A structural meta-analysis of welfare reform experiments and their impacts on children

Joseph Mullins *

May 15, 2020

Abstract

Using a model of maternal labor supply and investment in children, this paper synthesizes the findings from five separate welfare reform experiments across twelve sites. The proposed model maps variation in experimental design to parameters that define labor supply behavior, child care use, and the importance of time and money in the development of child skills. The estimation procedure amounts to a *structural meta-analysis* in which the model's parameters are estimated by matching predicted control and treatment group means to those found in publicly available evaluation reports. Thus, experimental impacts are aggregated to identify the model's key causal parameters, permitting sharply defined counterfactuals of policy interest. Estimates suggest that while family resources play a quantitatively significant role in shaping both academic and behavioral outcomes, the predicted effects do not exhibit persistence over time, nor do the available data support prior evidence that non-maternal care has negative impacts. Estimates also suggest that time limits and work requirements had a relatively small permanent impact on the labor force participation of welfare recipients.

*Dept. of Economics, University of Minnesota. Email: mullinsj@umn.edu The author wishes to thank Roy Allen, Mark Colas, Chris Flinn, and Jim Heckman for useful comments, as well as seminar participants at Washington University St Louis and UNC Chapel Hill. Excellent research assistance has been provided by Filip Babalievsky, Alexander Djanin and Nikhil Saxena.

1 Introduction

This paper uses an economic model of behavior to aggregate statistical information from multiple experiments. It follows the mission statement laid out by Frisch (1933) in the very first issue of *Econometrica*: to coordinate the accumulation of empirical evidence for policy and prediction using a theoretical framework[1]. While traditional meta-analyses approach this problem by specifying the average treatment effect (ATE) as a global parameter of interest —using linear statistical models to construct a weighted average estimate of this parameter —this paper alternatively suggests using variation in experimental results to identify underlying parameters of an economic model.

There are three main payoffs to taking a structural approach to meta-analysis. First, if the experimental setting is well-suited to the application of economic theory, then a model can provide an organized comparative interpretation of treatment effects across experiments. Second, a well-specified model can map structural primitives to a broad class of counterfactual scenarios, and can therefore be used in cases where the ATE itself does not define an explicit policy counterfactual of interest (Heckman, 1992; Heckman and Vytlacil, 2005). This mapping serves as an input either directly into policy, or into the cost-effective design of future experiments. Finally, an economic model allows researchers to make normative statements about how treatments (both observed and counterfactual) are valued by participants, by applying the lens of revealed preference.

This paper uses the method to study the design of cash assistance and its effect on child outcomes, using results from five experimental evaluations of welfare-to-work programs in the United States. These experiments, conducted between the years 1991 and 1998 by the Manpower Demonstration Research Corporation (MDRC) restuctured the conditions of cash assistance for participants with the goal of increasing labor force attachment and financial independence. Importantly, in these specific sites, MDRC also collected a set of developmental indicators for children, resulting in a set of control and treatment group means on earnings, welfare participation, and child outcomes for each site. To aggregate statistical information from these experiments, I develop a model of labor supply, welfare participation, childcare use, and investment in children. Variation in the design of treatment programs across sites is leveraged to identify the key causal parameters of the model, which determine the response of mothers' behavior and child development outcomes to differently designed welfare programs.

To the extent that changes to the policy environment shape two household resources, time and money, a growing literature has argued that these policies may have consequences for child development (Bernal, 2008; Dahl and Lochner, 2012; Mullins, 2019; Agostinelli and Sorrenti, 2018). The

[1] In his editorial introduction of the journal, Frisch (1933) wrote: "Statistical information is currently accumulating at an unprecedented rate. But no amount of statistical information, however complete and exact, can by itself explain economic phenomena. If we are not to get lost in the overwhelming, bewildering mass of statistical data that are now becoming available, we need the guidance and help of a powerful theoretical framework. Without this no significant interpretation and coordination of our observations will be possible."

impacts of these experiments provide an opportunity to learn about the effect of time and money on children's skill formation. Using a model to explore these impacts is useful, because technology parameters are not practically separable from behavioral responses, as can be illustrated by the following example. Consider the *ceterus paribus* effect of an additional \$1000 in household income, the preferred hypothetical of Dahl and Lochner (2012) and Duncan et al. (2011) to analyze the importance of income in skill formation. This is not the same as simply *giving* families an additional \$1000, since we should expect the transfer to have some effect on labor supply, shifting other resources within the household. In other words, the counterfactual used to define the technology parameter does not map to a plausible policy counterfactual, and vice versa. The model, when fit to observed control and treatment group means, simulates this counterfactual to find that a \$1,000 annual transfer to households leads to a 2% of a standard deviation increase in both academic and behavioral outcomes[2].

The estimated model represents the aggregated information from each experiment, and can be used for a rich set of causal and counterfactual statements, of which this paper only presents an illustrative subset. With respect to child outcomes, one surprising finding is that treatment effects do not support the hypothesis that non-maternal care has negative developmental impacts, in contrast to the findings of Bernal (2008); Bernal and Keane (2010), Agostinelli and Sorrenti (2018), and Mullins (2019). Other counterfactuals speak to long-standing questions about the labor force and welfare participation of mothers through this time period (Grogger and Michalopoulos, 2003; Grogger, 2002, 2003; Hoynes, 1996; Meyer, 2002; Chan, 2013). Counterfactual scenarios in which I introduce time limits and work requirements demonstrate a limited impact of time limits on labor force participation, and that the positive effect of work requirements is not persistent. Participation elasticities with respect to wages, on the other hand, are robust, heterogeneous, and positive, ranging across sites between 0.2 and 0.8. This suggests that increases in wages and financial incentives through taxes and welfare were a major source of historical growth in mothers' labor force participation.

Prior attempts to synthesize the findings of these experiments use more traditional techniques, and have focused on distinguishing between sites that provided changes in financial work incentives and those that relied only on work requirements (Clark-Kauffman et al., 2002; Gennetian et al., 2004). In one notable exception, Duncan et al. (2011) combine the data from these and other experiments[3], using treatment assignment as an instrumental variable to estimate the effect of income on child achievement. The method proposed here is complementary to these approaches, while being quite distinct in its goal: to identify parameters of a model in which parental behavior

[2]While smaller than other estimates in the literature, this number incorporates behavioral responses, which other findings typically do not

[3]The authors included data from the New Hope experiment, which included a health insurance program component, and the Canadian Self-Sufficiency Program, which included access to high quality child care.

and child outcomes are fully articulated with respect to a clearly defined welfare policy. This resolves the tension between merely illustrative and implementable counterfactuals.

The remainder of the paper is organized as follows. Section 1.1 below clarifies the method and its relationship to traditional meta-analysis. Section 2 describes the design of the experiments in question and presents preliminary observations from treatment effects, which motivate specific modelling choices. Section 3 introduces the model and shows the mapping between particular treatment components and parameters of interest. Section 4 discusses estimation of the model, while Section 5 explores the implications of these estimates using simple counterfactuals, thereby aggregating experimental results into specific lessons for policy.

1.1 Methodology

To fix methodological ideas, consider a vector $X_{k,N}$ of asymptotically normal statistics from experiment k, and let $\mathbf{X}_N$ be the stacked vector of statistics across K sites. Let the parameter vector γ index a class of models that generate a model equivalent $\mathbf{X}(\gamma)$. The suggested estimator $\hat{\gamma}$ is immediate:

$$\hat{\gamma} = \arg\min_{\gamma}(\boldsymbol{X}_N - \mathbf{X}(\gamma))'\mathbf{W}(\mathbf{X}_N - \mathbf{X}(\gamma) \tag{1.1}$$

with convenient asymptotic properties under the usual conditions (Ferguson, 1958)[4]. When each $X_{k,N}$ is comprised solely of estimates of treatment effects, and the model $\mathbf{X}(\gamma)$ is the single parameter defining the ATE, this framework nests the so-called "Fixed Effects" approach to meta-analysis, in which the estimator $\hat{\gamma}$ is simply a weighted average of site-level treatment effects (Higgins and Green, 2011). Alternatively, the mapping $\mathbf{X}(\gamma)$ may be given by an economic model in which the vector γ summarizes structural primitives of interest. Since the ambition is still to aggregate information from multiple experiments, this paper intoduces this latter class of estimator as a *structural meta-analysis*. Taking this approach provides one manner of response to address the critiques of randomization as a tool for policy evaluation articulated by Heckman (1992). In particular, the structural approach provides an explicit framework with which to handle the traditional bugbears of experiment design: non-compliance and substitution bias, two phenomena containing inherent information by revealed preference. When applied to single experiments, structural methods have been employed with success to interpret findings and construct ex-post improvements in policy efficiency (Todd and Wolpin, 2005; Duflo et al., 2012; Rodriguez, 2018), as well as to identify parameters of broader policy interest such as labor supply elasticities (Kline and Tartari, 2016) and discounting (Chan, 2017). This paper continues in the tradition of this literature with a more specific focus on the task of aggregation across multiple experiments.

A statistical literature has spawned from the challenge of comparing results across sites when

[4]This is a standard approach to structural econometrics, fitting model implied statistics to observed statistics (Gourieroux et al., 1993).

treatment populations, implementations, and intensities are not strictly comparable. Allowing for this heterogeneity necessitates augmenting the statistical framework found in (1.1). The "Random Effects" approach (Higgins and Green, 2011) and Bayesian hierarchical methods (Rubin, 1981; Gelman et al., 2013) are two such examples. On this issue, the structural approach presents some unique solutions. First, in this setting, many of the differences in implementation across sites, such as the design of incentives, have a natural parameterization within the model, and are in fact crucial for identifying behavioral parameters. Second, the model yields a set of predictions on control group behavior *in addition to* treatment effects. These predictions can be used to discipline heterogeneity across sites that arises due to differences in sample composition. The aforementioned Bayesian techniques, with uninformative priors over each parameter, are used to produce estimates of the model's parameters. This method exploits the approximate normality of the statistics $\mathbf{X}_N$ to derive a likelihood, which is combined with a hierarchical prior on site-level parameters. While this practice is more common in other fields (Gelman et al., 2013), Meager (2019) has employed these techniques to learn about the treatment effects of microfinance and advocates for the method based on desirable statistical properties.

Finally, the estimation exercise in this paper makes use of aggregated data only, in the form of treatment and control group means. However, the conceptual approach readily generalizes to the case in which rearchers have access to disaggregated experimental data. In this case, one could incorporate richer heterogeneity within sites as well as across sites, and potentially model more explicitly the sample selection processes that generate site-level differences. The goal of this paper is to show what can be achieved with this method using just public reports, in the spirit of traditional meta-analyses, with a model designed so that we may learn exclusively from averages. Despite this limitation, this approach still produces precise and meaningful estimates for most parameters of interest, such as the elasticity of labor supply and the price elasticity of child care use, as well as estimates of the technology of skill formation. However, in order to demonstrate the flexibility of the structural approach, Appendix D describes how estimation would proceed in this richer empirical setting.

2 Description of Experiments

The main analysis of this paper aggregates results from 5 experimental evaluations of welfare-to-work programs: Connecticut Jobs First (CTJF), the Family Transition Program (FTP), Los Angeles Greater Avenues for Independance (LAGAIN), the Minnesota Family Investment Program (MFIP), and the National Evaluation of Welfare-to-Work strategies (NEWWS). Each of these studies were conducted by the Manpower Demonstration Research Corporation (MDRC), who publish publicly available reports that include site-specific details regarding treatment components and sample design, as well as control and treatment group averages for outcomes of interest (Bloom et al., 2002, 2000; Freedman et al., 2000; Miller et al., 2000; Gennetian and Miller, 2000; Hamilton et al., 2001). Data used in this paper are collected from these reports. Table 2.1 summarizes the key features of each experiment, with some supporting discussion in this section.

Treatment Components

Each program offered some combination of the following four treatment components: (1) changes to benefit computation formulae, which increase the financial incentive to work, (2) a 30 hour work requirement featuring mandatory participation in employment services when not meeting the requirement, (3) time limits on welfare participation, and (4) increased access to child care subsidies. In the case of MFIP, the experiment was conducted with two treatment arms, one that provided only changes to financial incentives (the "Incentives Only" arm), and one that combines these incentives with mandatory work requirement and employment services (the "Full" program). At NEWWS sites, two versions of program services were modelled as part of the work requirement, one that focused on labor force attachment (LFA programs) and one that focused on Human Capital Development (HCD programs). This anlaysis considers only the treatment effects for the LFA group, since it appears that the HCD treatment does not have a clear comparison to work requirements in other sites.

