

Explaining and Forecasting Results of The Self-Sufficiency Project

Christopher Ferrall

Department of Economics, Queen's University, Kingston, ON K7L 3N6 CANADA

Email: ferrallc@queensu.ca

December, 2011

This paper studies the Self-Sufficiency Project (SSP), a controlled randomized experiment concerning welfare, by estimating a model of endogenous skill accumulation, multi-dimensional job opportunities, and time-varying opportunity costs of labor market time. Methods for estimating dynamic programming models with unobserved heterogeneity are extended to account for unexpected policy interventions and endogenous sample selection and initial conditions. Parameters are identified and consistently estimated by imposing optimal responses to the exact form of the SSP earnings supplement and the experimental program, which induces exogenous variation between treatment groups and within groups as treatment progresses. The estimated model tracks primary outcomes well in and out of sample, except for under-estimating trends in the sample of new welfare applicants. Predictions from counterfactual experiments run counter to non-structural results reported elsewhere, and they suggest that details of the SSP's design are critical for interpretation of results. The separate SSP Plus treatment may have longer lasting and more generalized impacts than the in-sample impacts suggest.

Keywords: Dynamic Household Behavior, Welfare Policy, Controlled Experiments, GMM

Research support from SSHRC and computing resources from HPCVL, Westgrid and SHARC-NET are gratefully acknowledged. Hartmut Schmider (HPCVL) and Martin Siebert (Westgrid) were especially helpful with computing issues. I have benefited greatly from extensive comments from three referees and the editor. Data access is based on a research contract with Social Research and Demonstration Corporation (SRDC). The micro data used to compute the moments are available within Canada Research Data Centres through a research contract with Statistics Canada.

1. Introduction

The Self-Sufficiency Project (SSP) was a controlled randomized experiment conducted in Canada designed to study whether long-term recipients of income assistance (i.e. welfare) respond to earnings subsidies. The main SSP treatment group, single parents on income assistance (IA) for at least one year, was offered a large supplement to earnings if a full-time job was acquired and the parent went off IA within one year. The hope was that inducing sustained full-time employment would generate skill accumulation, substantial wage gains, and ultimately self-sufficiency. This expectation was partially met. About one-third of the treatment group qualified for the supplement and at the peak twice as many worked full time than the controls. However, most of the impact disappeared soon after the supplement expired. The hoped-for self-sufficiency through endogenous wage gains failed to appear. Despite this, the careful and ambitious design of the experiment provides a unique opportunity to study labour market dynamics among low-income households.

This paper provides causal evidence on the effect of labour market policies on welfare dependency. The evidence is causal in two senses. First, welfare policy is parameterized while estimating the policy-invariant (structural) parameters of a model of household behaviour. In the model the experimental treatment causes single parents to alter their labour market activity which results in complicated patterns recorded in the SSP data. The restrictions from the forward-looking model identify time-varying and heterogeneous influences on low-income single parents. The evidence is also causal in the sense that it comes from exogenous variation in the parents' environment due to controlled random assignment to treatment.

Michalopoulos et al. (2002) and Ford et al. (2003) summarize the findings from the SSP for all measured outcomes. The experiment and the model developed here were designed primarily to explain what might be termed intermediate effects of income assistance policy. These intermediate effects relate static incentives to dynamic labour market outcomes over many months but not years or decades. Miller and Sanders (1997), Fang and Silverman (2004), and Swann (2005) estimate models of welfare persistence using non-experimental data emphasizing similar issues. Keane and Wolpin (2002) and Kennan and Walker (2010) also study longer term responses to welfare, including fertility and migration, respectively.¹

The SSP subsidy tries to create a path out of the "welfare trap." The strong early response followed by long-term lack of impact among the treated group indicates the depth of the trap is

¹ While assumed fixed for this analysis, location and household size are not treated as strictly exogenous, because the distribution of unobserved household types differs by location and household size. So policy experiments using the model can account for longer run impacts in location and family size by shifting the distributions in ways suggested from other studies.

difficult to measure since it depends on how current earnings potential relates to labour market history. That is, how much would a person earn now if they had worked more and more steadily in the past? The explanation offered here for the welfare trap combines factors present in previous empirical work but not considered jointly before and not estimated using controlled randomized variation. Four aspects of the data drive the explanation that arises from the estimated model.

1. Going on and off Income Assistance. The welfare trap is not as deep for some parents as others. A large fraction of households on long-term IA eventually leave without incentives. To explain this requires churning and friction. Some transitions come from layoffs, but single parents also classify many separations as quits. To explain this the model allows the parent's opportunity cost of labour market time to vary over time to approximate shifts in child care arrangements, commuting costs and other factors. Persistent shifts in their leisure-income tradeoff induce parents to change work hours or quit altogether and move back to IA. IA status also changes without any change in employment. In the model, parents can access private outside support that, if accepted, makes them ineligible for IA. As with the value of leisure, the value of outside support also shifts over time which can effect a change in IA status without a change in labour market status.

2. Gradual Response to Static Incentives. The SSP treatment subsidizes labour market attachment now in order to raise earnings and make work more attractive than transfers in the future. The response to the SSP incentive was gradual. Given a year to qualify for the SSP, parents did so steadily. Without labour market frictions a short-term response would happen immediately if it were to happen at all. Frictions in the form of costly job search and job-specific limits on work hours play an important role in the estimated model. Non-working parents given an incentive to work full-time must search for a suitable job. Some succeed, some do not. Not all parents working part-time at the start of the treatment immediately qualify for the supplement by increasing their hours. Work hours are not fully flexible and jobs differ in wages and flexibility.

3. Lack of Dynamic Response. After the gradual short-term response, self-sufficiency failed to happen. Wages grew at roughly the same rate in both the treatment and control groups. Once the subsidy ended the treated parents returned to being on IA at a similar rates as the controls. The model explains the short-term success and long-term failure of the SSP to alter behaviour. If the model automatically related wage growth to employment it could not combine a strong short-term response in working with no long term response in wages. So a third important element of the explanation is heterogeneity in the growth potential of jobs. Jobs are characterized by hours flexibility, an initial wage offer and whether wages on the job respond to skill accumulated through learning-by-doing. Parents can qualify for the subsidy by finding and accepting full-time jobs, but they do not discriminate enough between "dead-end" and "stepping-stone" jobs.

4. Expectations and Heterogeneity. So far broad patterns in the data have been related to key elements of the model as if the patterns were separable, but the average outcomes are surrounded by complex distributions. Heterogeneity across households is important for matching these distributions. In addition, forward-looking households do not view these responses as separate. Movement on and off IA, the gradual short-term response to treatment, and the minimal long-term response are interconnected. What glues them together is parent foresight and anticipation, captured as usual by a non-zero discount factor in a dynamic program. A myopic household facing short-run incentives to work full time will do so, even if it may take time to find such a job. But a myopic parent has weak incentives to hold out for a good job that will pull them out of the welfare trap years in the future. The SSP provides experimental variation to estimate a distribution of discount factors in the population. For most single parents the road to self-sufficiency paved by the SSP is too long for them to respond to treatment in ways that eliminate the welfare trap.

One of the motivations for developing a multifaceted explanation of the SSP is to apply it outside the experiment. Subsequent research can use the estimated model to study a larger set of counter-factual policies and experimental designs. The applications reported in this paper focus on the experiment itself. Efficient GMM estimates of the model from the first 36 months of experimental data are used to forecast out-of-sample for the last 18 months of data. Several experiments are run at the estimated parameters to ask whether the impacts observed in the SSP are robust to modest changes in the design of the experiment. In some dimensions the parameters of the experiment do not have a great effect on the model outcomes. In other dimensions the results are sensitive to the details.

This paper also considers the role experimental variation plays when using a behavioural model to explain the data. Such variation can be used for estimation or validation of the model after estimation. To quantify the tradeoff between these two uses of the data, standard errors are re-computed at the estimates as if data on only treated or control groups were available. The size, and in some cases the direction, of change in the standard errors is surprising. In the case of this model and the SSP experiment, not using the treatment group in estimation would have severely restricted identification of the parameters.

The next section describes the SSP experiment and key elements of the model with minimal notation. Details of the model are provided in section 2 and of the estimation procedure in section 3. Results and implications are described in section 4 followed by a conclusion in section 5. The extensive online Supplement provides additional details of the data, the model, the solution method, parameter identification, and outputs. Material in the supplement is indicated with an S prefix.

2. Overview

2.1 Sample Selection and Treatment in the SSP

The static tradeoffs between income and leisure created by the welfare system are illustrated in Figure 1. The details apply to Canadian income assistance programs in the 1990s but the static aspects are similar across other jurisdictions. A non-working single parent is eligible for income assistance benefits in amount $\$IAB$ as long as they do not take forms of non-government outside support, with value $\$OS$, that make them ineligible for welfare. For example, if the parent cohabitates or marries basic benefits fall substantially, possibly to $\$0$.² A working parent can set aside earnings up to $\$SA$ each month without a reduction in benefits. Thereafter benefits are replaced by earnings. The result is an incentive to work few or no hours (point A), especially if child care and commuting time are costly (a steep indifference curve).

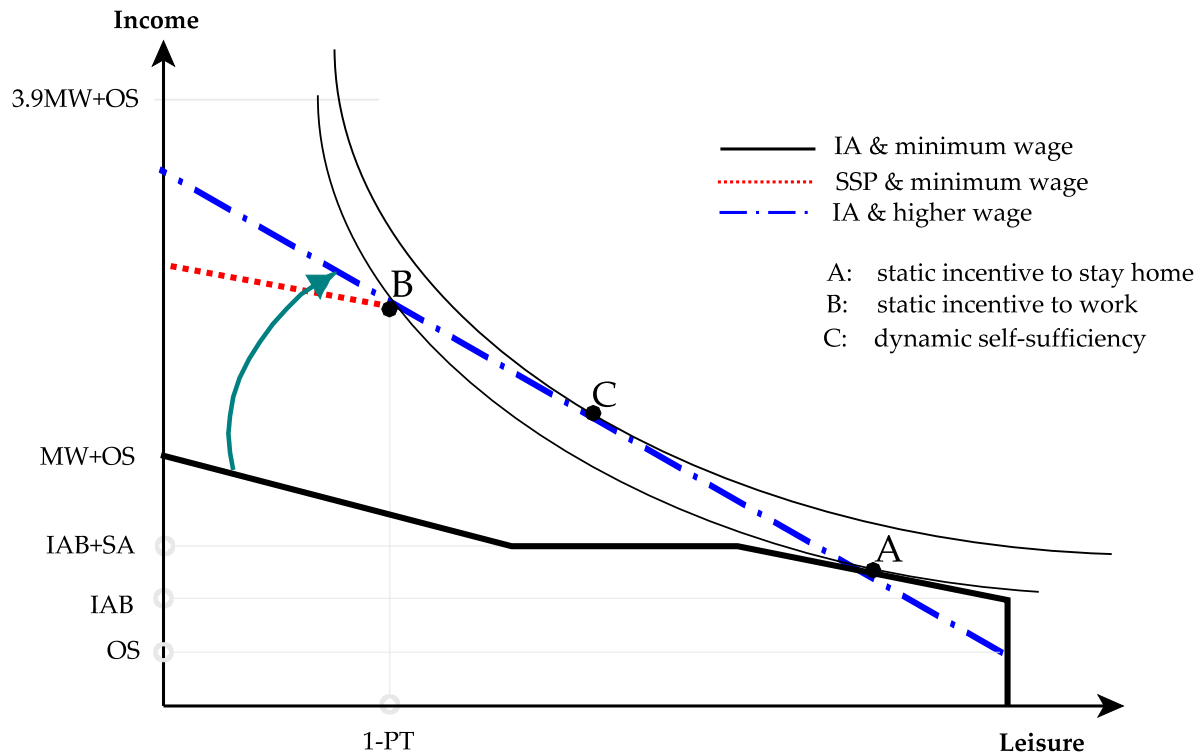
Figure 1 also displays the main SSP treatment. A treated parent retains the option of taking IA in any month, but this precludes receiving the SSP supplement. The SSP budget has slope equal to the wage until reaching full-time work hours. Then the supplement becomes available, shown as the dotted line segment and computed as half the vertical difference between earnings and 3.9 times full-time minimum wage earnings, $\$MW$. This particular parent is indifferent between staying on IA at point A (still a feasible choice during treatment) and working full time with the supplement at point B. This is the short-term response discussed in point 2 of the introduction.

The wage subsidy increases government transfers in order to subsidize full-time work with the premise that sustained employment induces wage growth through learning-by-doing. This dynamic response is shown as an arrow to a mixed (blue) line in Figure 1, the non-subsidized budget under a high wage job. After treatment ends the supplement budget disappears but the new optimal choice is C. The parent can balance work and home without assistance through part-time hours at a high wage rate. The temporary SSP treatment has eliminated the welfare trap through employment-induced wage growth.³ But if the wage underlying point B does not respond to experience then the shift to the new budget does not occur, and the parent moves back to A. Which households qualify for the SSP and which ones experience wage growth depends on

² Citing early U.S. research, Moffitt (1992) concludes, "Most exits from AFDC are not a result of an increase in earnings by the female head, but are instead the result of a change in marital status that results in the loss of AFDC eligibility."

³ Keane and Moffitt (1998) use an estimated model of income maintenance programs in the U.S. to predict that reforms without such a full-time work requirement would significantly increase total transfer payments to poor households and shift many away from full-time work.

Figure 1.
Earnings, Income, & Work Hours under IA and the SSP



Notes: IA equals $\$IAB$ when earnings are below $\$SA$, thereafter reduced 1-for-1, discouraging full-time work (tangency A). Dynamically, this inhibits skill growth and keeps wages low. The SSP treatment subsidizes full-time work (corner solution B). It requires a parent to forgo IA, which allows them to accept ineligible transfers at value $\$OS$. If the wage responds, the budget rotates out endogenously, making IA less attractive once treatment ends (tangency C). The estimated model allows the parent to hide a fraction of income from the authorities thereby increasing the slope of the budget. In the figure a full-time job is available, which is just one of several possible states in the model. Without a full-time job the SSP subsidy is unavailable; with no job the household's static budget is flat.

heterogeneity in skills, wage growth and patience (points 3 and 4 in the introduction).

