Explaining and Forecasting Results of
The Self-Sufficiency Project

Christopher Ferrall
Department of Economics
Queen’s University

November 5, 2007

Abstract:
This paper models the Self-Sufficiency Project (SSP), a controlled randomized experiment concerning welfare. The model of household behavior includes stochastic labor market skill, job opportunities, and value of non-labor market time. All the variation within and between treatment groups, jurisdictions (provinces), demographic groups, and sub-experiments is derived from four underlying sources: policy variation, endogenous selection into the experimental samples, the SSP treatments themselves, and different mixtures over 4 underlying types. Using the variation within the treatment group is quantitatively important for identifying the complex model: At the efficient GMM parameter estimates the standard errors of many parameters explode when based on only moments from the control groups. The model tracks the primary moments well except in the entry sample, and it matches out-of-sample outcomes not available for estimation. Predictions of the estimated model are computed for different welfare reform experiments. Counterfactuals suggest the SSP+ treatment has the most potential for generating long-run impacts.

Data access was based on a research contract with Social Research and Demonstration Corporation (SRDC). Research support from the Social Sciences and Humanities Research Council of Canada is also gratefully acknowledged. I thank Hartmut Schmider at HPCVL and Martin Siegert at Westgrid for help with resources and Doug Tattrie for comments on a preliminary draft. In addition I have had the benefit of comments and suggestions from too many audiences over the years to thank individually.
I. Introduction

The Self-Sufficiency Project was a controlled randomized experiment performed in two Canadian provinces designed to study whether long-term recipients of income assistance (i.e. welfare) respond to earnings subsidies. The main SSP treatment group, consisting of single parents on income assistance (IA) for at least one year, was offered a large supplement to earnings if and when a full-time job was acquired within one year (and the parent went off IA). Approximately 35 percent of the treatment group qualified for the supplement and at the peak that group had a 100 percent increase in full-time work and a 70 percent increase in earnings relative to the controls. However, most of the impact disappeared soon after the supplement expired (Michalopoulos et al. 2002). Because the SSP treatments induced employment increase did not result in substantial wage gains, the hoped-for self-sufficiency was not observed. Despite this, the careful and ambitious design of the experiment provides a unique opportunity to study labor market dynamics among low-income households.

This paper describes a forward-looking model of single parent households that predicts labor market outcomes before, during, and after the SSP experiment. The SSP creates exogenous variation in budget constraints and expected income that is used to identify a rich model. The model is estimated using GMM on the first 36 months of the (roughly) 54 months of experimental data. Results from the final 18 months are matched closely by out-of-sample predictions from the model.

With the model performing well in and out of the sample, counter-factual experiments are computed. The experiments are designed 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. But in other dimensions the long-term impact of slightly different enhancements to welfare

---

1 Most of the work analyzing the SSP is discussed and summarized by the authors themselves in SRDC 2006.
is much larger than in the data. It appears unwarranted to extrapolate the experimental impact to related welfare policies without reference to a model of behavioral response to the experiment.\textsuperscript{2} To determine the value of using the treatment for estimation, standard errors are computed within experimental groups. For many parameters of the model the standard errors explode when based only on the control samples, quantifying the notion that experimental variation helps identify a richer model than identified by control variation alone. The richer model, if validated in other ways, may then be applicable in environments farther than the sample than a weaker model. Surprisingly, re-scaled standard errors tend to fall when based only on the treatment groups. Thus, without accounting for the direct cost of offering the treatment, it appears to be more efficient to have a larger treatment group that control group when basing policy predictions on a model estimated from the experiment.

Low-income single parents face a number of constraints when trying to establish a long-term attachment to the labor market. The well-known static tradeoffs between income and leisure created by the welfare are illustrated in Figure 1. A household eligible for an amount IAB is precluded from IA if they take some forms of non-government outside support (denoted OS). In Canada in the 1990s earnings up to $200 per month could be set aside without a reduction in benefits. Thereafter benefits were replaced by earnings (in the figure 1-for-1). An indifference curve with optimal point A illustrates the disincentive to work more than part-time under this budget.

Figure 1 also displays the budget under the main SSP treatment as a solid red line. The parent always has the option of receiving IA, but receipt of the supplement requires they choose not to take IA. Thus, SSP income starts at OS and has a slope equal to the wage until the full-time work requirement is met. At this point, earnings are supplemented to be half the difference between actual earnings and 3.9 times full-time earnings on a minimum

\textsuperscript{2} A follow-up paper will use the estimated model to study a larger set of counter-factual policies and experimental designs.
The parent is indifferent between staying on IA at point A and working full time with the supplement at point B.

If the induced shift to full-time work generates wage growth (through channels such as learning-by-doing) then the untreated budget dynamically shifts up, shown as a mixed blue line in Figure 1. After treatment ends the supplement budget disappears and a new optimal choice is C. In this case the temporary SSP treatment can lessen the welfare trap.
through employment changes inducing sufficient wage growth.³

The depth of the welfare trap is difficult to measure because it depends on how a person’s current skill relates to their labor market history. An inference is required on what wages would be now if the parent had (counter-factually) worked more or worked more steadily in the past. Papers that address dynamic in welfare policy using non-experimental data include Miller and Sanders (1997), Swan (1998), Kennan and Walker (2003), Keane and Wolpin (2002), and Fang and Silverman (2003). This paper provides new evidence on what labor market policies can do to affect welfare dependency by confronting a model that captures many elements of previous papers with the large and complex variation in static and dynamic incentives created by the SSP. Exploiting this variation allows separate identification of the time-varying and heterogeneous effects of income and opportunity cost that underly patterns of welfare receipt. The experimental outcomes are used while simultaneously making a household’s eligibility for the experiment endogenous to the model. The results generated by the model are therefore applicable quite generally to policies that would fundamentally alter the duration and incidence of welfare receipt. The econometric specification is designed to be applied in other contexts without re-estimating most parameters and without requiring data that matches up to that available within the SSP.

II. The Environment

II.A States and Parameters

In the model a household’s situation each month outside the experiment is described

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.
by nine endogenous state variables:

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

(1)

The state variables are listed in Table 1 and described in more detail in the Appendix. Briefly the variables are indicators or indices for: the parent lost their previous job; the parent worked in the previous month; the earnings offer in the current job; the parent’s skill based on previous experience; the upper bound on working hours in the current job; the level of outside support; the opportunity cost of time spent outside the household; the observed demographic group, and the parent’s unobserved type (k). The first two variables, $l$ and $p$, do not affect the household’s decisions directly and are tracked as state variables in order to match SSP results on job loss and quits.

The endogenous state vector $\theta_{\text{end}}$ is contained in the overall state vector $\theta$, which concatenates five sub-vectors:

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

(2)

Each state variable belongs exclusively to one of these sub-vectors. The other sub-vectors are described as needed.

The demographic index $d$ varies across households but not over time for a given household. Characteristics treated as demographic in the SSP model are indicators for province of residence and whether the parent has two or more children. Each of the $D = 4$ demographic groups has a vector of policy parameters,

$$\Psi_p[d] \equiv (IAB_d \ SA_d \ CB_d \ MW_d).$$

(3)

The parameters are the maximum level of income assistance benefits, the income set-aside before benefits are clawed back, the claw-back rate on benefits, and full-time earnings on a minimum wage job. The parameter values were illustrated in Figure 1 and listed below in Table A.2.
The index for unobserved type, \( k \), is also fixed for a household and determines which of \( K = 4 \) vectors of exogenous parameters pertains to the household:

\[
\Gamma[k] \equiv ( \Upsilon_k \ \Pi_k \ \delta_k \ \rho_k ). \tag{4}
\]

Exogenous parameters that shift utility are contained in \( \Upsilon_k \). Parameters that shift the evolution of state variables are contained in \( \Pi_k \). The scalar \( \delta_k \) is the discount factor, and \( \rho_k \) controls smoothing of choice probabilities. The elements of \( \Upsilon_k \) and \( \Pi_k \) and the exact roles of \( \delta_k \) and \( \rho_k \) are described in the rest of this section.

Within demographic group \( d \) household type is distributed according to \( \Lambda[d] \), a vector with elements \( \lambda[d, k] \). For example, \( \lambda[1,2] \) is the second element of \( \Lambda[1] \) and equals the proportion of \( k = 2 \) households in group \( d = 1 \). Only the proportions and policy parameters \( \Psi_p[d] \) vary with \( d \). For example, no province dummy appears in the earnings-offer distribution. Instead, inter-provincial differences are generated by a different mixture of types across provinces. Also, the opportunity cost of labor market time does not depend directly on the number of children. Instead, one-child households can be a different mixture across unobserved types than 2+ households.

The exogenous vector contains all the estimated parameters:

\[
\theta_{\text{exog}} = ( \Lambda[1] \ \cdots \ \Lambda[4] \ \Gamma[1] \ \cdots \ \Gamma[4] ). \tag{5}
\]

There are \( N = 19 \) parameters in \( \Gamma[k] \) leading to a total of \( K(D + N) = 4(4 + 18) = 88 \) exogenous parameters. Free parameters are fewer because six parameters in \( \Gamma[k] \) are constrained to be equal across type. Accounting for this and the fact that elements of \( \Lambda[d] \) sum to 1 results in \( 6 + 4(19 - 6) + 3(4) = 70 \) parameters estimated from the data.

**II.B Actions**

Each month the parent chooses an action vector,

\[
\alpha \equiv ( m \ a \ i ) \in A(\theta), \tag{6}
\]
containing three variables: labor market hours, active job search, and acceptance of income assistance. The feasible set $A(\theta)$ imposes two restrictions. First, active job search while working is ruled out: $m > 0$ or $a = 1$, but not both. Second, the parent faces an upper bound on work hours: $m \leq u(b)$ where $b$ is a state variable specific to the current job. When the parent has no job, $b = 0$. With a part-time job they can only work less than $b = PT < 1$ hours relative to full-time. PT equals the value in the SSP treatment of 75%. When holding a full-time job ($b = 1$) the parent can chose to work fewer hours. A parent holding a job who does not work at all effectively quits and loses the option to work until a new job is found and accepted.

II.C Outcomes and Results

The combination of an action and a state, $(\alpha, \theta)$, is referred to as an outcome. The state next month, $\theta'$, is randomly determined by the transition $P\{\theta'|\alpha, \theta\}$ (fully described in the Appendix). Not all aspects of a household’s outcome can be observed by outsiders. Some states and some actions are unobserved. Understanding welfare and the incentive to work involves many hidden states, including skill, job quality, and leisure-income tradeoffs. The vector of measurements made from an outcome $(\alpha, \theta)$ is denoted $Y(\alpha, \theta)$. Only some endogenous state variables in the SSP model are directly available in $Y(\alpha, \theta)$: $l$, $p$, and $d$. The action variables $i$ and $m$ are observed but active job search ($a$) is not. The moments drawn from the data to estimate the model are based on $Y(\alpha, \theta)$ and are described later on.

II.D Utility

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

$$U(\alpha, \theta) = \text{Income}(\alpha, \theta) + \text{OS}(\alpha, \theta) - C(\alpha, \theta).$$  \hspace{1cm} (7)
In turn, income is the sum of earnings, income assistance payments, and SSP payments:

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

Under-reporting of income while on IA is allowed, so measured earnings are a fraction of TrueEarn. The components of (8) are defined in the Appendix. The second term in (7) is the sum of non-government transfers and additional utility (in dollar equivalent) from forgoing IA:

\[
\text{OS}(\alpha, \theta) \equiv (1 - i)s\xi IAB.
\] (9)

The transfer component of OS is support that, if accepted, disqualifies the parent for IA. 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 labor market status. A drop in s may push the parent back to receiving IA. Because OS includes foregone stigma (of the static form in Moffitt (1983)), maximum OS is expressed as a factor of IAB, the maximum amount of IA the household is entitled to. The parameter \(\xi\) is a positive exogenous parameter.