Of these, components (1) and (3) will have a natural parameterization in the dynamic model. In order to model components (2) and (4), the available evidence suggests some simplifying assumptions, which are outlined in section 2.1. This will amount to modelling work requirements as an additional ordeal (nonpecuniary cost) when participation is combined with the choice to not work, and modelling subsidies as a reduction in child care prices at the site by treatment arm level.

Randomization Design

At each site, randomization occurred when individuals arrived for either application to, or recertification for, welfare participation. This is important to consider, since one may wish to augment the analysis with an explicit model of the selection process, in order to make comparisons to a broader population of interest [5]. Even under this sample definition, there were some important deviations across sites that should be noted. First, in LAGAIN and NEWWS, individuals were assigned upon verification that they were eligible for the prevailing welfare-to-work program at the time. For NEWWS, this means the sample required an additional condition of compliance with an initial request to attend the Job Opportunities and Basic Skills Training program (JOBS) (Hamilton et al., 2001). For LAGAIN, the same sample selection procedure was used[6], however the sample of control and treatment group members were additionally drawn from subsample of applicants

[5] This extension is left for future work, though Appendix D provides a brief discussion of how to handle selection when disaggregated data are available

[6] In California the prevailing welfare-to-work program was known as Jobs First-GAIN

ranked by administrators' assessment of long-term recipiency risk (Freedman et al., 2000). While these differences in sample design may seem discouraging, the evidence below suggests that even among sites with comparable sampling strategies, there is significant heterogeneity in control group behavior. The chosen empirical strategy, therefore, will be to allow for latent heterogeneity at the site level, using control group behavior to discipline estimates. This heterogeneity is assumed to be generated not just by differences in sample design, but also by regional variation in preferences, skills, and opportunities.

Table 2.1: Description of Experiments

Site	Sources	Financial Incentives	Work Requirement	Time Limit	Child Outcome Measures	Location	Sample Selection
CTJF	Bloom et al. (2002)	Yes	Yes	21 months	Achievement, Suspension, BPI, PB	Manchester & New Haven, CT	Random assignment at application or recertification
FTP	Bloom et al. (2000)	Yes	Yes	24-36 months	Achievement, Grade Repetition, Suspension, BPI, PB	Escambia County, FL	Random assignment at application or recertification
LAGAIN	Freedman et al. (2000)	Yes	Yes	No	Achievement, Grade Repetition, Suspension	LA County, CA	Random Assignment to selected sample of applicants.
MFIP	Miller et al. (2000) & Gennetian and Miller (2000)	Yes	No/Yes⋆	No	Achievement, Grade Repetition, Suspension, BPI, PB	Hennepin, Anoka & Dakota Counties MN	Random assignment at application or recertification
NEWWS	Hamilton et al. (2001)	No	Yes	No	WJ-Math, WJ-Reading, Grade Repetition, Suspension	Atlanta, GA & Grand Rapids, MI & Riverside, CA	Random assignment at application or recertification among JOBS eligible

Notes: Achievement is parental evaluation of child's achievement in school on a 5 point Likert scale. BPI and PB are indices of behavioral problems and positive behaviors, taken from parental reports of child behavior. WJ-Math and WJ-Reading are math and reading scores from the Woodcock Johnson assessment.

⋆ MFIP consisted of two treatment arms, one that provided only additional financial incentices to work, and one that also included mandatory employment services.

2.1 Program Data

From the reports listed in Table 2.1, we collected sample sizes, along with control and treatment group means for labor force participation, welfare participation, earnings, and income. In addition to these variables, we also collect whatever treatment effects are available in these reports on child outcomes. The set of available measures across sites can be roughly divided into those measuring academic outcomes (a 5-point parental rating of achievement in school, rates of grade repetition, and Math and Reading scores from the Woodcock-Johnson aptitude test) and those measuring behavioral outcomes (indices of Positive Behaviors and Behavioral Problems, as well as rates of school suspensions). Table 2.1 summarizes which outcomes are available at each site, although it should also be noted that there is also variation across sites in the age range at which effects are reported, as well as the years in which outcomes were measured. Within the model, these differences in measurement are fully articulated, and are therefore not an issue for this analysis. Finally, to discipline the model's estimates of the effect of formal vs informal child care use, we draw from Gennetian and Michalopoulos (2003), who report the fractions of individuals receiving subsidies, the fraction of households that use paid care as their main arrangement, as well as current out of pocket costs across sites. While paid care and formal care are not equivalent categories, it is argued by Gennetian and Michalopoulos (2003) that this measure may be a reasonable proxy, and is provided in a consistent format. Other measures of rates of formality are available in individual reports, but these data are preferred given their consistency in terms of methods of measurement.

Before describing the structure of the model, it will help to first perform some preliminary examination of the experiment data, and establish which features of the model will be important to consider. Table 2.2 reports some of the characteristics of sample members across sites, establishing that while the overwhelming majority of sample members across sites are single mothers, there are other substantial differences in the characteristics of program participants across sites. The former observation leads to a model with a single decision-making agent (the mother), while the latter demands that the empirical framework allow for site-specific characteristics.

Figure 2.1 depicts the control and treatment group means for labor force participation, program participation, and earnings, and suggests the following observations. First, observe that treatment effects on work and welfare participation are small relative to the trends in both treatment and control. Looking at earnings, the figure suggests that these trends are being driven by strong growth in wages and employment opportunities over this time. A traditional meta-analysis that focuses exclusively on treatment effects would discard this information, however in the context of the model, the responsiveness of control groups to these changes in conditions provides additional discipline on estimates.[7] Second, the figure reaffirms that there is significant heterogeneity in

[7]The trends are compelling for an additional reason. Control group members were insulated from the effects of welfare reform, the introduction of which overlaps with the observation window of these experiments. Thus, trends in the control group also reveal a counterfactual no reform scenario, and suggest that the impacts of PROWRA were small compared

Table 2.2: Baseline Characteristics Across Sites

Program/Site	Earnings Last 12 Months	Employed at Random Assignment	White	Married w/ Spouse	Average Number of Children
CTJF-Manchester	62.3%	28.3%	64.9%	1.3%	1.6
CTJF-New Haven	42.6%	21.1%	28.8%	1.3%	1.8
FTP	46.2%	N/A	45.4%	N/A	2
MFIP-LR	32.1%	13.9%	52.8%	0.3%	N/A
MFIP-RA	74.8%	22.7%	65.1%	0.3%	N/A
LA-Gain	27.1%	9.5%	17.3%	6.8%	2
NEWWS-A	23.6%	6.9%	3.5%	1.4%	2.1
NEWWS-GR	46%	11.4%	50.1%	3.3%	1.8
NEWWS-R	40.7%	11.2%	49%	8.1%	2

Note: MFIP reports are split into two samples, long term recipients (MFIP-LR) and recent applicants (MFIP-RA).

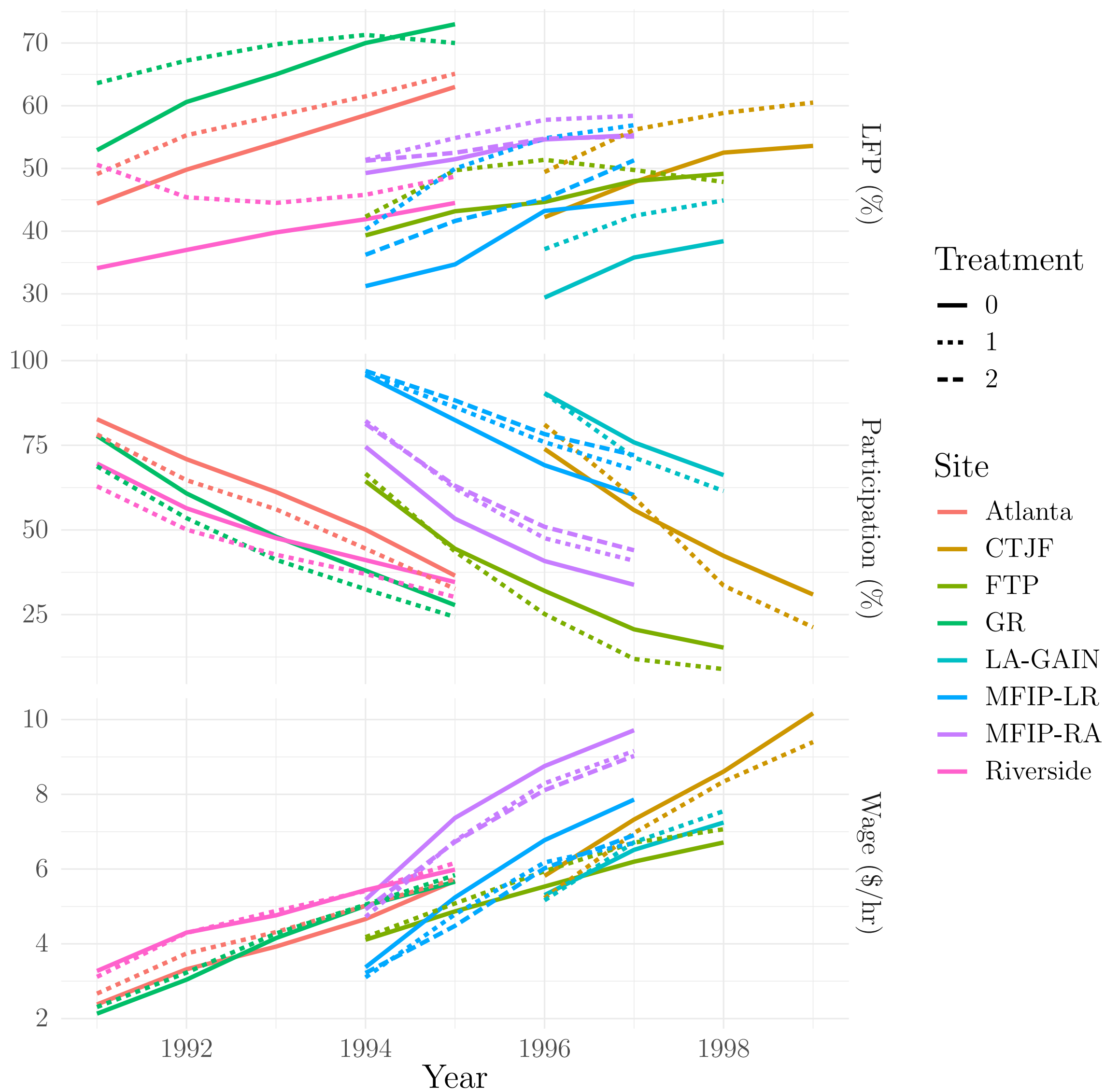
control group behavior across sites.

Table E.1 summarizes the evidence on experimental impacts for other outcomes that could, foreseeably, be affected by treatment. Importantly, there is little evidence of a significant impact on wages of participants, nor on hours worked conditional on working, nor on marital or fertility behavior. This evidence (or lack thereof) validates the following modelling simplifications. First, the model does not allow for any of the employment or training services across sites to have a positive impact on human capital. Second, there is no endogenous fertility and household formation in the model. Third, only a binary work decision is considered, with focus placed exclusively on the extensive margin of labor supply.

Finally, Figure 2.2 draws on the child care means reported by Gennetian and Michalopoulos (2003), and plots subsidy coverage rates against reported out of pocket costs, conditional on using paid care. As expected, the extent of subsidy coverage has a meaningful negative impact on the price paid. The reality of child care subsidies is quite complicated, with claims to subsidies through three main eligibility criteria: (1) if working while receiving welfare, or participating in services through welfare, (2) in the year or two years after transitioning off welfare, or (3) by meeting state income standards. Conditional on the program, subsidy formulae are complicated and idiosyncratic. Furthermore, programs that "expanded" subsidies did so mainly through reducing administrative and informational burdens (Bloom et al., 2000, 2002; Miller et al., 2000). In light of this, the model allows for the impact of subsidies as a single price, $p_{F,k,j}$ for formal care at site k and treatment arm j. Using data collected on subsidy coverage and out of pocket costs, a linear model imputes the effective out of pocket cost and captures the effect of subsidy coverage on price. Figure 2.3 plots this imputed price against rates of using paid care, showing that the imputed price has a meaningful relationship with demand for care. The rest of the paper uses the notion of formal and

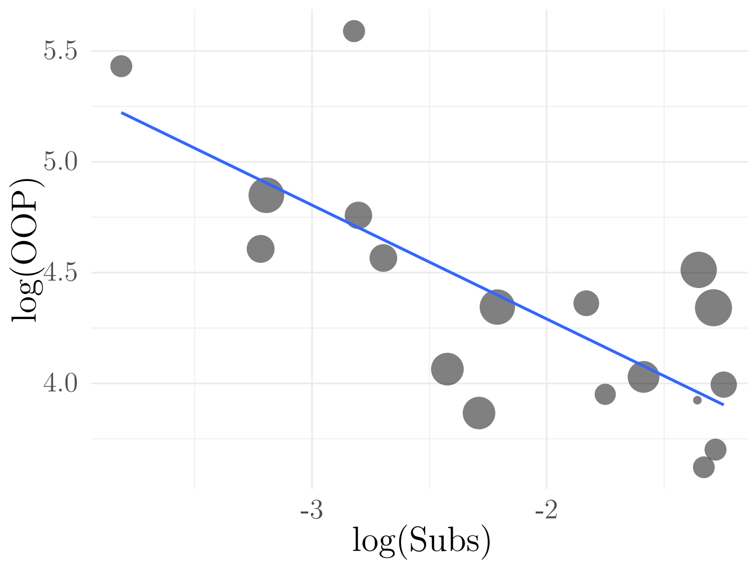
to the effect of economic trends.