Starting with the static treatment in Figure 1 several separate experiments occurred within the SSP. First, the SSP treatment depends on minimum wages which differed between provinces. IAB (in Figure 1) depends on province and household size, and we distinguish between one and two-or-more children households. Second, parents were selected in two different ways. The Recipient Study ran in both British Columbia (BC) and New Brunswick (NB). The Applicant Study ran only in British Columbia. Within the New Brunswick Recipient Study a related treatment called SSP Plus (or SSP+) included job search counseling and other support to leave IA. Since SSP and SSP+ share a single control group, there are 4 distinct treatment groups and 3 distinct control groups across studies. That brings the total number of distinct groups subject to policy variation equal to $7 \times 2 =$

14.

Treatment took place in three phases, as shown in Figure 2. To be eligible for random assignment in the Recipient Study the parent must have been on IA for twelve or more months, shown as a subset of single parent households.⁴ Treated households had twelve months to qualify for the supplement by accepting a full-time job and going off IA. Figure 2 puts the treated group on a timeline at month $t = 0$, the start of the twelve month qualifying phase. If the parent qualified they moved immediately to the eligible phase that lasted 36 months. A parent who had not gone off IA and started a full-time job within twelve months failed to qualify and their treatment ended. Otherwise, while eligible the household could return to IA and/or stop working full-time. They received no supplement in such months, and the 36-month clock continued to count down. But once working full time without IA receipt the supplement resumed. Thus the full SSP budget constraint in Figure 1 applies during the eligible phase, including the segment with IAB. After 36 months of eligibility the household returned to the status quo. The separate SSP+ treatment group followed the same schedule as the Recipient Study, so it shares the control group with the regular SSP.

A wage subsidy offered to long-term IA recipients might induce recipients to stay on IA longer so as to qualify. This entry effect would mitigate any impact observed in the experiment. Direct evidence for this effect was sought by conducting the Applicant Study. To be eligible for it a single parent had to initiate a new IA spell after six or more months away from IA. This is also graphically represented in Figure 2 as a subset of states for the same population. The Applicant and Recipient sets do not overlap because a household could not be eligible for both in a given month (at least six months separate eligible states for a single household). Treated households in the Applicant Study were told (truthfully) that they would enter the qualifying phase if their new IA spell lasted twelve months. They are assigned to phase 1: waiting to be allowed to qualify. Since the entry point for the Recipient Study was designated $t = 0$, applicants began treatment at $t = -11$. Treatment ended if they went off IA before time 0; otherwise for $t \geq 0$ treatment was the same as in the Recipient Study.

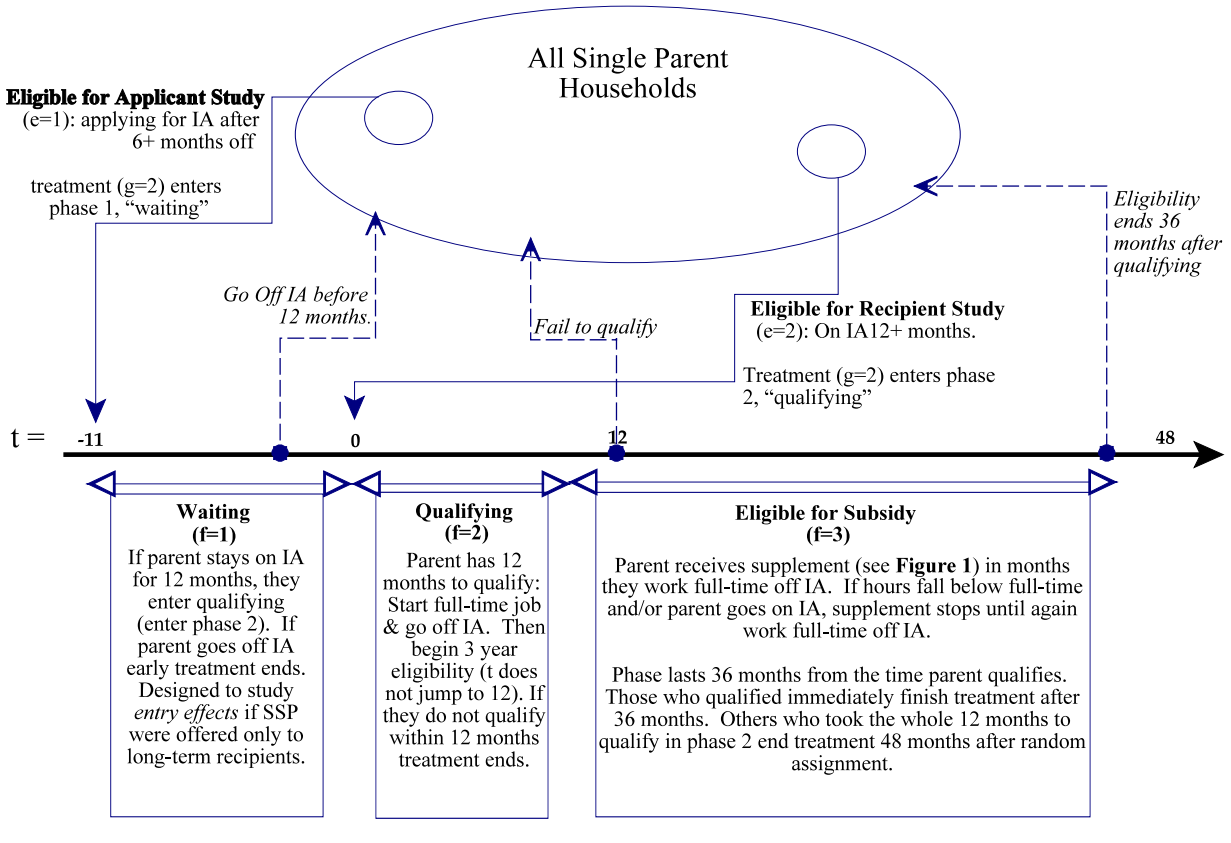
2.2 Innovations for Analyzing Welfare and Experimental Data

The rich variation from the experiment is used to sort among several factors associated with the welfare trap. In the model the parent chooses each month whether to accept IA or not. If

⁴ This simplifies the actual condition for the Recipient Study: on IA for 12 out of the last 13 months. Kamionka and Lacroix (2003) assess non-response bias in the SSP. A possible extension to this model makes participation a choice related to the expected value of treatment relative to control. This was not included here in order to devote computational resources to correcting for endogenous selection and modeling wages and job search as realistically as possible.

Figure 2.

Sample Selection and Program of Treatment



not, they forgo welfare stigma and can accept non-government transfers that preclude IA (\$OS in Figure 1). while employed the parent chooses hours of work including zero hours (quitting). While unemployed they choose whether to search actively for a job or not. Given choices and current state, the household's state next period is stochastic. It is affected by market shocks (appreciation and depreciation of skills and job offers), internal shocks (opportunity costs of labour market time), and external non-market shocks (the level of outside support).

Parents search for jobs that differ in hours flexibility, initial wages, and wage growth. Skills accumulate stochastically while working but they can also depreciate while inactive. Person-specific skill, job quality, and minimum wage levels interact in determining current wages and potential future wage growth (e.g. Gladden and Taber 2009). In short, there are multiple reasons why households end up on IA for long periods of time even while some apparently similar households come and go.

Unobserved heterogeneity is both permanent (parameters of the dynamic program) and transitory (hidden state variables). A finite mixture over types is estimated. As in Eckstein and Wolpin

(1999) types differ in many dimensions and there is no interaction between exogenous observed characteristics and type-specific utility parameters. As in Ferrall (1997) the mixture proportions differ across groups, and a discount factor specific to each type is estimated. Efficient GMM estimates are obtained from means of variables conditioned on exogenous values, including time since randomization, experimental group and exogenous household characteristics. As detailed in section 2.6, eligibility for random assignment is explained with the same model as post-assignment observations. To focus on month-to-month decision making while addressing selection and incomplete histories, the household's environment outside the experiment is assumed to be stationary.⁵ Within this stationary environment the finite phases of treatment create controlled exogenous variation in both budget constraints and *expectations*. For example, the "waiting" and "qualifying" phases in Figure 2 affect only expectations about contingent future utility. If single parents were completely myopic then the distributions for treated parents in these phases would be identical to their control counterparts. Instead, observed differences between these groups help identify the patience of single parents. Identification of discount factors is discussed further in section 3.2.

3. The Model

3.1 The Dynamic Program

The model of household behaviour is a discrete choice dynamic program. All parameters and state variables associated with the program are contained in a generalized state/parameter vector denoted θ . The current action of the parent is a vector α , chosen from feasible actions $A(\theta)$. The combination (α, θ) , referred to here as the outcome, is the argument for the one-period return or utility, $U(\alpha, \theta)$, and the stochastic transition to next month's state, $P\{\theta' | \alpha, \theta\}$. The combined state/parameter vector tracks different types of households assigned, unexpectedly, to a finite period of treatment.⁶ It also tracks all estimated parameters of the dynamic program. Specifically, θ is composed of sub-vectors that group parameters and state variables by their role in the analysis:

$$\theta \equiv (\theta_{\text{clock}} \quad \theta_{\text{exp}} \quad \theta_{\text{end}} \quad \theta_{\text{exog}} \quad \theta_{\text{pol}}). \quad (1)$$

⁵ *A priori* the SSP has no permanent effect in the model in the sense of lasting for eternity. But nothing stops the treatment effects from lasting, say, 40 years on average. This would be the stationary equivalent to a permanent lifetime effect. Further, several unobserved state variables are free to jump in value often or almost never. Lifecycle features such as children reaching school age can be approximated by a low probability of shifts in labour market time without tracking ages. That is, a shift in the leisure-income tradeoff is expected every sixty months this would be a stationary approximation to a five year period of child care.

⁶ Many of the technical elements of how the model is constructed are available in the Supplement. In addition, Ferrall 2003 provides a complete discussion of the underlying framework.

The **endogenous** state of the household outside the experiment is stored in θ_{end} . In a non-experimental model this would be the whole state vector. The two leftmost components of θ are added to track the experiment as illustrated in Figure 2. The vector $\theta_{\text{exp}} = (g \ e)$ tracks the household's **experimental** group using indicators for **entry point** (Applicant or Recipient) and **treatment group** (control, treated, SSP+ treated). For the treated groups $\theta_{\text{clock}} = (r \ f)$ **clocks** progress through **treatment phase** (f) and **months residing in the current phase** (r). **Exogenous** (estimated) parameters of the problem appear in θ_{exog} . **Policy** values (such as IAB and MW introduced earlier) appear in θ_{pol} .⁷

Using discount factor δ (an element of θ_{exog}), the value of a state θ and an outcome (α, θ) satisfy the stationary infinite horizon Bellman's equation:

$$\begin{aligned} \forall \alpha \in \mathbf{A}(\theta), \quad v(\alpha, \theta) &\equiv U(\alpha, \theta) + \delta E[V(\theta')] \\ &= U(\alpha, \theta) + \delta \sum_{\theta'} P\{\theta' | \alpha, \theta\} V(\theta') \end{aligned} \quad (2)$$

$$\forall \theta, \quad V(\theta) = \max_{\alpha \in \mathbf{A}(\theta)} v(\alpha, \theta). \quad (3)$$

The conditional probability of choosing a particular action is either 0 or 1 (ignoring exact ties) because the model does not include *ex ante* additive IID choice-specific shocks to utility, such as the convenient choice of extreme value shocks in work following Rust (1994). Here all randomness in the model relates to the transitions of non-IID state variables. Choice probabilities are therefore smoothed *ex post* using the logistic kernel in choice-specific values with parameter $\rho > 0$:⁸

$$\begin{aligned} \tilde{v}(\alpha, \theta) &\equiv \exp\left\{\rho[v(\alpha, \theta) - V(\theta)]\right\} \\ P\{\alpha | \theta\} &= \tilde{v}(\alpha, \theta) / \sum_{\alpha^* \in \mathbf{A}(\theta)} \tilde{v}(\alpha^*, \theta). \end{aligned} \quad (4)$$

Combining smoothed choice probabilities in (4) with exogenous outcome-to-state transitions (defined in supplement equation S12) generates the state-to-state transition, denoted $P_s\{\theta' | \theta\}$. This optimal transition is computed for status quo and for every stage of the experimental treatment.

3.2 State and Decision Variables

Within the framework described above, a household's situation outside the experiment is described by nine state variables contained in the endogenous vector:

$$\theta_{\text{end}} \equiv (l \ p \ n \ x \ b \ s \ h \ d \ k). \quad (5)$$

⁷ The values of policy parameters and the model of treatment phases are described in S6.2.3.

⁸ Eckstein and Wolpin (1999) solve a dynamic programming model using Monte Carlo on jointly normal IID error terms. They also smooth choice probabilities *ex post* with the logistic kernel.

Reading from right to left, k is the parent’s unobserved type and d its observed demographic group. These two values vary for a household over time. The next five variables are the key time-varying states. Briefly they are indices for: the opportunity cost of time outside the household; the level of available non-governmental or family support; the upper bound on hours in the current job (0 if not employed); the parent’s skill based on previous experience; and the earnings offer in the current job. The two leftmost variables, l and p , do not enter the utility or transition. They are extra state variables tracked solely to match SSP results and identify key parameters directly from the data. Namely they indicate the parent worked in the previous month and lost their previous job.

Each of the $D = 4$ demographic groups (province and household size) has a vector of policy parameters, $\Psi_p[d] \equiv (IAB_d \quad MW_d)$, illustrated in Figure 1. Within demographic group, unobserved type k is distributed across $K = 4$ types according to $\Lambda[d]$, a vector with elements $\lambda[d, k]$.⁹ For example, $\lambda[1, 2]$ equals the proportion of type 2 households in group 1 (New Brunswick, one child) households.

The action vector has three variables,

$$\alpha \equiv (m \quad a \quad i), \tag{6}$$

representing labour market hours, active job search, and acceptance of income assistance. The i and m choices are observed but active job search is not. The feasible set $\mathbf{A}(\theta)$ imposes two restrictions. First, active job search while working is ruled out: $m > 0$ or $a = 1$, but not both.¹⁰ Second, the parent faces an upper bound on work hours: $m \leq u(b)$ where b is specific to the current job. When the parent has no job ($b = 0$) work is not available, and $u(0) = 0$. With a part-time job ($b = 1$) they can only work less than $u(1) = \text{PT} = 75\%$ of full-time hours. When holding a full-time job ($b = 2$) the parent can chose to work any number hours ($u(2) = \text{FT} = 100\%$). A parent with a job who sets $m = 0$ quits and loses the option to work until a new job is offered and accepted.