Three possible sets of feasible work hours (depending on \(b\)) are shown in Figure 2 as ranges along the x axis starting from the right at zero work hours \((m = 0)\). The value of \(b\) changes from one month to the next for various reasons. A non-working parent finds a job with probability \(p_j(\alpha, \theta)\), which will have an upper bound on \(m\) of either PT or 1. A working parent loses a job permanently with probability \(\pi_l\) and results in an upper bound of 0 next month. A working parent can quit by setting \(m = 0\). A parent who quits or is laid-off can immediately engage in job search \((a = 1)\), but a job offered that month begins the next month. Thus leaving or losing a job is matched to cases where the parent experiences at least one month not working. Job-to-job transitions are treated as the same job, and the model attributes growth in full-time equivalent earnings between contiguous jobs as skill acquisition.
The cost of labor market time,
\[ C(\alpha, \theta) = \nu (m + \kappa a)^{c(h)} W_{\text{max}}, \]
is expressed as a fraction of \( W_{\text{max}}, \) maximum possible earnings (defined later). It depends on working hours and search time when not working. Job-search is converted to work time by the exogenous parameter \( \kappa. \) The cost of full-time work is then \( \nu W_{\text{max}}. \)

The curvature of costs is determined by \( c(h). \) It shifts with \( h \) as described in the Appendix. Figure 2 illustrates the cost of time generated by \( C(\alpha, \theta). \) The x-axis is non-
market time expressed as a fraction of full-time employment. The y-axis is dollars per month. The discrete values of $m$ are indicated by vertical lines. When not working the parent can choose to search actively for a job and incur cost $\nu c(h)$, which is shown on the graph along the mixed (red) line.

A shifting preference for full and part-time work hours is represented by three different costs depending on the state variable $h$, which jumps to a new value each month with probability $\pi_h$. Costs rise slowly with $m$ when, for example, children are in school and part-time work has a low opportunity cost. Costs rise quickly with $m$ when, for example, children are young or sick or part-time care arrangements break down. When the value of $h$ jumps to a new value a working parent may change hours or quit and drop out of the labor market. Either change may induce a change in welfare receipt. A non-working parent may respond to a change in $h$ by beginning or ending active search.

II.E Skill, Job Search and Wages

Skill is expressed as a fraction of full-skill: $x \in \{1/4, 1/2, 3/4, 1\}$. From month to month $x$ can increase or decrease by $1/4$ with a probability that depends on labor market status. While working, skills accumulate with a probability $m\pi_a$. While not working, skills decrease with probability $\pi_d$. When $\pi_a = \pi_d = 0$ endogenous skill accumulation and depreciation are eliminated and $x$ becomes a permanent random effect for the parent.

The Mincer earnings function that relates skill to accumulated labor market experience assumes $\pi_a = 1 = 1 - \pi_d$. That is, the stock of skill accumulates linearly with experience and does not depreciate while not working. In other cases of $\pi_a$ and $\pi_d$ welfare spells caused by transient conditions can last longer than those conditions. The longer a parent is out of work the more likely skill has fallen. Wage offers fall and become less valuable relative to time spent in the household. If a job were taken, $x$ would eventually increase. But in the presence of IA (even with forward-looking behavior) the rate of endogenous wage growth may be too slow to make work pay.
Wages are expressed as full-time equivalent monthly earnings, denoted $W(\theta)$. Jobs have two characteristics, $b$ described above and an earnings offer $n$ that takes on 6 values. The offer $n = 0$ is a “dead-end” job that does not depend on skill and pays MW regardless of skill. Such job offers are a fraction $\pi_m$ of all jobs. Job offers with $n > 0$ come from by a discretized log-normal distribution with log-mean $\mu$ and log variance $\sigma$.

The wage function provides for an interaction between the minimum wage, skills, the distribution of offers and the subsequent growth of earnings. To explain, start with the simple case of $MW = 0$. Then, $W()$ collapses to a familiar log-linear form:

$$\ln W^0(\theta) = \mu + \sigma \Phi^{-1}(n) + \eta \ln x, \quad n > 0.$$  \hfill (11)

The offers are percentiles of the distribution with skill shifting the distribution. Each offer is equally likely and the parameter $\eta$ corresponds to the return to experience. In the $MW=0$ case $n = 0$ is not a real offer. When $MW > 0$ it is assumed that regular offers ($n > 0$) are not each associated with its own level of earnings. Instead, for a given $x$ let $\phi_x$ denote the fraction of the underlying distribution below MW:

$$\phi_x = \Phi\left[\frac{\ln(MW) - \eta \ln(x) - \mu}{\sigma}\right].$$  \hfill (12)

For the lowest $x$ the lowest two regular offers produce a wage of MW. Each offer occurs with probability of $(1 - \pi_m)\phi_x/2$. For next skill level ($x = 2/4$) only the $n = 1/6$ offer is at MW 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:

$$M(n, x) = B[n = 0] + B[x \in \{1/4, 2/4\} \& n \in \{1/6, 2/6\}] + B[x = 2/4 \& n = 1/6].$$  \hfill (13)

For other combinations of $n$ and $x$ the wage exceeds the minimum wage. Each offer is equally likely given $x$. Let $\tilde{n}(x)$ be the number of offers above MW,

$$\tilde{n}(x) = 3 + B[x > 1/4] + B[x > 2/4].$$  \hfill (14)
We arrive at the general expression for full-time earnings:

\[ W(\theta) = M(n, x)MW + (1 - M(n, x)) \left( x^n \exp \left\{ \mu + \sigma \Phi^{-1} (\phi_x + (1 - \phi_x) / \tilde{n}(x)) \right\} \right) \]  

(15)

\[ W_{\text{max}} \equiv \exp \left\{ \mu + \sigma \Phi^{-1} (\phi_1 + (1 - \phi_1) / 5) \right\}. \]

For a low-skill parent, minimum wage jobs differ in their growth potential. For some offers they will wait for two increases in skill before pay increases. For others only one increase is required.\(^4\) For high-skilled workers only \( n = 0 \) offers start at MW. All other offers will increase wages with the first accumulation of skill. Skilled workers may choose to quit a low-offer job to search for a better one (depending on such things as the job offer probability and the risk of losing skill). Even a dead-end job is not really a dead-end since it is assumed skills still accumulate on them. This allows for a pattern in which people respond to the SSP subsidy in terms of employment but may not be on track to self-sufficiency because it does not create a strong incentive to low-wage jobs with growth potential versus those without.

**II.F Value and Choices**

To recap, the exogenous parameters that determine utility and transitions are gathered into two vectors,

\[ \Upsilon \equiv (\beta \ \eta \ \kappa \ \mu \ \nu \ \sigma \ \zeta \ \xi) \]  

(16)

\[ \Pi \equiv (\pi_j \ \pi_m \ \pi_f \ \pi_h \ \pi_i \ \pi_d \ \pi_l \ \pi_s \ \pi_v). \]

Where: \( \beta \) is the rate of income reporting; \( \eta \) is the curvature in skill; \( \kappa \) converts job search into work time; \( \nu \) is the (scaled) income-equivalent cost of full-time work; \( \mu \) and \( \sigma \) determine the location and spread of wage offers; \( \zeta \) determines the variance in the curvature of

\(^4\) This can (loosely) be interpreted as the employer over-paying a worker whose productivity is below MW and then eventually under-paying them once their skills increase to recoup the loss, but no explicit model of bargaining or contracting is present in the model. See Flinn 2006 for such an analysis of minimum wages.
time-costs over time; and $\xi$ is the factor on outside support. The $\Pi$ vector includes all parameters that enter the transition from one period to the next: $\pi_j$ is the probability that active job search generates a job offer (in the absence of job-finding support); $\pi_m$ is the proportion jobs that are true minimum wage jobs; $\pi_f$ is the proportion of job offers that are full-time jobs; $\pi_h$ is probability that the curvature in time-costs change; $\pi_a$ is the probability that skills accumulate while working; $\pi_d$ is the probability that skills decline while not working; $\pi_l$ is the probability that a working parent loses their job exogenously; $\pi_s$ is the probability that outside changes; and $\pi_+$ is the parameter that determines the effectiveness of the SSP Plus treatment described later on.

The value of an outcome and the value of a state satisfy Bellman’s equation:

$$v(\alpha, \theta) \equiv U(\alpha, \theta) + \delta E[V(\theta')] = U(\alpha, \theta) + \delta \sum_{\theta'} P\{\theta' | \alpha, \theta\} V(\theta') \quad (17)$$

$$\forall \theta \in \Theta, \quad V(\theta) = \max_{\alpha \in A(\theta)} v(\alpha, \theta). \quad (18)$$

State-contingent choice probabilities are smoothed by logistic kernel with parameter $\rho \geq 0$:

$$\tilde{v}(\alpha, \theta) \equiv B[\alpha \in A(\theta)] \exp\{\rho[v(\alpha, \theta) - V(\theta)]\}$$

$$P\{\alpha | \theta\} = \tilde{v}(\alpha, \theta) / \sum_{\alpha'} \tilde{v}(\alpha', \theta). \quad (19)$$

Given $\theta$ and the choice probabilities the expected result vector is

$$E[Y | \theta] = \sum_{\alpha \in A(\theta)} P\{\alpha | \theta\} Y(\alpha, \theta). \quad (20)$$

Combining endogenous choice probabilities with exogenous outcome-to-state transitions generates the state-to-state transition, $P_s\{\theta' | \theta\}$. Based on this transition there exists an ergodic (stationary) distribution over the endogenous variables, conditional on the non-ergodic values $d$ and $k$ (see Ferrall 2003). Let $P_\infty\{\theta\}$ denote this distribution, which is the starting point for modeling the selection into the experiment.

III. The SSP Experiment
This section provides an overview based on the schematic representation in Figure 3. The Appendix provides technical details. The oval represents the set of all outcomes \((\alpha, \theta)\) outside the experiment (the state space). This is phase 0 of the experiment, the real world before random assignment to treatment. The space is partitioned into households which receive IA \((i = 1)\) or not \((i = 0)\), the only endogenous outcome related to selection into the SSP samples.

### III.A Experimental Samples and Treatment

Two distinct samples were studied in the SSP: the Recipient Study \((e = 2)\) and the Applicant Study \((e = 1)\). Each experimental sample had a control group, denoted \(g = G = 3\). Within the Recipient Study there was a main treatment \((g = 2)\) and a smaller SSP Plus treatment \((g = 1)\). The Recipient Study selected single-parent households that had been on IA for 12 out of the last 13 months. This is simplified in the estimation and the figure to 12 consecutive months on IA. Graphically, any sequence of 12 outcomes in the on-IA partition is eligible for the Recipient Study. Households assigned to the control group remain in the real world, and their transition from pre- to post-assignment status is reflected in a change from phase 0 to phase 6, which is the real world after random assignment.

Eligible households assigned to treatment leave the outside world and enter the treatment program. The Recipient Study starts in phase 2. Unlike phase 0/6, the treatment program is non-stationary. This is represented in Figure 3 as rectangular areas with a timeline below it.

