive skills are positive, modest, and in the case of cognitive skills, only marginally bounded away from zero.

It is not straightforward to compare estimates of δ_x to other results in the literature because this model predicts nonlinear effects of income, while the most prevalently cited evidence is phrased as the effect of an additional \$1,000 in annual income. Still, a back-of-the-envelope calculation demonstrates how small these point estimates are relative to empirical precedent. Consider a household with \$6,000 in annual income, a number far below the median in this sample. According to these estimates, a \$1,000 increase in income for this household translates into a 1.4-1.5%sd increase in behavioral skills and a 0.2-0.6%sd increase in cognitive skills. Those numbers are an order of magnitude smaller than previous estimates (Dahl and Lochner, 2012; Akee et al., 2018). Nevertheless, the modest causal impacts

[18] Controlling for type in thise case is not essential, but allowing for interaction between policies and type strengthens the instruments, and makes the approach robust to any systematic correlation between policies and type.

[19] Measurement error in skills necessitates the use of a second measurement as an instrument for lagged skills in the outcome equation.

Figure 4: Estimates of Factor Shares



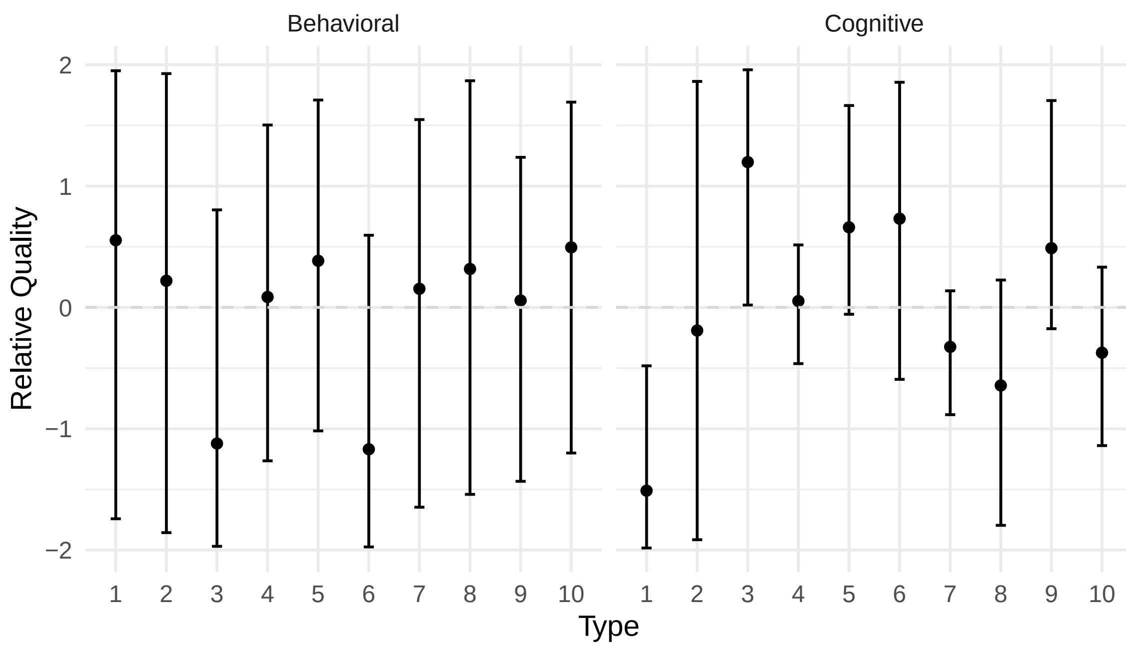
This figure shows posterior means and 95% credibility intervals of factor shares $(\delta_x, \delta_\tau, \delta_\theta)$ from the quasi-bayesian estimation procedure using control functions. Intervals are based on a MCMC chain of length 10,000. See Appendix E.5 for more details.

will still amount in Section 6 to quite substantial skill impacts as a result of large and persistent increases in household income.

Figure 5 plots the posterior means and 95% credibility intervals for the estimates of relative care quality by type. These combine with estimates of the factor shares $\delta_{\tau,a}$ to produce a ceterus paribus effect of additional work hours on skill outcomes. While the picture of behavioral skills is largely ambiguous, some interesting patterns emerge for cognitive skills. Estimates suggest likely negative impacts of work for Types (1), (7), and (8), and positive impacts for Types (3), (5), and (9). This widely varying heterogeneity has important implications for the distributional impacts of labor market policies that encourage or discourage labor supply.[20]

[20]Figure 14 will later demonstrate the importance of this heterogeneity.

Figure 5: Estimates of Relative Care Quality ($\log(\overline{\varphi}_k)$)



This figure shows posterior means and 95% credibility intervals for relative care quality, $\log(\overline{\varphi}_k)$, from the quasi-bayesian estimation procedure using control functions. Intervals are based on a MCMC chain of length 10,000. See Appendix E.5 for more details.

5.3 Net Present Value of Skills

To estimate the effect of skills on lifetime earnings, this paper follows Chetty et al. (2011) and Kline and Walters (2016) in assuming that the effect of skills on earnings is proportional at each age. If the percentage impact of skills on income at the beginning of the lifecycle can be estimated, this assumption allows a forecast to be constructed. As pointed out by Chetty et al. (2011), this assumption is somewhat conservative since it is likely that increases in human capital could increase the growth in earnings over the lifecycle. Analysis of the earnings of CDS children in young adulthood suggests that a standard deviation increase in cognitive skills leads to a 32% increase in earnings, while a standard deviation increase in behavioral skills leads to an 18% increase. A simple forecasting method from representative CPS samples suggests that the net present value of earnings (conditional on working) for this cohort is \$996,686, and that the CDS sample averages earnings (conditional on working) at around 73% of this representative sample, with a probability of positive earnings set to 0.57. Putting this together implies that a standard deviation increase for children in the CDS sample is worth about \$134,000 in earnings for cognitive skills and \$76,000 for behavioral skills. The exercises in this paper adopt more conservative numbers by subtracting one estimated standard deviation from each point estimate of the earnings return to skills, resulting in calibrated values of:

$$\tilde{\Gamma}_{18,c} = \$87,900 \qquad \tilde{\Gamma}_{18,b} = \$49,300.$$

Appendix E.6 provides more detail.

6 Optimal Policy Exercises

With the estimated model in hand, this section turns to the calculation of an optimal transfer policy, as ranked by the objective function (15). The static[21] policy $\boldsymbol{y}$ – with which the planner's objective function simplifies to (20) – serves as

[21]Recall that this function is "static" in that it simply maps earnings to net allocations irrespective of past decisions.

the benchmark and main policy of interest. The results for that exercise submit to a clear analysis in Section 6.1 below thanks to the theoretical results in Section 3. However, it is also natural to ask whether particular extensions that resemble current policy frameworks can improve upon the static policy. In this spirit, Section 6.2 below will evaluate the inclusion of time limits, welfare ordeals (as embodied by the participation cost parameters), and work requirements to see whether improvements on the planner's objective are possible in a dynamic environment.

Calibrating Benchmark Parameters All of the quantitative exercises below take the year 2000 as the reference year for a benchmark economy. In particular, the empirical distribution $\hat{\pi}_{2000}(s)$ from the PSID estimation sample provides an estimate for the distribution $\pi_{2000}(s)$ over states (s) and life-cycle stages (t). Practically, this amounts to calculating the planner's objective by averaging over individuals and potential states using posterior probabilities from estimation.

Equation (19) provides direction for setting the ratio $\boldsymbol{\mu}/\boldsymbol{\lambda}$, which measures the planner's taste for redistribution toward this population. Assuming that tastes for redistribution are not affected by the presence of children, equation (19) simplifies to:

$$\mathbb{E}[\boldsymbol{y}(E_j(s))] = \frac{\boldsymbol{\mu}}{\boldsymbol{\lambda}}$$

for the same population of women with no children in the household. Thus, the average generosity of non-child related programs (as measured by average net income) provides a measure for the planner's tastes for redistribution. In the model, this is average net income for households that can only participate in food stamps. The implied value for $\boldsymbol{\mu}/\boldsymbol{\lambda}$ is therefore

$$\frac{\widehat{\boldsymbol{\mu}}}{\boldsymbol{\lambda}} = \mathbb{E}[\boldsymbol{Y}_j(s)|\hat{\Theta}, \text{No Children}]$$

the right hand side of which is provided by the estimated model.

6.1 Optimal Static Policies

Figure 6 presents the solution to the optimization problem presented in equation (18) while restricting the allocation $\boldsymbol{y}$ to be either a piece-wise linear function (with a single discontinuity at $\boldsymbol{y}(0)$) or a piece-wise step function (with discontinuities at each grid point). For comparison, Figure 6 also plots the net income allocation for welfare participants and non-participants by averaging over the sample of PSID individuals in the year 2000.

The optimal policy exhibits two striking features. Most notably, allocations are far more generous at the optimum than in the benchmark. Under the optimal policy, households with no earned income universally receive a monthly transfer of around \$1800 per month. In the benchmark, households with no earned income received an average of around \$600 per month and *only if* they were participating in welfare.

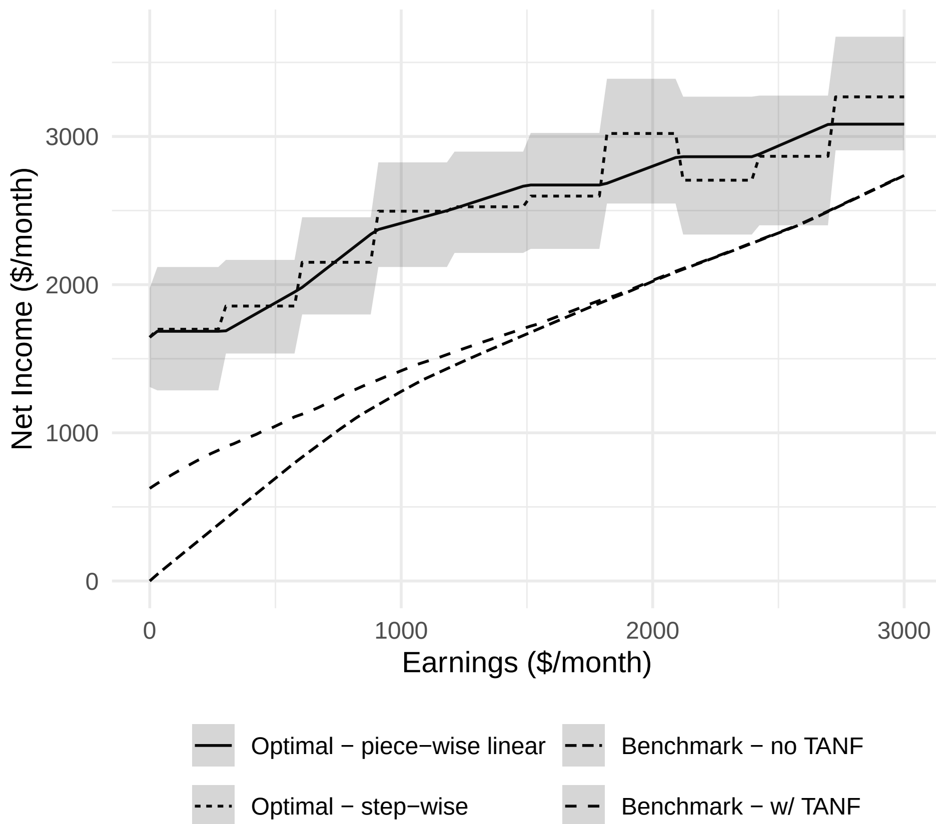
Table 3 quantifies average generosity under the benchmark and compares it to generosity at the optimum with and without skill formation considerations. Average net income under the optimal policy is roughly \$2030 per month as compared to about \$1300 a month in the benchmark, representing a 52% increase in overall generosity. Table 3 confirms that this difference comes directly from the contribution of money inputs to child skills by calculating optimal generosity when δ_x is set to zero. In this case, optimal generosity is in fact smaller than in the benchmark, suggesting that the calibration of the planner's taste for distribution is quite conservative.

Table 3: Optimal size of $\boldsymbol{y}$ with and without skill development

	Benchmark	Optimal - No Skill	Optimal
Average Net Income (\$/month)	1337.18	1033.70	2032.88
		(55.52)	(192.64)

The optimal policy also features stronger labor supply disincentives, especially among those with low potential

Figure 6: Optimal static policy $\boldsymbol{y}$



This figure shows the solution for the optimal allocation function $\boldsymbol{y}$ described in (18) using the empirical distribution of households in the year 2000. 95% bootstrapped credibility intervals are shown for the solution, where production parameters are drawn from the Markov Chain described in Section 5 and behavioral parameters are drawn from the estimated sampling distribution using the standard asymptotic approximation. "Benchmark - TANF" shows average net income for the sample when individuals are participating in cash welfare, while "Benchmark - no TANF" shows the average without participation. Dollar amounts are reported in year 2000 USD per month.

earnings. For example, an individual who earns \$1000 in a month faces a participation tax rate of 75%. This contrasts drastically with labor supply incentives in the benchmark, where mothers are encouraged to work through work requirements and employment subsidies (such as the Earned Income Tax Credit). Accordingly, the optimal policy results in large reductions in earnings. Figure 15 in Appendix A illustrates this by plotting the cumulative distribution of earnings in the benchmark and at the optimum.

Equation (20) facilitates an analysis of what determines the shape of this allocation. This decomposition applies for the optimal step-wise function but technically not the piece-wise linear function, but the close similarity of these optimal schedules ensures insight for both cases. Recall that (20) decomposes the allocation $\boldsymbol{y}(e)$ at earnings e into a first-best, the expected value of $\boldsymbol{w}(s)$ given earnings e, and a wedge term equal to:

$$\text{wedge}(e) = \mathbb{E}\left[\sum_i \eta_{ij}(s)\mathcal{R}_i(s)\middle| E_j(s) = e\right]$$

Figure 16 in Appendix A confirms that the conditional mean of $\boldsymbol{w}(s)$ does not vary much with earnings, such that changes in this wedge term with earnings explains most of the shape of $\boldsymbol{y}$.

There are competing determinants of the shape of the wedge, and three particular decompositions will help evaluate what matters quantitatively. First, the term $\mathcal{R}_i(s)$ is comprised of short and long-run revenue impacts, which can be separated to get:

$$\text{wedge}(e) = \mathbb{E}\left[\sum_i \eta_{ij}(s)(E_j(s) - \boldsymbol{y}(E_j(s)) + \boldsymbol{y}(0))\Big|E_j(s) = e\right] + \mathbb{E}\left[\sum_i \eta_{ij}(s)\mathcal{D}_i(s)\Big|E_j(s) = e\right]. \tag{22}$$

In the remainder of this section the "skill component" of the wedge refers to the second term in this equation. Second, behavioral responses can occur at the extensive (from zero hours to any positive hours) or intensive (across hours choices) margin. Accordingly, the semi-elasticity η_{ij} can be decomposed into an extensive and intensive marginal component to give:[22]

$$\text{wedge}(e) = \mathbb{E}\left[\sum_i \varepsilon_{ij}(s)\mathcal{R}_i(s)\Big|E_j(s) = e\right] + \mathbb{E}\left[\sum_i (\eta_{ij}(s) - \varepsilon_{ij}(s))\mathcal{R}_i(s)\Big|E_j(s) = e\right] \tag{23}$$

As above, let the "extensive margin" component of the wedge refer to the first term in this equation. Finally, the semi-elasticities η may covary with earnings and contribute to the shape of allocations. To evaluate this, define $\overline{\eta}_{ij}$ as the average semi-elasticity of i with respect to j, and take:

$$\text{wedge}(e) = \mathbb{E}\left[\sum_i \overline{\eta}_{ij}\mathcal{R}_i(s)\Big|E_j(s) = e\right] + \mathbb{E}\left[\sum_i (\eta_{ij}(s) - \overline{\eta}_{ij})\mathcal{R}_i(s)\Big|E_j(s) = e\right] \tag{24}$$

where the first term is the wedge "with fixed semi-elasticity".

Figure 7 plots the total wedge at each earnings grid point against: (1) the second component of (22), reflecting the skill component of the wedge; (2) the first component of (23), reflecting the extensive margin component; and (3) the first component of (24), reflecting the contribution of $\mathcal{R}_i(s)$ holding fixed the movements in $\eta_{ij}(s)$.

The top panel in Figure 7 shows that when semi-elasticities are fixed, the resulting wedge closely tracks the actual wedge, indicating that movements in semi-elasticities with earnings do not matter much for the shape of the optimal schedule. The middle panel shows that the extensive margin component explains all of the wedge at 0 earnings, but shows little movement thereafter. Thus, responsiveness of individuals at the intensive margin must explain the shape of $\boldsymbol{y}(e)$ for all positive earnings levels. Finally, the bottom panel in Figure 7 shows that the skill component of the wedge explains most of the total wedge (and therefore most of the shape of $\boldsymbol{y}$) for all earnings levels below \$2000/month. This final decomposition is particularly important, as it suggests that the additional accounting of skill formation in the planner's objective is the main driving force shaping labor supply incentives.