3.3 Utility, Skill and Wages

⁹ There is a computational tradeoff between more distinct types (greater K) and allowing more parameters to vary across types. Initially K was set to two and only a few parameters differed with k in order to locate reasonable values. An arduous process followed of alternately iterating on the objective then adding elements to the model to address major issues. For example, the bound on hours was added so that not all part-time workers are predicted to immediately qualify for the SSP by switching to full-time. A final specification was chosen that allowed nearly all parameters to vary by type. Only parameters related to unobserved household costs and outside support were held common across type. This flexibility dictated a low value of K , with 4 being at the limit of available computational resources.

¹⁰ Job-to-job transitions are treated as the same job, so passive search while working is allowed. The model attributes growth in full-time equivalent earnings between contiguous jobs as skill acquisition.

Utility equals income plus outside support minus the opportunity cost of labour market time:

$$U(\alpha, \theta) = \text{Income}(\alpha, \theta) + \text{OS}(\alpha, \theta) - C(\alpha, \theta). \quad (7)$$

In turn, income is the sum of earnings, income assistance payments, and SSP payments:

$$\text{Income}(\alpha, \theta) \equiv \text{TrueEarn}(\alpha, \theta) + \text{IA}(\alpha, \theta) + \text{SUP}(\alpha, \theta). \quad (8)$$

The components of income are defined as:

$$\begin{aligned} \text{TrueEarn}(\alpha, \theta) &\equiv mW(\alpha, \theta) \\ \text{Earn}(\alpha, \theta) &\equiv (1 - \beta i) \text{TrueEarn}(\alpha, \theta) \\ \text{IA}(\alpha, \theta) &\equiv i \max \left\{ \text{IAB} - \left(\text{Earn}(\alpha, \theta) - \min \{ SA, \text{Earn}(\alpha, \theta) \} \right), 0 \right\} \\ \text{SUP}(\alpha, \theta) &= \mathcal{B}[f = 3 \ \& \ i = 1 \ \& \ m > \text{PT}] \max \left\{ 0, \frac{1}{2} [3.9 \times \text{MW} - \text{TrueEarn}(\alpha, \theta)] \right\}. \end{aligned} \quad (9)$$

True earnings equal work hours (as a fraction of full-time work) times the full-time equivalent wage $W(\alpha, \theta)$ defined below. When a parent is on IA ($i = 1$) earnings in the data and reported to the welfare authorities are a fraction β of TrueEarn. Underreporting of income is encouraged by the 100% implicit tax above the set aside amount, and allowing for it in (9) helps explain the sizeable fraction of households working while receiving IA. $\text{SUP}(\alpha, \theta)$ includes the algebraic version of the supplement illustrated in Figure 1 with a factor in front that indicates the person is eligible. Here $\mathcal{B}[x]$ denotes the Boolean (0/1) value of x (often referred to as the indicator I_x). So the first term in SUP is one if the parent is in the eligible phase, is off IA, and is working fulltime.

Non-government transfers and additional utility (in dollar equivalent) from forgoing IA equals

$$\text{OS}(\alpha, \theta) \equiv (1 - i)s \left[\xi \text{IAB} \right]. \quad (10)$$

The transfer component of OS is support from others that, if accepted, disqualifies the parent for IA. Marriage and cohabitation (detected by the welfare authorities) are both implicitly included in OS. Outside support varies from month to month based on the endogenous variable s . When s changes the parent may go off welfare and rely on other sources of support with or without any change in labour market status. A drop in s may push the parent back to receiving IA.

As with many job search models, utility is linear in income and separable in leisure. But unlike basic job search models $C(\alpha, \theta)$ is time-varying and non-linear in hours, resulting in shifting static preferences for work hours. To keep time-costs, $C(\cdot)$, in the relevant range during estimation it is expressed as a fraction of maximum possible earnings for this type of person, W_{\max} , which is defined below. Specifically, the cost of labour market time,

$$C(\alpha, \theta) = W_{\max} \nu [m + \kappa a]^{c(h)}, \quad (11)$$

depends on work hours and search when not working, which is converted to work time by the exogenous parameter κ . These preferences explain changes in hours and related movement on and off IA as highlighted in the introduction. The curvature is determined by $c(h)$ which shifts with the state variable h . This functional form is explained in detail in S6.2.2 and can be viewed as monetary costs of child care, commuting, and other employment costs. Non-linearity allows for non-market options for child care including school time and care by relatives.¹¹

3.4 Skill, Job Search and Wages

$W(\theta)$ denotes full-time equivalent monthly earnings. Wages are determined by worker skill (x), job offer index (n), and minimum wage MW. Skill takes on four values, $x = 1/4, 1/2, 3/4, 1$. From month to month x either remains constant, accumulates with probability $m\pi_a$ while working or decreases with probability π_d while not working. If $\pi_a = \pi_d = 0$, endogenous skill accumulation and depreciation are eliminated, and x becomes a permanent random effect for the parent in addition to their fixed type k . The standard Mincer earnings function essentially assumes $\pi_a = 1$ and $\pi_d = 0$. That is, skill accumulates log-linearly and deterministically with experience and does not depreciate when not working. This rules out a suspected culprit for the welfare trap: the decay of potential wages due to a temporary absence from the labour force which then prolongs the absence. When $\pi_a < 1$ and $\pi_d > 0$ transient conditions can persist. Even when the parent is forward-looking and sees the skill investment value of work, the rate of wage growth and/or their degree of patience may be too slow to make work pay.

To see how $W()$ goes beyond Mincer in this and other ways, start with the case MW = 0, for which $W()$ collapses to $W^0()$:

$$W^0(\theta) \equiv \begin{cases} \exp\{\mu + \sigma\Phi^{-1}(n/6) + \eta x\} & \text{for } n > 0 \text{ with probability } (1 - \pi_m)/5 \\ 0 & \text{for } n = 0, \text{ with probability } \pi_m. \end{cases} \quad (12)$$

$\Phi()$ is the standard normal distribution. Discrete job offers are indexed by n , which takes on values 0 to 5. Conditional on an offer, real (positive) offers occur when $n > 0$ which happens with probability $1 - \pi_m$. The offers are derived from a log-normal distribution with type-specific parameters μ and σ . This distribution is discretized by associating n with the $100(n/6)^{th}$ quantile of the log-normal distribution. Worker quality also contributes to the wage through the parent's current acquired skill, x , and the estimated return η . The offers with index $n = 0$ are jobs that pay nothing absent a minimum wage, but it will be assumed if taken the parent could still increase x while working at

¹¹ The range of $C(\alpha, \theta)$ is $[0, W_{\max}]$ and for $OS(\alpha, \theta)$ is $[0, \xi IAB]$. The forms ensure that the state-dependent values stay within their relevant ranges while exogenous parameters are varied during estimation. Otherwise current values of ξ and ν can end up in regions where choice probabilities are flat.

such a job.¹² The relatively simple form in (12) does not account for the distortion of a minimum wage. It is generalized to move the mass point from 0 to $MW > 0$ and distort the underlying offer distribution. Loosely speaking, the fully wage function assumes workers whose productivity is below MW are first overpaid and then, as x increases, they are under-paid. This captures some elements of a bargaining or contracting model of minimum wages such as Flinn (2006). Certain low-quality job/worker pairs pay MW .

Offers of type $n = 0$ now pay MW regardless of skill. The existence of such jobs inhibits but does not rule out self-sufficiency.¹³ Let ϕ_x denote the fraction of the underlying log-normal offer distribution below MW :

$$\phi_x = \Phi \left[\frac{\ln(MW) - \eta \ln(x) - \mu}{\sigma} \right]. \quad (13)$$

For $x = 1/4$ it is assumed the lowest two regular offers ($n=1,2$) produce a wage of MW . Each offer occurs with probability $(1 - \pi_m)\phi_x/2$. For the next skill level ($x = 2/4$) only the $n = 1$ offer pays MW and it occurs with probability $(1 - \pi_m)\phi_x$. For greater skill levels no wages other than $n = 0$ are at the minimum wage. So $W(\theta) = MW$ if any of three mutually exclusive indicators are true at θ :

$$M(n, x) = \mathcal{B}[n = 0] + \mathcal{B}[x = 1/4 \ \& \ 1 \leq n \leq 2] + \mathcal{B}[x = 2/4 \ \& \ n = 1]. \quad (13)$$

That is, $M(n, x)$ indicates the parent currently works at minimum wage. Such jobs are heterogeneous in their growth potential based on both the job characteristic n and the person characteristic x . Otherwise the wage exceeds MW , and each such offer is equally likely for a given x . Using $\tilde{n}(x) = 3 + \mathcal{B}[x > 1/4] + \mathcal{B}[x > 2/4]$ to denote the number of offers above MW for skill level x , the general expression for full-time earnings that interacts skill, job quality and minimum wages is:

$$\begin{aligned} W(\theta) = & M(n, x)MW + \\ & (1 - M(n, x)) \left(x^\eta \exp \left\{ \mu + \sigma \Phi^{-1}(\phi_x + (1 - \phi_x)/\tilde{n}(x)) \right\} \right) \\ W_{\max} \equiv & \exp \left\{ \mu + \sigma \Phi^{-1}(\phi_1 + (1 - \phi_1)/5) \right\}. \end{aligned} \quad (14)$$

Although this formula may seem arbitrary, it is (ex post) easy to explain and fairly intuitive. For high x only $n = 0$ offers start at MW and all other offers provide wage growth with each increment of skill (until $x = 1$). The lower a person's skill while searching the more likely they are to be

¹² With nine state variables the number of values any particular variable can take on is limited. A balance was struck with n having the most distinct values (see Table). Otherwise the predicted wage distribution would be very coarse making it difficult to pin down μ and σ .

¹³ This assumption allows even low-skill jobs to transmit good habits such as showing up on time. But to take advantage of such learned skills the parent must quit the job and find another with $n > 0$. This feature helps explain why large fractions of all groups are on minimum wages throughout the sample period.

offered MW initially and the longer they will stay at MW even if employed steadily. Regardless of job quality, skills do accumulate (stochastically) and eventually steady work results in high skill. But some “dead-end” jobs provide no direct return to skill, and it may be optimal to quit them to search for a better job. In addition, by fixing the job indices that pay MW for each skill level but allowing offer probabilities to differ across n , (14) has the important property that it is continuous in the estimated parameters μ and σ .

3.5 SSP Plus and Offer Probabilities

The SSP Plus treatment ($g = 1$) offered the supplement plus employment services. This additional treatment is not represented in Figure 1 which presumes a job is available. The model allows that employment services enhance active search by raising the job offer probability. Ordinarily a month spent in active search ($a = 1$) generates an offer with probability π_j , a type-specific estimated parameter. The effectiveness of the SSP+ treatment is captured by another estimate parameter, π_p . Then, in general, the offer probability is a function of the outcome:

$$p_j(\alpha, \theta) = \begin{cases} a(\pi_j + \pi_p(1 - \pi_j)) & \text{if } g = 1 \text{ \& } 2 \leq f \leq 3 \\ a\pi_j & \text{otherwise.} \end{cases} \quad (15)$$

The treatment parameter determines how much the services improve the chance of a job offer. If $\pi_p = 1$ then any month of active search generates an offer ($p_j = 1$). Holding constant state-contingent choices, parents in SSP+ treatment receive more offers than in the regular treated group, but they can also respond to SSP+ employment services by searching actively in states and reject offers they would not have otherwise. The model is thus restrictive by tying all predicted differences between SSP and SSP+ treatment groups in New Brunswick to a single type-specific parameter.

3.6 Endogenous Eligibility for Random Assignment

If the SSP had selected single parents at random the initial conditions for both the treatment and control groups, the experimental results would reflect the distributions across unobserved states for the target population. Instead, as illustrated in Figure 2, both the Recipient and Applicant Studies selected parents conditional on past receipt of IA. The population eligible for random assignment is therefore endogenous to the policy and the behaviour being studied. The exogenous share of type k in group d , $\lambda[d, k]$, does not equal the share of k eligible for random assignment. And the distribution across endogenous states is different at random assignment than in the general population. Correcting parameter estimates so that they apply to the wider population of single-parent households, not just the select subsets illustrated in Figure 2, is essential for policy relevance.

To make controlling for endogenous eligibility feasible the environment is made stationary outside of treatment.¹⁴ The state-to-state transition, $P_s \{ \theta' | \theta \}$ in Supplement equation (S13),

¹⁴ To my knowledge no previous research has used a model of forward-looking agents to deal

combines the primitive transition and smoothed optimal choice probabilities. Following Ferrall (2003) the primitive transition is specified so that, outside of treatment, the state-to-state transition is ergodic. That is, a unique stationary distribution over θ_{end} exists for all exogenous parameters θ_{exog} in the interior of the parameter space. Let $P_{\infty}\{\theta\}$ denote this distribution. Having solved the infinite horizon problem (2)- (3), this distribution is computed exactly over the discrete state space for each combination of demographic group d and unobserved type k . From this distribution eligibility criteria for the Applicant and Recipient Studies are applied sequentially.

Consider first the Recipient Study, which required at least twelve months of IA receipt ($i = 1$). Only choices in the outside world with $i = 1$ keep a parent eligible, resulting in choice probabilities that generate a state-to-state transition different than $P_s\{\theta' | \theta\}$. Starting from $P_{\infty}\{\theta\}$ this transition is imposed, resulting in a new distribution across endogenous states, representing households in a cross-section that have been on welfare at least one month. This transition does not preserve the size of the population as it drops households off IA ($i = 0$). The amount of leakage is recorded, and then the mass is rescaled to 1 to continue selection (equation S15). The conditional transition is applied again to form the distribution of households on welfare at least two months. Repeated twelve times, this produces the distribution of households eligible for random assignment in the Recipient Study. This sequential conditioning is repeated for all k in group d . Using the total leakage over the 12 selection months the share of type k eligible for assignment to treatment is computed (equation S16) and used in generating post-assignment predictions (section 3.2). For the Applicant Study the procedure is carried out for its seven months of criteria and implied transitions: a parent must be off welfare ($i = 0$) for six months then initiate a new claim ($i = 1$).

Thus the model accounts not just for responses to treatment but for selection on unobservables, both time-varying and time-invariant. The distribution across skills, job offers and household costs is different than in the general population of single parents and unique to each study and demographic group. The mix of permanent types eligible for the experiment is not the same as the exogenous (and estimated) type proportions in the general population. No ad hoc auxiliary model is used to account for initial conditions. Smoothed choice probabilities and computation (not simulation) of the sequence of distributions across the finite state space maintains continuity of predictions in the estimated parameters.

