Evaluating the Effectiveness of Washington State Repeated Job Search Services on the Employment Rate of Prime-age Female Welfare Recipients*

July 29th, 2003

Abstract

This paper uses an unbalanced panel dataset to evaluate how repeated job search services and personal characteristics affect the employment rate of the prime-age female welfare recipients in the State of Washington. We propose a transition probability model to take account issues of sample attrition, sample refreshment and duration dependence. We also propose a nonlinear two-stage least square estimator to allow for selection due to unobservables and generalize Honoré and Kyriazidou's (2000) conditional maximum likelihood estimator to allow for state-dependent individual specific effects and slope coefficients. We then provide a conditional nonlinear two-stage least square estimator that allows for both unobserved individual heterogeneity and selection on unobservables. The specification tests indicate that the assumptions of no selection due to observables or no unobserved individual specific effects are not violated. Our findings indicate that job search services do have positive and significant impacts on the employment rate of those who are initially unemployed, but their impacts are insignificant for those who are employed. Furthermore, there are significant experience enhancing effects. These findings suggest that providing job search services to unemployed individuals to help them find jobs quickly may have a lasting impact on raising their employment rate.

*This research was supported in part by Employment Security Department, State of Washington, but does not necessarily represent an official position of Employment Security Department. We thank M. Kaiser and Geert Ridder for very helpful comments and suggestions. After the completion of the preliminary version of this paper, we have come across a paper by D'Addio and Honoré (2002) that proposes a similar conditional maximum likelihood estimator for a state dependent coefficients and individual effects logit model.

1 Introduction

The aim of this paper is to provide measurements of the effects of repeated job search services (JSS) on the employment rates of female welfare recipients who participated in the WorkFirst program of the State of Washington. The WorkFirst program is the implementation of the Federal Temporary Assistance for Needy Family (TANF) program in the State of Washington. Initiated in August 1997, its main goal is to help financially struggling families to find jobs, keep themselves off unemployment, and get better jobs. Emphasizing that getting a low-paying job now is better than waiting for a high-paying job in the future, the WorkFirst program has a process that focuses its main activities on job search services (JSS). Nevertheless, TANF recipients have been returning to welfare in great numbers. Over 70 percent of the entrants to welfare were former TANF recipients during the sample period (1998.II - 2000.IV). To efficiently allocate the limited resources, policy makers are particularly interested in finding out whether it is efficient to provide Job Search Services (JSS) to the same clients repeatedly.

In this study the population of interest are female TANF recipients between the age 25 - 35. However, our data set is not a balanced panel data as some of the early studies (e.g., Bassi (1984), Ashenfelter and Card (1985), Heckman and Hotz (1989)). A significant feature of our data is that clients entered and left the program at different time periods. For instance, only about 3.4 percent clients have the complete treatment history over the sample period 1998.II to 2000.IV. There is also the issue of right censoring because the data period ends at 2000.IV. If we restrict our attention to the subsample of clients that entered and left a training program in the same time periods, it would greatly reduce available observations. Moreover, misleading inference may occur if clients are not randomly selected from the whole population. If we choose to start our sample later than 1998.II to keep as many balanced panel observations as possible, participation histories would be incomplete for those who entered the program late, hence makes the estimation of repeated treatment effects difficult.

In this paper we propose a transition probability model of finding employment or staying on employment as a means to take account of issues arising from sample attrition, sample refreshment and duration dependence. Being state dependent, a transition probability model also allows one to accommodate dynamics in a simple format. We first estimate such a model by assuming that (1) participations of JSS are not endogenous so that there is no bias stemming from selection on unobservables; and (2) there are no unobserved individual specific effects. To check the validity of these assumptions, we also suggest three estimators that relax one or both of the above two assumptions: a nonlinear two-stage least square estimator that deals with the endogeneity of participation decisions, a generalized conditional maximum likelihood estimator that allows for state-dependent fixed effects and slope coefficients, and a conditional nonlinear two-stage least square estimator that allows for both unobserved individual heterogeneity and endogeneity of the participation decision. The Hausman (1978) specification tests indicate that the above two assumptions are not contradicted by our sample information, therefore, we shall discuss our empirical findings from the model treating participation decisions as exogenous and without the presence of unobserved individual specific effects.

Our findings show that the repeated job search services do have positive and significant effects on the employment rate of initially unemployed clients, but their effects are not significant for those who are already employed. Furthermore, the probability of employment is also influenced by the duration in employment or unemployment, family factors, education level, geographic and local labor market conditions as well as other welfare services.

Section 2 introduces the model and section 3 presents the estimation results. Diagnostic checkings are discussed in section 4. Conclusions are in Section 5. Detailed descriptions of our data are available upon request.

2 The Model

2.1 A Transitional Probability Model for the Outcomes

In this section we propose a transitional probability framework to take account issues of sample attrition, sample refreshment and duration dependence. Let y_{it} be the binary indicator that takes the value 1 if the ith client is employed and the value 0 if otherwise, $t = t_i,\ t_i + 1,\ ...T_i$, where t_i and T_i denote the first period and last period that client i is observed. We assume that y_{it} depends on the initial states, $y_{i,t-1}$, previous JSS treatments, and strictly exogenous socio-demographic variables, $\mathbf{x}_{it}$. We separate y_{it} into two groups depending on the value of $y_{i,t-1}$ being 0 or 1. Let y_{it}^s be those y_{it} where $y_{i,t-1} = s,\ s = 0,\ 1$. Let y_{it}^{s*} denote the potential state given $y_{i,t-1} = s,\ s = 0, 1$. We assume that

$$y_{it}^s = \begin{cases} 1, \text{ if } y_{it}^{s*} > 0, \\ 0, \text{ if otherwise.} \end{cases} \tag{1}$$

A client is considered to have taken JSS if records show that she has taken at least one JSS within one quarter. Since over 95 percent participants took no more than 3 JSS over the period of 1998.II - 2000.IV, we will focus on the evaluation of the effects of repeated JSS to be at most three. We consider the effects of previous period JSS treatment on current y_{it} since one JSS can last up to 12 continuous weeks. A client taking at most 3 JSS before period t is in one of the four possible potential states: (1) she has not taken any JSS; (2) she has taken one JSS; (3) she has taken two JSS; and (4) she has taken three JSS. Let $\tilde{d}_{it}^m$ be mutually exclusive dummies such that

$$\begin{aligned} \tilde{d}_{it}^m &= \begin{cases} 1, \text{ if exactly } m \text{ JSS has(have) been taken before period } t, \\ 0, \text{ otherwise,} \end{cases} \\ m &= 0, 1, 2, 3. \end{aligned} \tag{2}$$