6.2 Optimal Dynamic Policies

Can the planner improve on the static optimal policy by incorporating features of current welfare programs? This section considers three potential extensions to the previous section's static transfer policy: (1) time limits; (2) ordeals; and (3) work requirements. In each of these examples, the planner chooses two schedules $(\boldsymbol{y}_0, \boldsymbol{y}_1)$ that allocate net income as a function of earnings. In addition to their hours choice, individuals must now make a binary participation decision $A \in \{0, 1\}$ every period, receiving the schedule $\boldsymbol{y}_A$ as a result.[23] The quantitative results below continue to use the empirical distribution $\hat{\pi}_t(s)$ from the year 2000 as an input in the planner's objective. Since time limits re-introduce dynamics to the individual's decision problem, and to maintain consistency across comparisons, all of the results below use the planner's dynamic objective in equation (15) to calculate and rank optimal policies.[24]

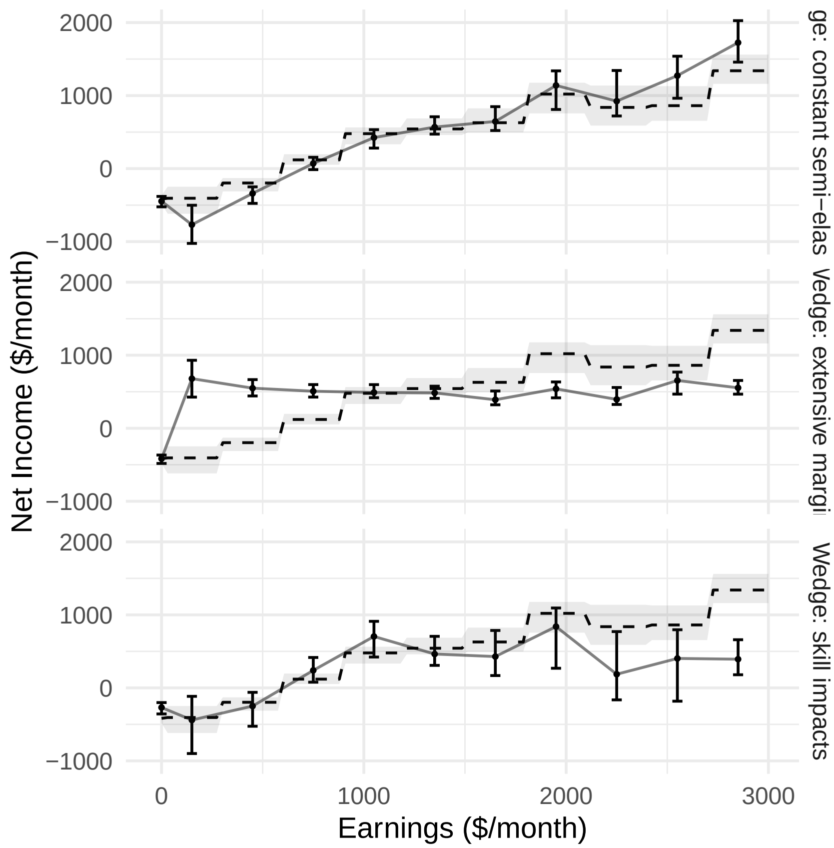
Time Limits In this extension, when an individual's cumulative periods of participation reach a pre-specified limit, Ω, they are no longer eligible for the program and face $\boldsymbol{y}_0$ thereafter. Since the planner is always free to set $\boldsymbol{y}_0 = \boldsymbol{y}_1$, this extension must weakly increase the planner's objective relative to the static policy. The size of this improvement

[22]See Appendix F.2.1 for this decomposition.

[23]To model the participation decision, assume once again the three-layered nested logit structure of the baseline model, with the same estimated dispersion parameters σ.

[24]To maintain a "fair" welfare comparison, results below model the static policy also as an opt-in policy with $\boldsymbol{y}_0 = \boldsymbol{y}_1$. This avoids spurious welfare improvements from simply adding variety, through extreme value shocks, to the choice set.

Figure 7: Decomposing the shape of optimal $\boldsymbol{y}$



This figure shows the three decompositions (22)-(24) of the wedge between optimal allocations and the first-best. See text for discussion. 95% bootstrapped credibility intervals are shown at the solution, where production parameters are drawn from the Markov Chain described in Section 5 and behavioral parameters are drawn from the estimated sampling distribution using the standard asymptotic approximation. Dollar amounts are reported in year 2000 USD per month.

Table 4: Change in Planner's Objective for Optimal Policy Extensions

	Static Policy	5 Year Time Limit	2 Year Time Limit	Ordeal	Work Requirements
Planner Objective	0.00	5.11	1.17	-2909.19	-1946.98
	(0.00)	(2.28)	(1.40)	(196.32)	(185.25)
Short Run Revenue	-12030.18	-12029.90	-12030.85	-12028.75	-12314.36
	(114.55)	(114.14)	(114.16)	(112.93)	(119.06)
Long Run Revenue	9476.45	9318.12	9399.60	9264.85	9409.84
	(1514.52)	(1508.07)	(1511.18)	(1504.99)	(1518.03)

This table presents the results of dynamic optimal policy exercises where the planner adopts time limits, welfare ordeals, and work requirements as potential extensions of the policy space. The first row presents changes in the planner's optimum, normalized by $\boldsymbol{\lambda}$, relative to the optimal static policy $\boldsymbol{y}$. Rows 2 and 3 report changes in short-run and long-run resources as a result of each policy, measured as the NPV of monthly year 2000 USD. Bootstrapped standard errors are reported in parantheses.

depends on the relative importance of ex-post risk compared to ex-ante differences in productivity. When the former is more important, a time-limited program can more effectively provide insurance by offering more generous transfers that individuals can choose to receive after negative labor market shocks. On the other hand, time limits will not allow for more effective redistribution, and so gains may be small if ex-post risk is small relative to ex-ante risk.

Ordeals If types that benefit most from welfare participation face lesser ordeals, this may result in more efficient redistribution toward those types (Nichols and Zeckhauser, 1982). To assess the potential for this mechanism, this extension re-introduces estimated participation costs[25] from the baseline model to see whether this can achieve any improvement to the planner's objective.

Work Requirements The final extension reintroduces work requirements through the estimated costs parameters α_R. Since work requirements have the potential to make employment less costly, this extension may also yield benefits to individuals (and hence the planner).

Table 4 presents the results from these three extensions. The first row calculates differences in the planner's objective at each optimum compared to the optimal static policy, dividing by $\boldsymbol{\lambda}$ to scale the results in planner's monthly NPV dollars. As expected, time limits do improve on the planner's objective, but the gains to this additional policy flexibility are very small. This result suggests that, overwhelmingly, redistribution is the case of large welfare gains to the planner and not insurance.

Neither ordeals nor work requirements offer any improvement in the planner's objective, and both extensions are strictly dominated by the static policy, suggesting that the way programs are currently designed does not result in any mprovements in targeting.

The second and third columns report the change, relative to the benchmark, in the planner's objective derived from changes to their NPV resource constraint. Results are quite comparable across policies since the most substantive change in all cases is the large increase in program generosity that the previous analysis of the optimal static policy emphasised. Although very costly in the short-run, this large investment is returned in the long-run in the form of improvements to the cognitive and behavioral skills of children. Standardized errors for the estimate of this long-run return are larger than the short-run costs, reflecting greater statistical uncertainty in the model's production parameters relative to behavioral parameters.

6.3 Impacts of Optimal Policy Reforms

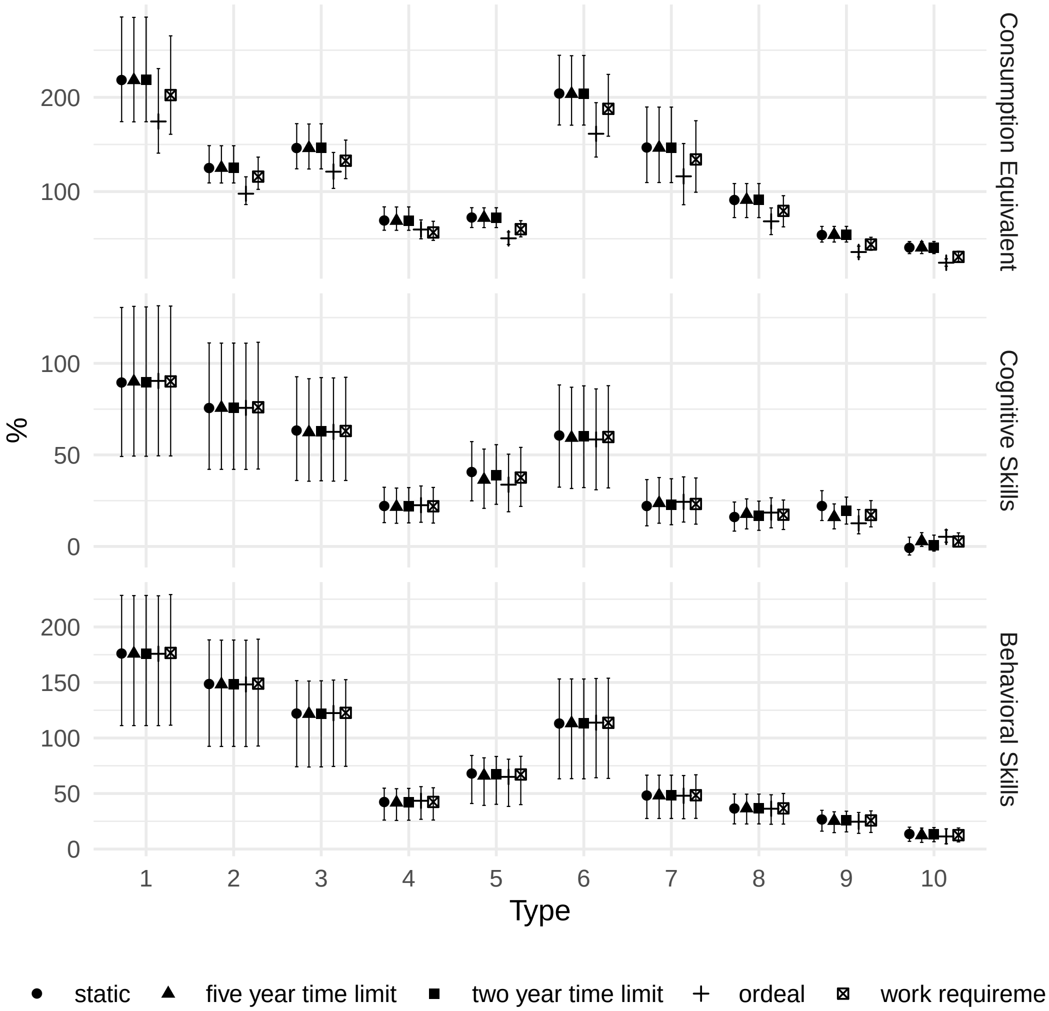
Each of the optimal policy reforms entail large increases in the net income of single mothers. This section explores the distributional welfare effects of each policy change and their long-run impacts on the skill outcomes of children. Figure 8 reports the welfare impact (as measured by an equivalent percentage increase in consumption) and the long-run skill impacts on children for each type in the model. To help understand the source of welfare and skill impacts Figure 9 plots – for the baseline and for the optimal static policy – average employment, average full-time employment, and average log-income for each type from the years 2000 to 2012.

Given the significant increase in the generosity of transfers, large impacts in welfare are not surprising. Figure 8 however also documents substantial heterogeneity in the magnitude of these reforms by type, ranging from 40 to 200% in consumption equivalent terms. In general, types that benefit the most are those that begin with the lowest average net income in the baseline (see Figure 9). However, there are some notable exceptions. For example Types 6 and 7 enjoy as much or more of a welfare gain compared to Types 2 and 3 despite experiencing smaller impacts on net income. This is because Types 6 and 7 have among the highest costs of employment and hours (see Table E.3) and are able to reduce their labor supply substantially in response to the reforms.

To highlight the importance of latent heterogeneity for properly understanding welfare impacts, Figure 18 in Appendix A plots a comparison of the welfare results for this policy reform to an estimated version of the model with

[25] For AFDC participation, the total cost is $\alpha_{A,k} + \alpha_{S,k}$.

Figure 8: Long-run welfare and skill impacts of optimal policy reforms

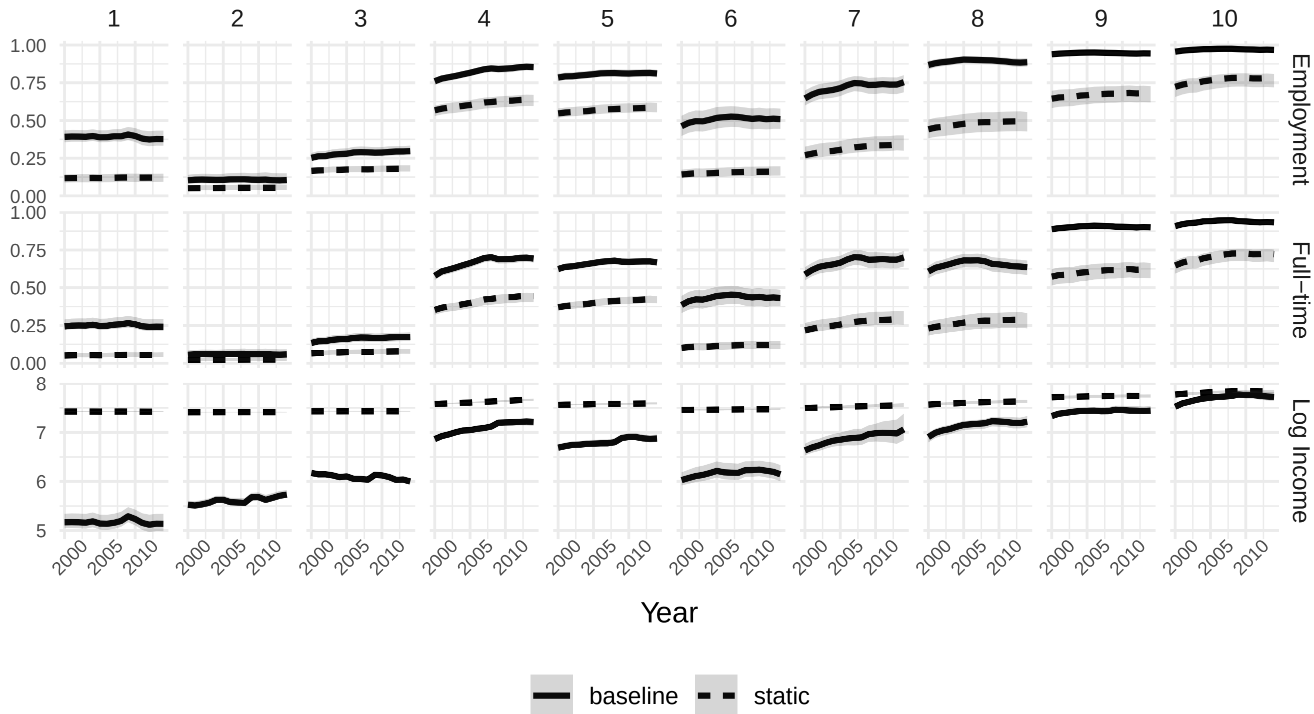


This figure shows the welfare (as measured by an equivalent percentage change in consumption) and skill impacts of the optimal dynamic policies that solve (15) using the empirical distribution of households in the year 2000. 95% bootstrapped credibility intervals are shown for the solution, where production parameters are drawn from the Markov Chain described in Section 5 and behavioral parameters are drawn from the estimated sampling distribution using the standard asymptotic approximation. Average skill impacts are calculated for the youngest child in each household and are expressed in % standard deviations.

$K = 2$ instead of $K = 10$. The two type model predicts much more even welfare effects and, in confusing ex-ante with ex-post heterogeneity, misallocates much of the distributional effects of the reform to insurance.

Although none of the extended policy exercises yielded improvements to the planner's objective function, the aggregated results in Table 4 provide no information about whether there may be trade-offs across types in which policies are preferred. Figure 8 confirms that, within the margins of statistical uncertainty, no such trade-off exists: all types prefer the optimal static policy to the alternatives. Since all optimal reforms involve large increases in transfers, it is natural to think that perhaps this large increase overpowers any desire for improvements in the efficiency of redistribution and insurance provision. To assess this, Figure 17 in Appendix A presents the results from optimal reforms that ignore the child development impacts by setting $\delta_x = \delta_\tau = 0$. In this case the preference for the static policy over alternatives is preserved. Given these results, the remainder of this discussion of long-run impacts will focus on results pertaining to the optimal static policy.

Figure 9: Dynamic impacts of optimal reform $\boldsymbol{y}$



This figure calculates the dynamic impacts of the optimal policy $\boldsymbol{y}$ (computed by solving (15) using the empirical distribution of households in the year 2000) on income and labor supply. 95% bootstrapped credibility intervals are shown, where production parameters are drawn from the Markov Chain described in Section 5 and behavioral parameters are drawn from the estimated sampling distribution using the standard asymptotic approximation.

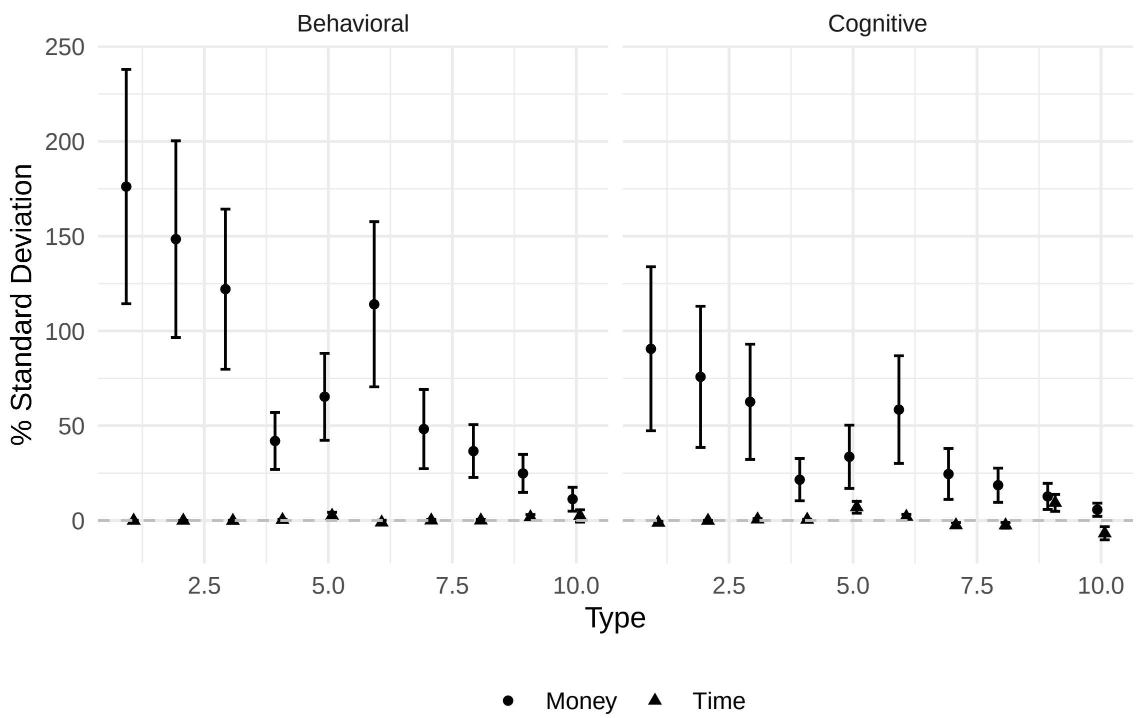
Turning to child skill impacts, these exhibit very similar patterns to the distributional effects of welfare in that the effect of the policy reform on log income explains most of the variation across types. The extent of this variation is quite substantial, ranging between 0 to 90% of a standard deviation in cognitive skills, and 13 to 175% of a standard deviation in behavioral skills. Note that due to the additivity of log income and hours in the skill outcome equation (14), it is straightforward to decompose skill impacts into the contribution of changes in log income and changes in hours of work. Figure 10 presents the results of this decomposition and confirms that, for this large expansion in the size of the program, the effect of income is very dominant in shaping skill impacts.