4. Estimation

with unobserved histories and endogenous selection. Attanasio et al. (2011) estimate a model of educational attainment using a cross-section of data from the Progres experiment. They control for initial conditions using a separate reduced-form model.

4.1 Estimated Parameters

All estimated parameters appear in θ_{exog} and can be grouped by which part of dynamic program they enter. Parameters of $U()$ are placed in a type-specific vector,

$$\Upsilon \equiv (\beta \quad \eta \quad \kappa \quad \mu \quad \nu \quad \sigma \quad \zeta \quad \xi). \quad (16)$$

Recapping the previous discussion, β is the rate of income reporting; η is the curvature in skill; κ converts search time into work time; ν is the (scaled) income-equivalent cost of full-time work; μ and σ determine location and spread of wage offers; ζ determines the variance in the curvature of time-costs over time; and ξ is the factor on outside support.

Transition parameters are also grouped together:

$$\Pi \equiv (\pi_j \quad \pi_m \quad \pi_f \quad \pi_h \quad \pi_i \quad \pi_d \quad \pi_l \quad \pi_s \quad \pi_p). \quad (17)$$

Referring to each probability by its subscript, j is the probability that active job search generates a job offer (in the absence of SSP+); m is the proportion jobs that are pure minimum wage jobs; f is the proportion of job offers that are full-time jobs; h is probability that the curvature in household costs changes; a is the probability that skills accumulate while working; d is the probability that skills decreases while not working; l is the probability that the parent is laid off exogenously; s is the probability that outside support changes; and p is the index of SSP+ effectiveness.

Then $\Gamma[k] \equiv (\delta_k \quad \rho_k \quad \Upsilon[k] \quad \Pi[k])$ contains all parameters that determine behaviour of type- k parents. The exogenous parameter vector contains those $K = 4$ vectors and the type proportions: $\theta_{\text{exog}} = (\Lambda[1] \quad \dots \quad \Lambda[4] \quad \Gamma[1] \quad \dots \quad \Gamma[4])$. For a household of type k in group d only $\Lambda[d]$ and $\Gamma[k]$ are relevant. There are $N = 19$ parameters in $\Gamma[k]$ leading to a total of $K(D + N) = 4(4 + 18) = 88$ exogenous parameters. Fewer parameters are free. Three parameters in $\Gamma[k]$ are constrained to be equal across type, on the presumption that they are the least likely to be identified by variation in observables and therefore most reliant on functional form assumptions. And the four elements of $\Lambda[d]$ must sum to 1 for all d . The result is $3 + 4(19 - 3) + 3(4) = 79$ parameters estimated from the data, all contained in θ_{exog} .

4.2 Measurements and Sources of Identification

Ideally the exogenous vector θ_{exog} would be estimated using a long panel on individuals and observing the full outcome (α, θ) starting from fixed initial states. In this case, applying endogenous choice probabilities sequentially along a single stochastic path will form the likelihood. However, two aspects of the SSP preclude this approach. First, full histories of the subjects are not available. Second, a realistic model of the welfare trap includes elements unobserved by the econometrician, such as skill, job quality, and time-varying leisure-income tradeoffs. When the

Table 1.

Demographic, Treatment, and Experimental Groups				
Vector	Index	Description	Subjects	% of Total
	d	Demographic Group		
θ_{end}	1	New Brunswick, 1 Child	1728	19%
	2	New Brunswick, 2+ Children	1217	14%
	3	British Columbia, 1 Child	3058	34%
	4	British Columbia, 2+ Children	2895	33%
	g	Treatment Group		
θ_{exp}	3	Control	4305	48%
	2	SSP Treatment	4300	48%
	1	SSP+ Treatment (NB only)	293	3%
	e	Experimental Group		
	2	Recipient Study	5682	63%
	1	Applicant Study (BC only)	3316	37%

Total observations used = 8898. Observations dropped due to invalid or missing age, high school attendance or number of children.

realized path is not full observed the likelihood requires multiple summation across states and choices weighted by time-varying endogenous probabilities. Here, estimation is based not on (α, θ) directly but on a vector of measurements denoted $Y(\alpha, \theta)$ and described in section 3.2. Some state and action variables are elements of Y and thus directly observed. Others, such as current skill, influence observables such as earnings but cannot be observed or inferred from Y . While not ruling out maximum likelihood estimation, this approach is a natural basis for GMM. The predicted expected are matched to observed average values of $Y(\alpha, \theta)$ conditioned on instruments. Four observed exogenous variables generate instruments, including three that are fixed for an individual: $\tilde{\theta}_{\text{cond}} \equiv (g \ e \ d)$. This vector includes the person's the demographic group d and the elements of θ_{exp} defined earlier and consisting of the entry (Applicant or Recipient) and treatment (control, SSP, SSP+) groups. There are 14 different values of $\tilde{\theta}_{\text{cond}}$. As shown in Table 1, 8,898 people who took part in the SSP experiment are included in the analysis here. Roughly two-thirds were sampled from British Columbia, because the Applicant Study was conducted in BC alone. The SSP Plus Study includes 292 people in New Brunswick assigned to treatment. Roughly one-half of the households had more than one child at the baseline.¹⁵

The fourth conditioning variable is time t , which appeared in Figure 2 but does not appear in the parent's state/parameter vector because it plays no direct role in the model solution only in the

¹⁵ More details and descriptive discussions of the data are available in the Supplement.

construction of predictions and observations. It is appended to the fixed conditional vector $\tilde{\theta}_{\text{cond}}$:

$$\theta_{\text{cond}} \equiv (t \ \tilde{\theta}_{\text{cond}}) = (t \ g \ e \ d) \quad (18)$$

to coordinate measurements across groups and track time during endogenous eligibility for assignment. The results here use 36 post-assignment months of data ($t = 1$ to $t = 36$) in the Recipient Study and 30 months ($t = -11$ to $t = 18$) in the Applicant Study. Province and household size are exogenous by assumption. Since the distribution of preference parameters is specific to each group, d can be interpreted as lagged endogenous and possibly sensitive to certain policy changes. After controlling for endogenous eligibility for randomization entry group e and treatment group g are randomly determined. Experimental time t increments each month regardless of decisions so it also is exogenous and acts as an instrument in two ways. First, the model predicts that control groups will re-approach their stationary distributions across states, P_{∞} . The drift in the data is matched to the model's prediction as correlated with t . Second, t is correlated to endogenous choices that determine the experimental clock, θ_{clock} , since each phase of treatment has a maximum finite length and explicit transition rules. The model relates behaviour directly to θ_{clock} and then indirectly to t through the distribution across states at each t .

The expected observation given the conditioning vector is written:

$$E[Y | \theta_{\text{cond}}, \theta_{\text{exog}}] = \sum_{k=1}^4 \lambda^*(k, \tilde{\theta}_{\text{cond}}) \left[\sum_{\theta_{\text{end}}, \theta_{\text{clock}}} \Omega\{\theta | k, \theta_{\text{cond}}\} \sum_{\alpha \in A(\theta)} P\{\alpha | \theta\} Y(\alpha, \theta) \right]. \quad (19)$$

The inner right-most summation is what the parent expects their measurement vector to be at the beginning of the month, as it conditions on θ and averages over state-contingent choice probabilities, $P\{\alpha | \theta\}$ defined in (4). In turn this is averaged over states using the distribution $\Omega(\cdot)$ defined in equation (S15), which accounts for endogenous eligibility, unobserved states, and the evolution of states since random assignment. The term $\lambda^*(\cdot)$ from (S16) is the fraction of eligible households in group $\tilde{\theta}_{\text{cond}}$ of type k . It is a reweighing of $\lambda[d, k]$ that accounts for the differences across unobserved types in being eligible for random assignment in entry group e . Parents in a control group act as they do outside the experiment (conditional on their current state). For them Ω differs from the stationary distribution P_{∞} because of non-random sampling. Post-assignment their choices lead Ω to converge back to P_{∞} . Among the treated the distribution can continue to diverge from its stationary value until all individuals are past the final stage of treatment.

Let $\hat{E}[Y | \theta_{\text{cond}}]$ denote the vector of average observed (empirical) results for θ_{cond} . The 12 contemporaneous variables chosen for Y are summarized in Table 2. Many of the moments were chosen because they are key outcomes of interest in the SSP and social assistance policy in general,

Table 2.
Experimental Results (Moments) Selected for Matching

Var.	Description	Model	Unit	Count	Mean	St.Dev	Note
earn	Reported Earnings	$(1-\beta_i)mW(\alpha,\theta)$	\$100	470	3.439	1.564	
earnsq	Earnings Sq.	earn ²	\$100 ²	470	62.59	43.470	
ia	IA Received	IA(α,θ)	\$100	466	5.966	1.869	Fwd.2 mth
iasq	IA Recv Sq.	IA ²	\$100 ²	466	57.78	27.593	Fwd.2 mth
gsu	SSP Suppl	SUP(α,θ)	\$100	240	1.530	0.600	Fwd.2 mth
onia	Received IA	i	0/1	470	0.708	0.161	
mwg	Min. Wage Job	$(n<6-\#n)(m>0)$	0/1	470	0.777	0.059	wage < MW+\$\$.10
leftjb	Left/quit a job	$p(l=0)(m=0)$	0/1	456	0.003	0.004	Excl. job-to-job
lossjb	Lost job	l	0/1	456	0.004	0.004	Excl. job-to-job
emft	Full Time	m>PT	0/1	470	0.223	0.088	
empt	Part-time	0 < m <= PT	0/1	470	0.130	0.023	
onXem	IA & Working	(ia) (m>0)	0/1	470	0.161	0.045	

Notes: The complete data used in estimation is available in the supplement. Count is number of cells. Total count = 5374. Mean and standard deviation are across cells not individuals. #n denotes the order of n in the feasible set. For example, #0 = 1, #1/6 = 2, etc. Fwd. 2 months means values in month t come from month $t + 2$ in the data to account for the lag in verifying status and then paying and recording the supplement.

including earnings, rates of IA receipt, full-time work, lay-offs, and quits. Others were chosen to help identify specific parameters of the model (see S6.3). The model makes a prediction for every combination of variables in θ_{cond} . So a 0/1 indicator for each value of the vector is itself a valid instrument in the usual sense. Thus the number of instruments is not three or four, based on the dimension of θ_{cond} , but rather the number of unique values of θ_{cond} in the data, over 400. Thus, there is a large amount of variation that is either strictly exogenous or conditionally exogenous. The measurement vector $Y(\alpha, \theta)$ contains twelve values that interact choices with endogenous states. Interacting measurements with θ_{cond} results in a total of 5374 total moments.

With the complex utility and transitions it is not possible to demonstrate formal identification of all parameters. Consider here only the identification of the discount factors δ_k from the experimental treatment. Rust (1994) shows the discount factor is non-parametrically unidentified in the case of infinite horizon discrete choice models. In many parametric applications the discount factor is still so poorly identified it is fixed *a priori*. However, Wolpin (1995) provides a formal proof that the discount factor is parametrically identified in a job search model with an exogenous finite horizon. Intuitively the change in behaviour leading up to the final decision period reveals the patience of the decision maker if other parameters are identified from other data (such as accepted wages). The fixed length of each phase of SSP treatment plays a role similar to an exogenous finite horizon.

Within a treatment phase utility shifts dramatically when the phase shifts, which occurs according to rules built into the model. The experimental clock θ_{clock} is counting down in each phase of treatment. Further, we observe identical control groups with no such finite deadlines, providing additional variation on expectations.

Finite phases of treatment made running the SSP tractable with the unintended benefit of providing exactly the kind of dynamic variation required to identify how patient the single parents are when making decisions. The small standard errors on δ_k and most other parameters reported later suggest the heuristic explanations are relevant for the sample. And when computing standard errors using data from within groups the change in standard errors confirms that identification of δ_k comes from the non-stationarity of the SSP treatment.

4.3 GMM Estimation

The exogenous vector is estimated by minimizing the weighted distance between observed and predicted values of the conditional moment vectors:

$$\Delta(\theta_{\text{cond}}) \equiv \hat{E}[Y | \theta_{\text{cond}}] - E[Y | \theta_{\text{cond}}, \hat{\theta}_{\text{exog}}]. \quad (20)$$

The objective in the first stage of the efficient GMM estimation procedure is

$$Z^1(\hat{\theta}_{\text{exog}}) \equiv \sum_{\tilde{\theta}_{\text{cond}}} \sum_{t=t_0(e)}^{t_{\text{max}}(e)} \frac{n(\theta_{\text{cond}})}{265159} \Delta(\theta_{\text{cond}})' \Sigma_0 \Delta(\theta_{\text{cond}}), \quad (21)$$

where Σ_0 is a 12×12 diagonal matrix.¹⁶ The terms t_0 and t_{max} denote the time of random assignment in entry group e and the maximum observed month, respectively.

Let $\hat{\theta}_{\text{exog}}^1$ denote the parameters that minimize Z^1 . Measurements for an individual are correlated across t . Across treatment, entry, and demographic groups the measurements are independent. So the population covariance of moments is block diagonal with the blocks defined by $\tilde{\theta}_{\text{cond}}$. $\Delta(\theta_{\text{cond}})$ is stacked across t to form $\tilde{\Delta}(\hat{\theta}_{\text{exog}}, \tilde{\theta}_{\text{cond}})$. The weighting matrices for these long vectors in the second stage are the inverse of the variance matrix of the moments, $\Sigma(\tilde{\theta}_{\text{cond}}, \hat{\theta}_{\text{exog}}^1)$, estimated by simulation using $\hat{\theta}_{\text{exog}}^1$.¹⁷ The second stage objective is

$$Z^2(\hat{\theta}_{\text{exog}}) = \sum_{\tilde{\theta}_{\text{cond}}} \tilde{\Delta}(\hat{\theta}_{\text{exog}}, \tilde{\theta}_{\text{cond}})' \Sigma(\tilde{\theta}_{\text{cond}}, \hat{\theta}_{\text{exog}}^1) \tilde{\Delta}(\hat{\theta}_{\text{exog}}, \tilde{\theta}_{\text{cond}}). \quad (22)$$

¹⁶ For the monetary values in Table 2 the weight is the inverse of the grand mean of the moment over conditioning states. For the binary variables a weight of $1/.5 = 2.0$ was chosen to avoid putting excessive weight on turnover values which are near 0 and noisy across months. The cell sizes $n(\theta_{\text{cond}})$ sum to 265,159 and are listed in Table S6.13.