Figure 2.1: Annual Means of Labor Force Participation, Program Participation, and Earnings



Notes: Labor Force Participation (LFP) and Welfare Participation are reported as annual averages of quarterly rates. Hourly wage is imputed using average quarterly earnings, assuming 30 hours worked per week.

Figure 2.2: Rates of Subsidy Coverage against Out of Pocket Child Care Costs



Notes: log(Subs) is the log of the fraction of individuals at each site and treatment arm receiving any child care subsidy. log(OOP) is log of the mean reported out-of-pocket costs in $ per week. Source: Gennetian and Michalopoulos (2003).

paid care interchangeably. This assumption is clearly not innocuous, and is relevant for the later discussion of results.

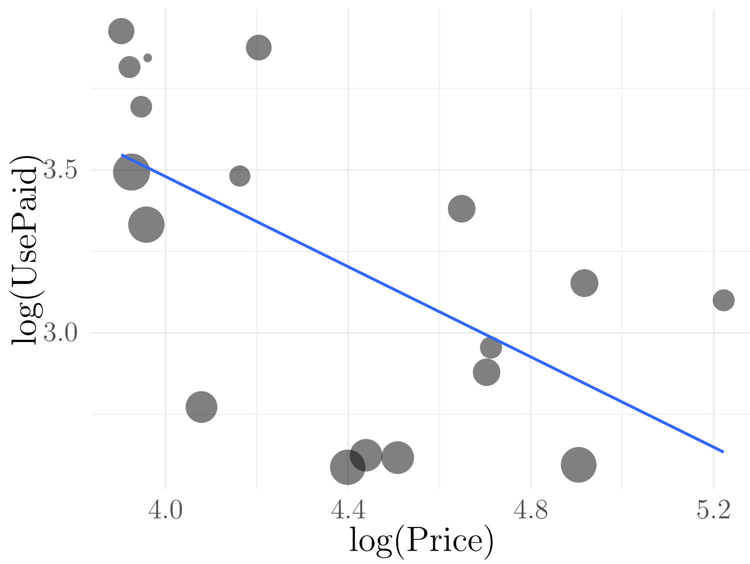
3 Model

This section describes a flexible empirical model of child development, designed to deliver outcome equations that allow a clear mapping to average treatment effects. This discussion serves dual purposes. First, it is helpful to see how the model's parameters map to predictions on treatment effects in labor supply, program participation, and child outcomes. Second, the inverse of this mapping clearly establishes identification of model parameters through control group means and experimentally induced changes to incentives. In the ensuing exposition, I suppress dependance of model outcomes on family size, and consider the evironment for a representative mother at experiment site k in treatment group j. Details on which other observables affect the problem are included in estimation and discussed in Appendix C.

3.1 Model Primitives

Time is discrete and indexed by t. The child's skills are malleable for T periods, at which point the investment problem ends. In this model, one period is equal to one year. We will model a

Figure 2.3: Rates of Subsidy Coverage against Out of Pocket Child Care Costs



Notes: log(UsePaid) is the log of the fraction of individuals at each site and treatment arm that report using paid care while working. log(Price) is log of the imputed price given the subsidy coverage rate in $ per week. Source: Gennetian and Michalopoulos (2003).

representative week[8] in this year, which consists of L hours. This is to allow for the fact that each mother may spend a portion of their week working, and will solicit external care for this part of the week. Accordingly, we set $L = 112$, to represent 112 hours in a week.

Choices and Preferences

In each period, mothers make a series of choices: (1) whether to participate in welfare, $A \in \{0,1\}$, (2) whether to work, $H \in \{0,1\}$, (3) conditional on working, whether to make a formal care arrangement for their child $F \in \{0,1\}$, (4) how much time to invest in their child, τ, and how much time to spend in housework, q, and (5) how much to spend on private consumption (C), and how much to spend on home investments on the child, x.[9] Let the vector $d \in \{0,1\}^3$ collect each of the discrete choices, $d = \{A, H, F\}$, and assume from here on that women who do not work maintain care of the child, and thus set $F = 0$ by default in this case.

Mothers value their private consumption, C_t, and their child's current stock of attributes, θ_t. Her utility in period t is given by:

$$U_k(C, d, \theta) = \alpha_C \log(C) + \alpha_\theta \log(\theta) - \alpha_{H,k} H - \alpha_{A,k} A + \alpha_{F,k} F + \epsilon_d$$

[8] This is an innocuous scaling assumption.

[9] x here is treated as an expenditure variable, since the price of child-specific investment goods is treated here as unobserved and must, therefore, be assumed to be constant across individuals.

where ϵ_d is a choice-specific random variable, which is distributed independently and identically across agents and over time. The parameters $(\alpha_{H,k}, \alpha_{A,k}, \alpha_{F,k})$ represent fixed, nonpecuniary utilities from working, participating in welfare, and soliciting paid childcare. They are permitted to be heterogeneous across locations k.

There is a terminal payoff at time T, when the child's development has concluded, equal to:

$$\alpha_{V,T} \log(\theta_T).$$

Mothers are forward-looking, discounting the future at rate β. Thus, she values future sequences of decisions as

$$V_{kt} = \mathbb{E}_t \left\{ \sum_{s=t}^{T-1} \beta^{s-t} U_k(C_s, d_s, \theta_s) + \alpha_{V,T} \log(\theta_T) \right\}$$

where $\mathbb{E}_t$ is her conditional expectation given information at time t. Preferences U are revealed by virtue of the assumption that mothers make their participation, time use, and investment decisions in order to maximize this expected discounted present value of utilities. As is typical in dynamic discrete choice problems, I assume that the taste shocks are drawn from a type I extreme value distribution, with nested shocks. In the first nest is the participation decision, with a work decision nested within each participation choice. When work is chosen, a third nest arrives containing the child care decision. I set variances for each nest equal to the triple $(1, \sigma_H^2, \sigma_F^2)$.

Technology and Constraints

There are three relevant technologies in the economy. First, the decision to work and participate in welfare generates income $Y_{kjt}(H, A)$, which depends on site location, k, treatment status j, and date t. This dependence is generated by location-specific wages and site-by-treatment specific policies. Second, time spent in housework can generate the consumption good, measured in dollars, inside the household at a linear rate, w_q. Third, the child's attributes evolve according to the technology of skill formation:

$$\theta_{t+1} = \theta_t^{\delta_\theta} I_t^{\delta_{I,t}}, \tag{3.1}$$

where the aggregate investment good I_t is a function of the mother's investment and childcare decisions:

$$I_t = \mathcal{I}_t(\tau_t, x_t, I_F; H_t, F_t)$$

Here the dependance of the function on t allows for age-sensitivity of development. If the mother works, then the child receives inputs I_F from the chosen form of care. Given the technology of home production, mothers face the following constraints on time and money:

$$\tau + q + 30H \leq L \tag{3.2}$$

$$C + x - p_{F,kj}F \leq Y_{kjt}(H, A) + w_q q \tag{3.3}$$

which can be combined into a single resource constraint:

$$C + x + w_q(\tau + 30H) - p_{F,kj}F \leq Y_{kjt}(H, A) + w_q L.$$

Two features of this constraint require exposition. First, notice that women are assumed to work 30 hours a week, conditional on working. Second, notice that mothers who go to work and make a formal care arrangement for their children must pay the cost of formal care, $p_{F,kj}$, which depends on site and treatment status.

Rather than fully specify the functional form of $\mathcal{I}$, I specify the corresponding expenditure function generated by agents solving the cost minimization problem:

$$e(I; H, F) = \min w_q \tau + x \qquad \text{s.t. } \mathcal{I}(x, \tau, H, F) \geq I$$

For analytical tractability, I specify a set of linear expenditure functions that depend on the care arrangement made for the child:

$$e(I; H, F) = \begin{cases} g_{0,t} I & \text{if } H = 0 \\ g_{1,t} I & \text{if } H = 1, F = 0 \\ g_{2,t} I & \text{if } H = 1, F = 1 \end{cases}$$

This expression can be written more compactly as

$$e(I, H, F) = g_{\kappa,t} I, \qquad \kappa = H + F$$

where $\kappa = H + F$ indexes the care arrangements made for the child. Here, the prices of investment $(g_{0,t}, g_{1,t}, g_{2,t})$ are a function of deeper economic primitives $(\mathcal{I}_t)$, however I follow Marschak (1953) and note that it is sufficient to estimate the prices themselves, subject to the assumption of policy invariance.[10] Examinining the cost minimization problem described above, we can see that invariance is a reasonable assumption for policies that do not affect the marginal value of time (w_q), the price of the money investment good, or the quality of inputs from informal and formal care arrangements.

3.2 Incorporating Welfare Policies

Thus far, cash transfer programs provided by governments feature only in the income function, Y_{kjt}, which has yet to be given an explicit formulation. This section unpacks this function in order to make it clear that the design of transfers will crucially shape participation and work behavior through the provision of incentives embedded in Y_{kjt}. The model also extends to incorporate three other crucial features of welfare policies: (1) time limits on participation, (2) work requirements, and (3) child care subsidies. While the first of these features has an unambiguous parameterization in the model, prior discussions of evidence inform specific assumptions on the latter two.

[10] In fact, we will see that only the ratios of the prices, $g_{1,t}/g_{0,t}$, and $g_{2,t}/g_{0,t}$ are necessary for counterfactual analysis of welfare policies.

Income Function

Income is the sum of earnings and receipt of transfers from the government. For control group members ($j = 0$), this is the sum of Aid to Families with Dependent hildren (AFDC, or welfare), and the Supplemental Nutrition Assistance Program (SNAP, or food stamps):

$$Y_{k0t}(H, P) = E_{kt}H + A\left(\text{AFDC}_t(E_{kt}H; Z_{A,kt}) + \text{SNAP}_t(E_{kt}H; Z_{S,kt})\right)$$

where welfare and food stamp payments are a function of earnings, $E_{kt}H$, as well as policy and person-level observables $(Z_{A,kt}, Z_{S,kt})$. These include the number of children in the household, benefit standards for both welfare and food stamps, and income disregards for both programs. Income disregards affect the deduction of earnings from total benefits, and therefore determine implicit marginal tax rates on earnings. For members of treatment groups, part of the treatment may involve changes to benefit computation formulae:

$$Y_{kjt}(H, P) = E_{kt}H + A\mathcal{T}_{kj}(E_{kt}H; Z_{A,kt}, Z_{S,kt})$$

Appendix B describes explicit formulae for each transfer function $\mathcal{T}_{kj}$ as well as control group policy rules. Explicit details are not necessary for the observation that assignment to treatment provides helpful randomization in marginal tax rates through changes to benefit computation.

Time Limits

Let Ω_{kj} indicate the lifetime limit on welfare participation, and let ω indicate the participant's stock of remaining periods of use. While $\omega > 0$, participation is not limited in any way. Once the limit is reached, $\omega = 0$, the AFDC portion of the benefit payment is removed, and the individual receives only food stamps. By default, the time limit on participation under AFDC ($j = 0$) is set to $\Omega_{k0} = \infty$.

Work Requirements

In all but one of the experimental sites, welfare participants were subject to a requirement that individuals either work for 30 hours a week or, if unemployed, participate in mandatory search, employment, or training services provided by the state. The design of these employment and training services were a focus of the original studies (Hamilton et al., 2001), and they differ in terms of their program offerings, tolerance for noncompliance, and strength of enforcement.