The Applicant Study was conducted in British Columbia alone. To be eligible for treatment the household had to apply for IA after being off IA for at least six months. A feasible history for this sample is represented in Figure 3 by six connected points in the off-IA partition followed by a month in the on-IA partition of the outcome space. If assigned to treatment they enter in phase 1.
Phase 1 and phase 2 do not treat utility, only expectations about future utility. Thus, if households were not forward looking these phases would be identical to phase 6 and measurements would be identical (in distribution) in the treatment and control groups. The treatment in phases 2-5 lasted 3 years and was discussed in the introduction and illustrated in Figure 1.5

The SSP Plus Sample was a small treatment group selected with the same criteria as the Recipient Study in New Brunswick. Subjects were given the same supplement under the same terms, but in addition they were offered a set of employment services. This additional treatment is not represented in Figure 1 which presumes a single wage is already available. Instead, the potential impact of these services is captured in a change in the job offer probability relative to otherwise identical households. The model assumes that these services enhance active search by raising the probability of a job offer each period:

$$p_j(\alpha, \theta) = a \left[ \pi_j + B \left[ g = 1 \right] \pi_+ (1 - \pi_j) \right].$$

This leads to a change in the decision to search actively for a job and it changes what is an acceptable job offer. Active job search is treated as unobserved, and no attempt is made to measure whether an eligible subject took advantage of the services. Thus, $\pi_+$ is a measure of how effective the offer of the services is not how effective the services are given they are used.

III.B Conditioning Variables, Impact and Predictions

Since $Y(\alpha, \theta)$ does not include the full outcome, the analysis must condition measurements on less information than households have. A standard impact analysis would condition only on variables that are not functions of past behavior (given eligibility). That

---

5 In the model, treatment phases 3-5 are identical and could be collapsed into a single longer phase. In the experiment they are indeed separate phases, because in at most two months per year the recipient could receive the supplement when hours fell below the full-time requirement. Modeling this facet would require an additional state variable to track months below full-time.
Figure 3. Sample Selection and Progress of Treatment

[Diagram showing the sample selection and progress of treatment with various stages and outcomes, including pre- and post-treatment outcomes space, eligible for the Applicant Study, assigned to treatment, treatment ends, and subsidized placements.]
is,
\[ \theta_{\text{cond}} \equiv (t \ g \ e \ d). \] (22)

The variable \( t \) is experimental time. Usually this would simply be the number of periods since random assignment, but the SSP has two different studies in which subjects enter the same program of treatment at different stages. Thus, to coordinate measurements, \( t \) is set to be 0 at the beginning of phase 2, the later of the two points of entry. The expected measurement is
\[ E[Y(\theta_{\text{cond}})] = \sum_{\theta} \lambda^* (k|\theta_{\text{cond}}) \Omega\{\theta | k, \theta_{\text{cond}}\} E[Y|\theta], \] (23)

where (20) defines \( E[Y|\theta] \) as the expected outcome conditional on the subject’s information and the Appendix defines \( \Omega(\cdot) \) as the distribution over states observed at \( \theta_{\text{cond}} \) among type \( k \) households and \( \lambda^*(\cdot) \) as the selected proportion of type \( k \).

Let \( \hat{Y}(\theta_{\text{cond}}) \) denote the vector of average observed (empirical) results conditional on the exogenous variables \( \theta_{\text{cond}} \). Observed impact is the difference in mean results between a treated group and its control:
\[ \hat{\Delta}(\theta_{\text{cond}}) \equiv \hat{Y}(\theta_{\text{cond}}) - \hat{Y}(\theta_{\text{cond}|G}). \] (24)

The notation \( |_G \) means replace \( g \) in \( \theta_{\text{cond}} \) with \( G (=3, \text{the control group}) \). The model’s predicted impact is simply
\[ \Delta(\theta_{\text{cond}}) \equiv E[Y(\theta_{\text{cond}})] - E[Y(\theta_{\text{cond}|G})]. \] (25)

Some insights can be drawn from these expressions without reference to the particular model or experiment. While undergoing treatment the transitions are different from the real world, so the treatment group drifts away from its control group. Selection on unobservables is important if \( \lambda^*(k|\theta_{\text{cond}}) \) differs significantly from \( \lambda[k,d] \). Control groups are drifting as well, but they continue to follow the same transitions as outside the experiment. Their state distribution converges back to \( P_\infty \) but only given the underlying (permanent)
type. Based on observables the control group outcomes converge to a different mean than outside the experiment due to selection on unobservables. Ultimately, treatment ends and treated households begin to converge to the same distribution as the controls. So the impact of treatment in a finite-lived experiment is relative to a non-stationary distribution that is converging to the same distribution as the treatment group but at a different rate.

IV. Experimental Outcomes

IV.A Measurements

As shown in Table 2, 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 Sample includes 292 people in New Brunswick. Roughly one-half of the households had more than one child at the baseline. 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. The result vector is computed for each value of demographic, experimental, and treatment group and by experimental time $t$.

The 12 contemporaneous variables chosen for study are summarized in Table 4. The monetary variables include mean monthly earnings, mean monthly IA benefits received, and mean monthly SSP supplements received (when applicable). The means of earnings squared and IA squared are also matched because higher moments of these distributions help identify the wage offer distribution. The remaining six results in $Y(\alpha, \theta)$ are indicators

---

6. Attrition from the sample after the baseline interview is treated as an exogenous result independent of the subject’s situation and the SSP treatment. According to this assumption it is valid to use either all individuals reporting results in a given month or use only those individuals who remained in the sample throughout the measurement period. Not all subjects entered the experiment in the same calendar month, so in the 36-month data file there are some observations beyond the 30 and 36 month cut-offs. For a cell’s values to be included in this analysis, there had to be at least 50 observations.

7. The complete list of moments and cell sizes is reported in Table B.

8. There is a lag in receiving SSP supplements and IA benefits. SSP benefits received and recorded in
for labor market outcomes. The mean values are therefore proportions of subjects in the
given situation, including receiving any IA in the month, earning a wage within $.10 of
the current provincial minimum wage, not working this period due to quitting a job last
period; not working this period due to losing a job last period, working full time (according
to the SSP minimum), working part-time, and an interaction between receiving IA and
working either full- or part-time.

**IV.B Experimental Impact**

Table 5 reports relative impacts \(\Delta (\theta_{\text{cond}}) / Y (\theta_{\text{cond}}|G)\) for selected variables at differ-
ent values of \(t\). At \(t_0 + 1\) relative impacts are small, as would be expected with random
assignment. The only impacts that appear sizeable one month after assignment are 25%
responses in earnings and full-time employment in the NB2+ and BC1 groups. By month
13 (one month after the qualification period ends) the earnings impact varies between
32% and 128%. By month 24 relative impacts are generally below the earlier maximum
impact, but in many groups is still larger than the initial values. The relative impact on
IA receipt is generally smaller than on earnings. By month 24 anywhere between 8%
and 32% fewer subjects in the treatment groups are on IA than in the control group. The
impact in the Applicant Study at month 13 is in the same range. The relative impact of
the SSP treatments on the proportion of jobs at the minimum wage is typically negative
and smaller than the other impacts. That is, conditional on working full or part time, a
smaller proportion of the treatment groups are working at or near the minimum wage
than in the control groups. The differences are small when compared to the impacts on
full-time work itself, which range from 52% to 146% in the Recipient Sample.

The impact of the SSP treatment is not limited to mean values of the measured results.
The co-relationship between the variables also differs across treatment groups. Table 6

---

\*month \( t = 2 \) are, for the most part, based on outcomes in month \( t = 1 \). For IA the lag is often two
months. For this reason SSP and IA results are forwarded by one and two months so that they are (roughly)
contemporaneous with the situation that generated them. This adjustment is not perfect, but it appeared to
be the best fixed rule.
reports the matrix of simple correlations in seven of the results. The SSP Plus Sample was excluded and the four demographic groups were combined, leaving four entry/treatment groups. The main purpose of Table 6 is to compare the same correlation between treatment and control groups. In other words, to compare entries across the diagonal. In each of the four quadrants the signs of the correlations follow similar patterns, which is not surprising given that earnings must be strongly related to work hours and negatively correlated with IA receipt. When comparing correlations across treatments and controls we see only small differences in the Applicant Study. For example, the correlation between earnings and IA benefits among the treated is -.356. Among controls the same correlation is -.360. The difference in the correlations is substantially larger in the Recipient Study, and the number of observations greater (however they are measured). For example, the same earnings/IA correlations are -0.409 and -0.317, respectively. This is consistent with the model since treatment is milder among applicants than recipients. For a minimum of twelve months there is no direct impact of treatment on utility for recipients. The impact is felt solely through the eventual opportunity to qualify for the supplement, and this forward-looking impact is the same as that felt in the Recipient Study from the start of their post-assignment period. But for the applicants the impact is discounted by δ and by the uncertainty of finding a job. Thus the applicant treatment group will on average appear closer to its control group than the recipient treatment group. The one caveat is that the two groups are created by nearly opposite criteria applied to IA receipt. As long as the underlying model exhibits positive correlation in IA receipt then the cross-treatment difference in correlations will indeed be smaller in the Recipient Study. The presence of skill accumulation and depreciation, along with persistence in the other household states and the IA rules themselves combine to ensure some measure of persistence in IA receipt.

Table 6 suggests that analyzing each measured result (and impact) separately is inefficient in a statistical sense. That is, earnings, IA, and full-time employment are not separate outcomes that each requires a separate sequence of impacts. More importantly, the SSP
treatment is associated with differences not just in mean results but also in correlations across contemporaneous results. Even when not using individual-level panel data, the different movements in mean results across variables through experimental time contains important information about treatment.

V. The Estimated Model

The model is estimated using Generalized Method of Moments by imposing the conditions that the observed and predicted values of the conditional moment vectors be equal:

\[ \hat{\Delta}(\theta_{\text{cond}}) = \hat{E}[Y | \theta_{\text{cond}}] - E[Y | \theta_{\text{cond}}] = 0, \]  

(26)

for all vectors \( \theta_{\text{cond}} \) post random assignment. The interaction of \( d, g, e, \text{and } t \) with the twelve contemporaneous results contained in \( Y(\alpha, \theta) \) results in \( \text{????} \) total moments. The Appendix describes the estimation procedure including computation of the optimal weighting matrix. It also discusses how variation across samples, treatments, provinces, experimental time and elements of the measurement vector \( Y(\alpha, \theta) \) contribute to the identification of parameters of the model.\(^9\)

V.A Parameter Estimates

Table 7.1 - Table 7.4 report the estimated parameter vector. Since there are no coefficients on observed variables included in the parameters (as in, say, a Mincer earnings function) many of the parameters are difficult to compare with other results. Many of the values are probabilities but judging their magnitude depends in part by the number of

\(^9\) One can treat a conventional impact analysis as estimating each difference between treatment and control, \( \hat{\Delta}(\theta_{\text{cond}}) \), with a free parameter (the predicted impact is the observed impact). Meanwhile, the estimated model generates impact as the difference between two of the model’s predictions without adding new parameters. Thus, the model estimated using GMM can be seen as a nested hypothesis within the unrestricted impact analysis. From this point of view, an impact analysis has as many parameters as moments and has no power to predict out of sample. The estimated model is parsimonious, with only 72 free parameters.
values the state variables take on. For these reasons the discussion of the parameters is short whereas as discussion of the model predictions is extensive.

The estimated mixing probabilities in Table 7.1 show that one or two of types are predominant in each demographic group. In fact, one could label type 2 the New Brunswick type. Types 1 and 3 are both common in the BC groups. Type 4 is only significant in NB. Within province the type proportions do not differ much between numbers of children. This implies that variance within province is explained by responses to the different welfare levels by family size (IAB). The dynamic programming parameters in Table 7.2 indicate that the types are very different in patience. A period is one month, so only the first two types $\delta_k$ close to 1 place a large amount of weight on the future. The other types make decisions close to a static manner: next year’s outcomes have essentially no impact on today.

Wage offer distributions differ greatly across types (Table 7.3) as does the stigma associated with welfare (captured by the coefficient on outside support, $\chi$). Full-time work has a very similar cost across type ($\nu$), but recall that this value is relative to maximal earnings for that type. This contrasts with the cost of active job search, which is only large and precisely estimated for type 1. Returns to skill and the convexity in household costs are difficult to interpret, but the income reporting parameter $\beta$ is straightforward. To rationalize the high fraction of work while on welfare (with low wage growth) the model has to have only 38% of earnings reported while on welfare.