6.4 Pursuing More Plausible Policy Reforms

The policies considered above all involve very large fiscal outlays in the short-run in order to receive the benefit of very large skill impacts in the long-run. An overarching lesson is that even when the causal impacts of time and money on skill outcomes are modest relative to existing estimates in the literature, these effects can scale up quite dramatically when sustained over every year of a child's development, especially for children facing extreme economic disadvantage.

The magnitude of the effects at play, however, put a strain on the plasubility of the policy exercise. Two concerns in particular motivate pursuit of a more moderate optimal policy calculation. First, the factor shares of time and money in the model find identification through comparably much smaller changes in policies and economic circumstances compared to the very large changes in the counterfactual. The extreme degree of extrapolation required by the exercise undermines, to some extent, the credibility of the causal parameters. Second, it is the short-run impact of time and money on skill outcomes that identifies factor shares. The long-run impact is then projected in the model through the

Figure 10: Decomposition of skill impacts of optimal reform $\boldsymbol{y}$



This figure decomposes the long-run skill impacts of the optimal policy $\boldsymbol{y}$ (computed by solving (15) using the empirical distribution of households in the year 2000) into changes in money and time inputs. 95% bootstrapped credibility intervals are shown, where production parameters are drawn from the Markov Chain described in Section 5 and behavioral parameters are drawn from the estimated sampling distribution using the standard asymptotic approximation. Average skill impacts are calculated for the youngest child in each household and are expressed in % standard deviations.

self-productivity (or "persistence") parameters δ_θ. This is also an interpolation of sorts that side-steps direct causal evidence of short-run changes in resources on long-run life-cycle outcomes.[26]

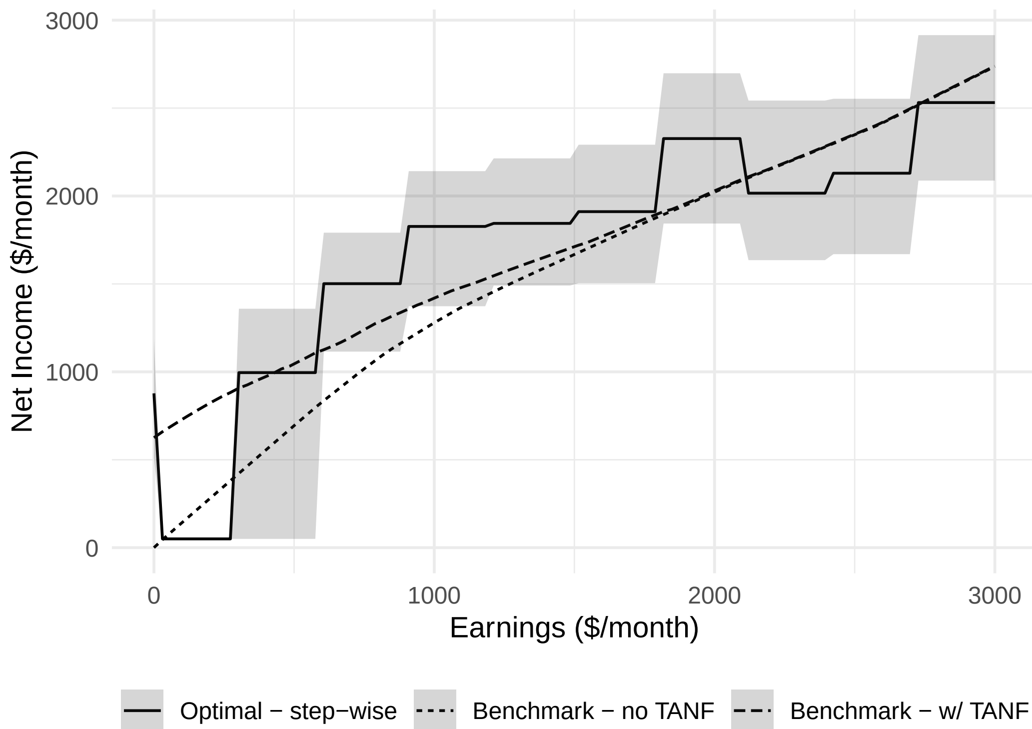
To address these concerns, this section calculates a more modest set of optimal policy reforms by increasing the marginal opportunity cost of an additional dollar to three times its originally calibrated value. This shrinks optimal generosity to a target net income of about $1,300 per month (compared to $1,337 per month in the benchmark).

Figure 11 depicts the optimal static policy for this more modestly sized reform. The size and shape of the policy exhibits slightly higher participation and marginal tax rates than the average benchmark, but the difference in the size of the transfers is no longer so dramatic. Most notably, the optimal schedule features a sharp decline when moving from zero to very small amounts of earnings, creating very strong labor supply disincentives at the bottom of the earnings distribution. Repeating the decomposition exercise from the previous section will help to understand this stark result.

Figure 12 compares the long-run (skill return) to short-run components of the wedge, as well as the extensive and intensive margin components in order to understand the determinants of Figure 11. First, comparing the long and short-run components of the wedge reveals that the skill formation channel ($\mathcal{D}_j(s)$) is again the dominant determinant of the shape of $\boldsymbol{y}$, as compared to short-run revenue considerations. Furthermore, Figure 12 indicates that it is intensive marginal responses that explain the large reduction in allocations near zero. In other words, the planner is seeking to discourage reductions in hours at the intensive margin rather than discourage small increases away from 0 earnings at

[26]The existence of external sources of evidence for long-run impacts also ought to allay these concerns somewhat (Barr et al., 2022; Bastian and Michelmore, 2018)

Figure 11: Optimal $\boldsymbol{y}$ with higher costs of public funds (moderate expansion)



This figure shows the solution for the optimal allocation function $\boldsymbol{y}$, calculated exactly as in Figure 6 but with $\boldsymbol{\lambda}$ scaled to three times its original value. 95% bootstrapped credibility intervals are shown for the solution, where production parameters are drawn from the Markov Chain described in Section 5 and behavioral parameters are drawn from the estimated sampling distribution using the standard asymptotic approximation. "Benchmark - TANF" shows average net income for the sample when individuals are participating in cash welfare, while "Benchmark - no TANF" shows the average without participation. Dollar amounts are reported in year 2000 USD per month.

the extensive margin. Given the dominant role played by skill impacts, this is likely due in part to the positive effect of additional hours on skill outcomes for some types.

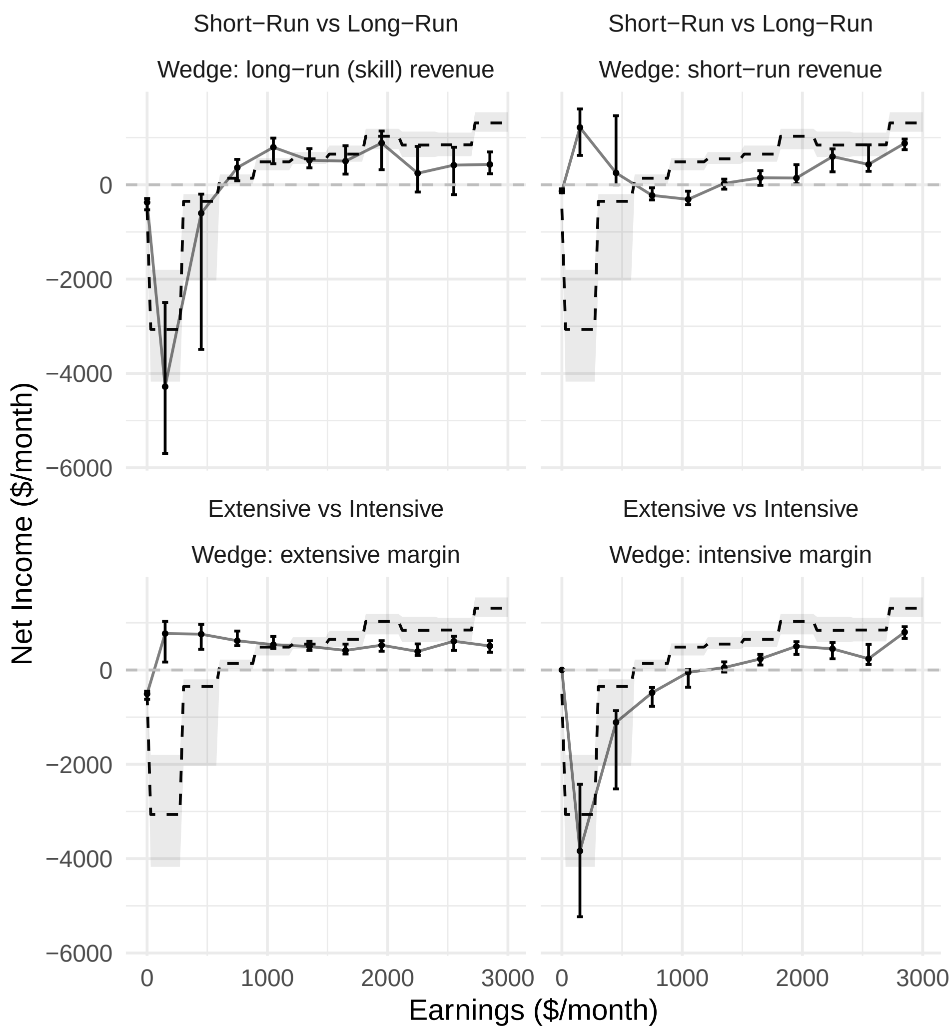
Figure 13 depicts the welfare and skill impacts for the same reform.[27] As with the larger optimal reform, welfare and skill impacts are widely varying in this case. One finding of note is that the optimal reform provides uniform welfare gains across types despite no longer exhibiting a large increase in generosity. This finding suggests fairly substantial welfare gains from insurance, even for higher earning types on average.

The same cannot be said for skill impacts, where a minor trade-off across types does emerge: children of Type (10) households experience unambiguous skill losses, and skill losses in some dimension are at least plausible for Types (7) through (9). Figure 14 decomposes skill effects into changes in time and money inputs to show that, unlike in the case of large reforms, changes in the quality of time inputs has a meaningful effect on skill outcomes for this reform. For some types these effects are off-setting, and for others they are enhancing the overall impact.

The moderate policy does still seek a trade-off between short-run costs and long-run returns. The program costs on average $3,690 additional dollars per household per month in NPV (standard deviation $106), and nets a long-run return of $4,145 dollars per month in NPV (standard deviation $653). Thus, this optimal expansion is a better-than-revenue-neutral reform that improves welfare for all types, and improves skills for children in most households.

[27]Results here derive from a re-calculated static policy that solves the dynamic planner's problem in (15) using the year 2000 benchmark.

Figure 12: Decomposing the shape of optimal $\boldsymbol{y}$ (moderate expansion)

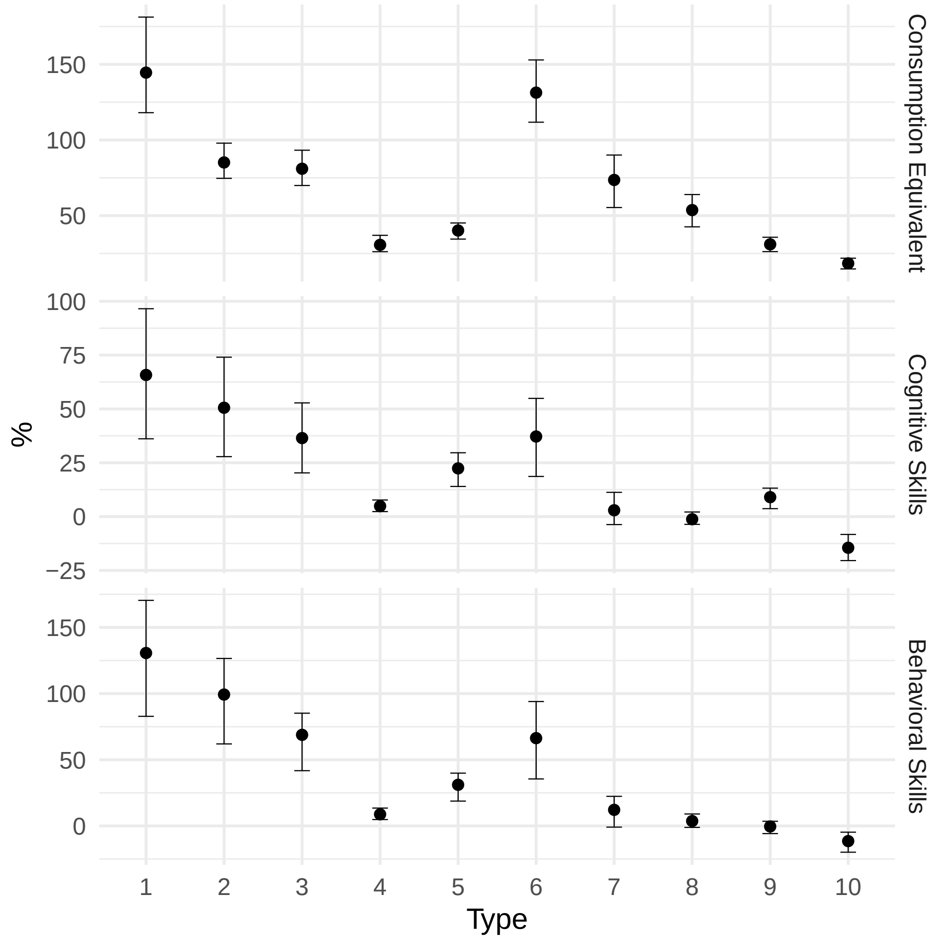


This figure shows the three decompositions (22)-(24) of the wedge between optimal allocations in Figure 11 and the first-best. See text for discussion. 95% bootstrapped credibility intervals are shown at the solution, where production parameters are drawn from the Markov Chain described in Section 5 and behavioral parameters are drawn from the estimated sampling distribution using the standard asymptotic approximation. Dollar amounts are reported in year 2000 USD per month.

7 Conclusion

The results of this paper demonstrate the substantial effect that properly accounting for future child skill development can have on the design of an otherwise standard policy aimed at providing social insurance. The results challenge researchers and policymakers to think about child skill development as a first order component when weighing the costs and benefits of social safety net policies. The relevance of both money and time for skill development has unique and quantitatively dramatic implications for the optimal size and shape of transfers, so it is especially important to have robust and externally valid estimates of these parameters. While this paper takes care to produce credible estimates that are internally consistent with the data and sample of interest, more work is required to reconcile evidence across studies and populations. While the data support meaningful differences in skill impacts through heterogeneity in the relative quality of non-maternal care, it is important to develop data and modeling approaches that can properly

Figure 13: Long-run welfare and skill impacts of optimal $\boldsymbol{y}$ (moderate expansion)



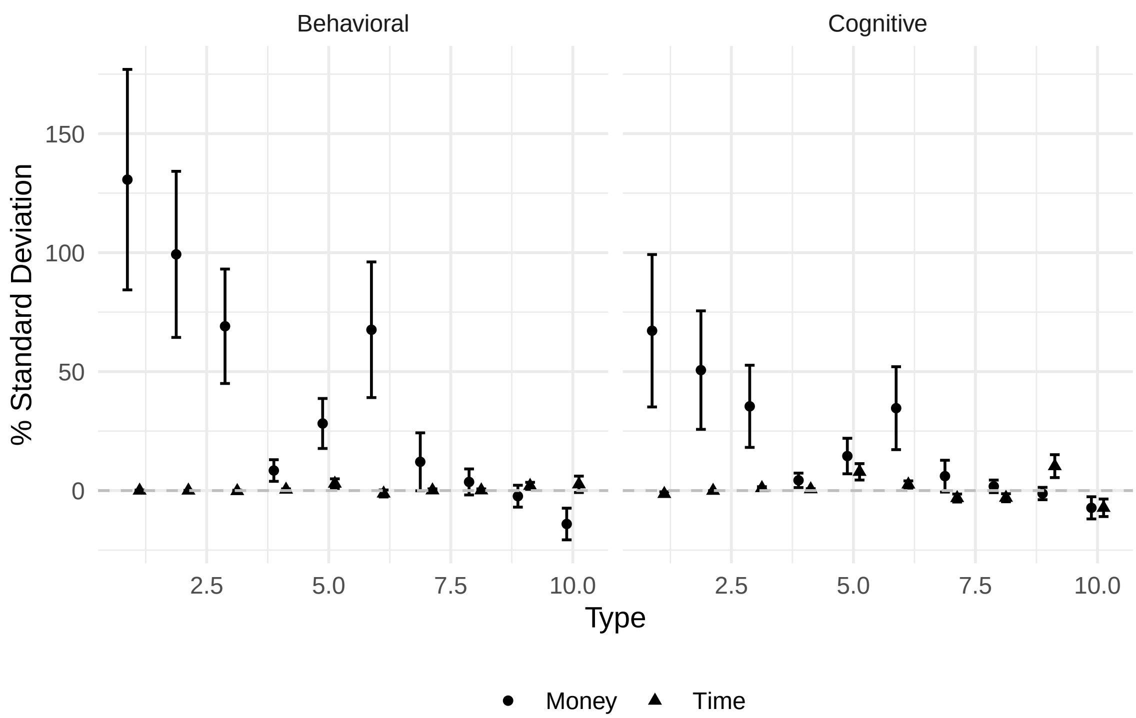
This figure shows the welfare (as measured by an equivalent percentage change in consumption) and skill impacts of the optimal $\boldsymbol{y}$, computed as in Figure 8 with $\boldsymbol{\lambda}$ scaled to three times its original value. 95% bootstrapped credibility intervals are shown for the solution, where production parameters are drawn from the Markov Chain described in Section 5 and behavioral parameters are drawn from the estimated sampling distribution using the standard asymptotic approximation. Average skill impacts are calculated for the youngest child in each household and are expressed in % standard deviations.

articulate, for the purposes of policy, these differences as the outcome of economic decisions.[28] In light of major shifts in the child care market since the the period of this study (1997-2007), this is especially important for updating the policy lessons of this paper to future cohorts.