However, given the limited evidence of any effect on wages or hours of work, conditional on working, I take a simple modelling approach: when participating in welfare, an additional ordeal or "hassle cost" must be paid by the recipient, when not working. Additionally, I allow for the potential for employment services to alter the disutility of work, which they may do by reducing the cost of finding employment. This feature can be incorporated by re-writing preferences as:

$$U_{kj}(C, d, \theta) = \alpha_C \log(C) + \alpha_\theta \log(\theta) - \alpha_{H,k}H + \alpha_{F,k}F - \alpha_{A,k}A - \mathcal{R}_{kj}(\alpha_{R,k}(1-H)A + \alpha_{R2,k}HA) + \epsilon_d$$

with the effect of the work requirement summarized by the additional pair of parameters $(\alpha_{R,k}, \alpha_{R2,k})$, and $\mathcal{R}_{kj} \in \{0, 1\}$ an indicator for whether work requirements apply in this setting.

Childcare Subsidies

As discussed previously, the effect of subsidies is summarised at the site (k) by treatment (j) level in the effective average price, $p_{F,kj}$ paid by all participants, which is imputed from data on out of pocket costs and subsidy coverage rates.

3.3 Key Outcome Equations and Identification

The model described above parameterizes a set of predictions on participation, labor supply, and child outcomes as a function of prevailing welfare policies and labor market opportunities. Naturally, these primitives also map to predictions on treatment effects. This section sketches out the model's empirical content by specifying these key outcomes under optimal behavior, which is given by the solution to a recursive dynamic program. Full details of the model's solution can be found in Appendix A, but a brief sketch of the solution is provided here.

First, dynamics are introduced through two mechanisms: (1) the technology of skill formation, which requires mothers to weigh the benefit of private consumption against future improvements in their child's traits, and (2) time limits, which require mothers to weigh the benefits of welfare use today against their use in the future. Phrased recursively, the functional forms of this model deliver an analytical solution to the investment problem, as in Del Boca et al. (2014) and Mullins (2019). This direct solution can be used to explicitly derive indirect utilities that depend on the age of the child, which effectively neutralizes the dynamics introduced by the investment problem. Therefore, in the absence of time limits, choice probabilities are delivered by a sequence of static problems using indirect utilities. In the presence of time limits, plugging in indirect utilities gives a discrete choice dynamic program, which permits a relatively parsimonious, albeit numerically computed, solution.

To aid in exposition, Table 3.1 provides a summary of the analysis below, specifying the relationship between model parameters and treatment group means, and points to the relevant equations that demonstrate this relationship.

Investment and Child Outcomes

First order conditions on investment yield a linear expenditure equation:[11]

$$g_{\kappa,t}I_t = \underbrace{\frac{\beta\delta_{I,t}\alpha_{V,t+1}}{\alpha_C + \beta\delta_{I,t}\alpha_{V,t+1}}}_{=\varphi_t}\left[Y_{kjt}(H,A) + w_q(L - 30H) - p_{F,kj}F\right]$$

[11]It is possible to relax the functional form of investment demand in this model, though in this setting we have only the child's outcome equation to discipline our assumptions on behavior.

where we define φ_t as the *marginal propensity to invest* in the child and $\alpha_{V,t}$ is a dynamic coefficient with a recursive representation:

$$\alpha_{V,t} = \alpha_\theta + \beta\delta_\theta\alpha_{V,t+1}.$$

Given the resource constraint, the investment equation implies

$$C_t = (1-\varphi_t)(Y_{kjt}(H,A) + w_q(L-30H) - p_{F,kj}F),$$

which means that we can solve for the indirect utility of the discrete choice, d, as:

$$\begin{aligned} u_{kjt}(d) = (\alpha_C + \beta\delta_{I,t}\alpha_{V,t+1})\log(Y_{kjt}(H,A) + w_q(L-30H) - p_{F,kj}F) \\ - \beta\delta_{I,t}\alpha_{V,t+1}\log(g_{H+F}/g_0) - \alpha_{H,k}H - \alpha_{A,k}A + \alpha_{F,k}F - \mathcal{R}_{kj}A[\alpha_{R2,k}H + \alpha_{R,k}(1-H)] \\ + \beta\delta_{I,t}\alpha_{V,t+1}(\log(\varphi_t) - \log(g_0)) + \alpha_C\log(1-\varphi_t) \end{aligned} \quad (3.4)$$

Notice that each term on the last line is a constant, unaffected by any choices. We can therefore ignore these terms in the remainder of this analysis. I will also, for convenience, define $\Gamma_t = \beta\delta_{I,t}\alpha_{V,t}$ and $\alpha_{C,t} = \alpha_C + \Gamma_t$.

Child Care

Define $P_{F,kjt}(A) = P[F=1|A,k,j,t]$ to be the probability of making a formal care choice, fixing participation, A. Conditional on working, the probability of choosing formal care can be written as:

$$\log\left(\frac{P_{F,kjt}(A)}{1-P_{F,kjt}(A)}\right) = \sigma_F^{-1}\left[\alpha_{C,t}\log\left(\frac{Y_{kjt}(1,A) + w_q(L-30) - p_{F,kj}}{Y_{kjt}(1,A) + w_q(L-30)}\right) + \alpha_{F,k} - \Gamma_t\log(g_{2,t}/g_{1,t})\right] \quad (3.5)$$

This expression shows that $\alpha_{F,k}$ will be identified by mean levels of formal care use, while σ_F determines the elasticity of formal care use with respect to the price of formal care. The latter parameter will therefore be identified by experimental variation in the effective price, $p_{F,kj}$, of child care through increases in subsidy generosity and availability.

Labor Supply

Define $P_{H,kjt}(A) = P[H=1|A,k,j,t]$ to be the probability of working, fixing participation, A. Then, integrating out the formal care preference shocks yields the expected value from working:

$$\alpha_{C,t}\log(Y_{kjt}(1,A) + w_q(L-30) - p_{F,kj}) + \alpha_{F,k} - \Gamma_t\log(g_{2,t}/g_{0,t}) - \sigma_F\log(P_{F,t}(A))$$

which yields the probability of work as:

$$\begin{aligned} \log\left(\frac{P_{H,kjt}(A)}{1-P_{H,kjt}(A)}\right) = \sigma_H^{-1}\Bigg[\alpha_{C,t}\log\left(\frac{Y_{kjt}(1,A) + w_q(L-30) - p_{F,kj}}{Y_{kjt}(0,A) + w_qL}\right) - \alpha_{H,k} \\ + A\mathcal{R}_{kj}(\alpha_{R,k} - \alpha_{R2,k}) + \alpha_{F,k} - \Gamma_t\log(g_{2,t}/g_{0,t}) - \sigma_F\log(P_{F,kjt}(A))\Bigg] \end{aligned} \quad (3.6)$$

Here we can see that $\alpha_{H,k}$ determines (and is identified by) mean levels of labor force participation, the joint term $(\alpha_R - \alpha_{R,2})$ is identified by experimentally induced variation in assignment to programs with work requirements $(\mathcal{R}_{kj})$, while σ_H is identified by the response of labor supply to experimentally induced changes in the incentive to work (through welfare computation formulae) as well as through changes over time in market wages (Y_{kjt}). Finally, w_q is identified by the labor supply response to changes in welfare generosity. Roughly speaking, σ_H and w_q determine the relative size of income and substitution effects in labor supply decisions. Finally observe that Γ_t is a function of δ and α_θ, thus α_θ is identified by differences in the levels of all variables by age of the child.

Program Participation without Time Limits

Define $\rho_{kjt}(\omega) = P[A = 1|t, k, j, \omega]$ to be the probability of program participation, given the current stock of remaining periods, ω. Without time limits, program participation is a static decision. By integrating out the distribution of shocks conditional on working and conditional on participation gives:

$$\log\left(\frac{\rho_{kjt}(\omega))}{1-\rho_{kjt}(\infty))}\right) = \alpha_{C,t}\log\left(\frac{Y_{kjt}(0,1)+w_qL}{w_qL}\right) - \sigma_H\log\left(\frac{1-P_{H,kjt}(1)}{1-P_{H,kjt}(0)}\right) - \mathcal{R}_{kj}\alpha_{R,k}. \quad (3.7)$$

This expression establishes that α_C is identified by the response of participation to experimentally induced changes in the generosity of welfare, and that $\alpha_{R,k}$ can be separately identified from $\alpha_{R,2,k}$ through participation responses to work requirements. Intuitively, if work requirements exclusively make working less costly, we should only see, *ceterus paribus*, reductions in welfare participation through mechanical deductions in eligiblity. However, any additional reduction in participation beyond this must be rationalized by $\alpha_{R,k}$.

Program Participation With Time Limits

So far, we have established identification of parameters either through levels $(\alpha_{H,k}, \alpha_{A,k}, \alpha_{F,k})$ or through treatment responses $(\sigma_F, \sigma_H, \alpha_C, \alpha_{R,k}, \alpha_{R2,k})$. To finish this section, I present an equation that shows us how discounting, β, is identified by the effect of time limits on welfare participation. In Appendix A, I show that one can exploit the finite dependance properties of the model (Arcidiacono and Miller, 2011) to express the effect of time limits on participation in period t in terms of its effect in period $t + 1$:

$$\log\left(\frac{\rho_{kjt}(\omega)}{1-\rho_{kjt}(\omega)}\right) - \log\left(\frac{\rho_{kjt}(\infty)}{1-\rho_{kjt}(\infty)}\right) = \beta\left[\log\left(\frac{\rho_{kjt+1}(\omega)}{1-\rho_{kjt+1}(\omega-1)}\right) - \log\left(\frac{\rho_{kjt+1}(\infty)}{1-\rho_{kjt+1}(\infty)}\right)\right]. \quad (3.8)$$

This equation decomposes the treatment effect on welfare use into two parts. The first term summarizes the effect of the incentive created to save limited welfare entitlements for periods beyond $t + 1$, while the second term summarizes the incentive to save one period of use today for use tomorrow.

Furthermore, by noting that $\rho_{kjt}(\omega) = \rho_{kjt}(\infty)$ for all pairs (ω, t) such that $T - t \leq \omega$, the treatment effects of time limits on all states can be expressed recursively as solely a function of choice probabilities without time limits, $\rho_{kjt}(\infty)$, and the discount factor β. The recursion begins with:

$$\log\left(\frac{\rho_{kjT-1}(1)}{1-\rho_{kjT-1}(1)}\right) - \log\left(\frac{\rho_{kjT-1}(\infty)}{1-\rho_{kjT-1}(\infty)}\right) = \beta\log(\rho_{kjT}(\infty))$$

From this point, each $\rho_{kjt}(\omega)$ can be constructed from $\rho_{kjt}(\infty)$, $\rho_{kjt+1}(\infty)$, $\rho_{kjt+1}(\omega)$ and $\rho_{kjt+1}(\omega - 1)$. Thus, the impact of time limits on this model are determined by β.

Child Outcomes

Given the current stock of traits, θ_t, child outcomes can be written as:

$$\log(\theta_{t+1}) = \delta_{I,t}\left[\log(Y_{kjt}(H, A) + w_q(L - 30H)) - \log(g_{\kappa,t})\right] + \delta_\theta \log(\theta_t).$$

Letting Δ denote an operator that compares choices d to some counterfactual d', then we get an effect on child outcomes of:

$$\Delta\log(\theta_{t+1}) = \delta_{I,t}\left[\Delta\log(Y_{kjt}(H, A) + w_q(L - 30H)) - \Delta\hat{g}_{\kappa,t}\right] + \delta_\theta\Delta\log(\theta_t) \tag{3.9}$$

where $\hat{g}_{\kappa,t} = \log(g_{\kappa,t}/g_{0,t})$. Equation (3.9) shows us how production parameters map experimental effects on choices to average treatment effects on child outcomes. Namely, $\delta_{I,t}$ will determine the magnitude of impacts when programs manipulate material resources directly (through Y_{kjt}) or indirectly. The relative log prices, $\hat{g}_{1,t} = \log(g_{1,t}/g_{0,t})$ and $\hat{g}_{2,t} = \log(g_{2,t}/g_{0,t})$, determine effects when mothers change their labor supply behavior and choice of care. A value of $\hat{g}_{\kappa,t}$ greater than zero indicates that investment is more costly when the child is placed in care option κ compared to when they remain at home, and would dictate negative impacts on outcomes, ceterus paribus, when mothers go to work. Finally, δ_θ determines the persistence of these impacts over time. Identification of these parameters relies on there being sufficient variation across sites in the extent to which treatments affect income, rates of labor force participation, and rates of formal care use, and the differences in the timeframe over which measurements are taken.