The last panel of the parameter estimates (Table 7.3) include the transition shifters. Here we see that type 1 is constrained by job offers but the other types are not. Most offers are full-time, so the high fraction of part-time work reflects a choice to work fewer hours than the job allows. In each group about 20% of job offers are true minimum wage jobs (with no on-the-job growth potential). Estimates of the home environment indicate that outside support is highly persistent ($\pi_s$ is small) but household costs of work and job search is not ($\pi_h$ is high). Skills exhibit fairly quick rates of accumulation while working.
For all types, skill $x$ is likely to reach its maximum within a year of full-time employment. Because average wages do not accumulate in the treatment group, this suggests that the return to skill ($\eta$) 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 impressionistic explanation 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 the large fraction of jobs have no growth potential and the low wage elasticity to skill accumulation.

**V.B Fit to Selected Moments**

*V.B.1 Earnings (earn)*

Figure 5a and Figure 5b

*V.B.2 Government Transfers (IA+SSP)*

Figure 6a and Figure 6b

**V.C Heterogeneity, Sample Selection and Policy Variation**

Figure 7a and Figure 7b

**V.D Using Treatment Outcomes for Identification vs. Validation**

The estimated standard errors reported in Table 7 suggest that many of the parameters are precisely estimated by the variation in moments generated by the experiment. The parameters are identified because the model places many 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. Todd and Wolpin (2006) and
Lise et al. (2003) follow this approach by estimating models of forward-looking agents on control groups within experiments (Progress and the SSP, respectively) and then using the experimental data for out-of-sample validation. A major advantage of this approach is that behavior under the treatment does not have to be solved repeatedly while estimating the parameters. The potential cost of this approach is that the model that can be estimated from the control group alone may 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 behavior in populations facing similar but not identical environments.

To see quantify this potential cost of not using the experimental variation for estimation, the standard errors for the parameters were re-computed using only the moments in the control groups. The results were re-scaled to mimic a sample of the original size. Table 8 reports the results. Standard errors based on all the data are compared to those from the control and treatment groups alone. The values in the “all” column are not quite the same as in Table 7 because standard errors for the transformed variables (that impose constraints on parameter ranges) are used. It is not surprising that throwing out the experimental variation increases the standard errors. However, for roughly half the parameters the standard error is more than ten times larger than when based on all the data. Included among these are key parameters for understanding dynamic behavior of low-income households: the discount factor (δ), the wage offer parameters (µ and σ), the return to skill (η) and many of probabilities that determine persistence in wages and other states. Thus, it seems likely that if the validation strategy had be used here the model that could have been estimated from the control data alone would have been much simpler in form and would have likely not captured some of the detailed patterns in the experimental outcomes.

Another result is revealed in Table 8 when the “all” column is compared to the “treat” column. In nearly all the cases the re-scaled standard errors are smaller when based on the
treatment groups alone. This counter-factual throws out the variation between treatments and controls and replaces it with more information on the experimental variation. The reduction is not trivial. In many cases the standard error is reduced by 25% or more. This has a perhaps surprising implication. When using social experiments solely to study impact ($\Delta (\theta_{\text{cond}})$) a control observation and a treatment observation have equal weight. Thus absence of other costs splitting the overall sample evenly is a reasonable design.

Table 8 suggests that this logic does not hold when experimental variation will be used to identify an underlying model. In this case impact is not the only outcome of interest and an additional treated observation may be more valuable than an additional control observation. In an experiment like the SSP a treated observation is expected to cost much more than a control observation. But when fixed costs of setting up experiments are considered it is quite likely than the optimal design of an experiment, for the purposes of estimating economic models, would have more treatment observations than control observations.

**V.E Out of Sample Predictions**

Figure 8 compares the model predictions to data from the end of the experiment that was not available for estimation. The micro data is not available any longer, and neither is the same demographic categories used in the estimation. So the model predictions are averaged to compute province-level predictions. Not surprisingly, the impacts seen to fade in the previous figures continue to fade toward zero as treatment ends and all subjects return to the status quo. Also unsurprising is the model’s similar prediction and its missing of the level of earnings in the applicant sample. However, one pattern that is intriguing is that the impact of the SSP+ continues to lie above the regular impact even after treatment ends. The impact decays more in the model but it also produces a lasting impact of the extra help in the SSP+ program. This is explored further in the next section.
VI. Policy Experiments and Other Implications

This section conducts experiments that explore the implications of the SSP for counter-factual policy questions. In the figures the results of the hypothetical changes are compared not with the data but with the model predictions based on the SSP experiment.

VI.A Experiment 1: Missing Samples

VI.B Experiment 2: Stock vs. Flow Sampling

How does model output differ for slightly different samples. For the Recipient Study we consider a sample of single parents who are on IA for exactly 6 months rather than 12 or more months. A practical reason for the or more clause is that it creates a large population to draw from and it includes long-term welfare recipients. On the other hand, if the SSP were implemented it would not be long until the people qualifying for it would only be on IA for twelve months. The stock of long-term recipients without the benefit of the SSP would no longer exist. Perhaps an experiment on the flow into the long-term recipient pool would more closely reflect results of a SSP policy after an initial transition period. Because the long-term response is so low in the recipient sample an entry condition of just six months on welfare is reported. Note that this is an out-of-sample change since many parents meeting this condition would not meet the twelve-month rule. A similar change is made to the Applicant Study. People newly applying to IA after one month or more off as opposed to six months or more. This is also outside the experimental sample since it includes people who would not be eligible for the SSP Applicant Study.

Figure 9

VI.C Experiment 3: Short-Lived, Large Flat Bonus

VI.D Experiment 4: SSP Treatment with 20% Reduction in IAB
VII. Conclusion
VIII. Technical Appendix

VIII.A Details of the Model

Components of Income

\[ \text{TrueEarn}(\alpha, \theta) \equiv mW(\alpha, \theta) \quad (A27) \]
\[ \text{Earn}(\alpha, \theta) \equiv (1 - \beta_i) \text{TrueEarn}(\alpha, \theta) \quad (A28) \]
\[ \text{IA}(\alpha, \theta) \equiv \max \left\{ \text{IAB} - \beta \text{CB} \min \left\{ \text{Earn}(\alpha, \theta) - \text{SA}, 0 \right\}, 0 \right\} \]
\[ Q(\alpha, \theta) = B \left[ i = 1 \& 2 \leq f \leq 5 \& m > \text{PT} \& W(\theta) \geq \text{MW} \right] \quad (A29) \]
\[ \text{SUP}(\alpha, \theta) = Q(\alpha, \theta) \max \{ 0, (1 - \text{TB}) \left[ \text{UL} \times \text{MW} - \text{Earn}(\alpha, \theta) \right]\} \quad (A30) \]

Endogenous Variables. To describe the transition for each variable, let \( q' = q^* (\bar{q}, \{\pi_j\}, \{Q_j\}) \) denote a discrete variable \( q \) that has a default value of \( \bar{q} \) next period and can then jump into one \( j \) different sets of values with probability \( \pi_j \) (not the same as the model parameter). Conditional on jumping into \( Q_j \) each element of the set is equally likely.

S1. Unobserved Type: \( k \in \{1, 2, 3, 4\} \)
\( \circ \) Role: index into \( \Gamma \) and the mixing distribution \( \Lambda \).
\( \circ \) Transition: \( k' = k^*(k, 0, \emptyset) \)

S2. Observed Type: \( d \in \{1, 2, 34\} \)
\( \circ \) Role: index into the policy vector \( \theta_{pol} \) and the mixing distribution \( \Lambda \).
\( \circ \) Transition: \( d' = d^*(k, 0, \emptyset) \)

S3. Household Time Cost: \( h \in H = \{1/4, 2/4, 3/4\} \)
\( \circ \) Role: determine the curvature of the time-cost function.
\( \circ \) Auxiliary Equations
\[ c(h) = -\zeta \ln (1 - h). \quad (A31) \]

The right hand-side is the inverse exponential distribution with decay rate \( 1/\zeta > 0 \). The value of \( c(h) \) determines the convexity of costs for labor market activity less than full-time. For values of \( c(h) < 1 \) the cost function is concave for feasible labor market time, creating a tendency to prefer part-time work. On the other hand, costs are convex when \( c > 1 \), which creates a tendency either to stay at home or work full time.
\( \circ \) Equation in Text: (10)
\( \circ \) Transition: \( h' = h^*(h, \pi_h, H) \)

S4. Outside Support Opportunities: \( s \in S = \{0, 1/3, 2/3, 1\} \)
\( \circ \) Role: determines the cash-equivalent amount of support available to the parent that, if accepted, disqualifies the parent from IA.
\( \circ \) Equations in Text: (9)
\( \circ \) Transition: \( s' = s^*(s, \pi_s, S) \)

S5. Upper bound on working hours: \( b \in \{0, \text{PT}, 1\} \)
○ Role: constraint on work hours in current job
○ Auxiliary Equations: See feasible actions below
○ Transition:
\[
\begin{array}{c|c|c}
\alpha, \theta & b' & P\{b'|\alpha, \theta\} \\
0 & 0 & B[m > 0] \pi_t \\
1 & (1 - p_j(\alpha, \theta))B[b < 2] + B[m = 0] \\
2 & p_j(\alpha, \theta)\pi_f \\
3 & p_j(\alpha, \theta)\pi_f \\
b & B[m > 0](1 - \pi_t). \\
\end{array}
\]
(A32)

### S6. Accumulated Skill:
\[ x \in \{1/4, 1/2, 3/4, 1\} \]
○ Role: level of earnings and future growth potential
○ Equations in Text: (11), (12), and (15)
○ Transition:
\[
x' = x^*(x, [m\pi_a + B[m = 0]\pi_d], [\min\{\max\{1/4, x + B[m > 0]/4 - B[m = 0]/4\}, 1\}])
\]
(A33)

### S7. Wage Offer:
\[ n \in \{0\} \cup N = \{1/5, 2/5, 3/5, 4/5\} \]
○ Role: search-sensitive component of wages
○ Equations in Text: (11)-(15)
○ Transition: with \( MW = 0 \),
\[
n'(\alpha, \theta) = n^*(n, \{ap_j\pi_m, ap_j(1 - \pi_m)\}, \{0\} N].
\]
With \( MW > 0 \)
\[
n'(\alpha, \theta) = n^*(n, \{ap_j\pi_m, ap_j(1 - \pi_m)\phi_x \ ap_j(1 - \pi_m)(1 - \phi_x)\], \{0\} \{1/6, \ldots, (5 - \tilde{n}(x))/6\} \{(6 - \tilde{n}(x'))/6, \ldots, 5/6\}].
\]
(A34)

Note that the distribution of \( n' \) depends on the contemporaneous state through the value of \( x' \). So between periods \( x' \) must be determined before \( n' \).