The three most important directions for future research are therefore to (1) expand the study to a broader population of interest and model selection into and out of these populations through endogenous household formation and fertility; (2) allow for childcare arrangements to be endogenous and consider policies that may adjust decisions along this margin (for example childcare subsidies); and (3) to reconcile evidence across studies on the short and long-run impacts of time and money resources on child outcomes.

[28] Griffen (2019) and Garcia-Vazquez (2024) are two recent papers that make important progress on this issue.

Figure 14: Decomposition of skill impacts of optimal reform $\boldsymbol{y}$ (moderate expansion)



This figure decomposes the long-run skill impacts of the optimal policy in Figure 13 into changes in money and time inputs. 95% bootstrapped credibility intervals are shown, where production parameters are drawn from the Markov Chain described in Section 5 and behavioral parameters are drawn from the estimated sampling distribution using the standard asymptotic approximation. Average skill impacts are calculated for the youngest child in each household and are expressed in % standard deviations.

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.

Aizer, A., H. W. Hoynes, and A. Lleras-Muney (2022): "Children and the US Social Safety Net: Balancing Disincentives for Adults and Benefits for Children," Working Paper 29754, National Bureau of Economic Research.

Akee, R., W. Copeland, E. J. Costello, and E. Simeonova (2018): "How Does Household Income Affect Child Personality Traits and Behaviors?" *American Economic Review*, 108, 775–827.

Arcidiacono, P. and R. A. Miller (2020): "Identifying dynamic discrete choice models off short panels," *Journal of Econometrics*, 215, 473–485.

Attanasio, O., C. Meghir, and E. Nix (2020): "Human Capital Development and Parental Investment in India," *The Review of Economic Studies*, 87, 2511–2541.

Bailey, M. J., H. W. Hoynes, M. Rossin-Slater, and R. Walker (2020): "Is the Social Safety Net a Long-Term Investment? Large-Scale Evidence from the Food Stamps Program," Working Paper 26942, National Bureau of Economic Research.

Bailey, M. J., S. Sun, and B. Timpe (2021): "Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency," *American Economic Review*, 111, 3963–4001.

Barr, A., J. Eggleston, and A. A. Smith (2022): "Investing in Infants: the Lasting Effects of Cash Transfers to New Families," *The Quarterly Journal of Economics*, qjac023.

Bastian, J. and K. Michelmore (2018): "The Long-Term Impact of the Earned Income Tax Credit on Children's Education and Employment Outcomes," *Journal of Labor Economics*, 36, 1127–1163.

Baum, L. E., T. Petrie, G. Soules, and N. Weiss (1970): "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *The annals of mathematical statistics*, 41, 164–171.

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 (2010): "Quasi-structural estimation of a model of childcare choices and child cognitive ability production," *Journal of Econometrics*, 156, 164–189.

——— (2011): "Child Care Choices and Children's Cognitive Achievement : The Case of Single Mothers," *Journal of Labor Economics*, 29, 459–512.

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. 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., K. Jochmans, and J.-M. Robin (2016): "Estimating multivariate latent-structure models," *The Annals of Statistics*, 44, 540–563.

——— (2017): "Nonparametric estimation of non-exchangeable latent-variable models," *Journal of Econometrics*, 201, 237–248.

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

Brilli, Y. (2022): "Mother's Time Allocation, Childcare, and Child Cognitive Development," *Journal of Human Capital*, 16, 233–272.

Bruins, M. (2019): "Taxes, Welfare and the Resources Parents Allocate to Children," *Working Paper.*

Caucutt, E. M. and L. Lochner (2020): "Early and late human capital investments, borrowing constraints, and the family," *Journal of Political Economy*, 128, 1065–1147.

Caucutt, E. M., L. Lochner, J. Mullins, and Y. Park (2020): "Child Skill Production: Accounting for Parental and Market-Based Time and Goods Investments," Working Paper 27838, National Bureau of Economic Research.

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

Chaparro, J., A. Sojourner, and M. J. Wiswall (2020): "Early Childhood Care and Cognitive Development," Working Paper 26813, National Bureau of Economic Research.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011): "How does your kindergarten classroom affect your earnings? Evidence from Project STAR," *The Quarterly journal of economics*, 126, 1593–1660.

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

Chetty, R., A. Guren, D. Manoli, and A. Weber (2013b): "Does indivisible labor explain the difference between micro and macro elasticities? A meta-analysis of extensive margin elasticities," *NBER macroeconomics Annual*, 27, 1–56.

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.

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.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society: series B (methodological)*, 39, 1–22.

Diamond, P. (1980): "Income taxation with fixed hours of work," *Journal of Public Economics*, 13, 101–110.

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.

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

Flood, S., M. King, R. Rodgers, S. Ruggles, J. R. Warren, and M. Westberry (2021): "Integrated Public Use Microdata Series, Current Population Survey: Version 9.0 [dataset]," .

Garcia-Vazquez, M. (2024): "The equilibrium effects of state-mandated minimum staff-to-child ratios," *Working Paper*.

García, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados (2020): "Quantifying the Life-Cycle Benefits of an Influential Early-Childhood Program," *Journal of Political Economy*, 128, 2502–2541.

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

Ge, H., K. Xu, and Z. Ghahramani (2018): "Turing: a language for flexible probabilistic inference," in *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*, 1682–1690.

Gennetian, L. A. and V. Knox (2003): "Staying Single: The Effects of Welfare Reform Policies on Marriage and Cohabitation," Tech. rep., Manpower Demonstration Research Corporation.

Griffen, A. S. (2019): "Evaluating the effects of childcare policies on children's cognitive development and maternal labor supply," *Journal of Human Resources*, 54, 604–655.

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.

Guner, N., R. Kaygusuz, and G. Ventura (2020): "Child-related transfers, household labour supply, and welfare," *The Review of Economic Studies*, 87, 2290–2321.

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

Heckman, J., R. Pinto, and P. Savelyev (2013): "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes," *American Economic Review*, 103, 2052–86.

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*.

Hendren, N. and B. Sprung-Keyser (2020): "A unified welfare analysis of government policies," *The Quarterly journal of economics*, 135, 1209–1318.

Ho, C. and N. Pavoni (2020): "Efficient Child Care Subsidies," *American Economic Review*, 110, 162–99.

Hoffman, M. D., A. Gelman, et al. (2014): "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." *J. Mach. Learn. Res.*, 15, 1593–1623.

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

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.

Hu, Y. and M. Shum (2012): "Nonparametric identification of dynamic models with unobserved state variables," *Journal of Econometrics*, 171, 32–44.

Kasahara, H. and K. Shimotsu (2009): "Nonparametric identification of finite mixture models of dynamic discrete choices," *Econometrica*, 77, 135–175.

Keane, M. P. and K. I. Wolpin (2010): "The role of labor and marriage markets, preference heterogeneity, and the welfare system in the life cycle decisions of black, hispanic, and white women," *International Economic Review*, 51, 851–892.

Kline, P. and C. R. Walters (2016): "Evaluating public programs with close substitutes: The case of Head Start," *The Quarterly Journal of Economics*, 131, 1795–1848.

McFadden, D. (1981): "Econometric models of probabilistic choice," *Structural analysis of discrete data with econometric applications*, 198272.

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.

Mogensen, P. K. and A. N. Riseth (2018): "Optim: A mathematical optimization package for Julia," *Journal of Open Source Software*, 3, 615.

Mullins, J. (2025): "A Structural Meta-Analysis of Welfare to Work Programs and their Impacts on Children," *Journal of Political Economy*.

Nichols, A. L. and R. J. Zeckhauser (1982): "Targeting transfers through restrictions on recipients," *The American Economic Review*, 72, 372–377.

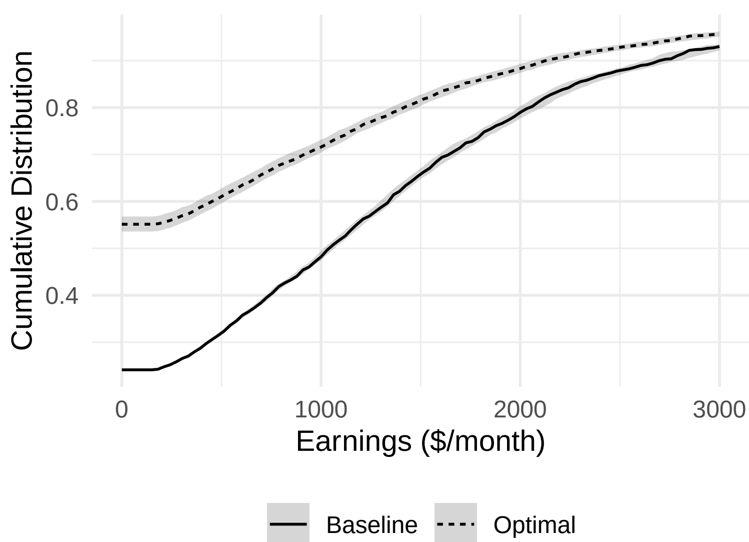
Revels, J., M. Lubin, and T. Papamarkou (2016): "Forward-Mode Automatic Differentiation in Julia," *arXiv:1607.07892 [cs.MS]*.

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.

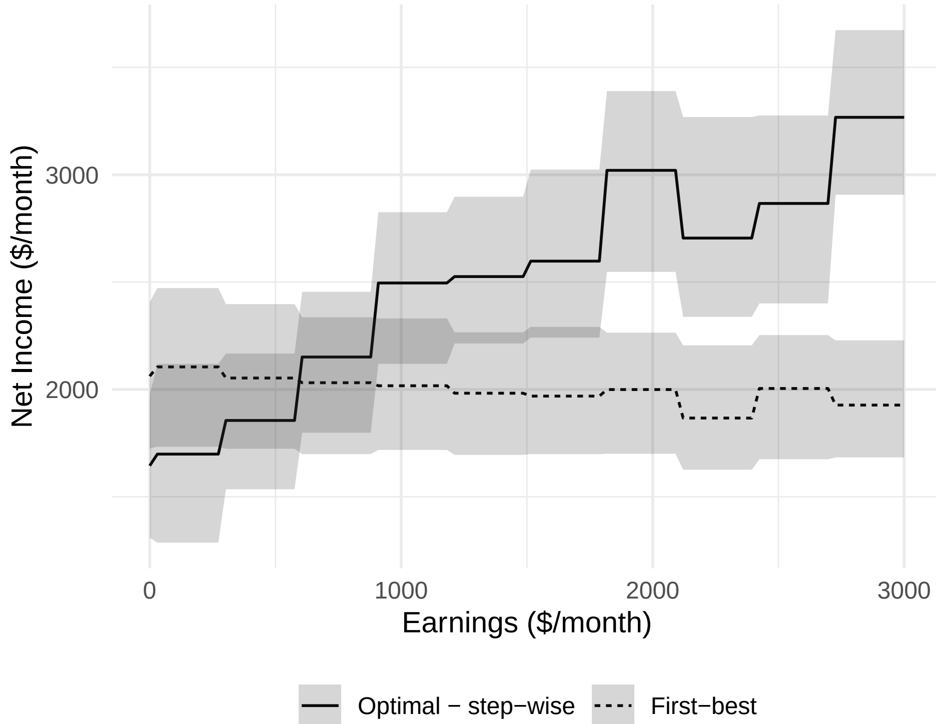
A Figures

Figure 15: Distribution of earnings: benchmark vs optimal $\boldsymbol{y}$



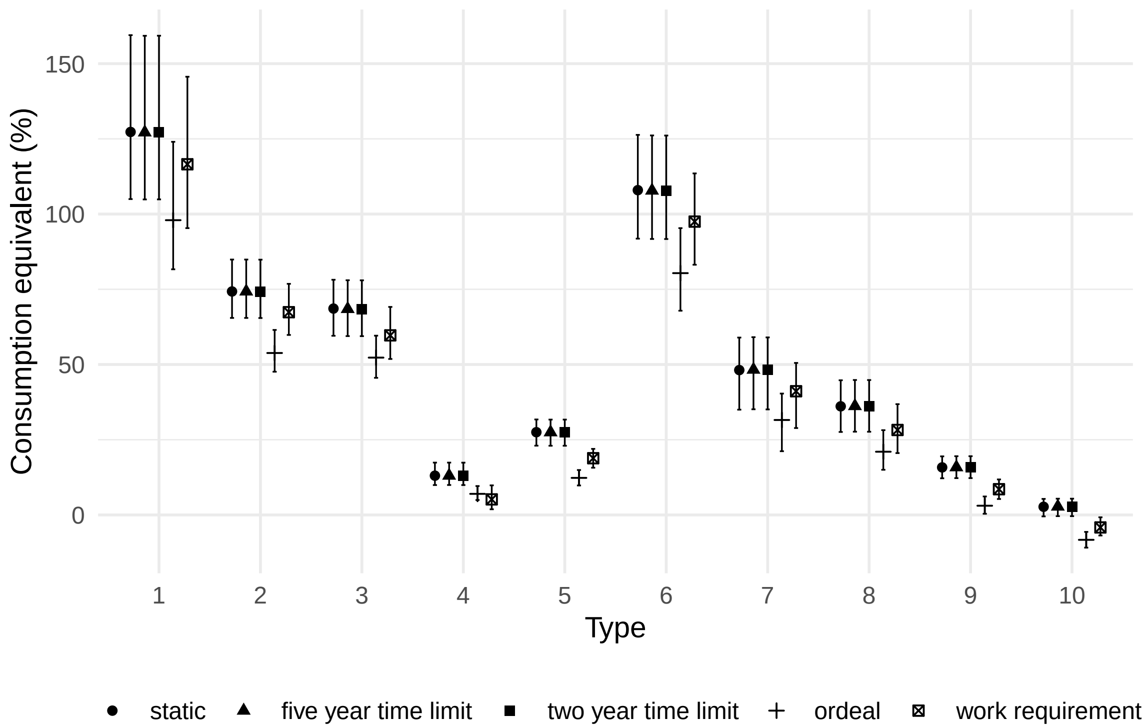
This figure depicts the distribution of earnings in the benchmark year (2000) and under the optimal policy $\boldsymbol{y}$ depicted in Figure 6. 95% bootstrapped confidence intervals are shown, behavioral parameters are drawn from the estimated sampling distribution using the standard asymptotic approximation.

Figure 16: Optimal $\boldsymbol{y}$ vs first-best allocations



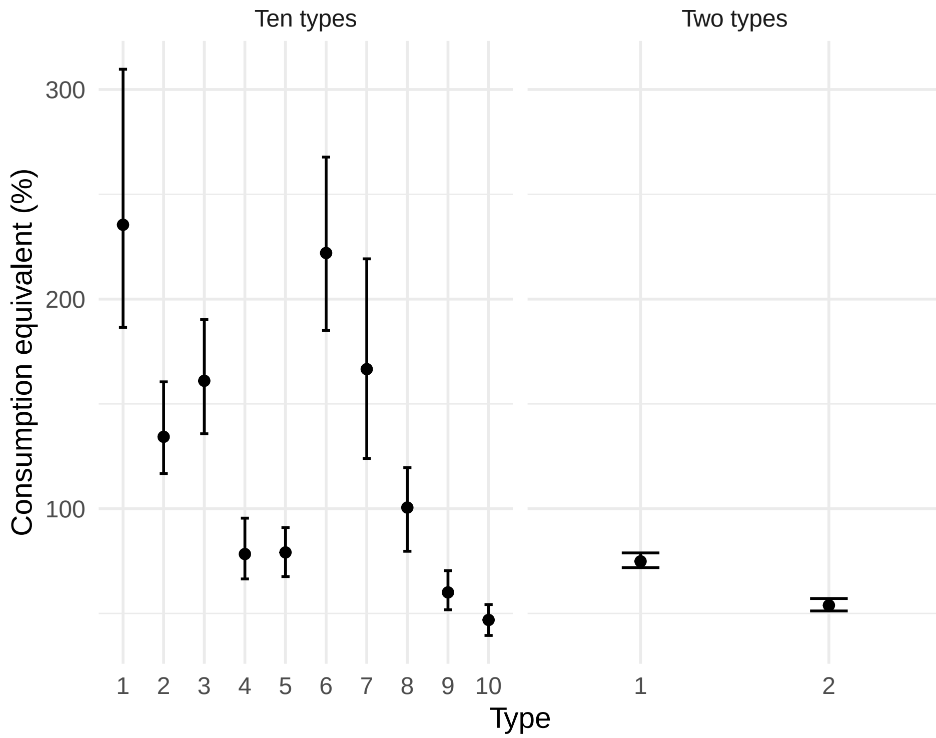
This figure decomposes compares the optimal step-wise allocation $\boldsymbol{y}$ from Figure 6 to the average first best allocation at each earnings level, $\mathbb{E}[\boldsymbol{w}(s)|E_j(s) = e]$. 95% bootstrapped credibility intervals are shown, where production parameters are drawn from the Markov Chain described in Section 5 and behavioral parameters are drawn from the estimated sampling distribution using the standard asymptotic approximation.