¹⁷ Note that to calculate the predicted moments in (19) simulation is not used. The full distribution across all states is stored for each t starting from the stationary distribution. This preserves smoothness of the objective in the estimated parameters. Since the weights are computed only once for one vector smoothness is not required. Details of the simulation are in S6.6.2.

Let $D(\hat{\theta}_{\text{exog}}, \tilde{\theta}_{\text{cond}})$ denote the matrix of gradients for the vector $\tilde{\Delta}$ with respect to the estimated parameters. The variance of the estimated parameter vector is computed as

$$\hat{V}ar[\hat{\theta}_{\text{exog}}] = \left\{ \sum_{\tilde{\theta}_{\text{cond}}} D(\hat{\theta}_{\text{exog}}, \tilde{\theta}_{\text{cond}}) \Sigma(\tilde{\theta}_{\text{cond}}, \hat{\theta}_{\text{exog}}^1) D(\hat{\theta}_{\text{exog}}, \tilde{\theta}_{\text{cond}})' \right\}^{-1}. \quad (23)$$

Standard errors for $\hat{\theta}_{\text{exog}}$ are based on (23). Two other sets of standard errors are reported below. In one, control groups ($g = 3$) are excluded from the sum over $\tilde{\theta}_{\text{cond}}$; in the other set treatment groups ($g = 1, 2$) are excluded. The standard errors are still evaluated at the parameter values found when including all the data, so these calculations measure how within- and between-group variation contributes to identification of the parameters.

5. Estimates and Implications

5.1 Parameters and Predicted Outcomes

Table 3 reports the efficient GMM parameter estimates.¹⁸ The estimated mixing probabilities show that two types predominate in BC and three types in NB, with NB1 primarily of one of those types. Type proportions vary more across provinces than between numbers of children. Types have very different levels of patience. A period is one month, so only for the first two types is δ_k close to 1.0. The other types make decisions close to a static manner: next year's outcomes have essentially no impact on today. The income reporting parameter β is straightforward to interpret. Three types are estimated to report approximately 40% of their income when on welfare. One type reports 95%.

Wage offer distributions differ across types as does the stigma associated with welfare (captured by the coefficient on outside support, χ). Full-time work has a very similar cost across type (ν), but recall that this value is relative to maximal earnings for a given type. This contrasts with the cost of active job search, which is only large and precisely estimated for type 1 (and to lesser extent type 4).

The bottom panel of Table 3 reports the transition shifters Π . Here we see that type 1 is constrained by a low job offer probability. Most offers are full-time, so the high fraction of part-time work reflects a choice to work fewer hours than the job allows. Between 13% and 53% of job offers are $n = 0$ jobs (with no on-the-job growth potential). Estimates of the home environment indicate

¹⁸ Since there are no coefficients on observed variables few of the magnitudes are easy to compare with results based on, say, a Mincer earnings function. The discussion of the estimates themselves is brief in favour of a longer discussion of the model's prediction. The transition parameters are probabilities, but those related to skill, job offers and household costs affect unobservables, and the magnitude depends on the number of values the corresponding state variable takes on. Adding points to make the discrete grid finer would make the expected change from a jump smaller and the estimated jump probability would go up.

Table 3.
 $\hat{\theta}_{\text{exog}}$: Estimated Parameters

Variable	Description	Type Index (k)				
		1	2	3	4	
Type Prop. Λ	dem. group	1 NB, One Child	0.0063	0.8216	0.0911	0.0810 -
		2 NB, Two+ Children	0.000003	0.2072	0.3253	0.4675 -
		3 BC, One Child	0.5228 *	0.00009	0.4771 *	0.0000009 -
		4 BC, Two+ Children	0.5736 *	0.00009	0.4263 *	0.00002 -
DP	Parameters	δ Discount Factor	0.9999 *	0.934 *	0.476	0.742 *
		β Income Reporting	0.399 *	0.413 *	0.955 *	0.391 *
		ρ Smoothing	36.220 *	99.192	11.507	6.103 *
Utility Shifters (Y)	Market	μ Job Offer Mean	-1.560 *	-0.072	0.025	0.020
		σ Job Offer St. Dev.	1.999 *	1.632 *	1.825 *	1.608 *
		η Return to Skill	1.355 *	2.964	31.685	7.070 *
	Home	χ Outside Support	1.427 *	1.080	1.535 *	0.825 *
		ν Cost of FT Work	0.346 *	0.409	0.447 *	0.487 *
		κ Cost of Job Search	0.461 *	0.0007	0.000002	0.081 *
		ζ 1 / Mean Convexity	2.997 *			
Transition Shifters (Π)	Market	j Job Offer ($b > 0$)	0.069 *	0.730	0.99994 *	0.831 *
		f Prop. Full Time	0.999996 *	0.889 *	0.99955 *	0.902 *
		m Prop. MW job ($n=0$)	0.131 *	0.439	0.531	0.249 *
		l Job Loss	0.021 *	0.0011	0.018 *	0.002
	p SSP Plus Effect	0.823 *	0.009	0.641	0.910 *	
	Home	s Support Change	0.029 *			
		h Costs Change	0.999989 *			
	Skills	a Accumulation	0.995 *	0.4787	0.1245	0.2601 *
d Depreciation		0.0006	0.0037	0.7654	0.1007 *	

Notes: Efficient GMM estimates. Value of Objective = 19.426. Standard errors reported in Table 4 "all" column. Estimates with t ratios significant at the 5% level indicated with *. Type 4 proportions sum to 1 across rows, so standard errors for $k=4$ not computed. Parameters with an estimate only in the $k=1$ column were estimated as common across types.

that outside support is highly persistent (π_s is small) but household costs of work and job search are not (π_h is high). Type 1 workers get a skill increment each period and have rapid on-the-job wage growth. For other types growth is intermittent. Type 3 is the only one with further skill accumulation expected after one year of working (based on the increment probability and four skill levels). Because average wages do not accumulate in the treatment group, this suggests that the return to skill (η) reported in the previous table is not large. Thus, parents achieve modest wage growth early in an employment spell but not sizeable long-term growth. Only for type 3 is depreciation of skills rapid while not working. Thus the impression the model gives for the SSP results is that the treatment requires long-term and persistent growth in skills. Skill persistence is much less of an issue than a predominance of jobs with no growth potential and a low wage elasticity to skill accumulation.

Figure 3 presents the observed and predicted results for earnings by demographic group. For each group the top panels show outcomes: data on the left, model predictions on the right. The difference between the left and right top panels enters the GMM objective as an element of $\tilde{\Delta}$. The bottom panels depict the observed and predicted impacts of the treatment. The model's vector-valued impact is defined for $g < 3$ as

$$E \left[Y \left(\theta_{\text{cond}}, \hat{\theta}_{\text{exog}} \right) \right] - E \left[Y \left(\theta_{\text{cond}} \Big|_G, \hat{\theta}_{\text{exog}} \right) \right], \quad (24)$$

where the notation $\Big|_G$ means replace g in θ_{cond} with $G = 3$, the control group, holding other elements of θ_{cond} fixed. Observed impact is similarly defined using the data. For both family sizes in New Brunswick the predictions track the data quite well. Selection and the evolution of state variables together generate the upward trend in the control groups as they return to the ergodic distribution.

The response to treatment generates an impact that mimics the data. The one aspect of the data that the model fails to capture qualitatively is the slope of change in the Applicant Study (shown in the left column of Figures 3-5). The starting level and impact are accurate but the selection effect is larger than the model predicts. The fraction of the BC 1 group on IA is shown in Figure 4. The match to the data is similar to that for earnings, although the mismatch in the Applicant Study is of a different form. For OnIA the model impact is too large before time 0. Figure 5 shows total transfers, IA + SSP, for the BC 2+ group. From the government budget perspective, the SSP is valuable if additional transfers during treatment result in lower transfers later on. Since the impact fades, the policy is a failure in total transfers. In all groups including those not shown and at each month the impact on transfers is non-negative. The subsidy never induces a substantial move to

Figure 3. Observed vs. Predicted Earnings

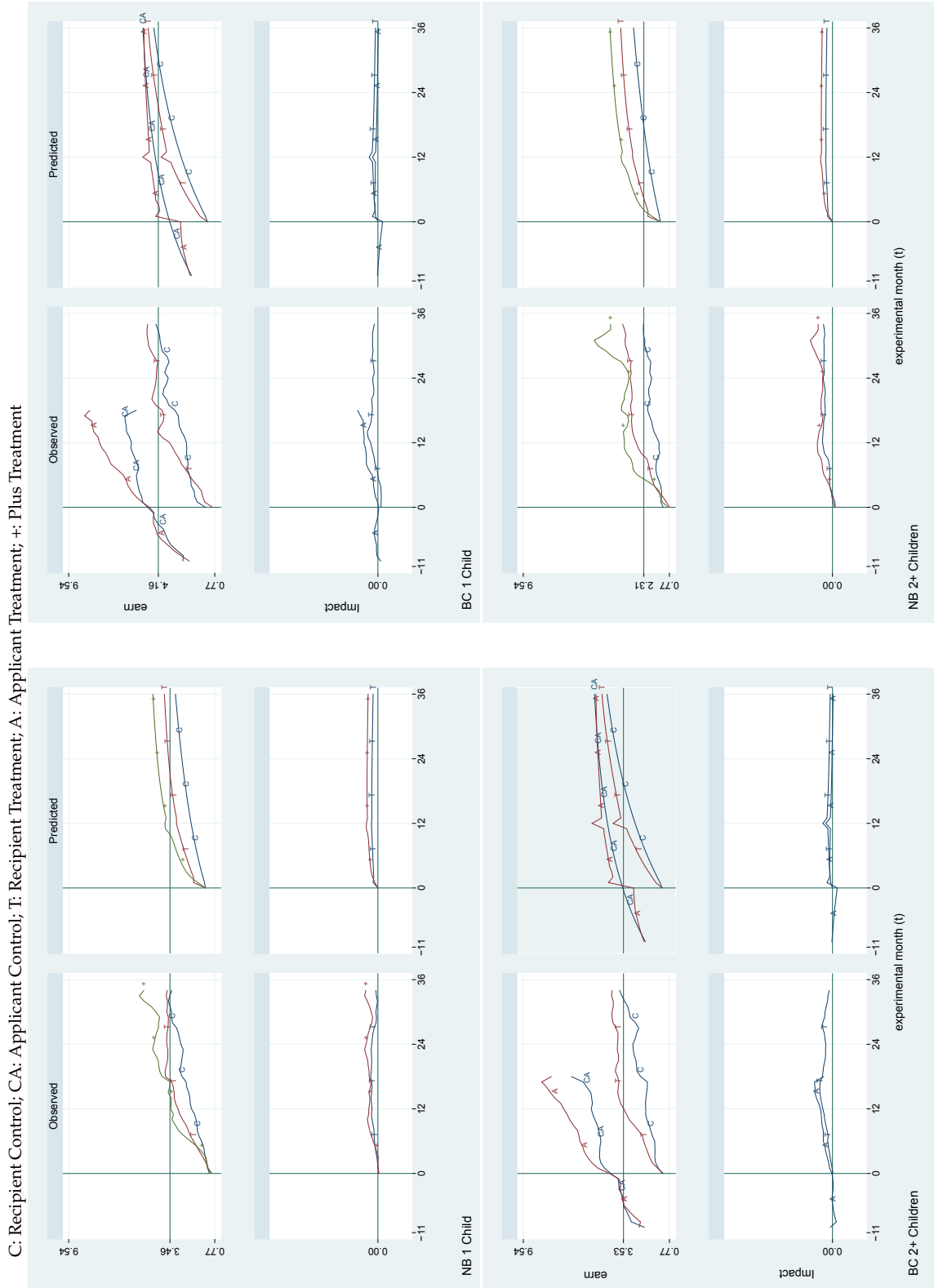
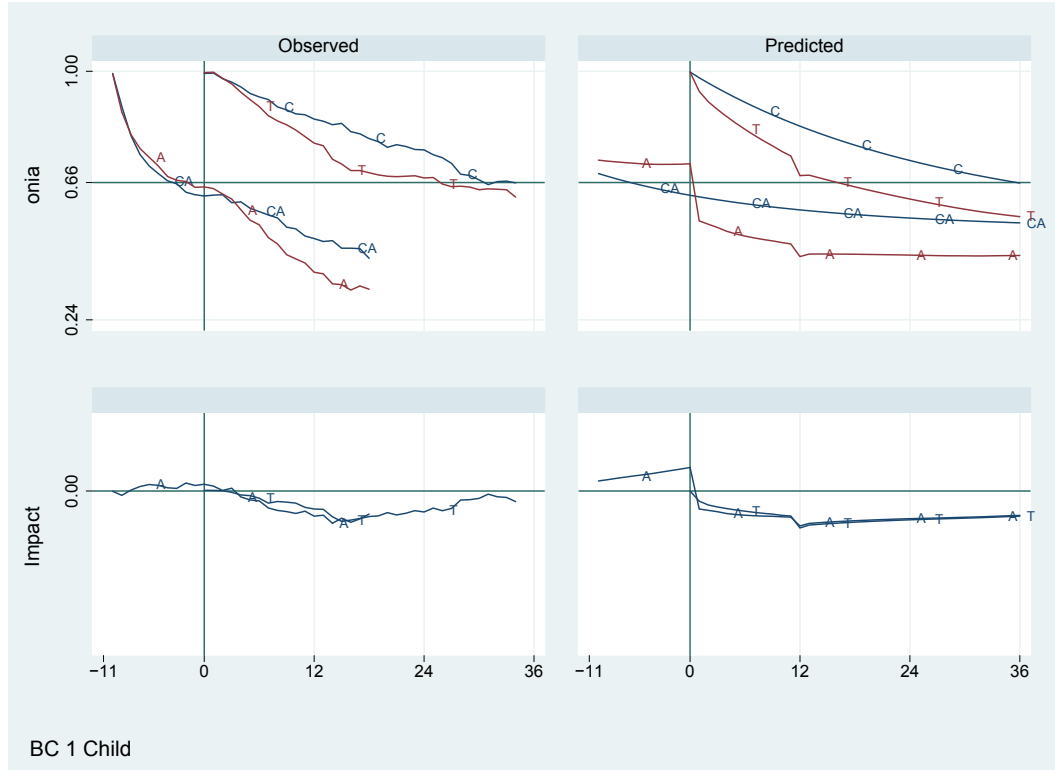


Figure 4.