### S8. Job Loss:
\[ l \in \{0, 1\} \]
○ Role: exogenous loss of job.
○ Transition: \( l' = l^*(0, B[m > 0][1, \{1\}] \)

### S9. Employed Previously:
\[ p \in \{0, 1\} \]
○ Role: tracks whether the person worked last period (with \( l \) can infer the parent quit).
○ Transition: \( p' = l^*(B[m > 0], 0, \emptyset) \)

**Actions.**

### A1. Labor market hours:
\[ m \in M = \{0, 1/4, 1/2, 3/4, 1\} \]
○ Equations in Text: (10)

### A2. Active Job Search:
\[ a \in \{0, 1\} \]
○ Equations in Text: (21), (10)
Table A.1. Program of Treatment

<table>
<thead>
<tr>
<th>f</th>
<th>phase name</th>
<th>R(f)</th>
<th>$f_n=0$</th>
<th>otherwise</th>
<th>default at $r=R(f)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>pre-random assignment</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>entry</td>
<td>12</td>
<td>$f_n=5$</td>
<td>$f_n=0$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>qualification for SSP</td>
<td>12</td>
<td>$i=1$</td>
<td>$f_n=3$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>year 1 of eligibility</td>
<td>12</td>
<td>automatic</td>
<td>$f_n=0$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>year 2 of eligibility</td>
<td>12</td>
<td>automatic</td>
<td>$f_n=0$</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>year 3 of eligibility</td>
<td>12</td>
<td>automatic</td>
<td>$f_n=0$</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>post-treatment</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**A3. Accept Income Assistance:** $i \in \{0,1\}$

○ Equations in Text: (9), (28), (29)

**Feasible Actions**

$$A(\theta) \equiv \{(m \ a \ i) \in \mathbb{M} \times \{0,1\} \times \{0,1\} : m < b \& ma = 0\}.$$  \(A35\)

**VIII.B The SSP Experimental Design**

A subject’s status in the treatment program is defined by the sub-vector $\theta_{\text{clock}} = (t \ r \ f)$, where $f$ is the current phase of treatment, $r$ is the number of periods the subject has resided in that phase, and $t$ is experimental time, which is defined below. The SSP program of treatment is defined by a vector of parameters, $$\Psi_t[g] = (R[1] \cdots R[5] \ f_n(y) \ PT \ TB \ UL).$$ \(A36\)

In the SSP experiment there are seven phases, numbered from 0 to $F = 6$. Both $f = 0$ and $f = 6$ correspond to the real, non-experimental world, before random assignment (0) and after treatment has ended (6). By definition, control groups ($g = 3$) transit immediately from phase 0 to phase 6. The treatment groups transit from phase 0 to the initial phase for their treatment group (listed in Table A.1). Ultimately they reach phase 6 as well. Phase 1 is the entry phase, where a parent must remain on IA for twelve months to get a chance to qualify for the SSP treatment. Phase 2 is the qualification period in which the parent becomes eligible for the SSP supplement if and when they begin a full-time job. They remain eligible for the supplement during phases 3 to 5. $R[f]$ is the maximum duration of treatment phase $f$. Since each phase of the SSP lasts at most 12 months, $R[f] = 12$ for $f = 1, 2, \ldots, 5$. The parameter $f_n(y)$ is shorthand for a set of deterministic transition rules for next period’s phase. In other words, it describes how the SSP treatment progresses. Table A summarizes the selection, assignment, and transition rules in the SSP.

The remaining elements of $\Psi_t$ are parameters that determine the value of the SSP supplement, $\text{SUP}(\alpha, \theta)$, which enters utility defined in (A7) through income defined in (A8). The full equation for $\text{SUP}(\alpha, \theta)$ appears in (A30). The red line in Figure 1 that passes through OS and 2.9MW+OS illustrates the effect of the supplement on the household budget.
The treatment variables $r$ and $f$ are not useful for coordinating observations across groups. For example, one parent may take 8 months to leave phase 2 while another may take only 4 months. After seven months the first parent’s clock would read $(7, 2)$, the second $(3, 3)$, and for all parents assigned to the control group it would read $(1, 6)$. And the values of $r$ and $f$ are meaningless for parents assigned to control groups. To make results generated by the model compatible across groups a separate data clock, $t$, tracks the experimental month at which a measurement is taken.

VIII.C The SSP samples

A subject’s treatment group in the SSP is indexed by the sub-vector $\theta_{\text{exp}} = (e, g)$, where $e$ is the experimental sample and $g$ is the randomly assigned treatment status within samples. The Recipient Study $(e = 3)$ includes parents that had been on IA for at least one year. The Applicant Study $(e = 2)$ includes parents initiating (applying for) a new spell of receiving IA after a period of at least six months without IA. The treatment variable $g$ takes on three values. Besides a control group $(g = 3)$ and a treatment group $(g = 2)$, the separate SSP Plus group $(g = 1)$ was offered job-search and employment services in addition to the SSP supplement. Each treatment group has associated with it an initial post-assignment clock setting, a pre-assignment selection period and a sequence of feasible histories.

\[ \Psi_x[e] = \left( \tilde{\theta}_{\text{clock}} T H[y; \theta_{\text{cond}}] \right). \]  

(A37)

The elements of $\Psi_x$ are listed in Table 2. To make measurements consistent across groups the experimental clock $t$ must be coordinated. The time $t_0$ corresponds to the point of random assignment in the group and is normalized to 0 in the group that enters the program of treatment last. Thus $t = 0$ at the beginning of the qualification phase $(f = 2)$ which is when the Recipient Study $(e = 2)$ is randomly assigned.

Prior to $t_0$ is the period of sample selection. For the Recipient Study this period is of length $T = 12$ and stretches back to $t_{\text{min}} = -11$. It requires the parent receive IA each period, so only outcomes with $i = 1$ are feasible during this time. The Applicant Study $(e = 1)$ is randomly assigned at $t_0 = -11$ and the selection period is $T = 7$ periods long, extending back to period $t_{\text{min}} = -17$. In the first six periods the feasible condition is $i = 0$, and the last period is the condition $i = 1$, the start of a new spell of receiving IA. One fine point is that after random assignment the Applicant sample has already spent one month on IA and requires only eleven more months to enter phase 2. Therefore the initial clock setting has $r = 2$. Formally the selection criteria can be represented several different ways. Table 2 represents them as a 0/1 indicator for a measurement vector $y$ that survives a period of selection. The indicator is denoted $H[y; \theta_{\text{cond}}]$ and it takes on either the $i$ component of the measurement vector or its complement $\sim i = 1 - i$ depending on time period and the entry sample.

With all of the policy vectors introduced the policy sub-vector defined as

\[ \theta_{\text{pol}} = (\Psi_p \quad \Psi_x \quad \Psi_t) \]  

(A38)

are summarized in Table A.2.

With all transitions defined, the primitive transition function as

\[ P \{ \theta' | \alpha, \theta \} = \prod_{q \in \theta} \left[ B[q' = \bar{q}] \left( 1 - \sum_j \pi_j \right) + \sum_j B[q' \in Q_j] \frac{\pi_j}{\# Q_j} \right]. \]  

(A39)
Table A.2. Policy Vectors Contained in $\theta_{pol}$.

<table>
<thead>
<tr>
<th>d</th>
<th>IAB</th>
<th>MW^a</th>
<th>SA</th>
<th>CB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$712</td>
<td>$650</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$755</td>
<td></td>
<td>$200</td>
<td>100%</td>
</tr>
<tr>
<td>3</td>
<td>$982</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$1,175</td>
<td>780</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>d</th>
<th>PT</th>
<th>TB</th>
<th>UL</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0%</td>
<td>0%</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>75%</td>
<td>50%</td>
<td>3.90</td>
</tr>
<tr>
<td>1</td>
<td>75%</td>
<td>50%</td>
<td>3.90</td>
</tr>
</tbody>
</table>

$\Psi_s[e]$ $\Psi_s[d]$ $H[y;\theta_{cond}]$

This notation means to take the product over all state variables $q$. Each state contributes the probability that it takes on the value in $\theta'$, denoted $q'$, conditional on $P\{\theta'|\alpha,\theta\}$. This is computed by finding the jump set that $q'$ is in (if any) and adding the default probability if $\tilde{q} = q'$ at $P\{\theta'|\alpha,\theta\}$.

VIII.D Conditional Distributions

To compute the selection into sample $e$ by group $d$, begin by setting $t = t_0 - T + 1$ and $g = 3$, which determines the value of the conditioning vector $\theta_{cond}$. Choose an unobserved type $k$ and use the corresponding ergodic distribution as the starting value:

$$\Omega\{\theta' | k, \theta_{cond}\} = P_{\infty}\{\theta\}.$$  

Initialize the selected weight of type $k$ to one:

$$\omega(k; \theta_{cond}) = 1.$$  

During selection the feasible choices are imposed on the choice probabilities.

$$P^*\{\theta'|\theta\} = \sum_\alpha P\left\{\theta'|_B[t=t_0]\tilde{\theta}_{clock} | \alpha, \theta\right\} \mathcal{H}[Y(\alpha, \theta); \theta_{cond}] P\{\alpha | \theta\}. \tag{A40}$$

The notation $|_x$ means to set elements of the state vector to $x$ holding other elements constant. The condition $B[t=t_0]$ means this only happens at time $t_0$. In other words, subjects make their last choice before random assignment ignorant of the experiment. Then during the transition to the next month’s state the experiment those in a treatment group have their clocks reset to the initial clock for that experimental sample. They
‘wake up’ in the program treatment with all other states determined by choices before the experiment.

Working recursively forward in time first compute the fraction of type \( k \) households that make it to the next period:

\[
\omega \left( k; \theta_{\text{cond}} \mid t + 1 \right) = \omega \left( k; \theta_{\text{cond}} \right) \sum_{\theta'} \sum_{\theta} P^* \left\{ \theta' \mid \theta \right\} \Omega \left\{ \theta \mid k, \theta_{\text{cond}} \mid t \right\}.
\] (A41)

The proportion of the unselected population that is eligible for assignment may become very small. Thus, the distribution across states is updated and re-normalized to sum to one:

\[
\Omega \left\{ \theta' \mid k, \theta_{\text{cond}} \mid t + 1 \right\} = \frac{\omega \left( k; \theta_{\text{cond}} \right)}{\omega \left( k; \theta_{\text{cond}} \mid t + 1 \right)} \sum_{\theta} P^* \left\{ \theta' \mid \theta \right\} \Omega \left\{ \theta \mid k, \theta_{\text{cond}} \mid t \right\}.
\] (A42)

Once \( t + 1 = t_0 \) we have the distribution eligible for random assignment. All of these calculations can be done independently (in parallel) across both \( d \) and \( k \). But once generating the predictions after random assignment the type-specific distributions must be adjusted:

\[
\lambda^* \left( k \mid \theta_{\text{cond}} \right) \equiv \lambda \left[ k, d \right] \frac{\omega \left( k; \theta_{\text{cond}} \right)}{\sum_{k'} \omega \left( k'; \theta_{\text{cond}} \right)}.
\] (A43)

Since the clock was set properly at \( t_0 \), the updating rules (40)-(42) apply for \( t > t_0 \) as well. Since all actions are feasible after random assignment in the SSP, \( \omega \left( k; \theta_{\text{cond}} \mid t + 1 \right) \) becomes constant and correction factor on \( \Omega \left\{ \theta' \mid k, \theta_{\text{cond}} \mid t + 1 \right\} \) becomes one. This assumes that attrition is uncorrelated with unobserved types (and unobserved states).

**VIII.E Solving the Model and Computing Predictions**

The size of the model and some technical details of the solution are listed in Table 1 and Table A. The size of the system is notable. Even though each endogenous state variable is restricted to a small set of values an individual subject can be in one of 2,304 states outside the experiment. The post-treatment infinite horizon problem requires convergence of the value function at these points, although some points in the state space are, from the subject’s point of view, redundant and do not require re-solving the maximization problem (A17). For example, the household is not affected by the values of \( l \) and \( p \), and a currently unemployed worker \(( b = 0 )\) does not care about values of \( n \).

Since a stationary distribution \( P_{\infty} \) over states is computed, 16 different linear systems of size 2,304 must be solved on each iteration of the model. The SSP program of treatment adds 60 additional values of \( f \) and \( r \). With the separate SSP Plus treatment and Applicant sample over 4 phases leads to 51,840 total states for an individual. In keeping track of all states while tracking experimental results a total of 6,672,384 different combinations are possible. Up to 12 actions are available at each state. When aggregating over all states (including demographic, unobserved, and equivalent variation) the result is an outcome space of size 80,068,608.