We assume that JSS participations influence the probability of employment through its impact on potential outcomes. Depending on the realization of $\tilde{d}_{it}^m$, y_{it}^{s*} can take one of the four possible forms, y_{it}^{sm*}, $m = 0, 1, 2, 3$, where y_{it}^{sm*} denotes the the corresponding potential outcome when $\tilde{d}_{it}^m = 1, m = 0, 1, 2, 3$. At time t, y_{it}^{s*} and y_{it}^{sm*} has the following relation,

$$\begin{aligned} y_{it}^{s*} &= \tilde{d}_{it}^1 y_{it}^{s1*} + (1 - \tilde{d}_{it}^1) \left\{ \tilde{d}_{it}^2 y_{it}^{s2*} + (1 - \tilde{d}_{it}^2) \left[\tilde{d}_{it}^3 y_{it}^{s3*} + (1 - \tilde{d}_{it}^3) y_{it}^{s0*} \right] \right\} \\ &= y_{it}^{s0*} + \tilde{d}_{it}^1 \tilde{\gamma}_{it}^{s1} + \tilde{d}_{it}^2 \tilde{\gamma}_{it}^{s2} + \tilde{d}_{it}^3 \tilde{\gamma}_{it}^{s3}, \end{aligned} \tag{3}$$

where $\tilde{\gamma}_{it}^{sm} = y_{it}^{sm*} - y_{it}^{s0*}$, $m = 1, 2, 3$, measures the cumulative effect of m JSS over no JSS for the ith individual at the tth time period, when the no-JSS state is treated as the base state, *i.e.*, $\tilde{d}_{it}^0 = 1$ when $\tilde{d}_{it}^1 = \tilde{d}_{it}^2 = \tilde{d}_{it}^3 = 0$, and $\tilde{\gamma}_{it}^{s0} = 0$.

There is a one-to-one relation between the cumulative effect and marginal effect. Rewrite equation (3) as

$$y_{it}^{s*} = y_{it}^{s0*} + \mathbf{D}_{it}\boldsymbol{\gamma}_{it}^{s}, \tag{4}$$

where $\mathbf{D}_{it} = [d_{it}^1, d_{it}^2, d_{it}^3]$, with $\mathbf{D}_{it} = [0,0,0]$ when $\tilde{d}_{it}^0 = 1$, $\mathbf{D}_{it} = [1,0,0]$ when $\tilde{d}_{it}^1 = 1, \mathbf{D}_{it} = [1,1,0]$ when $\tilde{d}_{it}^2 = 1$, and $\mathbf{D}_{it} = [1,1,1]$ when $\tilde{d}_{it}^3 = 1$, and $\boldsymbol{\gamma}_{it}^s = (\gamma_{it}^{s1}, \gamma_{it}^{s2}, \gamma_{it}^{s3})'$ with $\gamma_{it}^{sm} = \tilde{\gamma}_{it}^{sm} - \tilde{\gamma}_{it}^{s(m-1)} = y_{it}^{sm*} - y_{it}^{s(m-1)*}$, $m = 1,2,3$.

The value of γ_{it}^{sm} measures the marginal impact of the mth job search service on the ith individual at time t conditional on her last period employment status being s. The average treatment effect for the mth job search service conditional on last period employment status being s is $E\left(\gamma_{it}^{sm}\right)$.

However, y_{it}^{sm*} is not observable. We shall therefore define the treatment effects of the mth JSS on client i in terms of the changes in the probability of employment conditional on last period's employment status being s. For simplicity, we assume that y_{it}^{sm*} can be decomposed as the sum of the effects of observables $\mathbf{x}_{it}$, $g^{sm}\left(\mathbf{x}_{it}\right)$, and the effects of unobservables, u_{it}^{sm},

$$y_{it}^{sm*} = g^{sm}\left(\mathbf{x}_{it}\right) + u_{it}^{sm},\ m = 0,1,2,3. \tag{5}$$

Then $\Pr\left(y_{it} = 1 | y_{i,t-1}, \mathbf{x}_{it}\right) = \Pr\left(y_{it}^{s*} > 0 | y_{i,t-1}, \mathbf{x}_{it}\right)$, and the treatment effect of the mth JSS is defined as $\Delta_{it}^{sm} = \Pr\left(y_{it}^{sm*} > 0 | y_{i,t-1} = s, \mathbf{x}_{it}\right) - \Pr(y_{it}^{s(m-1)*} > 0 | y_{i,t-1} = s, \mathbf{x}_{it})$, $s = 0,1$, $m = 1,2,3$. The average treatment effect (ATE) of the mth JSS conditional on last period's employment status s is defined as $\Delta_{ATE}^{sm} = E\left[\Delta_i^{sm}\right] = E[\Pr(y_{it}^{sm*} > 0 | y_{i,t-1} = s, \mathbf{x}_{it}) - \Pr(y_{it}^{s(m-1)*} > 0 |\ y_{i,t-1} = s, \mathbf{x}_{it})]$. The treatment of the treated (TT) of the mth JSS conditional on last period's employment status s is defined as $\Delta_{TT}^{sm} = E\left(\Delta_{it}^{sm} | d_{it}^{m} = 1\right).$

ATE of the mth JSS is the mean impact of the mth JSS if clients are randomly assigned to JSS. It is of interest if one is interested in estimating the impact of the mth JSS on a randomly selected clients. TT of the mth JSS is the mean impact of the mth JSS on those clients who actually have taken the mth JSS. It is of interest if the same selection rule for treatment applies in the future.

Because it is impossible to simultaneously observe y_{it}^{sm*} and $y_{it}^{s(m-1)*}$, $m = 1, 2, 3$, we cannot directly estimate Δ_{ATE}^{sm} and Δ_{TT}^{sm} . In essence, to evaluate treatment effect is to deal with a missing data problem. If we approximate the no-JSS outcomes of the treated group by observed outcomes from the control group and calculate the treatment effect by

$$\frac{1}{N_m} \sum_{j \in \Phi_m} y_{jt} - \frac{1}{N_{m-1}} \sum_{j \in \Phi_{m-1}} y_{jt}, m = 1, 2, 3 \tag{6}$$

where N_m, N_{m-1} is the number of clients who have taken the mth and the $(m-1)$th treatment respectively, and Φ_m, Φ_{m-1} is the set that includes the corresponding client respectively, then equation (6) approaches

$$\begin{aligned}
& E\left(y_t = 1 | y_{t-1} = s, d_t^m = 1\right) - E\left(y_t = 1 | y_{t-1} = s, d_t^m = 0\right) \\
= \; & E\Big\{\Pr\left(y_t^{sm*} > 0 | y_{t-1} = s, d_t^m = 1\right) - \Pr(y_t^{s(m-1)*} > 0 | y_{t-1} = s, d_t^m = 1) \\
& + \Big[\Pr(y_t^{s(m-1)*} > 0 | y_{t-1} = s, d_t^m = 1) - \Pr(y_t^{s(m-1)*} > 0 | y_{t-1} = s, d_t^m = 0)\Big]\Big\} \\
= \; & \Delta_{TT}^{sm} + B_t^{sm},
\end{aligned}$$

where $B_t^{sm} = E([(\Pr y_t^{s(m-1)*} > 0 | y_{t-1} = s, d_t^m = 1) - \Pr(y_t^{s(m-1)*} > 0 | y_{t-1} = s, d_t^m = 0)])$ is the bias for estimating the effect of the mth treatment conditional on employment status s resulted from using control group to approximate the no-JSS state of the treated group. If $B_t^{sm} = 0$, then (6) provides a consistent estimate of Δ_{TT}^{sm}.