4 Estimation

As has been discussed, the central methodological proposition of this paper is motivated by the desire to estimate economic primitives that facilitate different policy counterfactuals, and may be used for treatment settings in which the average treatment effect is not an appropriately defined statistical object, let alone a parameter of policy relevance. However, a structural meta-analysis must deal with the same underlying issues faced by the traditional literature, chief among them being the existence of underlying site-level heterogeneity in impacts, sample selection, and treatment design.

Table 3.1: Guide to Identification of Model Parameters

Parameter	Source (relevant equation)
$\alpha_{F,k}$	Site level means in paid child care use (3.5)
$\alpha_{H,k}$	Site level means in LFP (3.6)
$\alpha_{A,k}$	Site level means in welfare participation (3.8)
$\alpha_{R,k}, \alpha_{R2,s}$	Experimental impacts on work and participation, (3.6) & (3.7)
σ_F	Response in paid care use to experimental change in price (3.5)
σ_H	Response in LFP to experimental change in financial work incentives (3.6)
w_q	Response in LFP to experimental change in generosity of welfare (3.6)
α_C	Response in participation to experimental change in generosity of welfare (3.7)
β	Response in participation to time limits (3.7)
δ_I	Response in child outcomes to mechanical and behavioral changes in income (3.9)
δ_θ	Differences in the timing of measurement of child outcomes (3.9)
$\hat{g}_{1,t}, \hat{g}_{2,t}$	Response in child outcomes to changes in work behavior and paid child care use (3.9)
Γ_t/α_θ	Differences in levels with respect to child age

This paper handles these challenges using three strategies, two of which are unique to the structural approach, and one that borrows from the existing literature. First, the model embeds predictions on control group means as well as treatment effects. Differences in control group behavior across sites are therefore informative about site-specific heterogeneity in preferences that may be due simply to regional variation, as well as differences in selection criteria across sites.

Second, the model suggests natural parameterizations of site-specific treatment components, such as child care subsidies, financial incentives to work, and time limits on participation. In contrast to the literature, these differences in treatment are crucially informative of important underlying parameters, not as complicating sources of variation that merely obscure the average treament effect. This marks an important and substantive deviation from traditional meta-analytic approaches.

Finally, one may wish to specify and estimate the distribution of primitives across sites, as a way to forecast the impacts of a treatment at a new site. Here, estimation follows the literature with a specification that uses hierarchical priors. This method is particularly useful for handling the effectiveness of work requirements across sites, something about which the written model has little to say. Bayesian methods are typically employed for estimating these hierarchical models, and have been applied in economic contexts by Meager (2019), who argues that they have desirable statistical properties (Gelman et al., 2013).

4.1 Data

The estimation procedure uses control and treatment group annual means of labor force participation, program participation, as well as surveyed rates of use of paid child care among those working. The vector X_k represents these compiled data for each site k. In addition, the procedure uses (when available) estimates of treatment effects on 4 academic outcomes (parental ratings of achievement, grade repetition, Woodcock-Johnson Math and Woodcock-Johnson Reading) and 3 behavioral outcomes (suspensions, behavioral problems, and positive behaviors). Let M_k denote this vector of treatment effects. In the case of MFIP, means are reported consistently in a way that distinguishes between long run recipients and recent applicants. Accordingly, these data are treated as separate sites in estimation (MFIP-LR and MFIP-RA).

MDRC's results tables do not report standard errors on treatment effects or means, which are necessary to quantify the uncertainty of estimates. Two strategies can solve this problem. First, for all binary outcomes $X_{j,k}$, standard errors are given by the formula: $s^2_{j,k} = X_{j,k}(1 - X_{j,k})/N_{j,k}$. Second, effect sizes, which divide treatment effects by the standard deviation of the control group outcome, imply sample standard deviations. Using sample size and an assumption on homoskedasticity, standard errors of the treatment effect can then be estimated. Let s_X and s_M denote the vectors of standard errors for the moments. Treatment effects are normalized by the sample standard deviation of the measure, so effects can be interpreted in fractions of a standard deviation. The sign on undesirable outcomes (repetition, suspension, and behavioral problems) has been reversed.

4.2 Details of Production

To simplify the age dependence of $\delta_{I,t}$ and $g_{\kappa,t}$, each parameter takes one value when the child is aged 0 to 9, and another value for children aged 10 and older. While this assumption is a stark restriction relative to what is known about developmental sensitivities (Morris et al., 2005), assumptions here are limited by the variation in age categories provided by public reports. Since multiple measures of academic skills, M_A, and behavioral skills, M_B, are available, a linear factor structure relates treatment effects on measure m at site k for treatment j:

$$M_{Z,m,k,j} = \lambda_{Z,m} \Delta \mathbb{E}[\log(\theta)|k,j] + \zeta_{Z,m,k,j}, \qquad Z \in \{A, B\}$$

where the variance of $\zeta_{Z,m,k,j}$ is estimated as previously described. The normalization $\lambda_{A,1} = \lambda_{B,1} = 1$ applies.

These assumptions yield the set of production parameters:

$$\gamma_\delta = (\delta_{I,0-9}, \delta_{I,10+}, \delta_\theta, \hat{g}_{1,0-9}, \hat{g}_{1,10+}, \hat{g}_{2,0-9}, \hat{g}_{2,10+}, \lambda)$$

which are estimated separately for the two different skill types.

4.3 Heterogeneity Across Sites

The model approached heterogeneity across sites by allowing for site-specific differences in average behavior, and specifying global values of parameters that govern the responsiveness of individuals to treatment. Following the identification discussion in Section 3.3, the location specific parameters are:

$$\gamma_k = (\alpha_{H,k}, \alpha_{A,k}, \alpha_{F,k}, \alpha_{R,k}, \alpha_{R2,k}).$$

These pin down control group levels in labor supply, welfare participation, formal care use, and response to work requirements. In addition, the following are global parameters:

$$\gamma_G = (\beta, \alpha_C, \alpha_\theta, w_q, \sigma_H, \sigma_F)$$

where β controls the responsiveness of participation to time limits, α_C controls the responsiveness of participation to program generosity, w_q controls the labor supply response to program generosity, σ_H controls the labor supply response to wages, and σ_F controls the responsiveness of formal care use to prices.[12]

I also specify a set of hyperparameters, γ_H, that govern the distribution of location-specific parameters, as follows:

$$\alpha_{x,k} \sim \text{LogNormal}(\alpha_x, \tilde{\sigma}_x^2) \qquad x \in \{H, A, F, R, R2\}$$

Finally, only very weakly informative priors, p, over global parameters, γ_G, and hyperparameters γ_H, apply. These are summarized in Appendix C.

Putting this together, the posterior likelihood as:

$$p(\gamma|X, M) \propto \prod_{k=1}^{K} \phi(X_k, M_k|s_{M,k}, s_{X,k}, \gamma_\delta, \gamma_G, \gamma_k)p(\gamma_k|\gamma_H)p(\gamma_H, \gamma_G, \gamma_\delta) \tag{4.1}$$

where $\phi(\cdot|s, \gamma)$ denotes the normal density with mean implied by model solution given γ and standard deviation s. A Markov Chain Monte Carlo method provides a sample from the posterior distribution, which Appendix C describes in more detail.

5 Estimates and Counterfactuals

5.1 Discussion of Preference Estimates

Table 5.1 reports estimates from the sample posterior distribution of γ_G and γ_H, while Table 5.2 reports estimates of site-specific parameters. Of primary interest is the behavioral implications of these estimates, which this section will explore, but some initial general comments are warranted.

[12] It should be noted that in practice, treatment effects are a function of all parameters in the model, and the above outline offers only a conceptual guideline to identification.

Table 5.1: Global Parameter and Hyperparameter Estimates

Parameter	2.5%	25%	Estimate	75%	97.5%
α_C	22.24	23.11	23.76	24.48	25.87
σ_H	8.61	8.98	9.11	9.38	9.70
σ_F	3.32	4.02	4.35	4.36	4.73
w_q	3.94	4.13	4.52	4.56	4.73
β	0.76	0.78	0.79	0.81	0.83
α_H	1.10	3.38	2.75	5.45	7.87
$\tilde{\sigma}_H$	2.68	3.53	4.12	5.12	8.13
α_A	3.82	4.38	4.56	4.86	5.47
$\tilde{\sigma}_A$	0.63	0.84	0.88	1.23	1.93
α_F	10.59	12.48	13.48	14.13	15.62
$\tilde{\sigma}_F$	1.25	1.73	1.82	2.59	4.16
α_R	-1.41	0.37	1.64	1.84	3.59
$\tilde{\sigma}_R$	1.89	2.55	3.85	3.78	6.10
α_{R2}	-3.01	-1.33	-0.71	0.15	1.91
$\tilde{\sigma}_{WR2}$	1.89	2.57	2.26	3.78	6.11

Notes: "Estimate" refers to the posterior estimate, the mode of the posterior distribution. Column $x\%$ refers to the x^{th} quantile of the posterior distribution, defining the boundaries of credibility sets.

First, the model yields quite reasonable estimates of β, with a 95% credibility interval that ranges from 0.76 to 0.83. Second, site-specific estimates of α_R and α_{R2} suggest that work requirements yield an effect both through the hassle cost of remaining unemployed while receiving benefits, as well as through reductions in the cost of work (although LAGAIN and NEWWS Atlanta are exceptions).

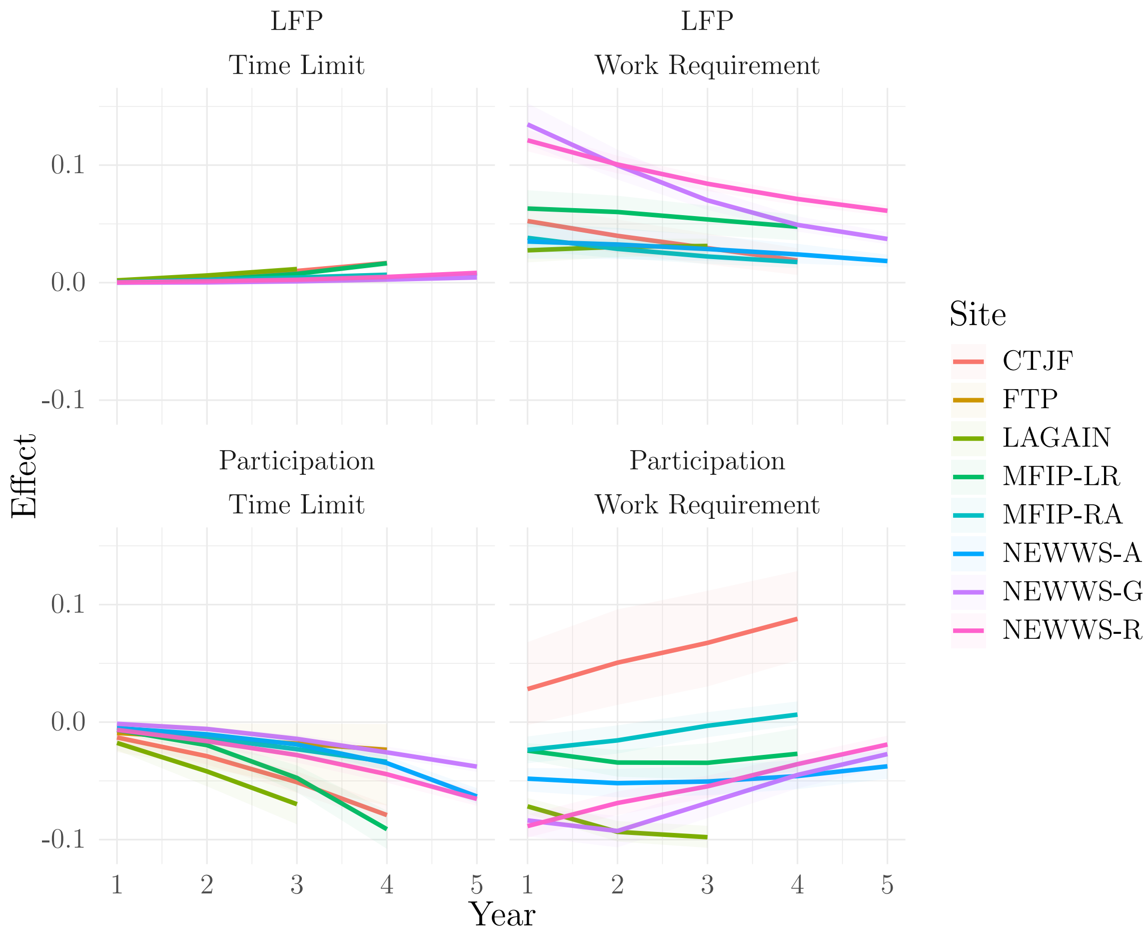
Figure 5.1 plots the labor force and participation impacts of two key program features when implemented in isolation on the control group: a work requirement and a 5 year time limit on welfare. Estimates suggest that time limits have, over the course of the experiment, a persistent and meaningful negative impact on participation between 5 and 10 percentage points, with little effect on labor supply. In constrast, work requirements appear to significantly affect both work and program participation, but these effects appear to fade over time. This finding is consistent with results at sites where work requirements were the only component of treatment (Hamilton et al.,