The value function (A18) is solved to a level of precision under the infinite horizon. Evaluating the model ‘from scratch’ takes a bit more than an hour using a single processor.
of a high-end server. The required time is sensitive to the size of the discount factor $\delta$. This cost can be cut by roughly $1/(16)$ through the use of 16 processors to solve in parallel the separate problems defined by $d$ and $k$. Further substantial savings occur when computing numerical gradients by taking account of the limited interactions across parameters implied by a finite mixture model (Ferrall 2005). These savings are essential to making the model feasible to solve. With the computing resources currently available a full iteration of the BFGS algorithm can completed in approximately an hour.

**VIII.F GMM Estimation Procedure**

**VIII.F.1 First Stage**

The weighted discrepancy between the data and the model used in the first stage is

$$Z^1 (\theta_{\text{exog}}) \equiv \sum_{d=1}^{4} \sum_{e=1}^{2} \sum_{g=1}^{3} \sum_{t = t_0(e)}^{t_{\text{max}}(e)} \frac{n(\theta_{\text{cond}})}{265159} \Delta (\theta_{\text{cond}}) \Sigma_0 \Delta (\theta_{\text{cond}})^\prime , \quad (A44)$$

where $\Sigma_0$ is a $12 \times 12$ diagonal matrix with elements listed in Table 4. For the monetary values the weights are the inverse of the grand mean of the moment over conditioning states. For the binary variables a weight of $1/0.5 = 2$ was chosen to avoid putting excessive weight on turnover values which are near 0 and noisy across months. The cell sizes $n_{\text{cond}}$ (in Table B.11) sum to 265,159 in (A44). The Appendix discusses how variation across samples, treatments, provinces, experimental time and elements of the measurement vector $Y(\alpha, \theta)$ contribute to the identification of parameters of the model.

Let $\hat{\theta}_{\text{exog}}^1$ denote the parameters chosen to minimize $Z^1$. From these estimates the covariance matrix of the moments was computed. Given the random assignment to groups and the assumption that demographic groups are different (exogenous) mixtures across types, the moments are uncorrelated across groups defined by $e$, $g$ and $d$. That is, the sequence of observed vectors $Y(\alpha, \theta)$ for an individual is correlated, but across entry, treatment and demographic groups the sequence of individual shocks are independent. The population covariance matrix of moments is block diagonal with non-zero entries only across $t$. There are ?? blocks varying in size between ?? and ???. To compute the covariance matrix individual paths of $Y$ were simulated from $\hat{\theta}_{\text{exog}}^1$ by following the model. First, initial states were drawn from the ergodic distribution. Then a choice was drawn from the choice probabilities and a transition was drawn from the primitive transition. Let $Y_r^t(\theta_{\text{cond}})$ denote the $r$th simulated path with vectors concatenated across experimental time $t$. Then the deviation of the path from the mean $E_t[Y | \theta_{\text{cond}}]$ was computed and weighted by the endogenous type proportion for the type within the sample. The outer product of the vector of deviations was computed and averaged across simulations. The resulting matrix is a consistent estimate of the covariance of the block of moments for the group $\theta_{\text{cond}}$. The inverse of the matrix was then computed for each block:

$$\Sigma(\theta_{\text{cond}}) = \left(\frac{1}{R} \sum_{r=1}^{R} \sum_{k=1}^{4} \lambda^*(\theta_{\text{cond}})(Y_r^t(\theta_{\text{cond}}) - E_t[Y | \theta_{\text{cond}}])(Y_r^t(\theta_{\text{cond}}) - E_t[Y | \theta_{\text{cond}}])^\prime \right)^{-1} .$$
Steps in Computation.
A0. Set $\theta_{\text{exog}} = \theta_{\text{exog}}^0$ and call an optimizer to minimize $W(\theta_{\text{exog}})$.
A1. To evaluate $W(\theta_{\text{exog}})$: Set $d = D$.
A2. Solve completely for one group $d$. Set $k = K$.
B0. Solve for behavior. Set $f = F$, $r = 1$, $g = G$, $e = E$.
   C0. Iterate on $V(\theta)$ in 18 to convergence.
   C1. Once converged, loop one more time over $\theta_{\text{end}}$ to compute choice probabilities ($P\{\alpha | \theta\}$ in 19) and $E[Y | \theta]$.
   C2. Solve the linear system that defines $P_{-\infty}$ for $k$ and $d$.
C3. Solve for the endogenous sample in entry group $e$. Set $t = t_{\text{min}}$.
   D0. From $P_{\infty}$, compute the first value of $\omega(k; \theta_{\text{cond}})$ and $\Omega\{\theta | k, \theta_{\text{cond}}\}$.
   D1. Increase $t$ by 1. Update $\Omega$ and $\omega$ by looping through all transitions.
   D2. Repeat previous step until $t = t_0$.
   D3. Store $\Omega$ to be used for all $g$ given $e, k, d$.
C4. Solve for behavior under treatment. If $g = G$ set $f = 0$ and skip this part.
   E0. Decrease $f$ and set $r = R[f]$.
   E1. Solve for $V()$, choice probabilities, and $E[Y | \theta]$.
   E2. Decrease $r$ by 1. Return to E1 until $r = 0$.
   E3. Repeat the previous two steps until $f = 0$.
C5. Compute expected outcomes given $k$. Set $t = t_0$ and restore $\Omega$.
   F0. Loop through $\theta_{\text{end}}$ and setting the clock to $\bar{\theta}_{\text{clock}}$. Compute $E[Y | k, \theta_{\text{cond}}]$ and update $\Omega$ for the next period.
   F1. Increase $t$ by 1. Repeat previous step until $t > t_{\text{max}}$.

B1. Decrease $g$ by 1. If $g > 0$ set $f = F$ and return to section E.
B2. Decrease $e$ by 1. If $e > 0$ then reset $g = 2$ and return to section D.
B3. Decrease $k$. If $k > 0$ return to B0.
B4. Compute empirical predictions. Set $e = E$, $g = G$, $t = t_0$, and $k = K$.
   G0. Loop over $k$ to compute the sample-selected mixture for values of $t$, $e$, and $g$ that apply for $d$.
   G1. Compute the contribution $\Delta(\theta_{\text{cond}})$ to the econometric objective as defined in (26).
   G2. Iterate on $t$ through $t_{\text{max}}$, then decrease $g$ and $e$ until 0.

A3. Accumulate $W(\theta_{\text{exog}})$. Decrease $d$. If $d > 0$ return to step A2.
A4. Use the optimizer to minimize the objective with respect to $\theta_{\text{exog}}$.
A5. Iterate on the weighting matrix $\Sigma$, return to previous steps to compute $\hat{\theta}_{\text{exog}}$. 

35
Based on this new weighting of the moments the parameter and fit changed a great deal. Therefore, the correlation matrices were computed on more time from the new parameter values. The second (or third) stage objective is:

\[ Z^2(\theta_{\text{exog}}) = \sum_{d=1}^{4} \sum_{e=1}^{2} \sum_{g=1}^{3} \Delta_t(\theta_{\text{cond}})' \Sigma(\theta_{\text{cond}}) \Delta_t(\theta_{\text{cond}}). \]

The GMM estimates are then

\[ \hat{\theta}_{\text{exog}} = \arg \min_{\theta_{\text{exog}}} Z^2(\theta_{\text{exog}}). \quad (A45) \]

Let \( D(\theta_{\text{cond}}) \) denote the matrix of gradients for the vector \( \Delta_t \) with respect to the estimated parameters. Then the variance matrix and standard errors of the estimates were computed using the standard formula

\[ \text{Var} [\hat{\theta}_{\text{exog}}] = \left( \sum_{d=1}^{4} \sum_{e=1}^{2} \sum_{g=1}^{3} D \Sigma D' \right)^{-1}. \quad (A46) \]

VIII.G Identification

Three sources of variation in Table B help identify parameters of the model:
- Controlled and time-varying (path of treatment and assignment to experimental group)
- Uncontrolled and time-invariant (variation in policy and demographic groups).
- Uncontrolled and time-varying (unobserved endogenous states and treatment status)

The first two sources of variation are captured by \( \theta_{\text{cond}} = (t \ g \ e \ d) \)

Different loadings on these three factors will produce different patterns within months (across contemporaneous moments), across months (progress of treatment and initial selection), across studies (differing selection and information), across treatment groups (impact), and across demographic groups (variation in the mixture across exogenous types). It is not possible to prove analytically that the estimated parameters are identified from data generated by the experiment. Instead, a heuristic argument is given. The sources of variation are appealed to roughly in the order given above.

Begin with identification in the case of no unobserved heterogeneity \( (K = 1) \) and the simplest parameter to identify, the job-loss probability \( \pi_l \). In the model job loss occurs exogenously and the SSP survey records reasons why a parent stop working. These were grouped into losses and quits as reported in Table 4. Thus the proportion of working parents losing a job each month is available in the data and is directly determined by the value of \( \pi_l \). Since the observed proportions differ across demographic groups it is feasible to consider unobserved heterogeneity in \( \pi_l \) with different mixtures across groups. Of course, the estimates of \( \pi_l \) enters into all other aspects of the model.

Parents in the control group receiving IA do not quit jobs unless the convexity parameter \( c(h) \) changes value. And some parents go on and off IA with no change in labor
market status, which occurs in the model only when the level of outside support changes. The measurement vector includes quits and IA status but not these conditional switch rates. However, the joint movement over time (within control groups) of IA, labor market status, and quits help identify the jump probability for \( h \) the jump probability for outside support, \( \pi_s \). How the quit rate correlates with labor market earnings helps identify the distribution of \( c(h) \) and thus \( \zeta \). Mean earnings and the square of mean earnings are included in \( Y(\alpha, \theta) \) so that two moments of the accepted distribution are available to match the mean and variance of the offer distribution. Wage growth and duration dependency in accepted starting wages identify the skill accumulation and depreciation parameters. The correlation between income and welfare benefits helps identify the income reporting rate.

In a stationary model estimated on non-experimental data, the job search parameters (cost of search, offer probability, proportion of full-time jobs) would have to be identified through the reservation wage and the proportion of households working part-time (along with parametric assumptions on the offer distribution already made). It is not guaranteed that they would be identified in such data. The SSP experiment, however, includes exogenous variation in the value of job search and the value of keeping a full-time job. For example, the change in the proportion of people working part-time in the first month of the SSP (relative to the controls) picks up the proportion of accepted jobs that are potentially full-time.

More subtle variation is provided when we begin to consider variation across the Applicant \((e = 1)\) and Recipient \((e = 2)\) samples. An impact studied focuses on differences between a treatment group and their matched control group. For the Applicant Study, this consists of those who know the SSP subsidy exists and can anticipate becoming eligible for it (i.e. they are in phase \( f = 1 \)), and those in the control group who cannot become eligible \((f = 6)\). The model makes clear predictions between the behavior of these two groups. The value of taking a job and/or leaving IA changes with the time spent in phase 1. As \( r \) (exogenously) approaches \( \mathcal{R}(1) \) the higher the value of continued receipt of IA becomes among the treated. The rate at which outcomes diverge across the two groups as \( r \) increases reflects this approach to the change in phase. The change in the value of IA across groups as \( \mathcal{R}(1) \) approaches is sensitive to the transition probabilities. For example, if offer probabilities are very high then the treatment group can afford to reject offers received early on and/or cease active job search. The pattern of impacts helps identify these probabilities, although there is no one observable difference that can be matched to each parameter.