When

$$u_{it}^{sm} \perp d_{it}^{m} | y_{i,t-1}, \bar{\mathbf{x}}_{it}, \tag{7}$$

where $\bar{\mathbf{x}}'_{it} = (\mathbf{x}'_{it}, \mathbf{x}'_{i,t-1}, ...\mathbf{x}'_{it_i})$, the treatment assignment is free from the influence of unobserved factors affecting y_{it}^{sm*} conditional on $(y_{i,t-1}, \bar{\mathbf{x}}_{it})$. Condition (7) is called Conditional Independence Assumption (CI) by Rosenbaum and Rubin (1983) or Ignorable Treatment Assignment Assumption by Heckman and Robb (1985) and Holland (1986). Then

$$\Pr(y_{it}^{s(m-1)*} > 0 | y_{i,t-1} = s, \mathbf{x}_{it}, d_{it}^{m} = 1) = \Pr(y_{it}^{s(m-1)*} > 0 | y_{i,t-1} = s, \mathbf{x}_{it}, d_{it}^{m} = 0).$$

In other words, under CI, conditional on $\mathbf{x}$, ATE($\mathbf{x}$) =TT($\mathbf{x}$).

For the unconditional bias B_t^{sm} to equal to zero we also need

$$\mathbf{x}_{it} \perp d_{it}^{m} | y_{i,t-1}. \tag{8}$$

When condition (8) does not hold, we say that the selection bias is due to observables or selection on observables. In other words, for $B_t^{sm} = 0$ unconditionally, we need that there is neither selection on observables nor unobservables.

Usually, the $\mathbf{x}_{it}$ that affects the potential outcome may also affect the participation decision.

It is hard to assume that $(\mathbf{x}_{it} \perp d_{it}^m | y_{i,t-1})$. Therefore, we shall analyze the treatment effects by simultaneously controlling the impacts of $\mathbf{x}_{it}$ and JSS. The nonparametric approach that takes account both the timing of treatments and interval between treatments involves many possible regimes. It is very complicated to analyze and requires huge number of observations (e.g., Gill and Robins (2001), Lechner (2001)). To extract information from finite sample, in this paper we shall take a parametric approach. Since our sample only covers a relatively short period from 1998 - 2000, for simplicity, we assume a linear structure for y_{it}^{s*},

$$y_{it}^{s*} = \mathbf{x}_{it}'\boldsymbol{\beta}^s + \mathbf{D}_{it}\boldsymbol{\gamma}^s + u_{it}^s. \tag{9}$$

We shall derive our model specification and inference under the CI assumption in this section. We then consider methods of testing CI assumption and methods of estimating ATE or TT conditioning on $\bar{\mathbf{x}}_{it}$ when CI is violated in section 5.

Let $P_{isk} = \Pr(y_{it} = k | y_{i,t-1} = s)$, $s, k = 0, 1$, be the transition probability that the ith individual is in state j in period $t-1$ and k in period t,

$$\begin{aligned} P_{is1} &= \Pr(y_{it} = 1 | y_{i,t-1} = s, \mathbf{x}_i) \qquad (10) \\ &= \Pr(\mathbf{x}_{it}\boldsymbol{\beta}^s + \mathbf{D}_{it}\boldsymbol{\gamma}^s + u_{it}^s > 0) \\ &= F(\mathbf{x}_{it}\boldsymbol{\beta}^s + \mathbf{D}_{it}\boldsymbol{\gamma}^s) = F_{it}^s, \ s = 0, 1, \end{aligned}$$

and $F(\cdot)$ denotes a certain cumulative distribution function. In this study, we assume F takes the logit form,

$$F_{it}^{s}=\frac{\exp\left(\mathbf{x}_{it}\boldsymbol{\beta}^{s}+\mathbf{D}_{it}\boldsymbol{\gamma}^{s}\right)}{1+\exp\left(\mathbf{x}_{it}\boldsymbol{\beta}^{s}+\mathbf{D}_{it}\boldsymbol{\gamma}^{s}\right)},\, s=0,1,\;\; t=t_i+1,...,T_i. \tag{11}$$

Since the lagged values of the initial employment status are not observed, the initial state y_{it_i} is approximated by an unconditional specification,

$$y_{it_i}=Q\left(\bar{\mathbf{x}}_i\right)+\varepsilon_{it_i},\;\; i=1,...,N, \tag{12}$$

where $\bar{\mathbf{x}}_i=\frac{1}{(T_i-t_i+1)}\sum_{t_i}^{T_i}\mathbf{x}_{it}$, and Q denotes a monotonic function. The means of the explanatory variables are used instead of $\mathbf{x}_i=(\mathbf{x}'_{it_i},...\mathbf{x}'_{iT_i})$ as it usually yields better finite sample results when the variation of $\mathbf{x}_{it}$ over time is limited (Hsiao (2003)).

Let $\mathbf{y}_i=\left(y_{it_i},y_{i,(t_i+1)},...y_{iT_i}\right)$ and $p_{it_i}=\Pr\left(y_{it_i}=1|\bar{\mathbf{x}}_i\right)$. Under the assumption that u_{it}^{1} and u_{it}^{0} are independent across individuals, the likelihood function for all N individuals takes the form

$$L=\prod_{i=1}^{N}\prod_{t=t_i+1}^{T_i}[P_{i11}^{y_{it}}P_{i10}^{(1-y_{it})}]^{y_{i,t-1}}[P_{i01}^{y_{it}}P_{i00}^{(1-y_{it})}]^{(1-y_{i,t-1})}p_{it_i}^{y_{it_i}}\left(1-p_{it_i}\right)^{(1-y_{it_i})}. \tag{13}$$

Equation (13) is similar to the likelihood functions of the binary qualitative response models. Because the log likelihood function is the sum of the log likelihoods for the job-holder group ($y_{i,t-1}=1$), for the job-seeker group ($y_{i,t-1}=0$) and for the initial states, $\left(\boldsymbol{\beta}^{1},\boldsymbol{\gamma}^{1}\right)$ and $\left(\boldsymbol{\beta}^{0},\boldsymbol{\gamma}^{0}\right)$ can be estimated separately using the data points from the job-holder group and those of the job-seeker group respectively.

Let $\boldsymbol{\theta}^{s}=(\boldsymbol{\beta}^{s\prime},\boldsymbol{\gamma}^{s\prime})'$ be an $m\times 1$ vector of unknown parameters, and $\boldsymbol{\theta}^{s}\in\Theta^{s}$. Let $\mathbf{w}_{it}=$

$(\mathbf{x}_{it}', \mathbf{D}_{it}')'$. The MLE of model (13) is consistent and asymptotically normally distributed under the assumptions that

A1 The parameter space Θ^s is an open bounded subset of the Euclidean m-space.

A2 $\{\mathbf{w}_{it}\}$ is uniformly bounded in i and t and $\lim_{n\to\infty} \frac{1}{n}\sum_i \sum_t \mathbf{w}_{it}\mathbf{w}_{it}'$ is a finite nonsingular matrix, where n denotes the total number of observations over i and t. Furthermore, the empirical distribution of $\{\mathbf{w}_{it}\}$ converges to a distribution function.

2.2 Evaluation of The Conditional and Unconditional Impacts

We are interested in two questions. First, how do repeated JSS and personal characteristics affect the employment rate of the job-seekers and the job-holders, respectively? Second, what are the unconditional impacts of JSS and other characteristics on the employment rate regardless of an individual's previous period employment status?