Observed vs. Predicted IA Participation (OnIA)

C: Recipient Control; CA: Applicant Control; T: Recipient Treatment; A: Applicant Treatment; +: Plus Treatment



self-sufficiency. In some groups the model generates a larger impact than the data, but it captures the rise and then near constant impact until month 36.

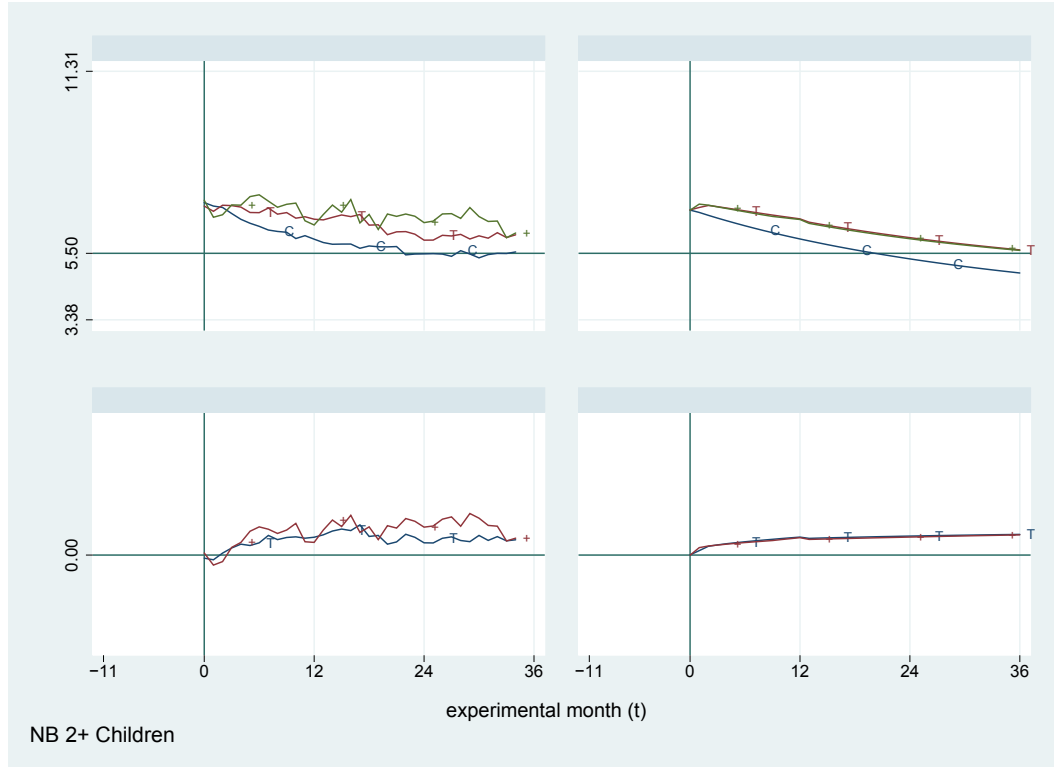
5.2 Variation from Policy, Selection and Heterogeneity

Figure 6 illustrates the combined effects of all sources of variation. Each panel shows the behaviour of a particular unobserved type in all four observed environments. The two most patient types, $k = 1$ and $k = 2$, are shown. Since preferences are held constant, the effect of policy variation is illustrated by comparing the four panels within each type. And since the SSP is based on a selected sample the trends in the control groups capture how distant the selected group is from the population average. The ergodic mean is shown as a triangle. For type 1 we see that all groups are well below the average in earnings. By month 36 the control group has nearly returned to the ergodic distribution. The most striking aspect of the top half of Figure 6 is the large response to SSP+, which is a combination of a large estimate of effectiveness (π_p) and a low job offer probability (π_j). Type 1 is a small fraction of the NB population so the modest additional impact of SSP+ is a

Figure 5.

Observed vs. Predicted Government Transfers (IA+SSP)

C: Recipient Control; CA: Applicant Control; T: Recipient Treatment; A: Applicant Treatment; +: Plus Treatment



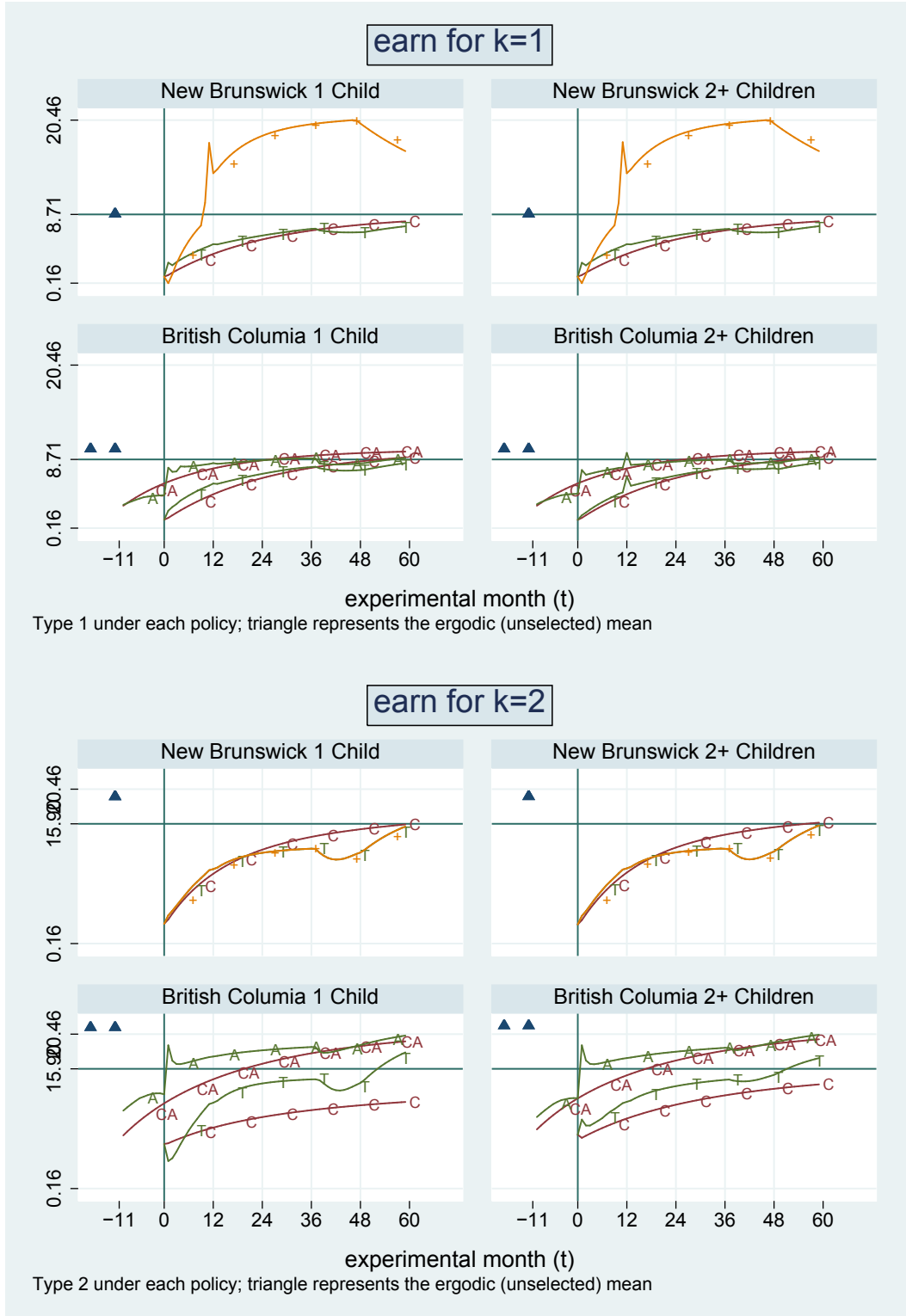
combination of a large individual response among a small part of the population. This same group is not particularly responsive to the SSP treatment; its households are constrained by a lack of job offers which the SSP+ alleviates.

The bottom half of the Figure shows type 2. For this type the selection effect is more extreme and even after 36 months the control group is still far from the stationary average. The impact under NB policies starts very small and then becomes negative. Apparently this group was induced to accept low wage jobs to qualify for the supplement while their control group counterparts held out for better jobs. Those who qualify tend to keep these jobs until the subsidy ends. This group illustrates one of the difficulties in designing incentive schemes for low-skill parents. The SSP encourages employment but not necessarily patience to wait for employment with high growth potential. The response of type 2 is itself heterogeneous, because the opposite pattern occurs under British Columbia policies. Here the expected impact in earnings appears and is long lasting. However, type 2 is estimated to be a vanishingly small fraction of the population in BC.

Figure 6.

Variation from Policy, Selection and Heterogeneity

C: Recipient Control; CA: Applicant Control; T: Recipient Treatment; A: Applicant Treatment; +: Plus Treatment



5.3 Treatment: Identification or Validation?

The marked estimates in Table 3 and the underlying standard errors reported in Table 4 indicate that many parameters are precisely estimated by the variation in moments generated by the experiment. The parameters are identified by restrictions on how the moments can vary across treatment groups, over time within a group, and across demographic groups. An alternative use of the exogenous variation generated by the experiment is to validate a model estimated only on the control group. Lise et al. (2005) and Todd and Wolpin (2006) follow this approach by respectively calibrating and estimating models of forward-looking agents within control groups of experiments (respectively SSP and Progresa). They then use the treatment group data for out-of-sample validation. A major advantage of this approach is that behaviour under the treatment does not have to be solved repeatedly while estimating the parameters. The potential cost is that the model that can be estimated from the control group alone may be not be as rich as one that can be estimated using the experimental data. Thus, the parameter estimates may be less applicable outside the sample and less reliable for understanding behaviour in populations facing similar but not identical environments.¹⁹

To quantify the cost of not using the experimental variation for estimation, the standard errors for the parameters were re-computed using only the moments within groups. Results were re-scaled to mimic a sample of the original size. Table 4 reports the results. Standard errors based on all the data are compared to those from the control and treatment groups alone. First consider the "Ctrl" column. It is not surprising that throwing out the experimental variation increases the standard errors. However, for nearly all the parameters the estimated standard error is eight times larger than when based on all the data. Included among these are key parameters for understanding dynamic behaviour of low-income households: the discount factor (δ), the wage offer parameters (μ and σ), the return to skill (η) and many probabilities that determine persistence in wages and other states. Thus, if the validation strategy had been used here, a model estimated from the control data alone would have been much simpler in form without the ability to capture some details in the experimental outcomes.

Another result is revealed in Table 4 when the "All" column is compared to the "Treat" column. This counter-factual throws out the variation between treatments and controls and replaces it with more information on the experimental variation. In nearly all the cases the re-scaled standard errors are *smaller* when based on the treatment groups alone and often the increased precision is not trivial.

¹⁹ Todd and Wolpin (2006) suggest that if the model is validated then one might re-estimate using the experimental data as well to increase precision of the estimates. Under this strategy, restrictions due to parameters unidentified from limited variation in the control group still limit the generality of the final model.

Table 4. Estimated Standard Errors: Based on All Data and Within Treatment Groups

Param. / Subscript	k=1 / common		k=2		k=3		k=4		
	All	Ctrl	All	Ctrl	All	Ctrl	All	Ctrl	
1	0.030	0.098	0.860	2.657	0.894	0.894	0.438	1.295	0.456
2	0.006	0.148	0.985	4.187	0.984	0.984	0.429	1.972	0.413
Λ	0.020	0.081	0.016	0.051	0.013	0.013	0.014	0.057	0.011
4	0.019	0.073	0.016	0.047	0.014	0.014	0.014	0.056	0.011
δ	0.0005	1.526	0.081	8.610	0.052	0.052	0.416	40.816	0.321
β	1.081	1114.74	143.937	6427.08	89.382	89.382	7.440	232.913	5.460
ρ	0.004	0.039	0.040	0.096	0.044	0.044	0.022	0.047	0.019
μ	0.045	11.413	0.154	673.94	0.045	0.045	0.020	0.896	0.014
σ	0.014	2.282	0.069	167.62	0.068	0.068	0.235	5.286	0.152
χ	0.037	1.316	3.665	146.915	2.508	2.508	0.023	0.718	0.017
ν	0.004	0.648	0.767	192.037	0.515	0.515	0.133	3.250	0.086
κ	0.036	41.937	0.350	524.13	0.222	0.222	0.002	0.322	0.001
η	0.155	371.760	30.406	2711.54	18.156	18.156	415.395	3913.08	313.367
ζ	0.069	3.380	0.050						
j	0.003	2.565	0.510	1520.55	0.351	0.351	0.060	0.274	0.048
f	0.0001	0.019	0.080	0.664	0.075	0.075	0.206	25.773	0.146
m	0.016	32.530	0.321	1174.60	0.221	0.221	0.512	11.121	0.385
l	0.001	0.006	0.007	0.020	0.007	0.007	0.006	0.009	0.006
s	0.002	0.004							
h	0.002	0.156							
p	0.015	-----	0.255	-----	-----	-----	1.060	-----	-----
a	0.038	39.199	0.898	40.614	0.679	0.679	0.071	2.743	0.056
d	0.001	0.019	0.021	1.000	0.015	0.015	1.066	9.138	0.870
							0.002	0.017	0.001
							0.028	2.257	0.019
							0.066	0.733	0.043
							0.030	2.227	0.021
							0.002	0.017	0.001
							0.036	-----	-----
							0.016	0.595	0.012
							0.009	0.407	0.007

Notes: "All" = standard error for the corresponding parameter in Table 3; "Ctrl" is the standard error using moments from the control group scaled by $\sqrt{1/2}$ to eliminate the sample size effect. "Treat" is the re-scaled standard error using only moments from the treatment groups. Bold indicates a standard error 4 times larger than "all". Italic indicates smaller than 3/4 of "all". SSP+ effect is unidentified from controls, so it was excluded from both "Ctrl" and "Treat" to make those columns comparable.

In many cases the standard error is reduced by 25% or more. The source of this extra precision is simply the experimental variation in incentives generated by the experimental design. Within the program of treatment the next month is quite different than the current month since one deadline or another is approaching. Within the control group no such deadlines exists.