Figure 17: Welfare impacts for policy with no skill development



This figure shows the welfare impacts (as measured by an equivalent percentage change in consumption) of an optimal $\boldsymbol{y}$, computed as in Figure 8 but with no skill development concerns (i.e. $\delta_x = \delta_\tau = 0$). 95% bootstrapped confidence intervals are shown at the solution, where behavioral parameters are drawn from the estimated sampling distribution using the standard asymptotic approximation.

Figure 18



This figure compares the welfare impacts (as measured by an equivalent percentage change in consumption) of an optimal $\boldsymbol{y}$, computed as in Figure 8, for two estimated models with $K = 10$ and $K = 2$ types.

B Tables

Table 5: Estimates of type selection coefficients $\Pi_K(\cdot|\mathbf{X})$

Type	Constant	Years Educ	Births ≥ 2	Births ≥ 3	Births ≥ 4	Age First Birth
2	2.17	-0.09	-0.33	-0.53	0.34	-0.03
	(2.44)	(0.13)	(1.20)	(0.71)	(0.62)	(0.06)
3	1.31	-0.14	2.06	-0.18	0.40	-0.06
	(11.29)	(0.10)	(11.15)	(0.58)	(0.47)	(0.04)
4	0.69	0.24	-0.92	-0.29	-0.10	-0.09
	(1.97)	(0.11)	(1.11)	(0.55)	(0.49)	(0.04)
5	0.44	0.05	0.03	-0.01	0.41	-0.02
	(1.95)	(0.10)	(1.16)	(0.58)	(0.47)	(0.04)
6	-0.11	0.34	0.15	-0.16	-0.60	-0.20
	(3.59)	(0.17)	(2.38)	(0.74)	(0.66)	(0.08)
7	1.47	0.14	-1.21	-0.25	-0.54	-0.09
	(3.02)	(0.16)	(1.22)	(0.70)	(0.73)	(0.07)
8	-1.89	0.32	0.11	-0.75	-1.45	-0.06
	(2.41)	(0.13)	(1.19)	(0.60)	(0.84)	(0.05)
9	-0.45	0.12	-0.29	-0.81	0.13	0.03
	(1.91)	(0.10)	(1.08)	(0.55)	(0.48)	(0.04)
10	-2.96	0.58	-0.20	-1.46	-0.21	-0.13
	(2.11)	(0.12)	(1.15)	(0.58)	(0.58)	(0.05)

This table reports maximum likelihood estimates of the coefficients of the multinomial logit type selection probability distribution $\Pi_K(\cdot|\mathbf{X})$. Standard errors in parentheses are estimated using the standard asymptotic approximation.

C Model Solution

A key simplification in the model is that the value function is additively separable in log-skills. Noting that terminal values are, by assumption, additively separable in log-skills, it is sufficient to show that this property is preserved by backward induction. Thus, letting

$$V_{t+1}(s,\theta) = \nu_{t+1}(s) + \alpha_{\theta,k}\left(\sum_{f:a(f,t)+1\geq 0} \Gamma_{a(f,t)+1,c}\log(\theta_{f,c}) + \Gamma_{a(f,t)+1,b}\log(\theta_{f,b})\right)$$

consider the definition of the value function at time t:

$$V_t(s,\theta) = \mathbb{E}_{\epsilon,\eta}\max_{j,c,l,x,\phi_\tau}\left[U_j(C,l,s,\theta,\epsilon) + \beta\mathbb{E}_{s',\theta'}\left[\nu_{t+1}(s') + \alpha_{\theta,k}\left(\sum_{f:a(f,t)+1\geq 0}\Gamma_{a(f,t)+1,c}\log(\theta'_{f,c}) + \Gamma_{a(f,t)+1,b}\log(\theta'_{f,b})\right)\Big|s,\theta,j,x,\phi\right]\right] \tag{25}$$

where utility is given by

$$U_j(C,l,s,\theta,\epsilon) = \log(C) + \alpha_l[(1-h_j)\log(l) + h_j\log(\ell_k)] + \alpha_{\theta,k}\left(\sum_{f:a(f,t)\geq 0}\sum_{i=c,b}\alpha_{\theta,i}\log(\theta_{f,i})\right) + \alpha_j(s) + \epsilon_j.$$

In order to take the conditional expectation of future skills, there are three cases. For all children f who are born at $t+1$ or after ($b_f \geq t+1$, and hence $a(t,f) < 0$):

$$\mathbb{E}[\log(\theta'_{f,i})|s,\theta,j,x,\phi_\tau] = \log(\theta_{f,i}) = \mu_{\theta,k,-1,i},\ i \in \{c,b\}.$$

For all children f who are 18 or older (whose skills are stationary):

$$\mathbb{E}[\log(\theta'_{f,i})|s,\theta,j,x,\phi_\tau] = \log(\theta_{f,i}),\ i \in \{c,b\}.$$

Finally, for children f who are still malleable to inputs ($0 \leq a(f,t) < 18$), the expectation is given by the production technology in (5):

$$\mathbb{E}_{\theta'}[\log(\theta'_{f,i})|s,\theta,j,x,\phi_\tau] = \mu_{\theta,k,a,i}+\delta_{x,a,i}\log(x_f)+\delta_{\tau,a,i}(1-h_j)\log(\phi_{\tau,f})+\delta_{\tau,a,i}h_j\log(\varphi_f(s))+\delta_{\theta_{c,i}}\log(\theta_{f,c})+\delta_{\theta,b,i}\log(\theta_{f,b}), \qquad i \in \{c,b\}$$

Substituting these rules, along with utility, into (25) yields:

$$\begin{aligned}
V_t(s,\theta) = \mathbb{E}\max_{j,C,l,x,\phi_\tau} \Bigg\{ & \log(C) + \alpha_{\theta,k}\sum_{f\in\mathcal{A}}\beta\left(\delta_{x,a,c}\Gamma_{a+1,c}+\delta_{x,a,b}\Gamma_{a+1,b}\right)\log(\phi_{x,f})+ \\
& (1-h_j)\left(\alpha_l\log(l) + \sum_{f\in\mathcal{A}}\beta\left(\delta_{\tau,a,c}\Gamma_{a+1,c}+\delta_{\tau,a,b}\Gamma_{a+1,b}\right)\log(\phi_{\tau,f})\right)+ \\
& h_j\left(\log(\ell_k) + \sum_{f\in\mathcal{A}}\beta\left(\delta_{\tau,a,c}\Gamma_{a+1,c}+\delta_{\tau,a,b}\Gamma_{a+1,b}\right)\log(\varphi_f)\right)+ \\
& \alpha_j(s)+\epsilon_j+\beta\mathbb{E}[\nu_{t+1}(S')|S,j]\Bigg\} \\
& +\alpha_{\theta,k}\sum_{f\in\mathcal{A}}\sum_{i\in\{c,b\}}\left(\alpha_{\theta,i}+\beta(\delta_{\theta,i,c}\Gamma_{a(f,t)+1,c}+\delta_{\theta,i,b}\Gamma_{a(f,t)+1,b})\log(\theta_{f,i})\right) \\
& +\alpha_{\theta,k}\left(\sum_{f\notin\mathcal{A}}\sum_{i\in\{c,b\}}\left(\mathbf{1}\{a(f,t)\geq 18\}\alpha_{\theta,i}+\beta\Gamma_{a(f,t),i}\right)\log(\theta_{f,i})\right)
\end{aligned} \tag{26}$$

subject to the resource constraints (10). Notice that terms that do not affect the optimization problem can be removed from the max operator. The first four lines of the equation define $\nu_t(S)$, as in (9), along with the recursive coefficients that embody the total value of money and time inputs at age a:

$$\begin{aligned}
\Gamma_{\tau,a} &= \beta\left(\delta_{\tau,a,c}\Gamma_{a+1,c}+\delta_{\tau,a,b}\Gamma_{a+1,b}\right) \\
\Gamma_{x,a} &= \beta\left(\delta_{x,a,c}\Gamma_{a+1,c}+\delta_{x,a,b}\Gamma_{a+1,b}\right).
\end{aligned}$$

The final two lines define the recursive coefficients in (8) that define the discounted present value derived from current child skills:

$$\Gamma_{a,i} = \begin{cases} (1-\beta)^{-1}\alpha_{\theta,i} & \text{for } a \geq 18 \\ \alpha_{\theta,i}+\beta\left[\delta_{\theta,i,c}\Gamma_{a+1,c}+\delta_{\theta,i,b}\Gamma_{a+1,b}\right] & \text{for } 0 \leq a < 18 \\ \beta\Gamma_{a+1,i} & \text{for } a < 0 \end{cases}$$

It is clear from this equation that the first order conditions for investment choices are:

$$\begin{aligned}
\frac{\alpha_l}{l} &= \frac{\alpha_{\theta,k}\Gamma_{\tau,a(f,t)}}{\phi_{\tau,f}},\ f\in\mathcal{A} \\
\frac{1}{C} &= \frac{\alpha_{\theta,k}\Gamma_{x,a(f,t)}}{x_f},\ f\in\mathcal{A}
\end{aligned}$$

which, when combined with the resource constraints in (10), produces the formulae for optimal input choices in (11) and (12). Substituting optimal input choices into (9) gives indirect utility:

$$\begin{aligned}
u_{k,j}(y,\boldsymbol{a}) = \log\left(\frac{1}{1+\alpha_{\theta,k}\Gamma_x(\boldsymbol{a})}\right) + (1-h_j)\log\left(\frac{\alpha_l}{\alpha_l+\Gamma_\tau(\boldsymbol{a})}\right) + h_j\log(\ell_k)+ \\
\alpha_{\theta,k}\sum_{f\in\mathcal{A}}\left(\Gamma_{x,a(f,t)}\log\left(\frac{\alpha_{\theta,k}\Gamma_{x,a(f,t)}}{1+\Gamma_x(\boldsymbol{a})}\right)+(1-h_j)\Gamma_{\tau,a(f,t)}\log\left(\frac{\alpha_{\theta,k}\Gamma_{\tau,a(f,t)}}{\alpha_l+\Gamma_\tau(\boldsymbol{a})}\right)+h_j\log(\varphi_f(s))\right)
\end{aligned} \tag{27}$$

where the pair of coefficients $(\Gamma_x(\boldsymbol{a}), \Gamma_\tau(\boldsymbol{a}))$ represent the total developmental importance of time and money for the household given the ages of each developing child:

$$\Gamma_x(\boldsymbol{a}) = \sum_{a\in\boldsymbol{a}} \Gamma_{x,a},\ \Gamma_\tau(\boldsymbol{a}) = \sum_{a\in\boldsymbol{a}} \Gamma_{\tau,a}.$$

Finally, collecting terms and including the simplification that $\varphi_f(s) = \overline{\varphi}_k\varphi_{\tau,f}$ results in equation (13).

C.1 Multiple Private and Public Investment Categories

Note that the key step in deriving indirect utility and the outcome equation was showing that investments are proportional to total resources, e.g. $x_f = \phi_f(s)y$. Suppose that each x_f is itself a Cobb-Douglas aggregate of multiple private and public investment categories:

$$\{x_l^{pu}\}_{l=1}^{L_1}, \{x_{l,f}^{pr}\}_{f=1,l=1}^{F,L_2}.$$

The first order conditions would be:

$$1/C = \frac{\beta\sum_f(\delta_{pu,l,c}\Gamma_{a(f)+1,c} + \delta_{pu,l,b}\Gamma_{a(f)+1,b}}{x_l^{pu}} \tag{28}$$

$$1/C = \frac{\beta(\delta_{pr,l,c}\Gamma_{a(f)+1,c} + \delta_{pr,l,b}\Gamma_{a(f)+1,b}}{x_{l,f}^{pr}} \tag{29}$$

and when combining with resource constraints we would still get proportional investment rules for each of these expenditure categories. Thus, the empirical content of the model is preserved in the presence of this model extension.

C.2 Model Complexity and Limitations

Two assumptions combine in this model to make it sufficiently tractable for estimation. The first is to assume that the full sequences of policy variables and birth outcomes are known. The vector of policy variables, Z_t, is very high-dimensional relative to what is typically computationally feasible[29], as is the space of potential birth histories, $\boldsymbol{b}$. By assuming that these sequences of variables are known, the dimension of the state space for any single agent is reduced to $\sum_{t=1}^{T}\mathcal{E}\times(\Omega+1)\times\mathbb{R}^2$, and any evaluation of an estimation criterion that uses all data points therefore requires a solution to the approximate order of $\sum_{m=1}^{M}\sum_{t=1}^{T_m}\mathcal{E}\times(\Omega_{m,t}+1)\times\mathbb{R}^2$. The combined dimension of the state space for Z_t and $\boldsymbol{b}$ can essentially be traded for sample size, M, which is much smaller.

The second key simplification comes from a pair of assumptions: that preferences over child skills take a log form, and child skills combine with aggregate investment in a Cobb-Douglas form to produce future skills. According to the previous section the value function is then additively separable in log-skills, further reducing the dimension of the state space for an individual problem to $\sum_{t=1}^{T_m}\mathcal{E}\times(\Omega_{m,t}+1)$, which delivers a sufficiently tractable case when the model must be solved to the order of sample size M for the purposes of estimation.

While these assumptions are necessary for tractability, they do impose some *a priori* empirical content on the model, mainly stemming from the additive separability between child skills and other dynamic state variables. In periods prior to welfare reform when time limits do not apply, all of the dynamics of the model are loaded into the recursive coefficients on consumption and leisure $(\Gamma_x(\boldsymbol{a}), \Gamma_\tau(\boldsymbol{a}))$, which implies that choices are unaffected by the agent's beliefs about future policy environments or births. Hence, there are no anticipatory effects from policy changes. Based on the model's fit of the data (discussed in Section 5), it does not appear that these anticipatory effects are necessary to fit observed time series patterns in labor supply and program participation.

Of course, anticipation effects would reappear with the introduction of additional ingredients that either augment the dynamics of the model, or break the additive separability result. With regards to preferences and technology, there is additional empirical content beyond the simplification of dynamics. On the developmental side, the elasticity of

[29] Recall that this vector includes (but is not limited to) the full set of federal and state marginal tax rates as well as income brackets, state poverty guidelines and benefit standards for up to 4 different family sizes, fixed and variable earnings disregards, and parameters determining income eligibility tests. All of these are continuous variables.

substitution between the stock of skills and aggregate investment is restricted to be one, which imposes a particular form of complementarity. However, previous research has shown that the Cobb-Douglas specification is difficult to reject in the data (Cunha et al., 2010; Attanasio et al., 2020). On the behavioral side, the model's implications are quite a bit stronger: it produces linear investment equations that do not depend on the child's current stock of skills. This implication is testable, and Appendix D uses within-family variation in skills to show that the data cannot reject it.

D Model Extensions and Specification Tests

D.1 Testing Investment Rules

While the combination of Cobb-Douglas technology and log preferences produces an important simplification of the model solution, it also places severe restrictions on mothers' investment behavior, which is testable. Recall that time investment for mother m on child f at time t is:

$$\phi_{\tau,f,t}(m) = \frac{\alpha_{\theta,k(m)}\Gamma_{a(m,f,t)}}{\alpha_l + \alpha_{\theta,k(m)}\sum_{f'\in\mathcal{A}(m,t)}\Gamma_{\tau,a(m,f',t)}}.$$

Taking a log-transformation of this specification, and allowing for measurement error in observed time use, time investment decomposes into an age-specific constant, $\gamma_{a(m,f,t)}$, a mother-year specific term $\mu_{m,t}$, and measurement error:

$$\log(\phi^o_{m,f,t}) = \mu_{m,t} + \gamma_{a(m,f,t)} + \epsilon_{m,f,t}$$

This specification embodies some of the strict empirical content of the model, namely that investment is a function of preferences and the age-composition of children only. It also suggests that one could test such a restriction using sibling pairs with mother-time and age fixed effects. The following regression:

$$\log(\phi^o_{m,f,t}) = \mu_{m,t} + \gamma_{a(m,f,t)} + \beta_1\text{LW}_{m,f,t} + \beta_2\text{BPE}_{m,f,t} + \epsilon_{m,f,t}$$

forms the basis of such a test. Under the null hypothesis that the model has been correctly specified, the estimands β_1 and β_2 are equal to zero. This approach to testing is appealing for two reasons. It has an exactly coherent interpretation within the model (the null hypothesis is structurally motivated), but it also produces the familiar sibling-pair design, which seeks to control for unobserved family level factors (in this case, the preference for child investment interacted with the age-composition of the family) by exploiting within-family variation in skill endowments. Table 6 reports estimates of this specification, using both the preferred measure of investment time, as well as the total time that the mother spends actively participating in activities with the child. In order to handle any attenuation bias caused by measurement error in skills, Table 6 also reports results where the Applied Problems (AP) score instruments for Letter-Word (LW) scores, and Internalizing Behavior scores (BPN) instrument for Externalizing Behavior Scores (BPE). These specifications do not find any strong evidence of an impact of cognitive or behavioral skills on time investment, and the null hypothesis that $\beta_1 = \beta_2 = 0$ is not rejected. Given evidence here and elsewhere (Cunha and Heckman, 2008; Cunha et al., 2010) that current skills are complementary with investments, this evidence suggests that the current model fits the data better than alternatives that do not allow for this separability.