We can calcuate the conditional impact of the mth JSS for job seekers and job holders from $E\left(y_t = 1 | y_{t-1} = s, d_t^m = 1, \mathbf{x}_t\right) - E\left(y_t = 1 | y_{t-1} = s, d_t^m = 0, \mathbf{x}_t\right), \quad s = 0, 1, \quad m = 1, 2, 3$. If $\mathbf{x}_t$ is randomly drawn, $E\left(y_t = 1 | y_{t-1} = s, d_{t-1}^m\right)$ can be approximated by the sample average of the predicted probabilities of those who have $d_{it}^m = d$, $d = 0, 1$. We can also evaluate the impact of one-unit change of x_{ij} on P_{is1} by $\partial P_{is1}/\partial x_{ij}$ if x_{ij} is continuous. The population impact of a one-unit change of x_j is $\int \frac{\partial P_{is1}}{\partial x_{ij}} dF\left(x_i\right)$. Assuming that x_{ij} is randomly distributed, this impact can be approximated by $\frac{1}{N}\sum_i \frac{\partial P_{is1}}{\partial x_{ij}}$.

The transitional probability framework also allows one to trace out an individual's dynamic path of P_{ist} from its initial state. However, for simplicity we shall only provide the unconditional impacts of JSS and other socio-demographic variables regardless of an individual's employment status. The equilibrium or marginal probability of employment is computed from $\pi_i^1 = F_i^0/(1 - F_i^1 + F_i^0)$, where F_i^0 is defined in (10) that are are evaluated at $\bar{\mathbf{x}}_i = \frac{1}{N}\sum_i \mathbf{x}_i$.

3 Findings

In this section we provide estimates of model (10) that considers the following social-demographic factors as explanatory variables: (i) participation of the WorkFirst program such as JSS, alternative services (AS) (for clients who could not participate JSS directly due to problems like drug abusement and family violence), and post-employment services (PS) (for clients who have got at least part time jobs) dummies; (ii) duration dependence such as number of quarters employed or unemployed; (iii) welfare history; (iv) family information such as number of adults, number of children, age of the youngest child, marital status; (v) race and ethnicity such as dummies for whites, blacks and Hispanic; (vi) language and education such as English speaking and grade 12 and above dummies; (vii) local economy such as local unemployment rate; and (viii) geographic and time dummies. A full description of these variables are presented in Table 1.

Table 2 and Table 3 columns 3 and 4 provide the estimates of the impacts of JSS and other socio-demographic factors on the probability of being employed for job-seekers and job-holders, respectively. For job-seekers we note that first, job search services have significant impacts on the probability of employment (at 5% level), with the first JSS having the biggest impact (0.32), and the second JSS and the third JSS having less but still statistically significant impacts (0.10 and 0.12 respectively); and second, the longer an individual stays unemployed, the less likely she will find a job (the estimated coefficient for duration of unemployment is -0.1240). These two results put together make a strong case for the state to provide job search services to unemployed individuals quickly to get them stay out of unemployment.

For job-holders it shows that quite a few variables that are significant for the job seekers turn out to be insignificant. None of coefficients of the job search services are statistically significant. The estimated coefficients for the first, the second and the third JSS are 0.04, -0.01 and 0.06, respectively, and the corresponding standard errors are 0.04, 0.06 and 0.08, respectively. Total number of previous participations in post employment services (PS) also has no significant impacts on employment rate, neither is marital status nor the race or ethnic dummy for black or for Hispanic. In addition, unemployment duration is a negative but insignificant factor in determining the employment rate of the job-holders. On the other hand, the longer an individual is employed, the higher the probability that she will stay employed next period. This seems to suggest that once a client is employed, what matters is not unemployment history, but her employment history.

The impacts of one-unit change in explanatory variables to the probability of being employed for job seekers and for job holders are reported in column 3 and 4 of Table 4, respectively. This table shows that for job seekers, the first JSS increases the probability of being employed by 8.45 percent on the treated group, the second JSS increases it by a further 2.15 percent, and the third JSS increases it by an additional 0.8 percent respectively. On the other hand, none of the three JSS have statistically significant impacts on job holders.

Column 5 of Table 4 presents the equilibrium impacts of the repeated JSS participations as well as other characteristics regardless of an individual's initial state. It shows that in the long run, the first job search service increases the probability of being employed by 3.7 percent, the second job search service increases it by a further 0.3 percent, and the third job search service increases it by an additional 0.7 percent.

Columns 3, 4 and 5 of Table 4 show that the longer one stays unemployed, the less chance one has for employment. On the other hand, the longer one stays employed, the higher the chance that she stays employed in the future. This information together with the finding that unconditionally, JSS have positive effects on the probability of employment may shed light on the debate between the education-first or the employment-first strategy.

4 Diagnostic Checking

The inference reported on section 3 requires the validity of the conditional independence assumption. It is also assumed that there is no unobserved individual specific effects that could be correlated with $\mathbf{w}_{it}$. If the above two assumptions do not hold, then our maximum likelihood estimates are biased. In this section we propose methods to relax the above assumptions and then check whether these assumptions are valid. We start with the possibility that there are unobserved characteristics that can affect the selection decision, that is, we allow for the possibility that selection is due to unobservables. Then we generalize Honoré and Kyriazidou's (2000) conditional maximum likelihood estimator to allow for the presence of unobserved individual specific effects in the main equation assuming no selection on unobservables conditional on $\mathbf{w}_{it}$ and the individual effects. Finally, a conditional nonlinear two-stage least square estimator is proposed to allow for both selection due to unobservables and unobserved individual heterogeneity. Hausman (1978) statistics are constructed to test for the validities of the CI assumption and the nonexistence of unobserved individual specific effects. Our sequential (or conditional) procedures allow for the use of more sample observations conditional on other assumptions being valid or relax the need for good instruments because consistent estimation methods that simultaneously relax both assumptions impose severe restrictions on the data that can lead to significant loss of sample information. Moreover, the conditional testing procedures are more powerful to detect the alternative if the conditional event is true. Pedagogically, it is also much simpler to show the validity of proposed procedures before presenting a simultaneous test of CI and the presence of unobserved individual specific effects.