Table 5.2: Site-Specific Parameter Estimates

	α_H	α_A	α_F	α_R	α_{R2}
CTJF	7.80	4.63	14.04	1.02	-1.96
	(0.39)	(0.08)	(0.54)	(0.27)	(0.20)
FTP	-0.55	5.71	9.59	-	-
	(0.36)	(0.38)	(0.38)	-	-
LAGAIN	8.20	3.92	15.50	0.90	0.01
	(0.38)	(0.08)	(0.56)	(0.08)	(0.15)
MFIP-LR	6.70	3.76	14.36	1.46	-1.45
	(0.37)	(0.09)	(0.57)	(0.14)	(0.17)
MFIP-RA	6.22	3.91	15.87	1.10	-0.81
	(0.38)	(0.08)	(0.61)	(0.16)	(0.16)
NEWWS-A	0.23	4.44	12.82	1.74	0.01
	(0.30)	(0.09)	(0.48)	(0.22)	(0.19)
NEWWS-G	-1.30	5.07	15.15	5.08	-1.90
	(0.41)	(0.08)	(0.58)	(0.43)	(0.18)
NEWWS-R	7.31	6.59	12.98	3.60	-3.90
	(0.33)	(0.11)	(0.46)	(0.21)	(0.23)

Notes: Reported estimates are the mode calculated from the sample posterior distribution. Bracketed numbers indicate standard deviation of the estimate in the posterior distribution. Estimates at FTP for α_R and α_{R2} are not identified, since work requirements applied to both groups.

Figure 5.1: Labor Market and Participation Impacts of Work Requirements and Time Limits



Notes: Figure shows the effect of a 5 year time limit and the imposition of a work requirement on labor force participation (LFP) and welfare use (Participation) across sites. Shaded region shows 95% credible interval, computed by Monte Carlo simulation.

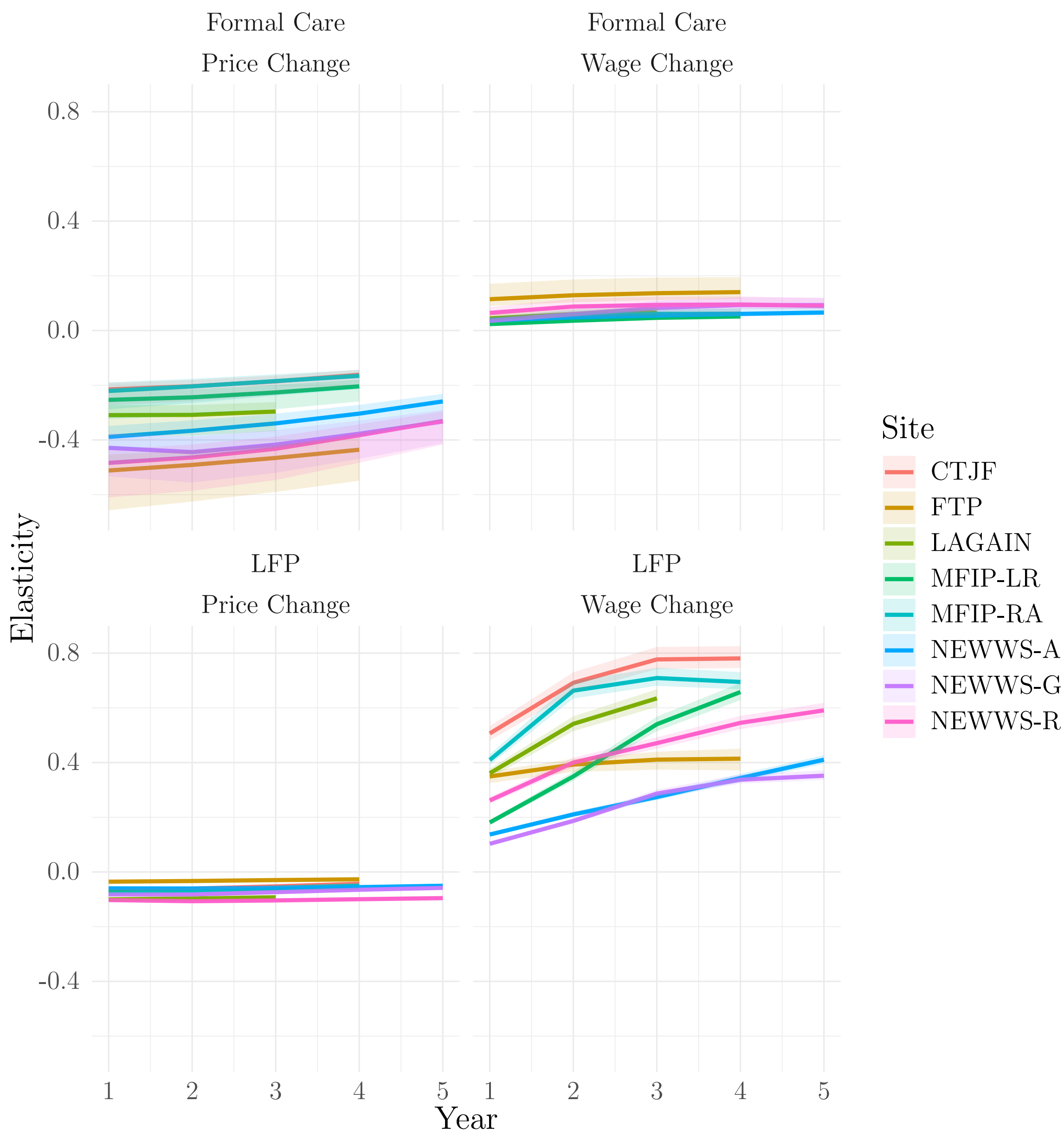
2001).

Figure 5.2 plots numerically computed marshallian elasticities of work and formal care use with respect to post-tax wages and prices. Own-price elasticities are quite significant across sites: between -.25 and -.5 for formal care, and between 0.2 and 0.8 for labor force participation. In the former case, evidence on price elasticities for care is not abundant, and so these findings provide fresh evidence that mothers are responsive in their use of paid care to changes in prices. In the latter case, the range of elasticities estimated is in line with previous findings in the literature.

5.2 Discussion of Production Estimates

Tables 5.3 and 5.4 show estimates of production parameters for Academic and Behavioral outcomes, respectively. Estimates suggest that a 1 log-point increase in resources leads to a 0.22 standard deviation increase in achievement scores (or a 0.25 standard deviation decrease in grade repetition),

Figure 5.2: Implied Elasticities to Child Care Price and Wage Changes



Notes: Figure shows the extensive marginal (Marshallian) elasticities of formal care use and labor supply with respect to changes in the price of care and post-tax wages. These are calculated by calculating counterfactual choice probabilities from a 10% increase in either price. Shaded region indicates 95% credible interval, computed by Monte Carlo simulation. Simulation for FTP missing, since work requirements imposed in both treatment and control.

and a 0.26 standard deviation decrease in behavioral problems (0.4 standard deviations for children aged 10 and above).

The quantitative significance of these magnitudes are, however, tempered significantly by estimates of δ_θ for both outcomes, which suggest that there is little evidence of any persistence in these effects. In the context of available data, δ_θ is only identified off variation in the timing of measurement and the length of exposure to treatment. It may be that richer data on the dynamics of skill formation could identify some persistence, but there is no evidence of this available in these effects.

Estimates of $\hat{g}_1$ and $\hat{g}_2$ are also quite striking. For academic outcomes, according to the posterior likelihood, there is little evidence to suggest any effect of paid or unpaid nonmaternal care on child outcomes relative to full-time maternal care. In the case of behavioral outcomes, the 95% credibility intervals encompass both negative and positive comparisons of the effects of care, with marginal evidence suggesting that both paid and unpaid care have favorable behavioral impacts compared to full-time maternal care. Finally, there is no evidence to be found in the sample treatment effects to suggest that paid care has any developmental advantages over unpaid care, with modal estimates suggesting the opposite story.

Taken together, the estimates on investment prices $\hat{g}$ marginally challenge (but do not strongly contradict) prior findings that, all else equal, maternal employment has negative impacts on development (Bernal, 2008; Agostinelli and Sorrenti, 2018; Mullins, 2019). These findings potentially warrant further investigation with individual level data from these experiments.

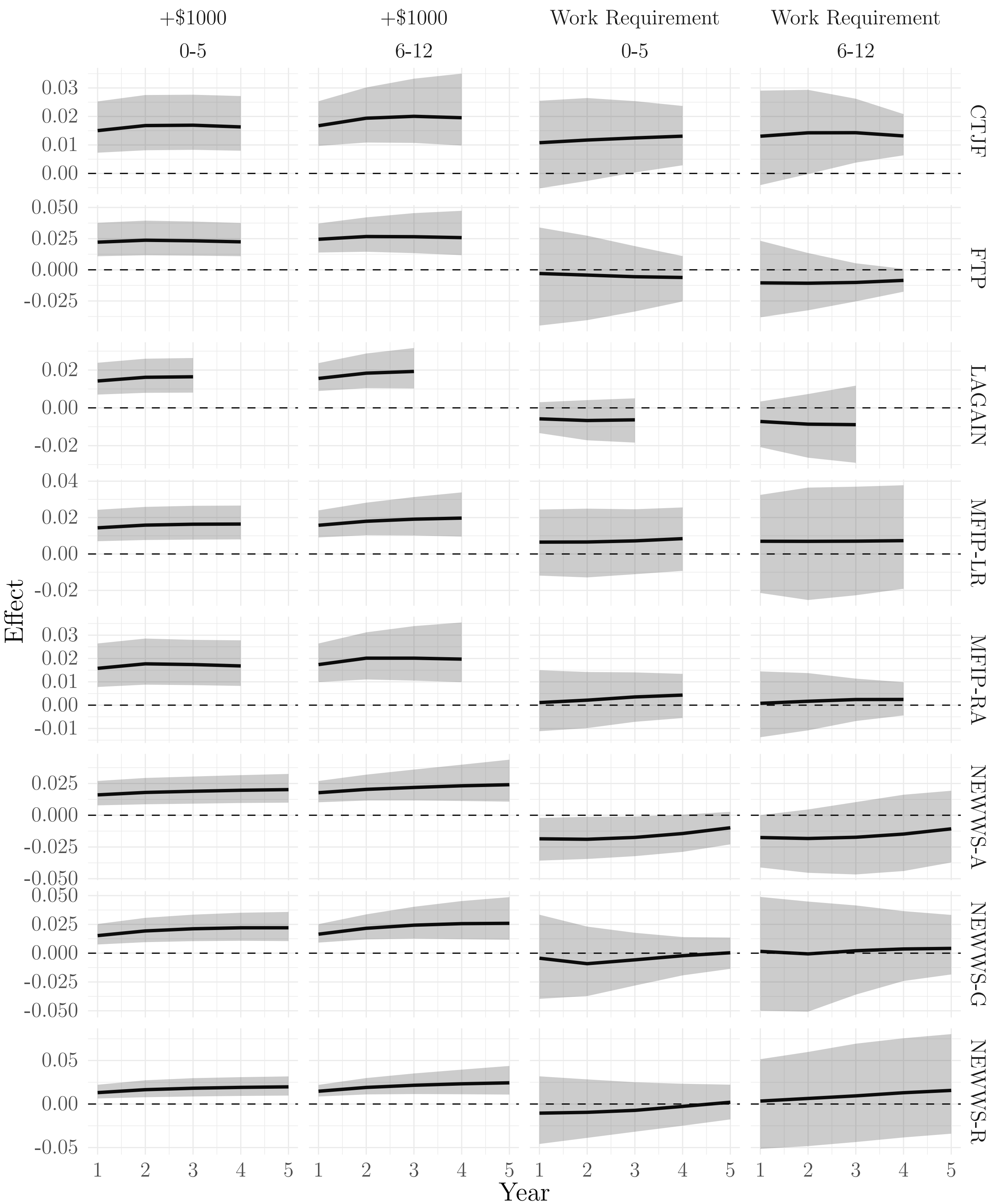
To finish analysis of these estimates, two counterfactuals can shed light on the important policy questions regarding child development that motivated this analysis. The first counterfactual simulates a \$1000 increase in family income in every period of the experiment. The second simulates again the introduction of a work requirement. The key result of interest in both scenarios is the predicted impact on child academic and behavioral outcomes. These exercises require all the parameters of the model, since both interventions are a mixture of direct effects and behavioral responses. Figures 5.3 and 5.4 show the results for children aged 0 to 5 at the beginning of the experiment, and children aged 6 to 12. The \$1000 intervention has unambiguous positive effects on academic and behavioral outcomes, though the magnitudes (around 2% of a standard deviation) are smaller than those found in prior studies (Duncan et al., 2011; Dahl and Lochner, 2012). A potential explanation for this that in constrast to prior results in the literature, these numbers incorporate a behavioral response in work and childcare behavior from mothers, since estimates here predict that, at least on average, the impacts of non-maternal care are favorable. Work requirements, on the other hand, show ambiguous effects on child outcomes. With the exception of Connecticut (where effects are positive) and Florida (where effects are negative), average impacts are not outside of 95% credibility sets. This underlying heterogeneity is a function of the fact that work requirements across sites have differential impacts on both work behavior and program par-