Table 6: Testing for effects of skill on investment behavior

Dependent Variables:	Investment Time		Total Active Time	
Model:	OLS	IV	OLS	IV
Variables				
Letter Word	-0.0466	-0.1835*	-0.0367	-0.0925
	(0.0468)	(0.1034)	(0.0360)	(0.0717)
BPE	0.0294*	0.0377	-0.0030	-0.0087
	(0.0159)	(0.0316)	(0.0100)	(0.0196)
Fixed-effects				
Mother×Year	Yes	Yes	Yes	Yes
Child Age	Yes	Yes	Yes	Yes
Fit statistics				
Observations	1,042	1,026	1,952	1,924
R^2	0.93718	0.93630	0.86562	0.86545

Clustered (Mother× Year) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

D.2 Childcare Quality

Consider an extension to the baseline model of Section 2 that allows for childcare quality to be purchased at an hourly price q rather than being an exogenous function of type. Holding fixed the rest of the model, the budget constraint can be augmented to:

$$C + \sum_f x_{f,t} + qh_j \sum_f \varphi_{f,t} \leq Y_j(s).$$

The first order condition for φ_f is:

$$\frac{q}{C} = \frac{\alpha_{\theta,k}\Gamma_{\tau,a(f,t)}}{\varphi_{f,t}}$$

while the first order conditions for the other investment variables are unaffected. This gives optimal money and childcare inputs as:

$$q\varphi_f = \frac{\alpha_{\theta,k}\Gamma_{\tau,a(f,t)}}{1 + \alpha_{\theta,k}\Gamma_x(\boldsymbol{a}) + \alpha_{\theta,k}h_j\Gamma_\tau(\boldsymbol{a})} \times Y_t$$

$$x_f = \frac{\alpha_{\theta,k}\Gamma_{x,a(f,t)}}{1 + \alpha_{\theta,k}\Gamma_x(\boldsymbol{a}) + \alpha_{\theta,k}h_j\Gamma_\tau(\boldsymbol{a})} \times Y_t.$$

Substituting these input rules into the outcome equation yields:

$$\log(\theta_{f,t+1}) = \mu_{k,a(f,t)} + \delta_x \log(Y_t) + \delta_\tau h_t \log(Y_t) + \delta_\theta \log(\theta_{f,t}) + g_k(h_j, \boldsymbol{a}) + \eta$$

where $g_k(h_t, \boldsymbol{a})$ is a function of type (k), hours (h_t), and the age of each child ($\boldsymbol{a}$).

Approximating this latter function as being linear in h_t, this equation suggests that the baseline model can be extended by adding an interaction term between net income and hours to the estimated outcome equation. In words: if childcare quality can be purchased out of income, then quality will be increasing in income. The importance of this quality for outcomes must increase with the amount of time spent in this case, and so the interaction will identify the factor share of this input on skills.

Figure 19: Estimates of the factor share of childcare quality in extended model

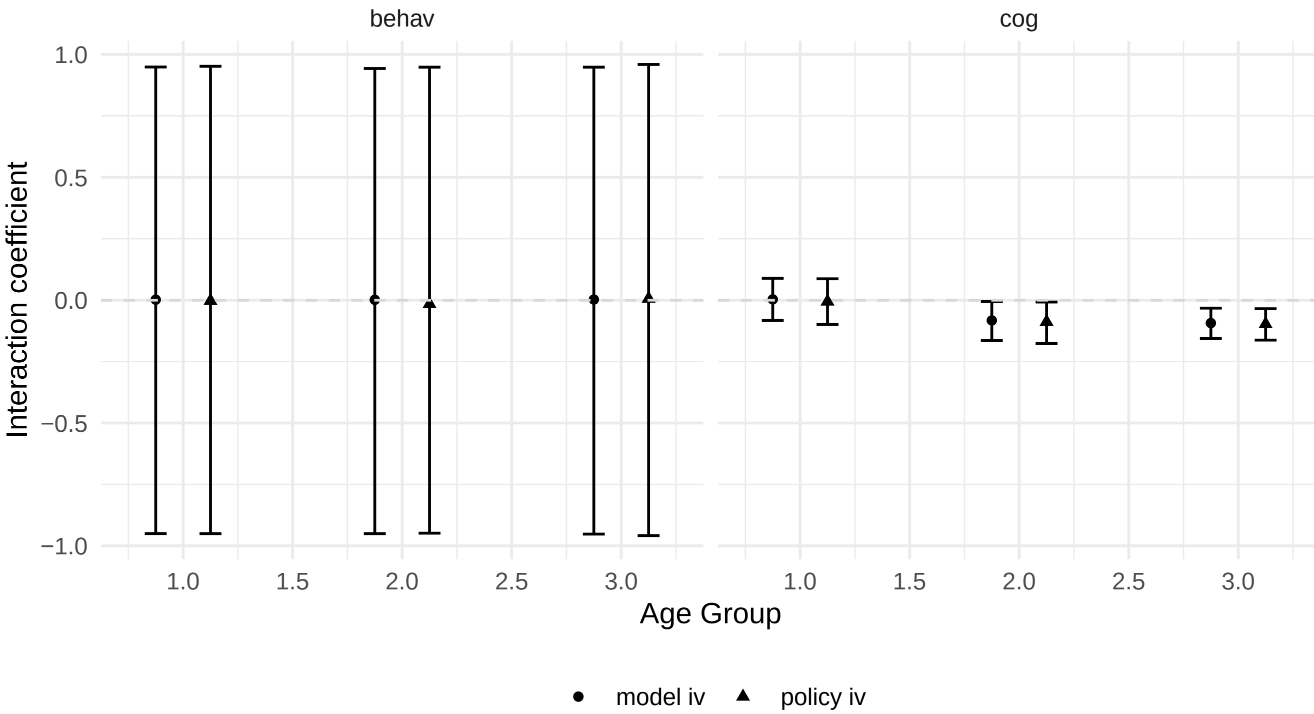


Figure 19 presents the estimates of this interaction term after augmenting the outcome equation to feature the childcare extension. The point estimates and credibility intervals do not support this model and, to the extent that estimates can be distinguished from zero, they suggest a *negative* interaction between hours and income rather than a positive one. These results informed a preference for the baseline model as the framework for analysis in this paper.

D.3 Mother's Human Capital

Consider now an extension of the model in which mothers' wages may appreciate while working and, conversely, depreciate when not working. In the extended model the wage shock, ε, moves up and down the grid space with with transition probabilies that depend on whether mother's work at least 30 hours per week.[30] To allow for the fact that returns to human capital may be heterogeneous, the extended model allows these probabilities to be type-dependent. All other aspects of the model are identical. Table 7 presents the maximum likelihood estimates of this transition process for a model with 6 types.[31]

Notice in Table 7 that, fixing the work decision, the probability of moving up or down the grid space is close to symmetric for each type. Thus, the estimates do not support a model where wages appreciate on average as a result of labor force attachment. This result echoes the findings of Blundell et al. (2016).

E Identification and Estimation

E.1 Measurement Assumptions

Additive measurement error contaminates observations of wages, time investment, and child skills in the model. The equations below summarize the measurement assumptions that inform estimation of behavioral and production param-

[30] Recall that in the baseline model a single probability π_W dictates these movements.

[31] The reduction in the number of types relative to the baseline is necessary due to computational constraints.

Table 7: Estimates of wage transition probabilities for extended model

	Move Up		Move Down	
	No work	Work	No work	Work
Type 1	0.25	0.07	0.20	0.11
	(0.08)	(0.17)	(0.09)	(0.13)
Type 2	0.15	0.13	0.14	0.14
	(0.04)	(0.12)	(0.03)	(0.08)
Type 3	0.15	0.13	0.14	0.14
	(0.02)	(0.02)	(0.03)	(0.02)
Type 4	0.21	0.13	0.17	0.13
	(0.07)	(0.09)	(0.06)	(0.08)
Type 5	0.22	0.09	0.21	0.08
	(0.03)	(0.02)	(0.03)	(0.02)
Type 6	0.11	0.03	0.15	0.05
	(0.02)	(0.01)	(0.03)	(0.01)

eters.

$$\log(W^o_{m,t}) = \log(W_{m,t}) + \zeta_{m,t,W}, \qquad \zeta_{m,t,W} \sim \mathcal{N}(0, \sigma^2_W) \tag{30}$$

$$\log(\phi^o_{\tau,m,t}) = \log(\phi_{\tau,m,t}) + \zeta_{m,t,\tau}, \qquad \zeta_{m,t,\tau} \sim \mathcal{N}(0, \sigma^2_\tau) \tag{31}$$

$$LW_{m,f,t} = \log(\theta_{m,f,t,c}) + \zeta_{m,f,t,LW}, \qquad \zeta_{m,f,t,LW} \sim \mathcal{N}(0, \sigma^2_{LW}) \tag{32}$$

$$AP_{m,f,t} = \lambda_c \log(\theta_{m,f,t,b}) + \zeta_{m,f,t,AP}, \qquad \zeta_{m,t,f,AP} \sim \mathcal{N}(0, \sigma^2_{AP}) \tag{33}$$

$$BE_{m,f,t} = \log(\theta_{m,f,t,b}) + \zeta_{m,f,t,BE}, \qquad \zeta_{m,t,f,BE} \sim \mathcal{N}(0, \sigma^2_{BE}) \tag{34}$$

$$BN_{m,f,t} = \lambda_b \log(\theta_{m,f,t,b}) + \zeta_{m,f,t,BN}, \qquad \zeta_{m,t,f,BN} \sim \mathcal{N}(0, \sigma^2_{BN}) \tag{35}$$

where each error term is independently distributed of mothers, children, and time periods. The noisy measures of skills are normalized to have mean zero. The presence of two measures per skill outcome is sufficient to guarantee identification of the measurement system according to the standard covariance restrictions (Cunha et al., 2010).

E.2 Identification of Choice Probabilities

Since preference parameters provide a low-dimensional parameterization of choice probabilites, $P(s)$, it is sufficient to establish non-parametric identification of those choice probabilities along with the objects that govern the initial distribution of latent states (Π_K) and transition probabilities (Π_ε). In this paper, transition laws for observed state variables are deterministic and known.

While the exact conditions required for identification of these models vary depending on the key sources of variation, a unifying theme is that they require a sufficiently large panel dimension. Hu and Shum (2012) for example show that models of this kind can be identified with a panel of length 5, which these data easily satisfy (see Section 4). The key requirement in Hu and Shum (2012) is a high level condition on the uniqueness of the spectral decomposition of a linear operator defined by the joint distribution of observable and unobservable variables over four periods. Kasahara and Shimotsu (2009) and Bonhomme et al. (2016) provide somewhat more intuitive conditions that rely, respectively, on sufficient variation in observable covariates that affect choice probabilities and on a sufficient number of conditionally independent measures of the latent state. Both papers establish identification of stationary choice probabilities with panels of length 3. This model provides variation in both dimensions, and so identification is established through either set of results[32].

With choice probabilities and transitions in hand, the parameters that govern utility for each type are identified by the parametric restrictions that map them to choice probabilities.[33]

E.3 Behavioral Parameters

E.3.1 Identification

Recall that total time investment is given by

$$\phi_{m,t} = \frac{\alpha_{\theta,k(m)}\Gamma_\tau(\boldsymbol{a}_{m,t})}{\alpha_l + \alpha_{\theta,k(m)}\Gamma_\tau(\boldsymbol{a}_{m,t})}$$

which identifies the vector Γ_τ for each age, as well as each $\alpha_{\theta,k}$ up to the scale of α_l. While this is not strictly necessary for identification, it implies that observations of time inputs provide an important source of information when inferring types and deciding on heterogeneity in the weight $\alpha_{\theta,k}$ that mothers place on child skill outcomes.

Next, for simplicity, assume a policy environment in which time limits and work requirements do not apply. Recall that the distribution F_ϵ of taste shocks is summarized by the vector σ of dispersion parameters. As is standard (McFadden, 1981), an inverse mapping G_σ exists such that:

$$\log\left(\frac{Y_j(s)}{Y_0(s)}\right) + \alpha_{j,k} = G_\sigma(P(s))$$

where $\alpha_{j,k}$ is the utility from discrete choice j that varies according to type. In this nonlinear transformation of the choice probabilities, we can see that $\alpha_{j,k}$ must adjust to pin down the levels of the probabilities for each type k, while the dispersion parameters σ determine the responsiveness of participation and labor supply with to financial returns, as measured by $\log\left(\frac{Y_j(s)}{Y_0(s)}\right)$.

E.3.2 Estimation

Recall that the triple $(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ indicates the set of observable exogenous variables, policy variables, and endogenous variables. The full log-likelihood of this triple is:

$$\begin{aligned}\mathcal{L}(\mathbf{Y}_m|\mathbf{Z}_m, \mathbf{X}_m, \Theta) = \log\Bigg(\sum_i \Pi_K(k|\mathbf{X}) \sum_{\varepsilon_m^T} \Pi_\varepsilon^o(\varepsilon_0) \prod_{t=0}^{T_m} \Pi_\varepsilon(\varepsilon_{t+1}|\varepsilon_t) \\ \times \prod_{j=0}^{J-1} P_j(s_{m,t};\Theta)^{\mathbf{1}\{j_{m,t}=j\}} \\ \times \left[\frac{1}{\sigma_W}\phi\left(\frac{\log(W_{m,t}^o) - \gamma_{W,k,0} - \gamma_{W,k,1}Age_{m,t} - \gamma_{W,k,2}Age_{m,t}^2 - \sigma_\varepsilon \varepsilon_t}{\sigma_W}\right)\right]^{\mathbf{1}\{H_{m,t}>0\}} \\ \times \left[\frac{1}{\sigma_\tau}\phi\left(\frac{\log(\phi_{\tau,m,t}^o) - \log\left(\frac{\alpha_{\theta,k}\sum_{f\in\mathcal{A}(m,t)}\Gamma_{\tau,a(m,f,t)}}{\alpha_l + \alpha_{\theta,k}\sum_{f\in\mathcal{A}(m,t)}\Gamma_{\tau,a(m,f,t)}}\right)}{\sigma_\tau}\right)\right]^{\mathbf{1}\{y(m,t)\in\{1997,2002\}\}}\Bigg).\end{aligned} \quad (36)$$

[32]The problem can be recast as identification of a finite mixture model with the unobserved state being $(k, \varepsilon_1, \varepsilon_2, \varepsilon_3)$ which has dimension $K \times K_\varepsilon^3$. In this case Kasahara and Shimotsu (2009) require that there be sufficiently many covariates in the state variable s that move choice probabilities across types. Recall that this model has a very high dimension of s, including benefit generosity, earnings disregards, the number and age of all children, and tax code parameters, all of which move choice probabilities differently across types, providing identification through this channel. Bonhomme et al. (2016, 2017) require three conditionally independent measures of the latent state, which in this case is provided by setting the state variable as (k, ε) and taking observations of earnings (E_{t-1}, E_t, E_{t+1}). Since earnings are continuously distributed, a partioning exists to satisfy rank conditions for each conditional earnings distribution.

[33]A previous version of this paper took a large panel approach to identification, showing sufficient within-panel variation in wages was sufficient to identify mother m's preferences as T_m grew large. This motivated the use of a score criterion with which a clustering approach was taken, following Bonhomme and Manresa (2015). Details are available for this method upon request.

where $\phi(\cdot)$ is the density of the standard normal. Observed hours choices are rounded to their nearest value on the grid $\{0, 15, 30, 45\}$ to determine (along with program participation) the observed discrete choice. Since evaluating this likelihood involves summing over all types, all potential sequences of the wage shock ε, and all potential choices j when $j_{m,t}$ is missing, it is not technically feasible to estimate directly. The Expectation-Maximization offers a practical alternative (Dempster et al., 1977). Each iteration computes the prior over types, wage shocks, and missing choices (the "expectation" step) using the Forward-Back algorithm (Baum et al., 1970).[34] The maximization step updates parameters in blocks using a small number of Newton iterations in `Optim.jl` (Mogensen and Riseth, 2018) using automatic differentiation with `ForwardDiff.jl` (Revels et al., 2016). All results in the paper use an estimate of the sampling variance of this estimator based on the standard asymptotic approximation using the inverse of the covariance of the score of the log-likelihood.

Figure 20: Extensive Marginal Labor Supply Elasticities by Type

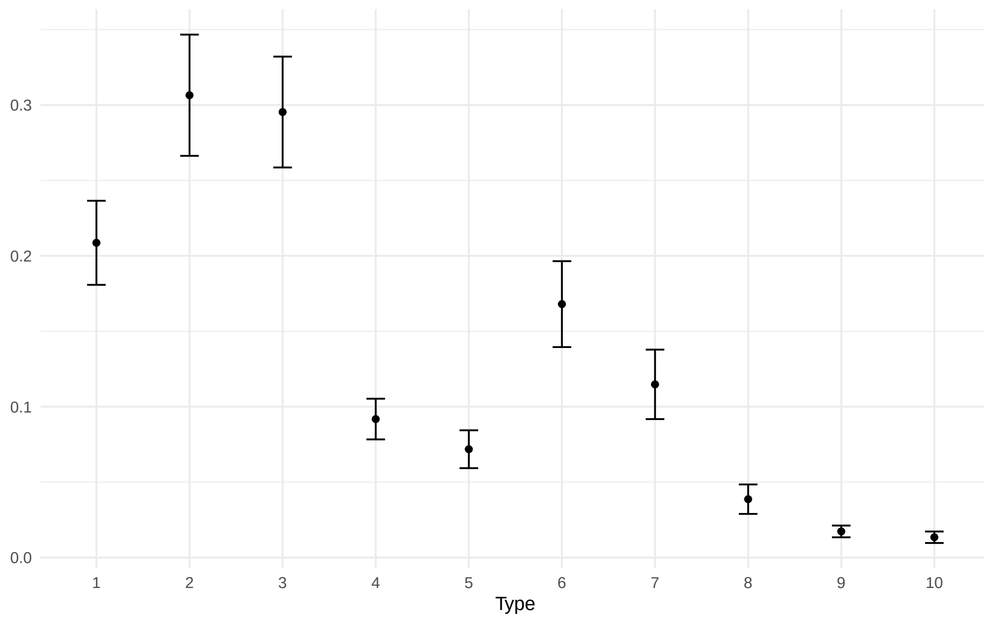


Figure 21: Caption

E.4 Elasticities and Type Characteristics

Figure 20 plots the average extensive marginal elasticity of labor supply by type, indicating substantial differences in the responsiveness of work behavior to incentives. These differences are driven largely by differences in mean rates of employment by group, with types that have very high rates of employment exhibiting lower elasticities. Excluding types with very high labor force attachment, the numbers are in the lower range of previous estimates (Chetty et al., 2013b). Figure 22 depicts estimated averages of specific demographic characteristics by type. It shows that there are indeed systematic correlations between latent types and fertility behavior (as measured by mother's age at first birth and total births), as well as education (as measured by High School and College graduation). Importantly however, these relationships are not monotonic in the endogenous variables of interest (labor force and program participation) which underscores the importance of controlling beyond observable heterogeneity.