4.1 Controlling for the Selection on Unobservables

In this subsection we consider the situation that the CI assumption does not hold. Instead of specifying a complete model of the interactions between potential or actual outcomes with sequential participation decisions that may depend on interim outcomes in a dynamic optimization framework, as long as the employment status is given by (9), we can use a limited information framework to take account of the issue that after controlling for the observed characteristics the treatment decision can still be correlated with the unobserved personal characteristics in the potential or actual state equation. We note that

$$y_{it}^s = F_{it}^s + \eta_{it}^s. \tag{14}$$

Then with probability F_{it}^s, $y_{it}^s = 1$ and with probability $\left(1 - F_{it}^s\right)$, $y_{it}^s = 0$, it follows that $E\left(\eta_{it}^s\right) = F_{it}^s\left(1 - F_{it}^s\right) + \left(1 - F_{it}^s\right)\left(-F_{it}^s\right) = 0$. Therefore, if we can find instruments $\mathbf{z}_{it}$ such that $E\mathbf{z}_{it}\eta_{it}^s = 0$ and $E\mathbf{z}_{it}\mathbf{w}_{it}'$ has full column rank, where $\mathbf{w}_{it}' = \left(\mathbf{x}_{it}', \mathbf{D}_{it}'\right)$ then we can apply Amemiya's (1974) nonlinear two stage least squares estimator to obtain a consistent estimator of $\boldsymbol{\theta}^s$ in a limited information framework. Let $Y = \left(\mathbf{y}_1', ..., \mathbf{y}_N'\right)'$, $\mathbf{y}_i' = \left(y_{it_i}, y_{i,(t_i+1)}, ...y_{iT_i}\right)$, $F^s = \left(\mathbf{F}_1^{s\prime}, \mathbf{F}_2^{s\prime}, ...\mathbf{F}_N^{s\prime}\right)'$,

$\mathbf{F}_i^{s\prime} = (F_{it_i}^s, ...F_{iT_i}^s)$, $Z = (\mathbf{z}_1, ...\mathbf{z}_N)'$ and $\mathbf{z}_i = (\mathbf{z}_{it_i}, ..., \mathbf{z}_{iT_i})$. The NL2SLS estimator of $\boldsymbol{\theta}^s$ is defined as the solution that minimizes

$$S^s = (\mathbf{Y} - \mathbf{F}^s)' Z(Z'Z)^{-1} Z (\mathbf{Y} - \mathbf{F}^s). \tag{15}$$

Theorem 1 *Under A1, A2 and the assumptions that*

A3 *If* $\theta^s \neq \theta^{*s}$, $\Pr[\Pr(y^s = 1|\mathbf{w}, \boldsymbol{\theta}^s) \neq \Pr(y^s = 1|\mathbf{w}, \boldsymbol{\theta}^{*s})] > 0$;

A4 $\lim n^{-1} \sum_i \sum_t z_{it} z'_{it_i}$ *exists and is nonsingular;*

A5 $\lim n^{-1} \sum_i \sum_t z_{it} \frac{\partial F^s}{\partial \boldsymbol{\theta}^{s\prime}}$ *converges in probability uniformly in* Θ^s*;*

A6 $plim n^{-1} \sum_i \sum_t z_{it} \frac{\partial F^s}{\partial \boldsymbol{\theta}^s}|_{\theta^s}$ *is full rank;*

A7 $\lim n^{-1} \sum_i \sum_t z_{it} \frac{\partial^2 F^s}{\partial \boldsymbol{\theta}_j^{s\prime} \partial \boldsymbol{\theta}^{s\prime}}$ *converges in probability to a finite matrix uniformly in* $\theta^s \in \Theta^s$*, where* θ_j^s *is the* j*-th element of the vector* θ^s*;*

the values of $\hat{\theta}^s$ *that minimizes(15) is consistent with asymptotic covariance matrix*

$$\begin{aligned} Asy \operatorname{Cov}\left(\hat{\boldsymbol{\theta}}_{NL2SLS}\right) &= [G'Z(Z'Z)^{-1} Z'G]^{-1} \times [G'Z(Z'Z)^{-1}Z'VZ(Z'Z)^{-1} Z'G] \\ &\quad \times [G'Z(Z'Z)^{-1} Z'G]^{-1}, \end{aligned}$$

where G *is the stacked matrix of* $\partial F^s / \partial \theta^{s\prime}$*, and* V *is a diagonal matrix with diagonal elements equal to* $F_{it}^s (1 - F_{it}^s)$*.*

We use sociodemographic variables that are excluded from $\mathbf{x}_{it}$ as instruments. Among them are regional dummies and number of children in the household. Using the MLE of $\boldsymbol{\theta}$ as initial values, we apply quadratic hill climbing procedure (Quandt (1983)) to the method of score to iterate until convergence. The convergence criteria is 0.001. Column 5 and column 6 of Table 2 and Table 3 present the NL2SLS estimates and the corresponding standard errors for initially unemployed clients and for those who already have jobs, respectively. These results show that the MLE and the NL2SLS are quite similar. Because the NL2SLS is consistent regardless of the validity of the CI assumption while the MLE is only consistent when the CI assumption holds, and also because the MLE is efficient under the CI assumption, a Hausman (1978) test statistic can be constructed to test the validity of the CI assumption. The calculated Hausman statistic is 2.66 for those who are initially unemployed and 0.54 for those who already have jobs. They are not significant at 5% level for a Chi square distribution with 16 degrees of freecom. In other words, the information of the data does not appear to contradict the CI assumption.

4.2 Controlling for Individual specific Effects

In this subsection we propose to generalize Honoré and Kyriazidou's (2000) conditional maximum likelihood estimator to allow for the presence of state-dependent individual specific effects as well as slope coefficients. We assume that the error term can be decomposed into two parts,

$$u_{it}^s = \alpha_i^s + \varepsilon_{it}^s, \ s = 0, 1, \tag{16}$$

where α_i^s denotes the unobserved individual specific effects given $y_{i,t-1} = s$, and ε_{it}^s is the error term that has zero mean and variance σ_ε^2. Assuming (16) holds, we combine y_{it}^{s*}, s = 0,1, into one equation:

$$y_{it}^* = \alpha_i^0 \left(1 + \delta_i y_{i,t-1}\right) + \mathbf{x}_{it}(\boldsymbol{\beta}^0 + \mathbf{b} y_{i,t-1}) + \mathbf{D}_{it} \left(\boldsymbol{\gamma}^0 + \mathbf{g} y_{i,t-1}\right) + \varepsilon_{it}, \tag{17}$$

where $\varepsilon_{it} = \varepsilon_{it}^1 y_{i,t-1} + \varepsilon_{it}^0 \left(1 - y_{i,t-1}\right)$, $\alpha_i^1 = \alpha_i^0 \left(1 + \delta_i\right)$, $\boldsymbol{\beta}^1 = (\boldsymbol{\beta}^0 + \mathbf{b})$, and $\boldsymbol{\gamma}^1 = \boldsymbol{\gamma}^0 + \mathbf{g}$.

If α_i^1 and α_i^0 are treated as randomly distributed, one can obtain the MLE provided their conditional distributions given $\mathbf{x}_i$ can be specified. However, the consistency of the estimated parameters depends on whether the conditional distributions of α_i^1 and α_i^0 are correctly specified. Moreover, even if the distribution assumptions of α_i^1 and α_i^0 are correctly specified, the estimation can be quite involved due to multiple integrations of α_i^1 and α_i^0 over $(T_i - t_i)$ period. On the other hand, if α_i^1, α_i^0 are treated as fixed, there is no need to specify their distributions conditional on $\mathbf{x}_i$ *a priori*. Therefore, we focus on fixed effect models. Assuming that conditional on $\mathbf{x}_{it}$, ε_{it}^1 and ε_{it}^0 follow a standard type I extreme value distribution, then