Treated households in the Applicant and Recipient Studies are in identical situations if and when they reach the qualifying phase of the experiment \((f = 2)\). From that point on, any difference between the behavior of the eligible households within the two groups is, within the model, forced to come from the difference in household states conditional upon reaching phase 2. In the Recipient Study reaching phase 2 is exogenous to the SSP and unexpected, whereas for the Applicant Study it is completely endogenous and can be expected and partially controlled up to one year in advance. Thus, the two samples provide experimental variation in \emph{unobserved} household states caused by lagged decisions made while anticipating different future opportunities. Many model parameters affect this cross-sample variation. For example, if job offers are rare then parents in the Applicant Study may not respond strongly to the information they have relative to the Recipient
Study before assignment. As argued above, other variation in the data contribute to identifying parameters like job offer rates. For purposes of this discussion, if we treat the other parameters as identified without comparing the entry and applicant treatment groups, then their comparison reveals the discount factor $\delta$.

The final parameter to discuss is the smoothing factor $\rho$. When $\rho = 0$ each feasible action has equal probability independent of the household’s state. This allows for a conclusion of completely ‘irrational’ behavior to be drawn from the data. The estimated model avoids this result because it is required to match the overarching patterns across groups and across experimental states that indicate choice probabilities vary systematically across states. For example, under complete irrationality, the proportion of households receiving IA each month would be the same no matter the assigned treatment group or how long ago random assignment occurred. Since statistically significant differences in choice probabilities exist across groups and experimental time, the estimated parameters will choose $\rho > 0$.

The point of the discussion so far is that each of the 19 exogenous parameters interacts with the design of the SSP experiment to affect specific aspects of the 12 matched results. The arguments account for the presence of many unobserved endogenous states, but they do not as yet account for unobserved exogenous parameters. Identification of unobserved heterogeneity in the parameters would be strengthen by applying the model to individual outcomes, because the likelihood or the predicted moments for a single individual would be conditioned on a single type. The computational cost of imposing these additional requirements is prohibitive.

Recall that demographic variation plays a restrictive role in the model. It determines the value of the policy parameters, such as the level of IA benefits, which are pre-determined and not free to explain variation in the data. The behavior of the unobserved types will respond to the differences in the policy parameters but there are no free parameters that directly control the influence of the demographic variables on predictions. That is, there is nothing like a ‘provincial coefficient’ in the wage offer distribution or a ‘number of children’ coefficient in the cost of time. Therefore, the model greatly restricts the freedom to calibrate responses in order to match the wide variation in experimental results across demographic groups. The only way for the estimates to gain more leverage in explaining the wide variation across demographic groups is to allow variation in the within-group proportions of each type. Thus it is likely (but not obvious how to demonstrate ahead of time) that the mixture parameters $\Lambda$ will be identified from the data along with differences in the underlying parameter vectors $\Gamma[k]$. 

38
IX. References


Ferrall, C. 2003. “Estimation and Inference in Social Experiments," working paper, Queen’s University, [Link to PDF].


Table 1. Endogenous Variables and Actions

<table>
<thead>
<tr>
<th>Item</th>
<th>Variable</th>
<th>Description</th>
<th>Num.</th>
<th>Values / Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lost job entering this month</td>
<td>2</td>
<td>{0,1}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>worked Previous month</td>
<td>2</td>
<td>{0,1}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>current earnings offered</td>
<td>6</td>
<td>{0,1/5,2/5,3/5,4/5}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>experience level</td>
<td>4</td>
<td>{1/4,1/2,3/4,1}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>outside Support</td>
<td>4</td>
<td>{0,1/3,2/3,1}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Opp. cost of time outside household</td>
<td>3</td>
<td>{1/4,2/4,3/4}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Demographic group</td>
<td>4</td>
<td>{1,2,3,4}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>unobserved type</td>
<td>4</td>
<td>{1,2,3,4}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Real states for individual (S1)</td>
<td>2,304</td>
<td>= 2^4<em>6</em>4<em>4</em>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All states given assignment (S2)</td>
<td>138,240</td>
<td>= S1 * 12 * 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All individual states (S3)</td>
<td>417,024</td>
<td>= S1 + S2 *3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All possible states (S)</td>
<td>6,672,384</td>
<td>= S3 * 4 * 4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>labor market work hours</td>
<td>5</td>
<td>{0,1/4,1/2,3/4,1}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>engage in active job search</td>
<td>2</td>
<td>{0,1}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>accept IA</td>
<td>2</td>
<td>{0,1}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feasible Actions (A)</td>
<td>12</td>
<td>= 6*2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Outcome Space</td>
<td>80,068,608</td>
<td>= S * A</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Demographic, Treatment, and Experimental Groups

<table>
<thead>
<tr>
<th>Vector</th>
<th>Index</th>
<th>Description</th>
<th>Subjects</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>d</td>
<td></td>
<td>Demographic Groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>New Brunswick, 1 Child</td>
<td>1728</td>
<td>19%</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>New Brunswick, 2+ Child.</td>
<td>1217</td>
<td>14%</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>British Columbia, 1 Child</td>
<td>3058</td>
<td>34%</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>British Columbia, 2+ Child.</td>
<td>2895</td>
<td>33%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>8898</td>
<td>100%</td>
</tr>
<tr>
<td>g</td>
<td></td>
<td>Treatment Groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Control</td>
<td>4305</td>
<td>48%</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>SSP Treatment</td>
<td>4300</td>
<td>48%</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>SSP+ Treatment (NB only)</td>
<td>293</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>8898</td>
<td>100%</td>
</tr>
<tr>
<td>e</td>
<td></td>
<td>Experimental Groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Recipient Study</td>
<td>5682</td>
<td>63%</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>Applicant Study (BC only)</td>
<td>3316</td>
<td>37%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>8998</td>
<td>100%</td>
</tr>
</tbody>
</table>

Observations dropped: invalid or missing age, high school attendance or number of children (from IA records).
Table 3. Experimental Results (Moments) Selected for Matching

<table>
<thead>
<tr>
<th>Var.</th>
<th>Description</th>
<th>Model</th>
<th>Unit</th>
<th>Count</th>
<th>Mean</th>
<th>St.Dev</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>earn</td>
<td>Rep. Earnings</td>
<td>(1-βi)mW(α,θ)</td>
<td>$100</td>
<td>470</td>
<td>3.460</td>
<td>1.541</td>
<td></td>
</tr>
<tr>
<td>earn²</td>
<td>Earnings Sq.</td>
<td>earn²</td>
<td>$100²</td>
<td>470</td>
<td>62.590</td>
<td>43.470</td>
<td></td>
</tr>
<tr>
<td>ia</td>
<td>IA Received</td>
<td>IA(α,θ)</td>
<td>$100</td>
<td>466</td>
<td>5.966</td>
<td>1.869</td>
<td>Fwd.2 mth</td>
</tr>
<tr>
<td>iasq</td>
<td>IA Recv Sq.</td>
<td>IA²</td>
<td>$100²</td>
<td>466</td>
<td>57.782</td>
<td>27.593</td>
<td>Fwd.2 mth</td>
</tr>
<tr>
<td>gsu</td>
<td>SSP Suppl</td>
<td>SUP(α,θ)</td>
<td>$100</td>
<td>240</td>
<td>1.530</td>
<td>0.600</td>
<td>Fwd.2 mth</td>
</tr>
<tr>
<td>onia</td>
<td>Received IA</td>
<td>i</td>
<td>0/1</td>
<td>481</td>
<td>0.692</td>
<td>0.191</td>
<td></td>
</tr>
<tr>
<td>mwg</td>
<td>Worked at MW</td>
<td>(n*&lt;6-#n)(m&gt;0)</td>
<td>0/1</td>
<td>470</td>
<td>0.777</td>
<td>0.059</td>
<td>hrly w &lt;= MW+$.10</td>
</tr>
<tr>
<td>leftjb</td>
<td>Left/quit a job</td>
<td>p(l=0)(m=0)</td>
<td>0/1</td>
<td>456</td>
<td>0.003</td>
<td>0.004</td>
<td>Excl. job-to-job</td>
</tr>
<tr>
<td>lossjb</td>
<td>Loss a job</td>
<td>l</td>
<td>0/1</td>
<td>456</td>
<td>0.004</td>
<td>0.004</td>
<td>Excl. job-to-job</td>
</tr>
<tr>
<td>emft</td>
<td>Full Time</td>
<td>m&gt;PT</td>
<td>0/1</td>
<td>470</td>
<td>0.223</td>
<td>0.088</td>
<td></td>
</tr>
<tr>
<td>empt</td>
<td>Part-time</td>
<td>0 &lt; m &lt;= PT</td>
<td>0/1</td>
<td>470</td>
<td>0.130</td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td>onXem</td>
<td>IA &amp; Working</td>
<td>ia * (m&gt;0)</td>
<td>0/1</td>
<td>470</td>
<td>0.161</td>
<td>0.045</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4915</td>
<td></td>
</tr>
</tbody>
</table>

A summary of the complete data listed in Table B.1-B.10. Count is the number of cells in Table B panel. 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.
Table 4. Relative Impacts on Selected Moments in Months -11,1,13,25

<table>
<thead>
<tr>
<th>Var.</th>
<th>t</th>
<th>NB / 1 Child Recipients</th>
<th>NB / 2+ Recipients</th>
<th>BC / 1 Child</th>
<th>BC / 2+</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SSP+</td>
<td>SSP</td>
<td>SSP+</td>
<td>SSP</td>
<td>SSP</td>
</tr>
<tr>
<td>Eam</td>
<td>-10</td>
<td></td>
<td></td>
<td></td>
<td>(0.15)</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.06)</td>
<td>(0.25)</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.51</td>
<td>0.39</td>
<td><strong>1.28</strong></td>
<td>0.89</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>0.58</td>
<td>0.30</td>
<td>0.74</td>
<td>0.70</td>
<td>0.20</td>
</tr>
<tr>
<td>OuiA</td>
<td>-11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>(0.24)</td>
<td>(0.19)</td>
<td>(0.29)</td>
<td>(0.19)</td>
<td>(0.21)</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td><strong>(0.32)</strong></td>
<td>(0.18)</td>
<td>(0.20)</td>
<td>(0.18)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>Mwgn</td>
<td>-11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.00</td>
<td>(0.01)</td>
<td>(0.04)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>(0.13)</td>
<td>(0.11)</td>
<td><strong>(0.15)</strong></td>
<td>(0.13)</td>
<td>(0.12)</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.12)</td>
<td>(0.07)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>Emft</td>
<td>-11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.16)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.11</td>
<td>(0.07)</td>
<td>(0.09)</td>
<td>(0.25)</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.97</td>
<td>0.76</td>
<td><strong>1.65</strong></td>
<td>1.46</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>0.90</td>
<td>0.64</td>
<td>0.86</td>
<td>0.90</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Difference between SSP and Ctrl columns in Table A divided by Ctrl column. Negative impacts in () and in red. Largest absolute impact within the table shaded for each moment.
Table 5. Contemporaneous Correlations Across Results