This result has a somewhat surprising implication. When using data solely to study mean differences between treatment groups, control and treatment observations have equal value as they enter linearly in the statistic of interest. Absent other costs, splitting the overall sample evenly minimizes the variance of the impact estimate. Table 4 suggests this is not true when experimental variation will be used to identify an underlying model. An additional treated observation may be more valuable than an additional control observation because their choices reflect more exogenous variation. In other situations it may be control group observations that contain more information for estimation.

5.4 Card and Hyslop Counterfactuals

Card and Hyslop (2005) model IA participation (one of the three elements of α) as a random effect probit in the Recipient Study excluding SSP+. ²⁰ They use their estimates to predict average IA participation by month t under two counterfactual programs of treatment. It is straightforward to replicate their experiment using the estimated model. It requires changing parameters of treatment (θ_{pol}), resolving behaviour among the treated, and computing new predicted values, $E \left[Y \mid \theta_{\text{cond}}, \hat{\theta}_{\text{exog}} \right]$.

The first counterfactual is lengthening the qualifying period (phase $f = 2$) by 3 months. Using R_f to denote the maximum length of phase f (Table S2.1), this is a policy shift from $R_2 = 12$ to 15. Forward looking behaviour implies that predictions can shift for all periods both before and after month 12. Parents who would have accepted any job to qualify in month 12 may now reject some jobs, expecting to get a better offer over the next three months. And parents who had searched actively since randomization may not start searching until closer to the new deadline. After month 15 the treatment is the same, but the distribution across permanent and transitory states may differ, leading to different trends. The second counterfactual considered by Card and Hyslop is lengthening the eligible period (here phase $f = 3$, in their notation $E_{it} = 1$), from $R_3 = 36$ to 48 months. In the model greater future potential income affects job search behaviour even during qualification before the supplement is payable. However, income 3 years hence is heavily

²⁰ Card and Hyslop estimate a suite of models and specifications. The focus here is on their joint model of treatment and control groups and their preferred specification, reported in the final column of Table VI in their paper, as well as the counterfactuals in section F.

discounted so the impact at the estimated parameters will be small early on. Once eligible, the longer supplement period increases the value of keeping jobs even before month 48. The impact would continue between months 37 and 49, but under the actual treatment eligibility would be ending, so a widening gap in net impact should be observed. At month 49 the longer eligibility periods start to expire and the patterns would mirror the pattern of qualifying during the first months.

There are important differences between these qualitative explanations and the Card and Hyslop econometric approach. Their panel probit for IA captures state dependence by including two lagged values of IA participation. Here state dependence and persistence are captured by assuming all state variables follow an estimated jump process. IA is itself not persistent, but the household and labour market conditions that lead to it as an optimal choice are. They include unobserved heterogeneity in coefficients, but unlike here the distribution is conditional on being eligible for randomization. That is, they do not correct for endogenous selection into the sample based on being a long-term welfare recipient under the status quo. Their specification includes a fourth-order polynomial in time-since-randomization (t) to capture drift or non-stationarity and an interaction between unobserved type and a quadratic trend in t . Here, drift and heterogeneity in drift is captured by the model with no additional free parameters. Parents in the control group continue to make choices as in the real world after $t = 0$, but since they were selected their distribution over states drifts back toward the stationary distribution. Parents in treatment drift for that reason as well but the treatment induces its own non-stationarity in choices. The same model and the same parameters capture (imperfectly) drift under two completely different selection criteria (applicant and recipient) as well as any other hypothetical selection criterion. With Card and Hyslop's method a different set of selection criteria (such as the Applicant Study) requires different coefficients on the time polynomial.

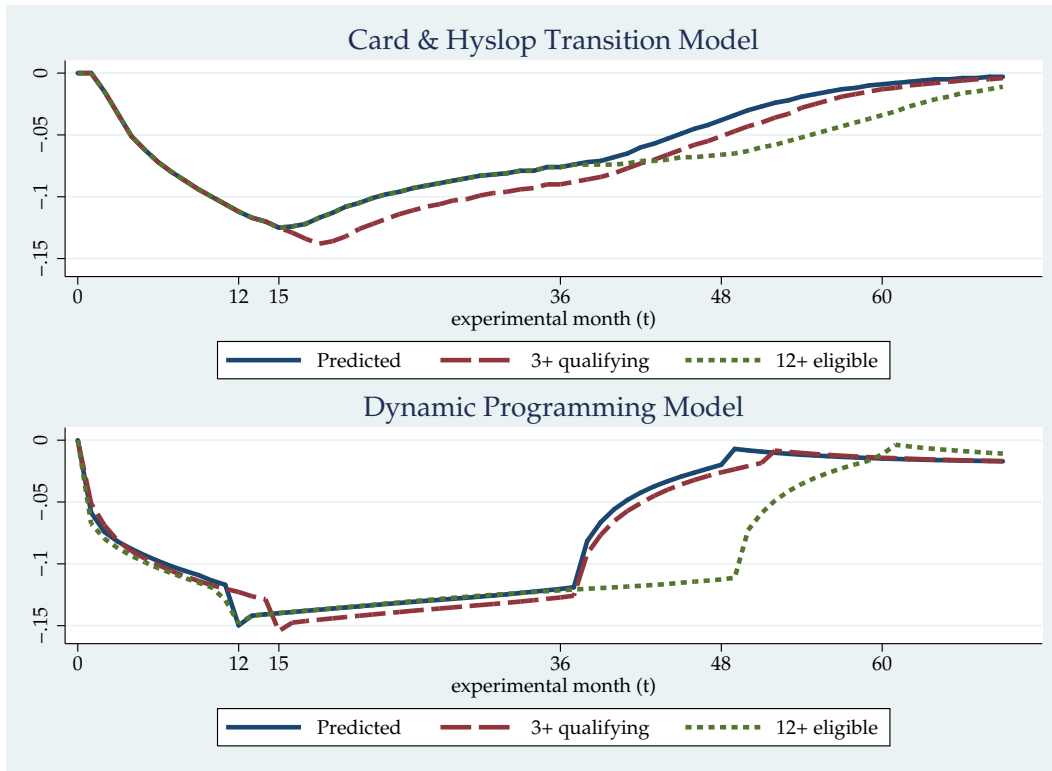
Card and Hyslop model the treatment group with separate coefficients whereas here all estimated parameters apply to both treatment and controls. They extend the panel probit by modeling qualifying (transiting from phase 2 to phase 3) with a normal hazard model. The hazard depends on time since randomization, with an implication that runs counter to the forward-looking model. Namely, the length of the qualification period (here R_2 , in Card and Hyslop T_e) has no effect on the probability of qualifying in months prior to R_2 . They include an arbitrary three month period of adjustment after qualifying and new state-dependency parameters that are specific to the actual treatment. As in the case of the qualifying deadline, the backward-looking state dependency in their specification means that the longer eligibility period in the second counterfactual has zero effect before month 36. Their prediction is based on extrapolating the polynomial trend terms for

greater values of t than in the sample. More generally, none of their models of IA account for income; neither current nor future potential income affect the chances of going on or off IA.

With these considerations in mind, the top panel of Figure 7 recreates Figure 11 of Card and Hyslop (2005). The bottom panel shows results from the estimated model and for the same two counterfactuals. Impact on onIA is shown. The data and model predictions for onIA impact were displayed in Figure 4 for one group. To match Card and Hyslop, which did not control for province or household size, impacts were averaged across all groups weighted by sample size. The in-sample predicted impacts are similar, which is not surprising since the models used the same data. Card and Hyslop's prediction matches the observed impact better, but their flexible form model was fitted to that one dimension, whereas the structural model was fitted to several thousand moments in twelve dimensions across all groups simultaneously. One artifact of their model is the precisely zero effect before months 13 and 37 for the two respective changes. Their impact turns at month 15 because of the ad hoc adjustment period. The estimated model makes the prediction that the net impact of the first change is negligible (but not zero) early on. In months 13-15 it is non-negligible as the "take any job to qualify" decision is delayed. Parents still trying to qualify react to three extra months. From month 15 on, the Card and Hyslop net impact of longer eligibility is larger than the estimated model. Together their prediction before and after month 12 is that a longer eligibility period dominates the shorter one, in terms of this outcome and the goal of moving long-term recipients off IA. A policy maker might conclude from their analysis that a longer qualification period is strictly preferred. But the estimated model shows the impact is quite small after month 15 and is offset by the opposite impact in months 12 to .

In response to the second counterfactual of adding 12 months of subsidy, the Card and Hyslop model predicts a widening impact on IA in months 36 to 48 then a gradual decay. The dynamic programming estimates differ. Card and Hyslop argue (p. 1766) their predictions are lower bounds for the effect of this counterfactual, because it raises the value of the subsidy and would encourage more take-up, an effect their method cannot capture since their model does not include income. Here discounting implies the impact of extended eligibility during qualification is nearly but not exactly zero, suggesting the Card and Hyslop lower bound would be very tight. However, the large fraction of parents who qualified in month 1 and 2 now stay on the subsidy in months 37 to 48. So the impact shifts to the right for twelve months. The Card and Hyslop approach does not impose this type of restriction across time in their predictions. That is, when they extrapolate outside the sample, the rate at which the impact changes in month t is not constrained by the actual take-up rate in month $t - 36$ or $t - 48$. The dynamic program automatically tracks the fraction of households ending eligibility each t , and therefore impact is consistent over time. This means the Card and

Figure 7.
Structural Versus Non-Structural Counterfactuals



Notes: Predicted treatment impact on IA participation (OnIA). The top panel reproduces Figure 11 in Card and Hyslop (2005). The bottom panel shows the estimated model’s impact, averaged by month across demographic groups.

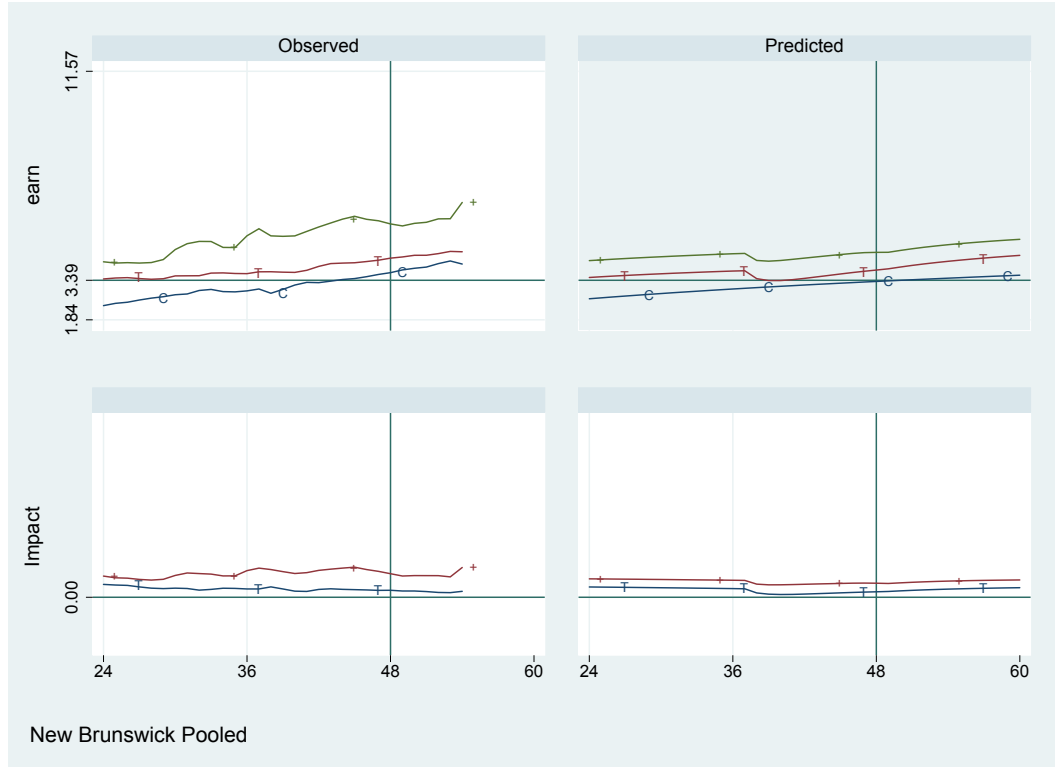
Hyslop impact is not a tight lower bound in months 36 to 48 as would be expected given the small effect of future earnings.

As discussed earlier, skill accumulation is negligible for most of the population, so the extra year of steady work induced by the longer subsidy has little marginal effect on skills and wages. Once eligibility expires outcomes for the treated approach the controls at essentially the same speed as in the actual experiment, just 12 months later. Notice that in months 54 to 70 the Card and Hyslop prediction is not a lower bound; it over-estimates the impact on IA participation. It also misses a slight rebound impact of the model at month 60. This corresponds to parents who took minimum wage jobs to qualify at month 12. They push back the impact to nearly 0 by returning to IA. However, some now find other better paying jobs and go off IA again creating a modest rebound. Together the lower panel in Figure 7 implies much greater total transfer payments under extended eligibility and no improvement in self-sufficiency. The Card and Hyslop results for onIA, while providing no prediction for transfers or earnings, suggests a milder response. This is indeed

Figure 8.

Forecast Earnings Out of Sample

C: Recipient Control; CA: Applicant Control; T: Recipient Treatment; A: Applicant Treatment; +: Plus Treatment



a smaller overall impact, but over time their predictions do not bound the impacts either way.

5.5 Other Predictions

Figure 8 compares the model predictions to data on earnings for New Brunswick from the end of the experiment that was not available for estimation. The averages come from SRDC and are available only by province so the predictions for 1 and 2+ children are combined. Since self-sufficiency did not emerge there is little movement in the impact and it is not surprising that the model is able to capture that. The SSP+ continues to have a modestly larger impact because of the better quality jobs that it allowed parents to hold out for. The figures show a continual drift in all groups which the model does reasonably well capturing and that it attributes to selection and then return to the stationary distribution.

Figure 9 displays the results of two separate counterfactual experiments. Results of both experiments project through month 60, well beyond the estimation sample. First, the Applicant Study was conducted only in British Columbia and the SSP+ treatment was conducted only in New Brunswick on the Recipient sample. Using the model the missing experiments can be run: the Applicant Study in NB and the SSP+ treatment in both provinces and on both studies. The left column shows total transfers (IA+SSP) for the missing samples in the NB 1 Child and BC 2+ Children demographic groups. These counterfactuals can account for the selected populations not in the data because the estimates correct for selection while allowing each demographic group to have its own unselected mixture over parameters.