$$\Pr\left(y_{it} = 1 | \mathbf{w}_{it}, \alpha_i^0, \delta_i, y_{it_i}, ... y_{i,t-1}\right) = \frac{\exp\left[\alpha_i^0 \left(1 + \delta_i y_{i,t-1}\right) + \mathbf{w}_{it}(\boldsymbol{\theta}_0 + \mathbf{c} y_{i,t-1})\right]}{1 + \exp\left[\alpha_i^0 \left(1 + \delta_i y_{i,t-1}\right) + \mathbf{w}_{it}(\boldsymbol{\theta}_0 + \mathbf{c} y_{i,t-1})\right]}, \tag{18}$$

where $\mathbf{w}_{it}' = [\mathbf{x}_{it}' , \mathbf{D}_{it}']$, $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}^{0\prime}, \boldsymbol{\gamma}^{0\prime})'$ and $\mathbf{c} = (\mathbf{b}', \mathbf{g}')'$. When $\delta_i = 0$ for all i, and $\mathbf{c} = (\gamma, 0, ..., 0)$, equation (18) becomes

$$\Pr\left(y_{it} = 1 | \mathbf{w}_{it}, \alpha_i^0, \delta_i, y_{it_i}, ... y_{i,t-1}\right) = \frac{\exp\left[\alpha_i^0 + \mathbf{w}_{it} \boldsymbol{\theta}_0 + \gamma y_{i,t-1}\right]}{1 + \exp\left[\alpha_i^0 + \mathbf{w}_{it} \boldsymbol{\theta}_0 + \gamma y_{i,t-1}\right]}, \tag{19}$$

which is of the same form as equation (6) in Honoré and Kyriazidou (2000). Therefore, the model considered in Honoré and Kyriazidou (2000) can be treated as a special case of this model (17) .

For ease of exposition, we shall only present the case that $T_i - t_i = 3$ and consider two events

$$\begin{aligned} A &= \{y_{it_i}, y_{i(t_i+1)} = 0, y_{i,(t_i+2)} = 1, y_{i,(t_i+3)}\}, \\ B &= \{y_{it_i}, y_{i(t_i+1)} = 1, y_{i,(t_i+2)} = 0, y_{i,(t_i+3)}\}. \end{aligned}$$

To simplify notations, we denote $y_{i(t_i+m)} = y_{im}$, $\mathbf{x}_{i(t_i+m)} = \mathbf{x}_{im}$, $m = 0, 1, 2, 3$ in the remaining of this section. Under the assumptions that $\mathbf{w}_{i2} = \mathbf{w}_{i3}$ and $y_{i0} = y_{i3}$,

$$P\left(A|A \cup B, \mathbf{w}_{it}, \alpha_i^0, \delta_i\right) = \frac{1}{1 + \exp\left[\left(\mathbf{w}_{i1} - \mathbf{w}_{i2}\right)\boldsymbol{\theta}_0 + \left(\mathbf{w}_{i1}y_{i0} - \mathbf{w}_{i3}y_{i3}\right)\mathbf{c}\right]}. \tag{20}$$

The conditional probability no longer depends on α_i^0 and δ_i when $\mathbf{w}_{i2} = \mathbf{w}_{i3}$ and $y_{i0} = y_{i3}$ hold. Therefore, we propose to estimate $\boldsymbol{\theta}_0$ and $\mathbf{c}$ by maximizing the objective function

$$\begin{aligned} &\sum_{i=1}^{N} 1\left(y_{i1} + y_{i2} = 1\right) \cdot 1(y_{i0} - y_{i3} = 0) \cdot \mathbf{1}\left(\mathbf{D}_{i2} - \mathbf{D}_{i3} = 0\right) \cdot K\left(\frac{\mathbf{x}_{i2} - \mathbf{x}_{i3}}{\sigma_n}\right) \\ &\times \ln\left(\frac{\exp[\left(\mathbf{w}_{i1} - \mathbf{w}_{i2}\right)\tilde{\boldsymbol{\theta}}_0 + \left(\mathbf{w}_{i1}y_{i0} - \mathbf{w}_{i3}y_{i3}\right)\tilde{\mathbf{c}}]^{y_{i1}}}{1 + \exp[\left(\mathbf{w}_{i1} - \mathbf{w}_{i2}\right)\tilde{\boldsymbol{\theta}}_0 + \left(\mathbf{w}_{i1}y_{i0} - \mathbf{w}_{i3}y_{i3}\right)\tilde{\mathbf{c}}]}\right) \end{aligned} \tag{21}$$

with respect to $\tilde{\boldsymbol{\theta}}_0$ and $\tilde{\mathbf{c}}$ over the parameter space, where $\mathbf{1}\left(A\right)$ denotes the indicator function, $K(\frac{\mathbf{x}_{i2} - \mathbf{x}_{i3}}{\sigma_n})$ denotes a kernel density function that gives more weight to those observations whose $\mathbf{x}_{i2}$ is closer to $\mathbf{x}_{i3}$, and σ_n is a bandwidth that shrinks toward 0 as n increases.

Theorem 2 *Let* $\mathbf{q}_i = [\mathbf{w}_{i1} - \mathbf{w}_{i2} \quad \mathbf{w}_{i1}y_{i0} - \mathbf{w}_{i2}y_{i3}]$, $\boldsymbol{\psi} = (\boldsymbol{\theta}_0', \mathbf{c}')$,

$$h_i(\boldsymbol{\psi}) = 1(y_{i1} + y_{i2} = 1) \cdot 1(y_{i0} - y_{i3} = 0) \cdot \mathbf{1}(\mathbf{D}_{i2} - \mathbf{D}_{i3} = 0) \times \ln\left(\frac{\exp(\mathbf{q}_i\boldsymbol{\psi})^{y_{i1}}}{1 + \exp(\mathbf{q}_i\boldsymbol{\psi})}\right). \quad (22)$$

Let the following assumptions hold:

C1 $\{(y_{i0}, y_{i1}, y_{i2}, y_{i3}, \mathbf{w}_{i1}, \mathbf{w}_{i2}, \mathbf{w}_{i3}), i = 1, ..., N\}$ is a random sample of N observations from a distribution that satisfies (18).

C2 The true value of the parameters of interest, $\boldsymbol{\psi}_0$, is in the parameter space Ψ, which is a compact subset of the Euclidean K-space(R^K), where $K = k + q + 1$.

C3 (i) The random vector $\mathbf{x}_{i2} - \mathbf{x}_{i3}$ conditional on $\mathbf{D}_{i2} = \mathbf{D}_{i3}$ is absolutely continuously distributed with density function $f(\cdot)$. $f(\cdot)$ is bounded from above, strictly positive and has support in the neighborhood of zero. (ii) $\Pr(\mathbf{D}_{i2} = \mathbf{D}_{i3}) > 0$.

C4 $E[||\mathbf{w}_{i1} - \mathbf{w}_{i2}||^2|\mathbf{A}]$ and $E[||\mathbf{w}_{i1}y_{i0} - \mathbf{w}_{i3}y_{i3}||^2|\mathbf{A}]$ are bounded on their supports, where $\mathbf{A} = [(\mathbf{x}_{i2} - \mathbf{x}_{i3}) = \mathbf{0}, \mathbf{D}_{i2} = \mathbf{D}_{i3}]$ for assumptions (C4) - (C6).

C5 The function $E(h(\boldsymbol{\psi})|\mathbf{A})$ is continuous in a neighborhood of zero for all $\psi \in \Psi$..