Table 5.3: Production Parameter Estimates - Academic Outcomes

Parameter	2.5%	25%	Estimate	75%	97.5%
$\delta_{I,0-9}$	0.10	0.17	0.22	0.27	0.42
$\delta_{I,10+}$	0.12	0.20	0.28	0.34	0.54
δ_{θ}	0.00	0.03	0.09	0.12	0.38
$g_{1,0-9}$	-0.73	-0.25	0.00	0.26	0.75
$g_{1,10+}$	-1.92	-1.24	-0.79	-0.40	0.59
$g_{2,0-9}$	-1.22	-0.66	-0.39	-0.11	0.38
$g_{2,10+}$	-0.82	-0.24	0.04	0.33	0.95
Achievement	-	-	-	-	-
Grade Repetition	0.60	0.89	1.13	1.34	2.07
WJ-Math	1.04	1.73	1.31	2.83	4.73
WJ-Read	1.05	1.64	1.14	2.66	4.09

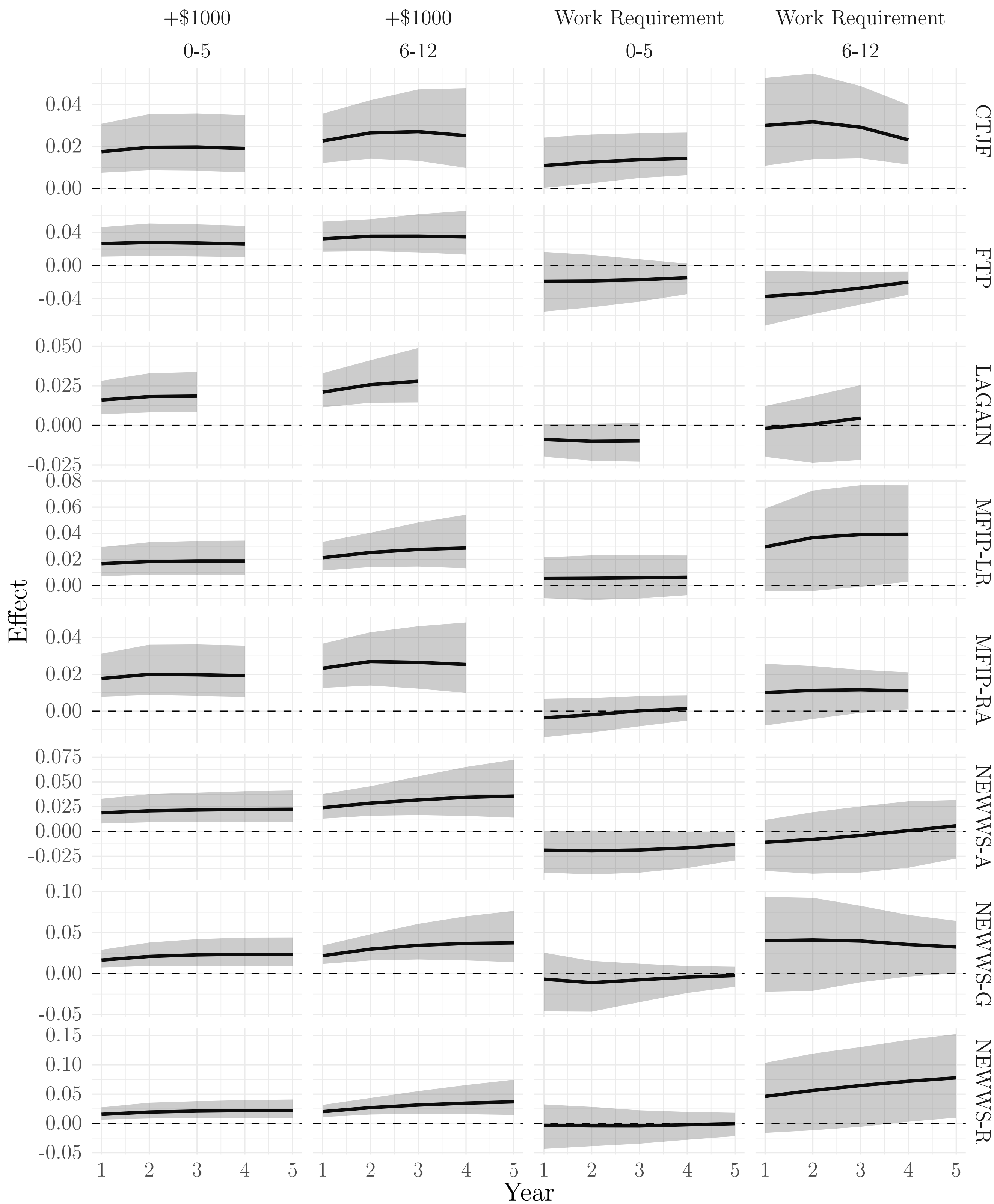
Notes: "Estimate" refers to the posterior estimate, the mode of the posterior distribution. Column x% refers to the x^{th} quantile of the posterior distribution, defining the boundaries of credibility sets.

Figure 5.3: Effects of Two Counterfactuals on Child Academic Outcomes



Notes: Shaded region indicates 95% credible interval, computed by Monte Carlo simulation. Simulation for FTP missing, since work requirements imposed in both treatment and control.

Figure 5.4: Effects of Two Counterfactuals on Behavioral Outcomes



Notes: Shaded region indicates 95% credible interval, computed by Monte Carlo simulation. Simulation for FTP missing, since work requirements imposed in both treatment and control.

Table 5.4: Production Parameter Estimates - Child Behavior

Parameter	2.5%	25%	Estimate	75%	97.5%
$\delta_{I,0-9}$	0.11	0.20	0.26	0.32	0.48
$\delta_{I,10+}$	0.14	0.27	0.39	0.47	0.81
δ_{θ}	0.00	0.02	0.08	0.10	0.30
$g_{1,0-9}$	-1.13	-0.64	-0.45	-0.23	0.13
$g_{1,10+}$	-2.87	-1.81	-1.36	-0.87	0.02
$g_{2,0-9}$	-0.48	-0.05	0.13	0.32	0.65
$g_{2,10+}$	-1.69	-1.08	-0.78	-0.47	0.03
Behavioral Problems	-	-	-	-	-
Positive Behaviors	0.99	1.51	2.15	2.60	4.24
Suspension	0.67	1.00	1.27	1.50	2.14

Notes: "Estimate" refers to the posterior estimate, the mode of the posterior distribution. Column $x\%$ refers to the x^{th} quantile of the posterior distribution, defining the boundaries of credibility sets.

ticipation. Hence, the answer to whether work requirements can increase household resources (and hence child outcomes), depends on site-level primitives that determine how these requirements are enforced, and the returns to work for individuals at each site.

6 Conclusion

This paper introduces a new approach to aggregating the results of comparable experiments by writing a structural model that parameterizes differences in treatments across sites. Taking this structural approach to meta-analysis entailed a number of conceptual advantages, including a means to explicitly leverage differences in treatment components and control group behavior for identification.

As both a proof of concept as well as a setting of inherent policy interest, this paper applied the method to study the effect of several welfare-to-work experiments on labor supply, program participation, and child outcomes. By aggregating results to estimate the model, we learned that treatment effects in the data map to quantitatively significant price elasticities in child care and labor supply use. Counterfactuals allowed the estimated the model to speak to the relative efficacy of common welfare-to-work strategies such as work requirements, time limits, and changes in financial work incentives. Of these, only changes in financial incentives (in tandem with wage growth), exhibit any permanent impact on labor force participation.

Regarding the development of children, results from counterfactuals also suggested that while household resources have significant impacts on academic and behavioral outcomes, these effects do not exhibit persistence in the data, though this persistence is hard to identify given available variation. Finally, estimates from this exercise do not suggest any detrimental effect of nonmaternal care.

The counterfactuals performed here, while sufficient to demonstrate the utility of the structural framework, represent only a relatively small subset of those made possible by the estimated model. In addition to the analysis here, the estimated model can be employed to study normative questions, such as the welfare consequences of particular program features, or the design of optimal cash assistance, as in Mullins (2019).

References

Agostinelli, F. and G. Sorrenti (2018): "Money vs. Time: Family Income, Maternal Labor Supply, and Child Development," *Working Paper*.

Arcidiacono, P. and R. A. Miller (2011): "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity," *Econometrica*, 79, 1823–1867.

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.

Bloom, D., J. J. Kemple, P. Morris, S. Scrivener, N. Verma, and R. Hendra (2000): "The Family Transition Program: Final Report on Florida's Initial Time-Limited Welfare Program," Tech. rep., Manpower Demonstration Research Corporation.

Bloom, D., S. Scrivener, C. Michalopoulos, P. Morris, R. Hendra, D. Adams-Ciardullo, and J. Walter (2002): "Jobs First Final Report on Connecticut's Welfare Reform Initiative," Tech. rep., Manpower Demonstration Research Corporation.

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

——— (2017): "Welfare dependence and self-control: An empirical analysis," *The Review of Economic Studies*, 84, 1379–1423.

Clark-Kauffman, E., G. J. Duncan, and P. Morris (2002): "How Welfare Policies Affect Child and Adolescent Achievement," *American economic review*, 93, 14–19.

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.

Duflo, E., R. Hanna, and S. P. Ryan (2012): "Incentives Work: Getting Teachers to Come to School," *American Economic Review*, 102, 1241–78.

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.

Ferguson, T. S. (1958): "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities," *Ann. Math. Statist.*, 29, 1046–1062.

Freedman, S., J. T. Knab, L. A. Gennetian, and D. Navarro (2000): "The Los Angeles Jobs-First GAIN Evaluation: Final Report on a Work First Program in a Major Urban Center," Tech. rep., Manpower Demonstration Research Corporation.

Frisch, R. (1933): "Editor's Note," *Econometrica*, 1, 1–4.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013): *Bayesian data analysis*, Chapman and Hall/CRC.

Gennetian, L. A., G. Duncan, V. Knox, W. Vargas, E. Clark-Kauffman, and A. S. London (2004): "How Welfare Policies Affect Adolescents' School Outcomes: A Synthesis of Evidence From Experimental Studies," *Journal of Research on Adolescence*, 14, 399–423.

Gennetian, L. A. and C. Michalopoulos (2003): "Child Care and Employment: Evidence from Random Assignment Studies of Welfare and Work Programs," Tech. rep., Manpower Demonstration Research Corporation.

Gennetian, L. A. and C. Miller (2000): "Reforming Welfare and Rewarding Work: Final Report on the Minnesota Family Investment Program Volume 2: Effects on Children," Tech. rep., Manpower Demonstration Research Corporation.

Gourieroux, C., A. Monfort, and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, 85–118.

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

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

Grogger, J. and C. Michalopoulos (2003): "Welfare Dynamics under Time Limits," *Journal of Political Economy*, 111, 530–554.