<table>
<thead>
<tr>
<th>Applicants (e=1)</th>
<th>earn</th>
<th>ia</th>
<th>onia</th>
<th>mwg</th>
<th>left</th>
<th>emft</th>
<th>Group (g) ; Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>earn</td>
<td>-0.360</td>
<td>-0.314</td>
<td>-0.678</td>
<td>-0.029</td>
<td>0.655</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ia</td>
<td>-0.356</td>
<td>0.733</td>
<td>0.360</td>
<td>-0.002</td>
<td>-0.375</td>
<td></td>
<td></td>
</tr>
<tr>
<td>onia</td>
<td>-0.303</td>
<td>0.720</td>
<td>0.302</td>
<td>-0.011</td>
<td>-0.338</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mwg</td>
<td>-0.668</td>
<td>0.379</td>
<td>0.303</td>
<td>0.047</td>
<td>-0.677</td>
<td></td>
<td></td>
</tr>
<tr>
<td>left</td>
<td>-0.029</td>
<td>0.008</td>
<td>0.004</td>
<td>0.050</td>
<td>-0.041</td>
<td></td>
<td></td>
</tr>
<tr>
<td>emft</td>
<td>0.640</td>
<td>-0.400</td>
<td>-0.353</td>
<td>-0.672</td>
<td>-0.044</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Recipients (e=2)</th>
<th>earn</th>
<th>ia</th>
<th>onia</th>
<th>mwg</th>
<th>left</th>
<th>emft</th>
<th>Group (g) ; Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>earn</td>
<td>-0.317</td>
<td>-0.294</td>
<td>-0.550</td>
<td>-0.011</td>
<td>0.564</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ia</td>
<td>-0.409</td>
<td>0.692</td>
<td>0.330</td>
<td>-0.018</td>
<td>-0.369</td>
<td></td>
<td></td>
</tr>
<tr>
<td>onia</td>
<td>-0.396</td>
<td>0.733</td>
<td>0.267</td>
<td>-0.021</td>
<td>-0.324</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mwg</td>
<td>-0.608</td>
<td>0.402</td>
<td>0.376</td>
<td>0.024</td>
<td>-0.576</td>
<td></td>
<td></td>
</tr>
<tr>
<td>left</td>
<td>-0.018</td>
<td>-0.009</td>
<td>-0.027</td>
<td>0.031</td>
<td>-0.020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>emft</td>
<td>0.638</td>
<td>-0.503</td>
<td>-0.511</td>
<td>-0.611</td>
<td>-0.030</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Red (italic) values indicate control. Black (bold) values indicate SSP treatment. Correlations across selected moments summarized in Table 5. SSP+ treatment group is excluded. SSP supplement is excluded because it does not vary within control groups. All observations are pooled across demographic and post random-assignment experimental months.
Table 6. Summary of the Estimation

<table>
<thead>
<tr>
<th>Item</th>
<th>Value</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size of linear system to compute ergodic distn.</td>
<td>2,304</td>
<td>See Table 1</td>
</tr>
<tr>
<td>Number of Type-Specific Parameters (N)</td>
<td>13</td>
<td>Table 7.1-8.4</td>
</tr>
<tr>
<td>Number of Common Parameters (C)</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Number of free exogenous parameters</td>
<td>70</td>
<td>D*(K-1)+K*N+C</td>
</tr>
<tr>
<td>CPU Time to Evaluate Objective (min.)</td>
<td>14</td>
<td>16 X UltraSPARC-III</td>
</tr>
<tr>
<td>Value of Objective ($Z^2$)</td>
<td>29.431</td>
<td></td>
</tr>
</tbody>
</table>
Table 7.1. $\hat{\theta}_{\text{exog}}$: Estimated Type Proportions ($\Lambda[d]$)

<table>
<thead>
<tr>
<th>d</th>
<th>Description</th>
<th>Type Index (k)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>NB, One Child</td>
<td>0.0324</td>
<td>0.6388</td>
<td>0.1409</td>
<td>0.1879</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.270)</td>
<td>(0.272)</td>
<td>(0.022)</td>
<td>(0.025)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>NB, Two+ Children</td>
<td>0.0380</td>
<td>0.5695</td>
<td>0.2147</td>
<td>0.1777</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.107)</td>
<td>(3.175)</td>
<td>(0.009)</td>
<td>(0.020)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>BC, One Child</td>
<td>0.5025</td>
<td>0.0014</td>
<td>0.4884</td>
<td>0.0076</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.175)</td>
<td>(1.542)</td>
<td>(0.018)</td>
<td>(0.019)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>BC, Two+ Children</td>
<td>0.5564</td>
<td>0.0032</td>
<td>0.4396</td>
<td>0.0008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

GMM estimates of $\lambda[d, k]$ based on (45). See paragraph above equation (5) for description. Estimated standard in parentheses is the square root of diagonal elements of (46).
Table 7.2. $\hat{\theta}_{\text{exog}}$ : Estimated Dynamic Programming Parameters ($\delta_k$ and $\rho_k$)

<table>
<thead>
<tr>
<th>Var</th>
<th>Description</th>
<th>Type Index (k)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Discount Factor</td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Smoothing</td>
<td>3.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.090)</td>
</tr>
</tbody>
</table>

GMM estimates based on (45). See (5) for description and and for roles of the parameters.

Estimated standard in parentheses is the square root of diagonal elements of (46).
Table 7.3. $\hat{\theta}_{\text{exog}}$: Estimated Utility Shifters (τ)

<table>
<thead>
<tr>
<th>Var.</th>
<th>Description</th>
<th>Type Index (k)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Job Offer Mean</td>
<td>-1.564</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.857)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Job Offer St. Dev.</td>
<td>2.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.176)</td>
</tr>
<tr>
<td>$\chi$</td>
<td>Outside Support</td>
<td>1.264</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.014)</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Cost of FT Work</td>
<td>0.346</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.005)</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>Cost of Job Search</td>
<td>0.353</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.047)</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Return to Skill</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\zeta$</td>
<td>1 / Mean Convexity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>Income Reporting</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GMM estimates of utility parameters summarized in (16) based on (45). Estimated standard in parentheses is the square root of diagonal elements of (46).
Table 7.4. $\hat{\theta}_{\text{exog}}$: Estimated Transition Shifters (II)

<table>
<thead>
<tr>
<th>Sub.</th>
<th>Description</th>
<th>Type Index (k)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><strong>Market</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>j</td>
<td>Job Offer ($b&gt;0$)</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
</tr>
<tr>
<td>f</td>
<td>Prop. Full Time</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.017)</td>
</tr>
<tr>
<td>m</td>
<td>Prop. MW job ($n=0$)</td>
<td>0.162</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.030)</td>
</tr>
<tr>
<td>l</td>
<td>Job Loss</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001)</td>
</tr>
<tr>
<td><strong>Home</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>Support Change</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
</tr>
<tr>
<td>h</td>
<td>Prob. Costs Change</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.031)</td>
</tr>
<tr>
<td><strong>Skills</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>SSP Plus Effect</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td>a</td>
<td>Accumulation</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.727)</td>
</tr>
<tr>
<td>d</td>
<td>Depreciation</td>
<td>0.0020</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001)</td>
</tr>
</tbody>
</table>

GMM estimates of parameters summarized in (16) based on (45). Estimated standard in parentheses is the square root of diagonal elements of (46).
Table 8. Using Treatment Outcomes for Identification vs. Validation: Estimated Standard Errors by Group

<table>
<thead>
<tr>
<th>Par. or Sub.</th>
<th>k=1 / common</th>
<th>k=2</th>
<th>k=3</th>
<th>k=4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>all</td>
<td>ctrl</td>
<td>treat</td>
<td>all</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.255</td>
<td>0.946</td>
<td>0.232</td>
<td>2.937</td>
</tr>
<tr>
<td>d=1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d=2</td>
<td>0.255</td>
<td>0.994</td>
<td>0.232</td>
<td>2.977</td>
</tr>
<tr>
<td>d=3</td>
<td>0.021</td>
<td>0.067</td>
<td>0.018</td>
<td>0.009</td>
</tr>
<tr>
<td>d=4</td>
<td>0.024</td>
<td>0.081</td>
<td>0.021</td>
<td>0.020</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.001</td>
<td>0.021</td>
<td>0.001</td>
<td>0.034</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.044</td>
<td>0.477</td>
<td>0.034</td>
<td>0.083</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.779</td>
<td>14.954</td>
<td>0.597</td>
<td>3.183</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.161</td>
<td>2.993</td>
<td>0.124</td>
<td>0.810</td>
</tr>
<tr>
<td>$\mu_m$</td>
<td>0.029</td>
<td>15.639</td>
<td>0.022</td>
<td>0.509</td>
</tr>
<tr>
<td>$\mu_a$</td>
<td>0.723</td>
<td>12.181</td>
<td>0.527</td>
<td>1.732</td>
</tr>
<tr>
<td>$\mu_d$</td>
<td>0.001</td>
<td>0.021</td>
<td>0.001</td>
<td>0.086</td>
</tr>
<tr>
<td>$\mu_j$</td>
<td>0.002</td>
<td>1.178</td>
<td>0.002</td>
<td>0.360</td>
</tr>
<tr>
<td>$\mu_l$</td>
<td>0.001</td>
<td>0.008</td>
<td>0.001</td>
<td>0.013</td>
</tr>
<tr>
<td>$\mu_h$</td>
<td>0.014</td>
<td>0.197</td>
<td>0.011</td>
<td>0.778</td>
</tr>
<tr>
<td>$\mu_f$</td>
<td>0.017</td>
<td>0.095</td>
<td>0.015</td>
<td>0.107</td>
</tr>
<tr>
<td>$\mu_v$</td>
<td>0.004</td>
<td>0.071</td>
<td>0.003</td>
<td>0.057</td>
</tr>
<tr>
<td>$\eta$</td>
<td>1.701</td>
<td>85.699</td>
<td>1.349</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.004</td>
<td>0.019</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>1.914</td>
<td>49.956</td>
<td>1.410</td>
<td></td>
</tr>
<tr>
<td>$\mu_s$</td>
<td>0.002</td>
<td>0.004</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>$\mu_h$</td>
<td>0.030</td>
<td>0.141</td>
<td>0.028</td>
<td></td>
</tr>
<tr>
<td>$\mu_\zeta$</td>
<td>0.080</td>
<td>1.159</td>
<td>0.062</td>
<td></td>
</tr>
</tbody>
</table>

"all" is the standard error for the corresponding parameter as computed for Table 7 except $\pi+$ is excluded from the computations because it is unidentified from the control group only. "ctrl" is the standard error computed using on 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 10 times the size of the "all" column. *Italic* indicates a standard error less than 3/4 of the "all" column.
Figure 5a. Result: Earnings, New Brunswick

\begin{tabular}{|c|c|}
\hline
\textbf{Observed} & \textbf{Predicted} \\
\hline
\end{tabular}

\begin{tabular}{|c|c|}
\hline
\textbf{Impact} & \\
\hline
\end{tabular}

\begin{tabular}{|c|c|}
\hline
\textbf{Observed} & \textbf{Predicted} \\
\hline
\end{tabular}

NB 1 Child

\begin{tabular}{|c|c|}
\hline
\textbf{Observed} & \textbf{Predicted} \\
\hline
\end{tabular}

\begin{tabular}{|c|c|}
\hline
\textbf{Impact} & \\
\hline
\end{tabular}

\begin{tabular}{|c|c|}
\hline
\textbf{Observed} & \textbf{Predicted} \\
\hline
\end{tabular}

NB 2+ Children

\begin{tabular}{|c|c|}
\hline
\textbf{Observed} & \textbf{Predicted} \\
\hline
\end{tabular}

\begin{tabular}{|c|c|}
\hline
\textbf{Impact} & \\
\hline
\end{tabular}
Figure 5b. Result: Earnings, British Columbia
Figure 6a. Result: Total Government Transfers (IA+SSP)

NB 1 Child

NB 2+ Children
Figure 6b. Result: Total Government Transfers (IA+SSP)

BC 1 Child

BC 2+ Children
Figure 7a. Heterogeneity, Sample Selection and Policy Variation

△ marks the ergodic (unselected) mean value.
△ marks the ergodic (unselected) mean value.
Figure 8. Forecast of Earnings by Province
Applicant sample conducted in NB; SSP+ conducted in BC and on the applicant sample.

Control groups not shown.
Figure 11. Experiment 2: On IA with Flow Sampling
Figure 10. Experiment 3: SSP and a 20% Cut in IAB

BC 2+ Children
XI. Data Appendix

Table B. Empirical Results (Moments) Used to Estimate the Model \( Y(\alpha, \theta) \)

Tables B.1-B.13 report the data used to estimate the model. If they do not begin on the next page they can be downloaded from