C6 The functions $E[(\mathbf{w}_{i1} - \mathbf{w}_{i2})'(\mathbf{w}_{i1} - \mathbf{w}_{i2})|\mathbf{A}]$ and $E[(\mathbf{w}_{i1}y_{i0} - \mathbf{w}_{i2}y_{i3})'(\mathbf{w}_{i1}y_{i0} - \mathbf{w}_{i2}y_{i3})|\mathbf{A}]$ have full column rank in the neighborhood of zero.

C7 $K : R^k \rightarrow R$ is a function of bounded variation that satisfies:(i) $\sup_{v \in R}|K(v)| < \infty$, (ii) $\int |K(v)| dv < \infty$, and (iii) $\int K(v) dv = 1$.

C8 σ_n is a sequence of positive numbers that satisfies: $\sigma_n \rightarrow 0$ as $n \rightarrow \infty$.

Let $\hat{\boldsymbol{\psi}}$ be the solution to the problem

$$\max_{\psi\in\Psi}\sum_{i=1}^{N}K(\frac{\mathbf{x}_{i2}-\mathbf{x}_{i3}}{\sigma_n})h_i(\boldsymbol{\psi}), \tag{23}$$

then $\hat{\boldsymbol{\psi}}\xrightarrow{p}\boldsymbol{\psi}_0$.

Assumptions (C2) to (C5) are the regularity conditions required for the objective function to converge to a nonstochastic limit which is uniquely maximized at $\boldsymbol{\psi}_0$ by a law of large numbers. Assumption (C6) is required for the identification of $\boldsymbol{\theta}_0$ and $\mathbf{c}$. Assumptions (C7) and (C8) are standard for kernel density estimation. The above assumptions are quite similar to those imposed in Honoré and Kyriazidou (2000), except that we separate the continuous explanatory variables and discrete variables in moment conditions and in kernels. Similar to that of Honoré and Kyriazidou (2000), the convergence rate is much slower than root-n. It is at rate $\left(n\sigma_n^k\right)^{1/2}$.

The assumptions for asymptotic normality are similar to those imposed in Honoré and Kyriazidou (2000) except that in addition to conditioning on $\mathbf{x}_{i2}=\mathbf{x}_{i3}$, we also need to condition on $\mathbf{D}_{i2}=\mathbf{D}_{i3}$. The proof of consistency and asymptotic normality follows straightforwardly to those in Honoré and Kyriazidou (2000).

Whether individual specific effects need to be controlled is critical for the adequacy of the model presented in section 2. The conditional MLE remains consistent when individual specific effects are not present. Significant information loss, however, occurs as the conditional MLE greatly restricts data points: only about 10 percent of the observations used in MLE satisfies the key condition that $\mathbf{x}_{i,s+1}\approx\mathbf{x}_{i,t+1}$ for the consistency of the conditional MLE. Furthermore, very few clients have taken the third JSS. The conditional MLE fails to converge when the third JSS dummy is included.

Table 5 presents the estimated coefficients for the first two JSS from MLE (model 1) and conditional MLE (model 2). The estimated coefficients and standard errors are in columns 3 and 4 for model 1, and in columns 5 and 6 for model 2. For initially unemployed clients, the estimated coefficients of the first JSS are both significant at 5% level; further, they are very close (0.32 versus 0.346). The second JSS has significant impact in model 1 but not so in model 2, probably due to the reduction of available observations that satisfy the consistency conditions of the conditional MLE. For those who are already employed, the estimated impacts of the first two JSS are insignificant both in model 1 and in model 2. The Hausman statistic for misspecification is merely 0.18 for the job-seeker group, which is not significant at a chi-square distribution with two degrees of freedom (5.99 at 5% level). These results appear to suggest that there is no evidence of the presence of significant unobserved individual heterogeneity conditioning on observed clients characteristics. They appear to further confirm that using model 1 is adequate to evaluate the effectiveness of repeated JSS.

4.3 Controlling for both Selection Bias and Individual Specific Effects

We have presented the nonlinear two-stage least square estimator that allows for selection due to unobservables but no unobserved individual heterogeneity, and the conditional maximum likelihood estimator that controls for unobserved individual specific effects but no selection due to unobservables. If both selection due to unobservables and unobserved individual heterogeneity are present, the above two estimators remain biased. In this subsection we propose a conditional nonlinear two-stage least square estimator that allows for unobserved individual heterogeneity as well as selection on unobservables:

$$\min \; S = \sum_{i=1}^{N} K_i L_i \left(y_i - F_i\right) z_i' \left(\sum_{i=1}^{N} K_i L_i z_i z_i'\right)^{-1} z_i \left(y_i - F_i\right) \tag{24}$$

where $L_i = 1\left(y_{i1} + y_{i2} = 1\right) \cdot 1(y_{i0} - y_{i3} = 0) \cdot 1\left(\mathbf{D}_{i2} - \mathbf{D}_{i3} = 0\right)$, $K_i = K(\frac{\mathbf{x}_{i2} - \mathbf{x}_{i3}}{\sigma_n})$, $F_i = \frac{\exp(\mathbf{q}_i \psi)}{1+\exp(\mathbf{q}_i \psi)}$ as in equation (22), $z_i' = (\tilde{z}_i', \mathbf{q}_i')$ and $\tilde{z}_i'$ is vector of instruments for individual i.

Theorem 3 *Under A1 - A7 and C1 - C8, the conditional nonlinear two-stage least squares estimator is consistent.*

The conditional nonlinear two-stage least square estimator not only greatly reduces the number of available observations, but also needs good instruments. Hence, it is unlikely to yield accurate estimates. The MLE estimates are efficient but not consistent when both the no selection on unobservables assumption and the no inidividual heterogeneity assumption hold but are inconsistent if either or both are violated. Again we use a Hausman test to check whether the null hypothesis of no selection on unobservables assumption and no inidividual heterogeneity hold. The resulting Hausman test statistic is merely of magnitude of 0.1, hence does not reject the null.

5 Conclusion

In this study we have evaluated the effects of repeated job search services and individual characteristics on the employment rates of the prime-age female TANF recipients in Washington State. We have suggested a transition probability framework to deal with the complicated issues of sample attrition, sample refreshment and duration dependence. We estimated conditional and unconditional impacts of the repeated job search services for job seekers and those who are already employed. We also proposed a nonlinear two-stage least square method to allow for selection due to unobservables, a generlized conditional maximum likelihood estimator to allow for state-dependent fixed effects and slope coefficients, and conditional nonlinear Two-Stage Least Square estimator to allow for both unobserved individual heterogeneity and selection due to unobservables. The specification tests indicated that the CI assumption and the no individual specific effects assumption conditional on included socio-demographic variables were not contradicted by the information of the data.

Our findings show that Job Search Services do have positive and significant impacts on the employment rates of those who are initially unemployed, with the first JSS increases the probability of being employed by 8.45 percent, the second JSS increases it by a further 2.15 percent, and the third JSS increases it by an additional 0.8 percent for the treated group. But the impacts of JSS are insignificant for those who are already employed. Combining these findings with the finding that for each additional quarter that an individual stays unemployed, her chance of being employed is decreased by 0.19 percent, it makes a strong case for the state to introduce job search services quickly to those who are unemployed so as to mitigate the self-enhancing effect of unemployment.