An Applicant study in NB (top left panel in Figure 9) would have had a very large short-run impact but once treatment ended those who qualified would quit work and return to IA. The pattern makes it clear that in NB the predominant types can easily find a full-time job before time 0, stay on IA to remain eligible then at time 0 start collecting the subsidy. Since this figure extends to month 60 it also reveals an implication not shown earlier in Figure 5. Namely, the model produces a very modest negative impact on government transfers in the NB 1 Child group. It occurs only near month 48 when all supplements are ended. Offering Plus treatment to the Applicants has no additional effect beyond the regular treatment.

An interesting difference is predicted to emerge if SSP+ were run in BC where job offers are a major constraint (bottom left panel of Figure 9).²¹ The extra job-finding help would not only have a major impact, but it is negative almost from the start, meaning that total government transfers are cheaper under the SSP than welfare. In BC good jobs (with low supplements) are available but hard to find. Extrapolating the effects of Plus to the BC population (through the type-specific parameter π_p) the improved offer probability is a significant component. Further, because good jobs are accepted the impacts are long-lasting. This prediction is out-of-sample and is possibly an artifact of the specification in (15). Perhaps the impact of SSP+ in job offers would not be so high in BC. Further, if it were feasible to estimate allowing for more types ($K > 4$) the concentrated effect in NB may not be shared by as large a fraction of the BC population as these estimates indicate. This highlights the difficulty of drawing inferences from experiments out of sample whether it is based on a model or an impact analysis. Together with the modest negative impact in NB it also shows that the hoped-for impacts of the SSP and SSP+ are present in the model, but when accounting for all the outcomes the responsive households are not common enough in the population to generate a large and long-lived impact.

²¹ Note that job offer rates are type-specific not province specific. So job offers are a constraint in BC indirectly because the predominant types in BC have low job offer probabilities.

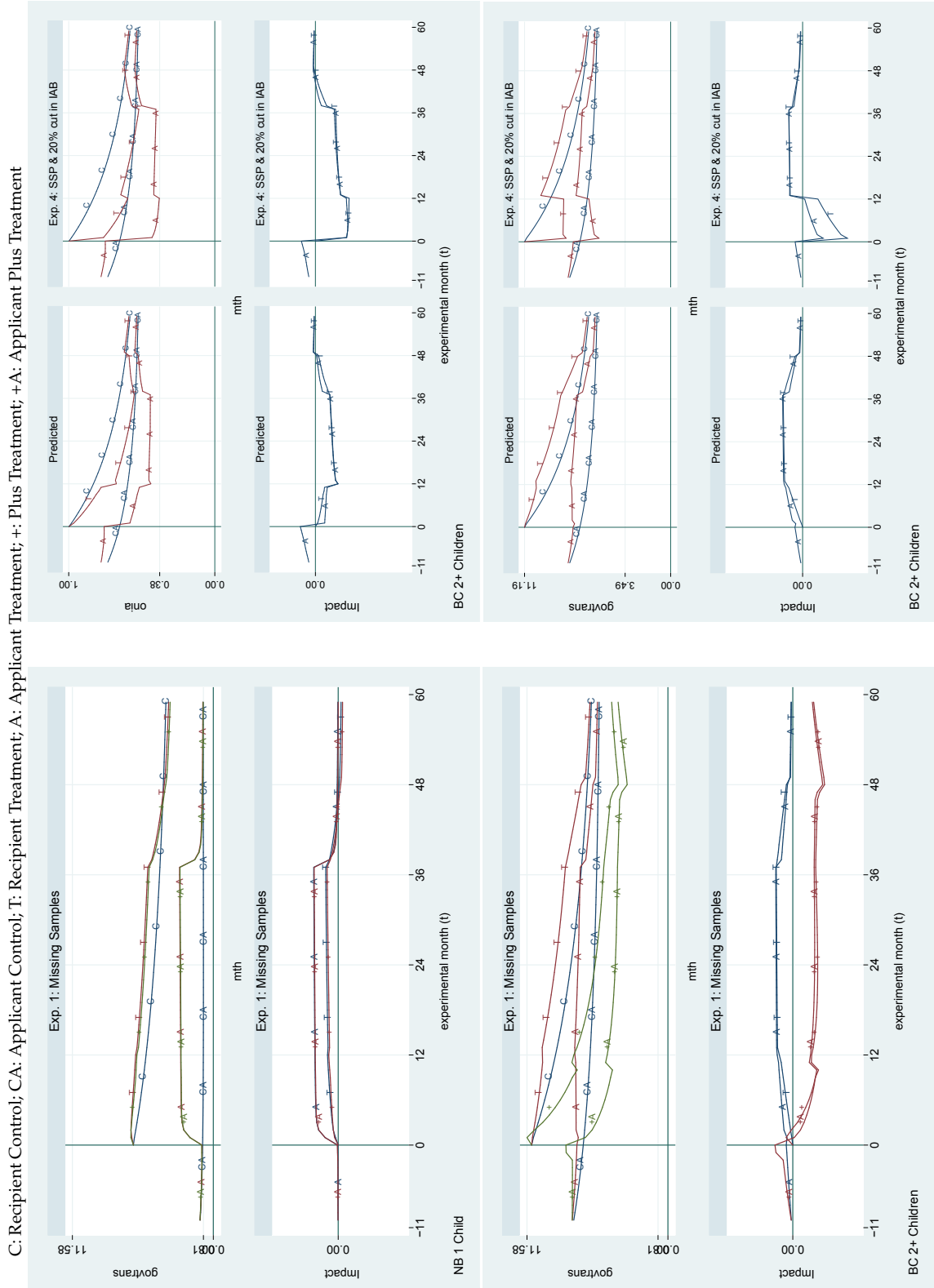
Next, consider the second counterfactual and the model's prediction under an alternative to the SSP treatment shown on the right side of Figure 9. Consider an experiment that would be difficult to run but may reflect a policy that is ultimately behind most reforms to welfare. Namely, suppose offering the SSP treatment while cutting IAB by 20% among the treated. Many parents who do not anticipate finding a job will be worse off in this treatment, but real policy changes might likely combine the carrot of the SSP with a stick of reduced IA levels. The figure shows the effect for BC 2+ for OnIA (top right) and total transfers (bottom right). The counterfactual results are shown on the right while the model's predictions for the actual study are shown on the left in order to display the differential impact. Recipients respond strongly to the cut in benefits. Rates on IA are much lower during the qualifying and eligible phases. And unlike the actual treatment the impact on total transfers are negative during the qualifying phase. But as implemented, those who failed to qualify leave treatment and return to regular IA benefits. Rates and transfers return to roughly what we see in the experiment.

6. Conclusion

Social experiments are designed to guide decisions based on a particular policy (the treatment). As a by-product they create exogenous variation which can be used to infer behavioural responses to other similar policies. That inference depends on a model, of course. This paper has found that results from the SSP can be modeled in a comprehensive way. During treatment the SSP generated sizeable impacts in key outcomes that the model captures quite well, but it failed to induce any obvious long-run move to self-sufficiency. Out-of-sample prediction of the model are validated on this score as well. The model confirms the difficulty in affecting long-term outcomes for low-income households due to lack of job market opportunities, slow transitory skill acquisition, and short decision horizons generated by low discount factors in some parts of the population. Counter-factual experiments confirm that related policies could induce greater short-run response. Only in the case of the SSP+ treatment is there any hint of lasting impacts among a fraction of the population. These are parents who are forward-looking and can acquire skills but have trouble securing employment. The model predicts stronger results may have been detected if the SSP+ had been run in British Columbia where the population mix contains a higher proportion of responsive parents.

Beyond the topic of welfare reform, this paper has explored an alternative approach to combining models and experiments. Estimated standard errors computed after removing groups demonstrate quantitatively that intra-group variation generated by the treatment is critical for identifying a rich and presumably more general model of household behaviour. The literature emphasizes either

Figure 9. Experiments



inter-group variation without any model of behaviour or a reliance on variation within the control group for identification. In the case of the SSP either of these strategies is an inefficient use of the costly output generated by the experiment. Also, counterfactual implications drawn from a non-structural model of the SSP diverge notably from the ones produced by this model. When trying to explain social experiments allowing for endogenous selection and forward-looking behaviour is challenging, but it is shown that in this case it matters quantitatively for making inferences about policy.

There are several limits to the analysis in this paper. First, in order to focus on medium-run causes and solutions to the welfare trap, equilibrium responses to welfare reform are not considered. A targeted earnings subsidy such as the SSP would also affect ineligible low-skill workers in the same labour market if it were implemented as policy. Lise et al. (2005) consider exactly this issue with an equilibrium search model calibrated using SSP data. Second, the utility function is linear in income and additively separable in leisure. As with many job search models linearity allows us to ignore consumption smoothing and assets, but a more general model may suggest that the low discount factor estimated for some types reflects borrowing constraints and not fundamental time preferences. A non-separable utility function is more general and may be able to match some aspects of the data better. Finally, the papers in SRDC (2006) demonstrate that single parents responded to the SSP in dimensions other than job search, work hours and welfare receipt, the ones modeled here. Among these would be longer term responses to welfare in terms of education, fertility and mobility across regions. Keane and Wolpin (2002) and Kennan and Walker (2010) have analyzed these issues using forward-looking models estimated on non-experimental data. Finally, the results can be applied to answer more general questions than posed in this paper. Although the SSP treatment was not very successful in eliminating the welfare trap, it did inspire Canadian provinces in subsequent reforms. Quebec implemented a reform of IA closely related to the SSP (Lacroix 2010). Applying the model estimated here to data from the Quebec experience using methods described in Ferrall (2003) is in development.

REFERENCES

- ATTANASIO, O., MEGHIR, C. and SANTIAGO, A. 2011. "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to evaluate Progresa," accepted for publication, *Review of Economic Studies* .
- CARD, D. and HYSLOP, D. R. 2005. "Estimating The Effects Of A Time-Limited Earnings Subsidy for Welfare-Leavers," *Econometrica* 73, 6, 1723-1770.
- ECKSTEIN, Z. and WOLPIN, K. I. 1999. "Why Youths Drop out of High School: The Impact of Preferences, Opportunities, and Abilities," *Econometrica* 67, 6, 1295-1339.
- FANG, H. and SILVERMAN, D. 2004. "On the Compassion of Time-Limited Welfare Programs," *Journal of Public Economics* 88, 7-8, 1445-1470.
- FANG, H. and KEANE, M. P. 2004. "Assessing the Impact of Welfare Reform on Single Mothers," *Brookings Papers on Economic Activity* 35, 1, 1-116.
- FERRALL, C. 1997. "Unemployment Insurance Eligibility and the Transition from School to Work in Canada and the United States," *Journal of Business and Economic Statistics* 15, 2, 115-129.
- FERRALL, C. 2003. "Estimation and Inference in Social Experiments," working paper, Queen's University.
- FERRALL, C. 2005. "Solving Finite Mixture Models: Efficient Computation in Economics under Serial and Parallel Execution," *Computational Economics*, 25, 4 (June), 343-379.
- FLINN, C. 2006. "Minimum Wage Effects on Labor Market Outcomes under Search, Matching, and Endogenous Contact Rates," *Econometrica* 74, 4, 1013-1062.
- FORD, R., GYARMATI, D., FOLEY, K., and TATTRIE, D. 2003. "Can Work Incentives Pay for Themselves? Final Report on the Self-Sufficiency Project For Welfare Applicants," research report, Social Research and Demonstration Corporation.
- GLADDEN, T., and TABER, C. 2009. "The Relationship between Wage Growth and Wage Levels," *Journal of Applied Econometrics* 24, 6, 914-932.
- KAMIONKA, T. and LACROIX, G. 2003. "Assessing the Impact of Non-Response on the Treatment Effect in the Canadian Self-Sufficiency Experiment", CIRANO Working Papers 2003s-62.
- KEANE, M. P. and MOFFITT, R. 1998. "A Structural Model of Multiple Welfare Program Participation and Labor Supply," *International Economic Review* 39, 3, 553-589.
- KEANE, M. P. and WOLPIN, K. I. 2002. "Estimating Welfare Effects Consistent with Forward Looking Behavior, Part II: Empirical Results," *Journal of Human Resources* XXXVII, 3, 600-622.
- KENNAN, J. and WALKER, J. 2010. "Wages, welfare benefits and migration," *Journal of Econometrics* 156, 1, 229-238.
- LACROIX, G. 2010. "Assessing the Impact of a Wage Subsidy for Single Parents on Social Assis-

- tance," CIRANO Working Paper, 2010s-19.
- LISE, J., SEITZ, S. and SMITH, J. 2005. "Equilibrium Policy Experiments and the Evaluation of Social Programs," Queen's University, Department of Economics Working Paper, 1076.
- MILLER, R. A. and SANDERS, S. 1997. "Human Capital Development and Welfare Participation," *Carnegie-Rochester Conference Series on Public Policy* 46.
- MICHALOPOULOS, C., Tattrie, D., Miller, C., Robins, P.K., Morris, P. Gyarmati, D., Redcross, C., Foley, K. and Ford, R. 2002. "Making Work Pay: Final Report of the Self-Sufficiency Project for Long-Term Welfare Recipients," research report, Social Research and Demonstration Corporation.
- MOFFITT, R. 1983. "An Economic Model of Welfare Stigma," *American Economic Review* 73, 5, 1023-35.
- MOFFITT, R. 1992. "Incentive Effects of the U.S. Welfare System: A Review," *Journal of Economic Literature* , 1-61.
- RUST, J. 1994. "Structural Estimation of Markov Decision Processes," in *Handbook of Econometrics*, 4, R. Engle and D. McFadden (eds.), North Holland.
- SOCIAL RESEARCH DEMONSTRATION CORPORATION 2006. "Making Work Pay Symposium November 15-16, 2005," research report, available at www.srdc.org.
- SWANN, C. 2005. "Welfare Reform When Agents are Forward-Looking," *Journal of Human Resources* 40, 1, 31-56.
- TODD, P. and WOLPIN, K. I. 2006. "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, 5, 1384-1417.
- WOLPIN, K. I. 1995. *Empirical methods for the study of labor force dynamics*, Harwood Academic.
- ZABEL, J., SCHWARTZ, S. and DONALD, S. 2006. "An Econometric Analysis of the Incremental Impact of SSP Plus," SRDC working paper.