[34] The M-Step of this estimator is

Figure 22: Mean characteristics by type

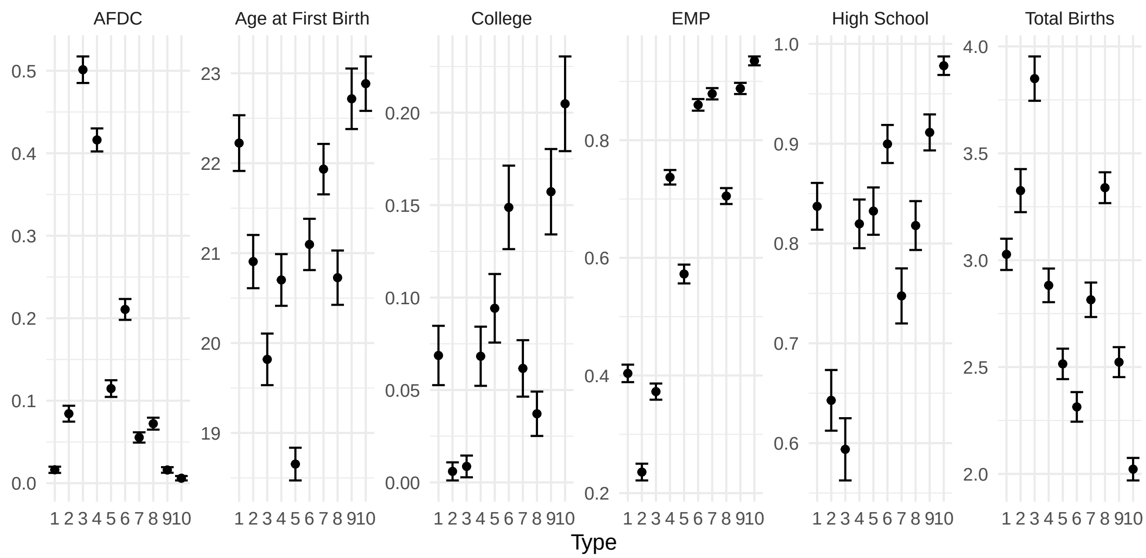


Figure 23: Caption

E.5 Estimation of Production Parameters

To avoid notational clutter, this section suppresses the dependence of variables on the mother m and child f. Assume for now that the type k of an individual is observable. Recall the outcome equation (14):

$$\log(\theta_{t+1}) = \mu_{\theta,k,a} + \delta_{x,a}\log(Y_{m,t}) + \delta_{\tau,a}\log(\overline{\varphi}_k)h_{m,t} + \delta_\theta \log(\theta_t) + e_t + \eta_t$$

where e is a function of type k and the age of each child in the household. Since skills are observed every five years in the CDS, iterate this equation forward for five periods to get:

$$\log(\theta_{t+5}) = \sum_{s=0}^{4} \delta_\theta^{4-s}\left(\mu_{\theta,k,a+s} + \delta_{x,a}\log(Y_{t+s}) + \delta_{\tau,a}\log(\overline{\varphi}_k)h_{t+s} + e_{t+s} + \eta_{t+s}\right) + \delta_\theta^5\log(\theta_t)$$

Let $S_{1,t} = [LW_t,\ BE_t]'$ be the first set of measures of each skill, and let $S_{2,t} = [AP_t,\ BN_t]'$ be the second set of measures. Let $\zeta_{1,t} = [\zeta_{LW,t},\ \zeta_{BE,t}]'$ and $\zeta_{2,t} = [\zeta_{AP,t},\ \zeta_{BN,t}]'$ be their respective measurement errors. Substituting in these scores according to the measurement equations (32)-(32) gives:

$$S_{1,t+5} = \sum_{s=0}^{4} \delta_\theta^{4-s}\left(\mu_{\theta,k,a+s} + \delta_{x,a+s}\log(Y_{t+s}) + \delta_{\tau,a+s}\log(\overline{\varphi}_k)h_{t+s} + e_{t+s} + \eta_{t+s}\right) + \delta_\theta^5 S_{1,t} - \delta_\theta^5\zeta_{1,t} + \zeta_{1,t+4} \quad (37)$$

$$S_{2,t+5} = \Lambda^{-1}\left(\sum_{s=0}^{4} \delta_\theta^{4-s}\left(\mu_{\theta,k,a+s} + \delta_{x,a+s}\log(Y_{t+s}) + \delta_{\tau,a+s}\log(\overline{\varphi}_k)h_{t+s} + e_{t+s} + \eta_{t+s}\right) + \delta_\theta^5 S_{1,t} - \delta_\theta^5\zeta_{1,t}\right) + \zeta_{2,t+4} \quad (38)$$

$$\Lambda = \begin{bmatrix} \lambda_{AP} & 0 \\ 0 & \lambda_{BN} \end{bmatrix} \quad (39)$$

Collecting terms yields:

$$S_{1,t+5} = \tilde{\mu}(k,a,\boldsymbol{a}) + g(\boldsymbol{\delta},\boldsymbol{x}_t) + \delta_\theta^5 S_{1,t} + \xi_{1,t} \quad (40)$$

$$S_{2,t+5} = \Lambda^{-1}\left(\tilde{\mu}(k,a,\boldsymbol{a}) + g(\boldsymbol{\delta},\boldsymbol{x}_t) + \delta_\theta^5 S_{1,t}\right) + \xi_{2,t} \quad (41)$$

where $\boldsymbol{x}_t = \{Y_{t+s}, h_{t+s}\}_{s=0}^4$ collects all inputs and the control function

$$\tilde{\mu}(k,a,\boldsymbol{a}) = \sum_{s=0}^{4} \delta_\theta^{4-s}\left(\mu_{\theta,k,a+s} + e_{t+s}\right)$$

absorbs the effect of differential productivities by type ($\mu_{\theta,k}$) as well as differences in the marginal propensities to invest (e_t) that are a function of the age of the child (a) and the age of other children in the household ($\boldsymbol{a}$). The error terms ($\xi_{1,t}, \xi_{2,t}$) collect all of the error terms in the above system:

$$\xi_{1,t} = \sum_{s=0}^{4} \delta_\theta^{4-s} \eta_{t+s} + \zeta_{1,t+4} - \delta_\theta^5 \zeta_{1,t}$$

$$\xi_{2,t} = \Lambda^{-1} \left(\sum_{s=0}^{4} \delta_\theta^{4-s} \eta_{t+s} - \delta_\theta^5 \zeta_{1,t} \right) + \zeta_{2,t+4}$$

Notice that according to the model's assumptions $\mathbb{E}[\xi_{j,t}|\boldsymbol{x}] = 0$ for $j \in \{1,2\}$. Hence, subject to dealing with measurement error in lagged skills, this condition can form the basis of an estimator by directly using the variation in inputs $\boldsymbol{x}$ after controlling for type and household demographics. The remainder of this section spells out a control function estimation routine for the weaker pair of conditions:

$$\mathbb{E}[\xi_{j,t}|\mathbf{Z}] = 0, \qquad \mathbb{E}[\xi_{j,t}|S_{2,t}] = 0, \qquad j \in \{1,2\}.$$

In order to derive the quasi-likelihood for estimation, let us decompose the test score $S_{1,t}$ and each input in $\boldsymbol{x}$ into orthogonal components:

$$\log(Y_t) = \gamma_{1,1} + \gamma_{1,2}\mathbb{E}[\log(Y)t|\mathbf{Z}] + \nu_{Y,t} \tag{42}$$

$$h_t = \gamma_{2,1} + \gamma_{2,2}\mathbb{E}[h|\mathbf{Z}] + \nu_{h,t} \tag{43}$$

$$LW_t = \gamma_{3,1} + \gamma_{3,2} AP_t + \nu_{LW,t} \tag{44}$$

$$BE_t = \gamma_{4,1} + \gamma_{4,2} BN_t + \nu_{BE,t} \tag{45}$$

Now augment the outcome equations with the vector of control variables $\nu_t = [\nu_{Y,t},\ \nu_{h,t},\ \nu_{LW,t},\ \nu_{BE,t}]'$:

$$S_{1,t+5} = \tilde{\mu}(k, a, \boldsymbol{a}) + g(\boldsymbol{\delta}, \boldsymbol{x}_t) + \rho'\nu + \epsilon_{1,t} \tag{46}$$

$$S_{2,t+5} = \Lambda^{-1}\left(\tilde{\mu}(k, a, \boldsymbol{a}) + g(\boldsymbol{\delta}, \boldsymbol{x}_t) + \rho'\nu\right) + \epsilon_{2,t}. \tag{47}$$

The system of equations (42)-(47) form the basis of a control function estimation approach, where specifying that the vectors of errors (ν, ϵ) are normally distributed yields a pseudo-likelihood.[35] To help with precision, estimation imposes some priors. This is especially useful because the vector for production parameters $\boldsymbol{\delta}$ includes a vector of type-specific relative quality coefficients ($\{\overline{\varphi}_k\}_{k=1}^{K}$). Table 8 reports the priors that the MCMC procedure imposes while sampling. The final sample that generates estimates and credibility intervals throughout the text is a chain of 10,000 draws from a No-U-Turn Hamiltonian Monte-Carlo chain (Hoffman et al., 2014) implemented in `julia` using `Turing.jl` (Ge et al., 2018).

E.6 Net Present Value of Skills

The proportionality of skill returns on earnings at each age implies the following model for the net present value of earnings:

$$\mathcal{Y}(\theta) = (1 + \tilde{\gamma}_c \log(\theta_c) + \tilde{\gamma}_b \log(\theta_b)) \times \sum_{a=18}^{65} \beta^{a-18} \gamma_{Y,a}$$

To forecast the age effects $\gamma_{Y,a}$ this paper uses the 1997-2018 waves of the Annual Social and Economic Supplement of the Current Population Survey (Flood et al., 2021) to estimate a model with cohort effects and a quadratic age trend:

$$Y_i = \kappa_c + \tilde{\gamma}_0 a_i + \tilde{\gamma}_1 a_i^2 + \epsilon_i$$

where κ_c are cohort effects for birth cohorts in the set: $\{\leq 1950, 1951 - 1960, 1961 - 1970, 1971 - 1980, \geq 1985\}$. This model provides a forecast of average earnings (conditional on positive earnings) for the 1985-1997 cohort. Figure 24 shows estimates of this projection as well as the age profile of earnings for the different cohorts in the sample.

[35] To implement the "model-based" IV approach, one can simply drop the control variables (nu_Y, ν_H) from the outcome equation.

Figure 24: Lifecycle Earnings Estimates by Cohort

This figure shows average earnings by age for each of the 5 cohorts from the 1997-2018 waves of ASEC. It also shows the resulting forecast of average earnings for the cohort born after 1985 when using estimates from the quadratic model described in the text.

Table 8: Table of prior distributions used in estimation

σ_ϵ	Std. deviation of errors in outcome equation	$U[0,10]$
λ	Factor loadings for AP and BN in measurement equation	$U[0,2]$
γ	Coefficients in (42)-(45)	$U[-10,10]$
ρ	Coefficients on control variables	$U[-2,2]$
σ_ν	Std deviation of control variables ν	$U[0,10]$
μ	Coefficients on type and age of child in outcome equation	Improper (diffuse) prior on $\mathbb{R}$
δ_x	Factor share of money in outcome equation	$U[0,0.5]$
δ_τ	Factor share of time in outcome equation	$U[0,2]$
δ_θ	Factor share of skills in outcome equation	$U[0,1]$
$\log(\varphi)$	Log of relative care quality in outcome equation	$U[-2,2]$

The earnings model also identifies the $\gamma_{Y,a}$ up to a constant of proportionality when using the 1985-1997 cohort effect:

$$\hat{\gamma}_{Y,a} = \frac{\mu_{CDS}}{\mu_{USA}}(\kappa_{\geq 1980} + \tilde{\gamma}_0 a + \tilde{\gamma}_1 a^2)$$

where the the constant of proportionality μ_{CDS}/μ_{USA} reflects the fact that average skills in the selected CDS sample is different from the representative ASEC sample. A comparison of mean earnings (conditional on working) between the ages of 23 and 27 (for which observations of the CDS sample are available) provides an estimate of this ratio. Table 9 shows the results and demonstrates the resulting calculation of the ratio μ_{CDS}/μ_{USA} used to calculate the net present value of earnings (conditional on working) for CDS sample children.

This approach delivers an estimated net present value of lifetime earnings (conditional on working) for the ASEC sample of \$996,686 in year 2000 dollars, using a chosen value of β of 0.98. [36] Based on observed positive earnings for the CDS estimation sample, a value of just under three quarters this number (\$731,700) is chosen for anchoring the earnings returns to skills. This number is scaled down by the observed probability of earning a positive amount in the CDS sample.

With this number in hand we can estimate the effect of cognitive and behavioral skills on lifetime earnings by estimating their proportional effect on earnings in early adulthood. Using the most recent available measure of skills for each child, I log annual earnings on the Letter-Word and Externalizing Behaviors scores, using the Applied Problems and Internalizing Behaviors as instruments to account for measurement error. Table 10 shows the results using all available observations of CDS children (column 1), as well as using only children from the chosen sample of single mothers (columns 2 & 3). Coefficients on cognitive and non-cognitive factors can be interpreted as the effect of a one standard deviation increase in that factor. Table 10 also reports the implied percentage impacts of a standard deviation increase in each skill, which can be interpreted as the pair $(\tilde{\gamma}_c, \tilde{\gamma}_b)$ and can be combined with estimates of the net present value of positive earnings and the probability of positive earnings to get the net present value of one standard deviation in each skill reported in Table 10. The quantitative exercises in the paper use a more conservative estimate by subtracting one estimated standard deviation from the log-returns in earnings to each skill.

[36] In Chetty et al. (2011) the authors use a discount rate of 3%, finding a net present value of \$522,000 for unconditional earnings in 2010 dollars when discounted to age 12. Making an equivalent calculation using my method delivers a comparable present value of \$599,000.

Table 9: Calculation of NPV of Lifecycle Earnings for CDS Sample

	Sample	
Age	PSID-CDS	CPS (1985-)
23	13502	16058
24	15138	18603
25	16408	22998
26	17727	24173
27	16295	25873
Mean (23-27)	15814	21541
$\widehat{\mu_{CDS}/\mu_{USA}} \approx 0.73$		
$\widehat{NPV}_{USA} = \$996,686$		
$\widehat{NPV}_{CDS} = \$731,705$		

This table reports mean earnings (conditional on positive earnings) for individuals aged 23-27 from the selected PSID-CDS sample used in this paper and from the 1997-2018 waves of the ASEC. Calculations for the latter case include only individuals born 1985 or after. The lower panel shows calculations of the net present value of earnings for the CDS sample by taking the estimated NPV from the ASEC samples and multiplying by the ratio of average earnings in the two samples (see text for more details).

F Optimal Policy

F.1 Deriving the Planner's Objective

The planner's objective is

$$\boldsymbol{W}_y = \int (\boldsymbol{\mu} V_t(s,\theta) + \boldsymbol{\lambda} R_t(s,\theta)) d\pi_y(s).$$

Let $\boldsymbol{W}_t(s,\theta)$ be the contribution to the planner's objective of a household in state s with child skills θ in period t. Letting $E_j(s) = H_j W(s)$ be the earnings derived from choice j, $\boldsymbol{W}$ is defined as:

$$\begin{aligned}\boldsymbol{W}_t(s,\theta) &= \mu V_t(S,\theta) + \boldsymbol{\lambda}\mathbb{E}\left[\sum_{r=t}^{T}\beta^{r-t}(E_j(S_r) - \boldsymbol{Y}_j(S_r)) + \sum_{f:a(f,t)\leq 18}(\tilde{\Gamma}_{18,c}\log(\theta_{f,c,t_f^*}) + \tilde{\Gamma}_{18,b}\log(\theta_{f,b,t_f^*}))\big|\theta, S\right] \\ &= \mu V_t(S,\theta) + \boldsymbol{\lambda} R_t(\theta, S)\end{aligned}$$

where $t_f^* = b_f + 18$ is the period in which child f matures. This is the period in which the net present value of their skills to the planner are realized and accounted for in their objective. The function $R_t(s,\theta)$ is the expected net present value of short and long-run contributions to the planner's resource constraint, where expectations are formed conditional on the current state s and current child skills θ.

The remainder of this section will demonstrate that this function is log-additive in child skills, which simplifies the solution to the planner's problem. This allows for analytical expressions when considering static policies in Section F.2.

The logic of the derivation is almost identical to the derivation of the value function in Appendix C. To begin, observe that

$$R_T(s,\theta) = \mathbb{E}[E_j(s) - \boldsymbol{Y}_j(s)] + \sum_{f:a(f,T)=18}(\tilde{\Gamma}_{18,c}\log(\theta_{f,c}) + \tilde{\Gamma}_{18,b}\log(\theta_{f,b}))$$

Table 10: Effect of Cognitive and Behavioral Skills on Earnings

	(1)	(2)	(3)
Cognitive	0.3924***	0.3200***	0.3364***
	(0.0591)	(0.1103)	(0.1148)
Behavioral	-0.1247***	-0.1821***	-0.1862***
	(0.0466)	(0.0643)	(0.0642)
Age			0.0898**
			(0.0382)
Mother's Ed			-0.0051
			(0.0254)
Observations	1,028	397	397
	Probability of Positive Earnings		
	0.56	0.57	0.57
	NPV of 1 s.d. ($1000 USD year 2000)		
Cognitive	-	134.14	-
Behavioral	-	76.34	-
	Conservative NPV of 1 s.d.		
Cognitive	-	87.9	-
Behavioral	-	49.3	-

*p<0.1; **p<0.05; ***p<0.01

This table reports estimates of the relationship between earnings in young adulthood for CDS children and measures of cognitive and behavioral skills. Estimates are calculated by two stage least squares using the Letter-Word score and Externalizing Behaviors in the outcome equation and Applied Problems and Internalizing Behaviors as instruments. Specification (1) shows results for all non-missing observations of CDS children, while Specifications (2) and (3) use only the chosen sample for the paper. The lower panel shows net present value calculations for the value of each skill by taking implied percentage effects on earnings, multiplying by the estimated net present value of positive earnings in thousands of year 2000 US dollars, and multiplying by the sample fraction of observations with positive earnings.