Hamilton, G., S. Freedman, L. Gennetian, C. Michalopoulos, J. Walter, D. Adams-Ciardullo, and A. Gassman-Pines (2001): "How Effective Are Different Welfare-to-Work Approaches? Five-Year Adult and Child Impacts for Eleven Programs," Tech. rep., U.S. Department of Health and Human Services Administration for Children and Families, Office of the Assistant Secretary for Planning and Evaluation.

Heckman, J. J. (1992): *Evaluating Welfare and Training Programs*, chap. Randomization and Social Policy Evaluation.

Heckman, J. J. and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation1," *Econometrica*, 73, 669–738.

Higgins, J. P. and S. Green (2011): *Cochrane handbook for systematic reviews of interventions*, vol. 4, John Wiley & Sons.

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

Kline, P. and M. Tartari (2016): "Bounding the labor supply responses to a randomized welfare experiment: A revealed preference approach," *American Economic Review*, 106, 972–1014.

Marschak, J. (1953): *Economic Measurements for Policy and Prediction*, John Wiley & Sons.

Meager, R. (2019): "Understanding the average impact of microcredit expansions: A Bayesian hierarchical analysis of seven randomized experiments," *American Economic Journal: Applied Economics*, 11, 57–91.

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

Miller, C., V. Knox, L. A. Gennetian, M. Dodoo, J. A. Hunter, and C. Redcross (2000): "Reforming Welfare and Rewarding Work: Final Report on the Minnesota Family Investment Program Volume 1: Effects on Adults," Tech. rep., Manpower Demonstration Research Corporation.

Morris, P., G. Duncan, and E. Clark-Kauffman (2005): "Child Well-Being in an Era of Welfare Reform: The Sensitivity of Transitions in Development to Policy Change," *Developmental Psychology*, 41.

Mullins, J. (2019): "Cash Transfers in the Presence of Children's Human Capital Formation," .

Rodriguez, J. (2018): "Understanding the Effects of a Work-Based Welfare Policy on Child Human Capital," Working Paper.

Rubin, D. B. (1981): "Estimation in Parallel Randomized Experiments," *Journal of Educational Statistics*, 6, 377–401.

Todd, P. E. and K. I. Wolpin (2005): "Assessing the Impact of a School Subsidy Program in Mexico : Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility," *American Economic Review*, 96.

A Details of Model Solution

The recursive expression for $\alpha_{V,t}$ can be found by starting with

$$\alpha_{V,T-1} = \alpha_\theta + \beta\delta_\theta\alpha_{V,T}$$

and working backwards. Using this, it is simple to take first order conditions with respect to investment and derive indirect utility, $u_t(d)$. The optimization problem can then be written recursively as:

$$V_t(\omega) = \mathbb{E}\left\{\max_d u_t(d) + \epsilon_d + \beta V_t(\omega - 1)\right\}$$

where the budget function $Y_{kjt}(H, A)$ determines consumption. The nested logit specification ensures that integration of the taste shocks can be performed analytically. When time limits do not apply, $\omega = \infty$ and the problem reduces to a repeated static choice problem.

Let $\overline{u}_t(A) = \mathbb{E}_{H,A}[\max_d u_t(d) + \epsilon_d | A]$ be the expected flow utility of making participation choice A, integrating out the conditional distribution of shocks. Define:

$$v_t(A, \omega) = \overline{u}_t(A) + \beta V_t(\omega - A)$$

as the choice-specific value. We can write:

$$v_t(1, \omega) = \overline{u}_t(1) + \beta[\overline{u}_{t+1}(0) + \beta V_{t+2}(\omega - 1) - \log(1 - \rho_t(\omega - 1))$$

and similarly

$$v_t(0, \omega) = \overline{u}_t(0) + \beta[\overline{u}_{t+1}(1) + \beta V_{t+2}(\omega - 1) - \log(\rho_t(\omega - 1))$$

Differencing these expressions and rearranging expressions for choice probabilities gives us the expression found in the main text.

B Details of Experimental Site Budgets

I will assume that program participation involves receipt of transfer payments from the program, as well as food stamps. When the age of the child is greater than 17, I assume that all only SNAP payments are made and all other policies are turned off.

Let n be the numer of children, g the state in which the program takes place, y the year, and ω the current usage of time limits. E is monthly earnings.

B.1 Control Groups

For AFDC I will simplify by assuming that eligibility coincides with the point at which benefits are equal to zero:

$$\text{AFDC}(g,n,y,E) = \max\{B(g,n,y) - (1-0.33)\max\{E-120,0\},0\}$$

here $B(g,n,y)$ is the benefit standard. Food stamp payments are:

$$\text{SNAP}(g,n,y,E) = G(g,n,y) - 0.3(0.8E + \text{AFDC} - 134)$$

B.2 MFIP

:

$$\text{MFIP}(g,n,y,E) = \max\{\min\{1.2(B(g,n,y)+G(g,n,y)) - (1-0.38)E, B(g,n,y)+G(g,n,y)\},0\}$$

where B and G are the benefit standards and maximum food stamp payment in Minnesota for each number of kids (n) and year (y).

B.3 CT-Jobs First

Here we have to model eligiblity, since no income is disregarded. Let $PG(n,y)$ be the poverty guideline for n kids in year y. This is the same as the poverty level for a family of size $1+n$. The payment is:

$$\text{CTJF}(g,n,y,E) = \mathbf{1}\{E < PG(n,y)\}(B(g,n,y) + G(g,n,y))$$

CTJF included a time limit of 12 months, which I will round to two years. I will assume that individuals who reach the time limit collect food stamps only, and are subject to the same rules as the control group.

B.4 FTP

Payments in this case are:

$$\text{FTP}(g,n,y,E) = \max\{B(g,n,y) - 0.5\max\{E-200,0\},0\}$$

Food stamps are the same as the control group.

While there is some variation in the effective time limit imposed on individuals in FTP, I assume a time limit of two years (24 months) and assume that individuals who have reached the time limit receive food stamps only, and the rules that apply are the same as the control case.

B.5 LAGAIN

Welfare payments are:

$$\text{LAGAIN}(g,n,y,E) = \max\{\min\{B(g,n,y), NS(g,n,y) - (1-0.33)*\max\{E-120,0\}\},0\}$$

where $NS(g,n,y)$ is the need standard.

C Details of Estimation

C.1 Setting Site Conditions

To calculate predicted means given a choice model parameters γ, I first specify an distribution π_0 over family size and the initial age of the youngest child, which I use as the only relevant child in the model. I use reported fractions in Gennetian and Michalopoulos (2003) for each site to set these distributions. I construct average earnings conditional on working by dividing reported unconditional means of earnings by rates of labor force participation, and feed these measures for each year into the budget functions described above in order to evaluate Y_{kjt}. In this model, means can be calculated numerically without simulation.

C.2 Specification of Priors

I specify the following prior distributions over θ_G and θ_H. For global parameters:

- β takes a uniform prior over the interval $[0.,1]$.
- α_C, σ_H, and σ_F each take a log normal prior with mean 0 and standard deviation 20.
- w_q takes a log normal prior with mean $\log(2)$ and standard deviation 1.

For hyperparameters, I specify that each $\alpha_x \sim \mathcal{N}(0,50)$ and $\log(\sigma_x) \sim \mathcal{N}(0,50)$ for $x \in \{H,A,F,R,(R,2)\}$. For production parameters I specify priors of:

- A normal with mean 0 and standard deviation 50 for all relative log-prices, $\hat{g}_\kappa$.
- Each δ_I has log normal prior with mean 0.5 and standard deviation 0.5.
- δ_θ takes a uniform prior over $[0,1]$.

C.3 Estimation Routine

To initialize the markov chain I first find the value of γ that maximizes the log of the posterior in (4.1). Sampling was performed using a Gibbs sampling algorithm in blocks over preference parameters, and I experimented with tuning of proposal parameters to get acceptance rates between 0.2 and 0.5. Estimates discussed in the paper are drawn from a posterior sample of length 50,000 after discarding the first 20,000 observations.

D Extending the Structural Method to Disaggregated Data

In this section, I demonstrate how the meta-analysis can be enriched when disaggregated experimental data is available. In this case, assume that instead of a panel of treatment and control group means at the site by treatment level, this panel is available at the individual level. Let $Y_i = \{Y_{it}\}$ denote this panel for mother i in site s and treatment group j. Instead of site-level heteregeneity in preferences and wages, we can specify a finite mixture model, with global parameters $\gamma_G = (\gamma_\delta, \beta, \alpha_C, \sigma_H, \sigma_F)$, and type-specific parameters $\gamma_k = (\alpha_{H,k}, \alpha_{A,k}, \alpha_{F,k}, \alpha_{R,k}, \alpha_{R,k2})$. Using choice probabilities outlined in Section 3.3, a log-likelihood can be written across sites s, treatment groups, j, as:

$$\mathcal{L} = \sum_s \sum_j \sum_{i=1}^{N_{s,j}} \log \left(\sum_k \pi(k|s) l(Y_i|\gamma_k, \gamma_G) \right)$$

where $\pi(k|s)$ is proportion of type k individuals at site s. Estimating this type selection at the site level allows us to explicitly estimate differences in selection across sites. Identification here follows from the availability of panel data with which to discern individual's likely type based on repeated observation of behavior. This framework can even be used to extrapolate to general populations of interest as long as representative panel data with equivalent information is available. This would be the case, for example, with data from the PSID or the NLSY. For the external sample of interest, estimates $\hat{\gamma}_G, (\hat{\gamma}_k)_{k=1}^K$ can be used to estimate type proportions π_k within this sample.

E Tables

Table E.1: Evidence for Program Impacts on Other Variables

Site	Variable	Significance Test	Significant Δ	Source
CTJF	Married with spouse	Yes	No	Bloom et al Table 5.1
	New Child or Pregnancy	Yes	No	Bloom et al Table 5.1
	0-19 Hours if Employed	Yes	No	Bloom et al Table 4.5
	20-29 Hours if Employed	Yes	No	Bloom et al Table 4.5
	30-39 Hours if Employed	Yes	No	Bloom et al Table 4.5
	40+ Hours if Employed	Yes	No	Bloom et al Table 4.5
FTP	Married with spouse	Yes	No	Bloom et al Table 4.1
	New Child or Pregnancy	Yes	No	Bloom et al Table 4.1
	Hours if Employed	No		
MFIP LR	Married	Yes	No	Miller et al Table 4.7
	Married or living with partner	Yes	No	Miller et al Table 4.7
	New Child or Pregnancy	No		
	1-19 Hours if Employed	Yes	No	Miller et al Table 4.3
	20-29 Hours if Employed	Yes	Yes	Miller et al Table 4.3
	30-34 Hours if Employed	Yes	No	Miller et al Table 4.3
	35-44 Hours if Employed	Yes	Yes	Miller et al Table 4.3
	45+ Hours if Employed	Yes	No	Miller et al Table 4.3
MFIP RA	Married	Yes	No	Miller et al Table 5.6
	Married or living with partner	Yes	No	Miller et al Table 5.6
	New Child or Pregnancy	No		
	1-19 Hours if Employed	Yes	No	Miller et al Table 4.3
	20-29 Hours if Employed	Yes	No	Miller et al Table 4.3
	30-34 Hours if Employed	Yes	No	Miller et al Table 4.3
	35-44 Hours if Employed	Yes	Yes	Miller et al Table 4.3
	45+ Hours if Employed	Yes	No	Miller et al Table 4.3
LA Gain	Married with spouse	Yes	No	Freedman et al Table 6.3
	New Child or Pregnancy	Yes	No	Freedman et al Table6.3
	0-20 Hours if Employed	Yes	Yes	Freedman et al Table 4.4
	20-29 Hours if Employed	Yes	No	Freedman et al Table 4.4
	30-39 Hours if Employed	Yes	No	Freedman et al Table 4.4
	40 Hours if Employed	Yes	No	Freedman et al Table 4.4
	$>$ 40 Hours if Employed	Yes	No	Freedman et al Table 4.4
Atlanta LFA	Married with spouse	Yes	No	Hamilton et al Table 9.2
	New Baby	Yes	No	Hamilton et al Table 9.3
	Hours if Employed	No		
Grand Rapids LFA	Married with spouse	Yes	No	Hamilton et al Table 9.2
	New Baby	Yes	No	Hamilton et al Table 9.3
	Hours if Employed	No		
Riverside LFA	Married with spouse	Yes	No	Hamilton et al Table 9.2
	Childbearing	Yes	Yes	Hamilton et al Table 9.3
	Hours if Employed	No		

Notes: Significance tests on hours categories were performed manually, by first creating probabilities conditional on working, then estimating standard errors of these sample proportions using sample size. Tests were performed at 95% significance.