Given this additive structure in the terminal period, it is sufficient to show that log-additivity is preserved by backward induction. Suppose that:

$$R_{t+1}(s,\theta) = r_{t+1}(s) + \sum_{f:a(f,t)+1\leq 18} (\tilde{\Gamma}_{c,a(f,t)+1}\log(\theta_{f,c}) + \tilde{\Gamma}_{b,a(f,t)+1}\log(\theta_{f,b}))$$

Since s is sufficient to forecast the stochastic components of s' in the next period, Rt can then be written recursively as:

$$R_t(s,\theta) = \mathbb{E}\left[E_j(s) - \mathbf{Y}_j(s) + \beta\mathbb{E}\left[r_{t+1}(s')|s,j\right] + \beta\mathbb{E}\left[\sum_{f\in\mathcal{A}(t)} \tilde{\Gamma}_{a(f,t)+1,c}\log(\theta'_{f,c}) + \tilde{\Gamma}_{a(f,t)+1,b}\log(\theta'_{f,b})\Big|s,\theta,j\right]\Big|s,\theta\right]$$

Next, using the optimal input formula derived in Appendix C, expectations of future skills can be written as:

$$\mathbb{E}[\log(\theta'_{f,i})|S,\theta,j] = A_t + \delta_{x,a,i}\log(\boldsymbol{Y}_j(S)) + \delta_{\tau,a,i}h_j\log(\overline{\varphi}_k) + \delta_{\theta,c,i}\log(\theta_{f,c}) + \delta_{\theta,b,i}\log(\theta_{f,b})$$

where A_t is a constant that depends on the mother's unobserved type, preferences, technology parameters, and the ages of all children in the household, but does *not* depend on any choices or policies. For simplicity then, disregard this term for the remainder of the analysis.

Substituting expected future skills into the recursive formula above yields the following set of recursive coefficients:

$$\tilde{\Gamma}_{a,i} = \beta\left(\delta_{\theta,i,c}\tilde{\Gamma}_{a+1,c} + \delta_{\theta,i,b}\tilde{\Gamma}_{a+1,b}\right), \qquad i\in\{c,b\} \tag{48}$$

$$\tilde{\Gamma}_x(s) = \beta\sum_{a\in\boldsymbol{a}}\left(\delta_{x,a,c}\tilde{\Gamma}_{a+1,c} + \delta_{x,a,b}\tilde{\Gamma}_{a+1,b}\right) \tag{49}$$

$$\tilde{\Gamma}_\tau(s) = \beta\sum_{a\in\boldsymbol{a}}\left(\delta_{\tau,a,c}\tilde{\Gamma}_{a+1,c} + \delta_{\tau,a,b}\tilde{\Gamma}_{a+1,b}\right) \tag{50}$$

which simplifes the formulation of $r_t(s)$ to:

$$r_t(s) = \mathbb{E}\left[E_j(s) - \mathbf{Y}_j(s) + \tilde{\Gamma}_x(s)\log(\mathbf{Y}_j(s)) + \tilde{\Gamma}_\tau(s)h_j\log(\overline{\varphi}_k) + \beta\mathbb{E}[r_{t+1}(s')|s,j]\Big|s\right]$$

For optimal policies, it is useful to write the outer expectation explicitly using choice probabilities, $P_j(s)$:

$$r_t(s) = \sum_j P_j(s)\left[E_j(S) - \mathbf{Y}_j(s) + \tilde{\Gamma}_x(s)\log(\mathbf{Y}_j(s)) + \tilde{\Gamma}_\tau(s)h_j\log(\overline{\varphi}_k) + \beta\mathbb{E}[r_{t+1}(s')|s,j]\right]$$

F.2 Optimal Transfers in a Static Setting

F.2.1 Choice Probabilities

Recall that $P_j(s)$ denotes the probability of choice j in state s, under any transfer policy $\mathbf{Y}$. In the static setting, these depend on the vector of indirect utilities, $\{u_{j,k}(s)\}_{j=0}^{J-1}$. In turn, choice probabilities are invariant to a location normalization $\Delta_{j,t} = u_{j,k}(s) - u_{0,k}(s)$. Now, the change in choice probability i with respect to a marginal change in income from choice j, $\mathbf{Y}_j(s)$, is given by:

$$\frac{dP_i(s)}{d\mathbf{Y}_j(s)} = \frac{\Gamma_{x,t}(s)}{\mathbf{Y}_j(s)}\frac{dP_i(s)}{d\Delta_j(s)}$$

where the first term is the marginal indrect utility of net income. Furthermore, given the constraint that $\sum_j P_j(s) = 1$:

$$\frac{dP_0(s)}{d\mathbf{Y}_j(s)} = -\sum_{i>0}\frac{dP_i(s)}{d\mathbf{Y}_j(s)}. \tag{51}$$

Define $\eta_{ij}(s)$:

$$\eta_{ij}(s) = \frac{1}{P_j(s)}\frac{dP_i(s)}{d\Delta_j(s)}$$

as the semi-elasticity of the probability of choice i with respect to the normalized utility of choice j.

When the policy of interest is a simple allocation function $\boldsymbol{y}$, the set of discrete choices simplifies to the four hours choices with the bottom two layers of the nested logit. Let $P_0(s)$ be the probability of not working ($H_0 = 0$). Using the structure of the nested logit gives:

$$\eta_{ij}(s) = \underbrace{\sigma_{H,0}^{-1} \frac{P_0(s)}{1 - P_0(s)} P_i(s)}_{\varepsilon_{kj}(s)} + \sigma_{H,1}^{-1} \frac{P_i(s)}{P_j(s)} \left(\mathbf{1}\{j = k\} - \frac{P_j(s)}{(1 - P_0(s))} \right)$$

where the first term, $\varepsilon_{ij}(s)$, is the change in the probability of working positive hours given a change in the utility from choice j, and the second component is the change in the probability of making choice i conditional on working positive hours. Thus, $\eta_{ij}(s)$ decomposes into an extensive and intensive marginal response, where the dispersion parameters $\sigma_{H,0}$ and $\sigma_{H,1}$ dictate responsiveness at the extensinve and intensive margins, respectively.

F.2.2 Optimal Policy for unrestricted functions of earnings

To simplify expressions, let $\mathcal{R}_j(s)$ collect all contributions to the NPV resource constraint *relative to choice* 0:

$$\mathcal{R}_j(s) = E_j(s) - \boldsymbol{y}(E_j(s)) + \boldsymbol{y}(0) + \mathcal{D}_j(s)$$

where $\mathcal{D}_j(s)$ is defined in the main text as the NPV contribution of child skills from choice j relative to choice 0:

$$\mathcal{D}_j(s) = \tilde{\Gamma}_x(s) \log \left(\frac{\boldsymbol{y}(e)}{\boldsymbol{y}(0)} \right) + \tilde{\Gamma}_\tau(s) \log(\overline{\varphi}_k).$$

Consider a marginal increase in the allocation $\boldsymbol{y}(e)$ at earnings level $e > 0$. Applying the envelope theorem for discrete choice models (McFadden, 1981), and substituting the constraint in (51) yields:

$$\int \sum_{j>0} \mathbf{1}\{E_j(s) = e\} P_j(s) \left[\frac{\boldsymbol{\mu}(1 + \alpha_{\theta,k}\Gamma_x(s))}{\boldsymbol{y}(e)} - \boldsymbol{\lambda} + \boldsymbol{\lambda} \frac{\tilde{\Gamma}_x(s)}{\boldsymbol{y}(e)} + \frac{\boldsymbol{\lambda}}{\boldsymbol{y}(e)} \sum_{i>0} \eta_{ij}(s) \mathcal{R}_i(s) \right] d\pi(s) = 0 \tag{52}$$

where $\eta_{ij}(s) = P_j(s)^{-1} \partial P_i(s) / \partial \Delta_j$ is the semi-elasticity of the probability of choice i with respect to the normalized utility of choice j. Similarly, a marginal change in the allocation $\boldsymbol{y}(0)$ at $e = 0$ must satisfy:

$$\int P_0(s) \left[\frac{\boldsymbol{\mu}(1 + \alpha_{\theta,k}\Gamma_x(s))}{\boldsymbol{y}(0)} - \boldsymbol{\lambda} + \boldsymbol{\lambda} \frac{\tilde{\Gamma}_x(s)}{\boldsymbol{y}(0)} - \frac{\boldsymbol{\lambda}}{\boldsymbol{y}(0)} \sum_{j>0} \frac{P_j(s)}{P_0(s)} \sum_{i>0} \eta_{ij}(s) \mathcal{R}_i(s) \right] d\pi(s) = 0 \tag{53}$$

Multiplying (52) through by $\boldsymbol{y}(e)/\boldsymbol{\lambda}$ and integrating over all potential earnings levels gives:

$$\int \sum_{j>0} P_j(s) \left[\frac{\boldsymbol{\mu}(1 + \alpha_{\theta,k}\Gamma_x(s))}{\boldsymbol{\lambda}} + \tilde{\Gamma}_x(s) - \boldsymbol{y}(E_j(s)) + \sum_i \eta_{ij}(s) \mathcal{R}_i(s) \right] d\pi(s) = 0 \tag{54}$$

which when combined with (53) (multiplied through by $\boldsymbol{y}(0)/\boldsymbol{\lambda}$) gives an expression for optimal generosity:

$$\mathbb{E}[\boldsymbol{y}(E_j(s))] = \mathbb{E} \left[\frac{\boldsymbol{\mu}(1 + \alpha_{\theta,k}\Gamma_x(s))}{\boldsymbol{\lambda}} + \tilde{\Gamma}_x(s) \right] \tag{55}$$

where the expectation is taken over states and choices. Define $\boldsymbol{w}(s) = \boldsymbol{\mu}(1 + \alpha_{\theta,k}\Gamma_x(s))/\boldsymbol{\lambda} + \tilde{\Gamma}_x(s)$ as the first best allocation: the allocation the planner would give to a household in state (t, s) in the absence of any behavioral responses.

Now looking to equation (52), define the conditional measure over states (s, t) and choices (j) as:

$$\pi_t(s, j|e) \propto \mathbf{1}\{E_j(s) = e\} P_j(s) \pi(s)$$

such that (52) can be rewritten as:

$$\boldsymbol{y}(e) = \mathbb{E}[\boldsymbol{w}(s)|e] + \mathbb{E} \left[\sum_i \eta_{ij}(s) \mathcal{R}_i(s) \middle| e \right]$$

where the conditional expectation is taken using the measure $\pi_t(s, j|e)$. Unpacking the term $\mathcal{R}_i(s)$, gives the expression in the main text:

$$\boldsymbol{y}(e) = \mathbb{E}[\boldsymbol{w}(s)|e] + \mathbb{E} \left[\sum_i \eta_{ij}(s) \left(e - \boldsymbol{y}(e) + \boldsymbol{y}(0) + \mathcal{D}_i(s) \right) \middle| e \right]. \tag{56}$$

F.2.3 Optimal Policy for parametric functions of earnings

Now consider the case where the transfer function $\boldsymbol{y}_\gamma(e)$ is parametric and indexed by a finite-dimensional vector, γ. Note that the first order condition with respect to any member of γ is simply a linear combination of the first order conditions derived above, weighted by $d\boldsymbol{y}_\gamma(e)/d\gamma$. Thus, the first order condition with respect to γ is given by:

$$\int \sum_{j>0} \frac{\partial \boldsymbol{y}_\gamma(E_j(s))}{\partial\gamma} P_j(s) \left[\frac{\boldsymbol{\mu}(1+\alpha_{\theta,k}\Gamma_x(s))}{\boldsymbol{y}_\gamma(E_j(s))} - \boldsymbol{\lambda} + \boldsymbol{\lambda}\frac{\tilde{\Gamma}_x(s)}{\boldsymbol{y}_\gamma(E_j(s))} + \frac{\boldsymbol{\lambda}}{\boldsymbol{y}_\gamma(E_j(s))}\sum_{i>0}\eta_{ij}(s)\mathcal{R}_i(s)\right] d\pi(s) +$$
$$\int \frac{\partial \boldsymbol{y}_\gamma(0)}{\partial\gamma} P_0(s) \left[\frac{\boldsymbol{\mu}(1+\alpha_{\theta,k}\Gamma_x(s))}{\boldsymbol{y}_\gamma(0)} - \boldsymbol{\lambda} + \boldsymbol{\lambda}\frac{\tilde{\Gamma}_x(s)}{\boldsymbol{y}_\gamma(0)} - \frac{\boldsymbol{\lambda}}{\boldsymbol{y}_\gamma(0)}\sum_{j>0}\frac{P_j(s)}{P_0(s)}\sum_{i>0}\eta_{ij}(s)\mathcal{R}_i(s)\right] d\pi(s) = 0 \quad (57)$$

The formula (55) for optimal generosity will hold quite generally for parametric allocation functions, as long as the function belongs to a class that can be scaled up proportionally. To see this, substitute $\partial\boldsymbol{y}_\gamma(e)/\partial\gamma = \boldsymbol{y}_\gamma(e)$ for all e in the equation above, and one immediately arrives back at (56).

In general, the formula for (56) for optimal shape does not hold for parametric functions. It does hold however for step-wise functions. Let $\boldsymbol{y}_\gamma(e)$ be parameterized as:

$$\boldsymbol{y}_\gamma(e) = \mathbf{1}\{e=0\}y_0 + \sum_{g>0}\mathbf{1}\{e \in (e_{g-1}, e_g]\}y_g.$$

The parameters γ are the allocations $\{y_g\}$ and the grid points $\{e_g\}$ with $e_0 = 0$. Notice that $\partial\boldsymbol{y}(e)/\partial y_g = \mathbf{1}\{e \in (e_{g-1}, e_g]\}$. Making this substitution into the first order condition above results in an analog to equation (56):

$$y_g = \mathbb{E}[\boldsymbol{w}(s)|E_g(s) \in (e_{g-1}, e_g]] + \mathbb{E}\left[\sum_i \eta_{ij}(s)\left(E_j(s) - y_g + y_0 + \mathcal{D}_i(s)\right)\Big| E_g(s) \in (e_{g-1}, e_g]\right]. \quad (58)$$

The decompositions of the optimal step-wise function in Section 6 use this formula with a uniform step-size of \$300/month.

G Transfers

This section describes the computation of the transfer functions (T^S, 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:

$$T^A_{mt} = \mathrm{El}^A_{mt} \times \mathrm{Ben}^A_{mt} \quad (59)$$

Where

$$\mathrm{El}^A_{mt} = \mathbf{1}\{\omega_{mt} \leq \Omega_{mt}\} \times \mathbf{1}\left\{E_{mt} + N_{mt} < r_{A,g,mt}e_{A,mt}\right\} \times \mathbf{1}\left\{(E_{mt} - D_{Ae,mt})(1 - R_{Ae,mt}) + N_{mt} < r_{An,mt}e_{A,mt}\right\}. \quad (60)$$

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, $e_{A,mt}$ which is inflated by some rate $(r_{Ag,mt}, r_{Ae,mt})$. Second, the computation of net income involves a fixed disregard on earnings, $D_{Ae,mt}$ and a percentage disregard. Benefit computation follows similarly:

$$\mathrm{Ben}^A_{mt} = \max\{G_{A,mt} - (E_{mt} - D_{Ab,mt})(1 - R_{Ab,mt}) - N_{mt}, 0\} \quad (61)$$

The payment standard $G_{A,mt}$ sets the generosity of the program when no other sources of income are reported, while the dollar and percentage disregards $(D_{Ab,mt}, R_{Ab,mt})$ 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, mt. In reality, individuals may be subject to asset tests for eligibility, which are not modelled here. Time limits in some states are also *periodic* in the sense that in addition to the total limit, there are shorter limits on the number of months of consecutive use. These periodic time limits are also not modelled due to tractability concerns.

Food Stamps

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

$$T^A_{mt} = \text{El}^S mt \times \text{Ben}^S_{mt} \tag{62}$$

Where

$$\text{El}^S_{mt} = \mathbf{1}\left\{E_{mt} + N_{mt} < 1.3e_{S,mt}\right\} \times \mathbf{1}\{\underbrace{0.8E_{mt} + N_{mt} + \text{Ben}^A_{mt} - 134}_{=\text{Net}^S_{mt}} < e_{S,mt}\}. \tag{63}$$

In the above expression, $e_{S,mt}$ 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 $G_{S,mt}$, the benefit calculation is:

$$\text{Ben}^F_{mt} = \max\left\{G_{F,mt} - 0.3\text{Net}^F_{mt}, 0\right\}. \tag{64}$$

Data Sources for Program Rules

To summarize, the parameter vector Z^A_{mt} can be written as

$$Z_{mt} = \{r_{Ag,mt}, r_{Ae,mt}, e_{A,mt}, D_{Ae,mt}, R_{Ae,mt}, G_{A,mt}, D_{Ab,mt}, R_{Ab,mt}, \mathcal{L}_{mt}\},$$

while Z^S_{mt} can be summarized as

$$Z^S_{mt} = \{e_{S,mt}, G_{S,mt}\}.$$

Parameters on welfare that comprise Z^A_{mt} and Z^S_{mt} were collected from the Urban Institute's TRIM3 simulation database[37] 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 [38]. 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, 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, mt), TAXSIM computes $T^T_{mt}(e)$ for any given earnings level. Thus, for each mt in my sample, I compute $T^T_{mt}(e)$ for earnings levels e on a grid, using increments of \$100, between \$0 and \$100,000[39]. Using this grid, the tax function is approximated using linear interpolation between these grid points.

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

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

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