References

[1] Amemiya, T. (1974), "The Nonlinear Two-Stage Least-Squares Estimator", *Journal of Econometrics*, 2, pp. 105 - 110.

[2] Ashenfelter, O.C. and Card C.(1985), "Using the Longitudinal Structure on Earnings to Estimate the Effect of Training Programs", *The Review of Economics and Statistics*, 67(4), pp. 648 - 60.

[3] Bassi, L. J. (1984). "Estimating The Effect of Training Programs with Non-Random Selection", *The Review of Economic and Statistics* 66(1): 36 - 43.

[4] Gill, R.D. and Robins, J.M. (2001). "Causal inference for complex longitudinal data: the continuous case", *Annals of Statistics,* Vol. 29, No. 6.

[5] Hausman, J. A., (1978) "Specification Tests in Econometrics", *Econometrica*, 6, 1251 - 1271.

[6] Heckman, J. J. and V. J. Hotz (1989). "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training." Journal of American Statistical Association 84(408): 862 - 874

[7] Honoré, B. E., and Kyriazidou, E. (2000). "Panel Data Discrete Choice Models with Lagged Dependent Variables", *Econometrica*, Vol. 68, No. 4, 839 - 874.

[8] Heckman, J., and Robb R. (1985). "Alternative methods for Evaluating the Impact of Intervention", Howard Wainer eds, *Longitudinal Analysis of Labor Market Data* Eds. Heckman J. and Singer B.

[9] Holland, P. (1986). "Statistics and causal inference", *Journal of the American Statistical Association*, Vol.

[10] Hsiao, C. (2002). *Analysis of Panel Data*, Cambridge: Cambridge University Press.

[11] Lechner, M., and Miquel, R. (2001), "A Potential Outcome Approach to Dynamic Programme Evaluation: Nonparametric Identification", *memio*, University of St. Gallen.

[12] Quandt, R. E., (1983) " Computational Problems and Methods", *Handbook of Econometrics*, Vol. I, Edited by Z. Griliches and M.D. Intriligator, North-Holland Publishing Company.

[13] Rosenbaum, P.R. and D. B. Rubin (1983). "The central role of propensity score in observational studies for causal effects", *Biometrika*, 70, 41 -55.

Table 1 Variable Definitions

Variable Category	Variable Name	Definitions
WorkFirst Participation	LJSS1	Indicator for whether the first Job Search Services (JSS) had been taken before period t.
	LJSS2	Indicator for whether the second JSS had been taken before period t.
	LJSS3	Indicator for whether the third JSS had been taken before period t.
	Ltotal_AS	Total number of Alternative Services (AS) before period t.
	Ltotal_PS	Total number of Post employment Services (PS) before period t.
Employment History	lunemploycount	Total unemployed quarters before period t.
	lemploycount	Total employed quarters before period t.
Welfare history	lafdcnow	Total quarters in AFDC and/or TANF before period t. (AFDC is the predecessor of TANF).
Family	num_adlt	Number of Adults in the Assistance Unit.
	num_chld	Number of Children in the Assistance Unit.
	Age_youngest	Age of the youngest child in the Assistance Unit. Calculated based on the first quarter that WorkFirst began, 1997.IV.
	Married	Marital status. 1 indicates married.
Race	Whites	Race indicator. 1 indicates client is white.
	Blacks	Race indicator. 1 indicates client is black.
	Hispanics	Race indicator. 1 indicates client is Hispanics.
Language and Education	English	Language indicator. 1 indicates client can speak English.
	grade12	Education indicator. 1 indicates client's highest grader higher than 12.
Geographic Information	region1	Location indicator. 1 indicates client is from Region 1.
	region2	Location indicator. 1 indicates client is from Region 2.
	region3	Location indicator. 1 indicates client is from Region 3.
Local economy	Unemployrate	The unemployment rate of the county that client is in.
Time	year98	Year indicator. 1 indicates the record is in year 1998.
	year99	Year indicator. 1 indicates the record is in year 1999.
	quarter1	Quarter indicator. 1 indicates the record is in quarter 1.
	quarter2	Quarter indicator. 1 indicates the record is in quarter 2.
	quarter3	Quarter indicator. 1 indicates the record is in quarter 3.

Table 2 MLE and NL2S Estimations for Initially Unemployed Clients
($y_{i,t-1} = 0$)

Category	Variable	MLE Estimation		NL2S Estimation	
		Coefficient	Standard Error	Coefficient	Standard Error
	Intercept	-1.288755	0.0857847	-1.288699	0.3209854
WorkFirst Participation	LJSS1	0.3208265	0.0336228	0.3210588	0.8014488
	LJSS2	0.0974651	0.0460288	0.0971165	2.6166248
	LJSS3	0.127637	0.0623509	0.1276976	6.2028023
	Ltotal_AS	-0.063875	0.0136483	-0.064106	0.0418666
	Ltotal_PS	0.1732402	0.054055	0.1733488	0.0783051
Employment History	Lunemploycount	-0.121042	0.0100561	-0.121247	0.0244949
	Lemploycount	0.1768895	0.0291352	0.1769557	0.076339
Family	Num_adlt	-0.184392	0.0407723	-0.184336	0.0531489
	Married	-0.135805	0.0450123	-0.135767	0.0479468
Race	Whites	-0.139019	0.0401963	-0.138914	0.0682148
	Blacks	0.0886408	0.0526743	0.0886086	0.0757805
	Hispanics	0.1162661	0.0488666	0.1163025	0.0514206
Education	English	0.5122075	0.0576963	0.5122995	0.1662142
	Grade12	0.2229153	0.0390711	0.2229834	0.0423111
Time	Year98	0.2476005	0.0370489	0.2476764	0.1510935

Hausman Test Statistic = 5.476 < chi-2 (16)= 26.30 at 5% significance level. The Hausman test therefore does not reject the MLE estimates.

Table 3 MLE and NL2S Estimations for Initially Employed Clients
($y_{i,t-1} = 1$)

Category	Variable	MLE Estimation		NL2S Estimation	
		Coefficient	Standard Error	Coefficient	Standard Error
	Intercept	1.3670399	0.107082	1.3669012	0.3064295
WorkFirst Participation	LJSS1	0.0397998	0.0383007	0.0407879	1.515584
	LJSS2	0.0124084	0.0592132	0.0126616	4.0119876
	LJSS3	0.0677785	0.080673	0.0679399	4.9785368
	Ltotal_AS	-0.107349	0.0197488	-0.107377	0.0206929
	Ltotal_PS	0.034416	0.0271343	0.034242	0.0342981
Employment History	Lunemploycount	-0.058854	0.0316035	-0.058877	0.0406632
	Lemploycount	0.1060398	0.0126613	0.1062231	0.0169491
Family	Num_adlt	-0.107615	0.0472479	-0.107726	0.0478171
	Married	0.051051	0.0503581	0.0510632	0.0557685
Race	Whites	-0.079711	0.0483444	-0.079871	0.057891
	Blacks	0.0056359	0.0590434	0.0056498	0.089708
	Hispanics	0.0175155	0.0542377	0.0173954	0.0719317
Education	English	-0.429121	0.0833468	-0.429282	0.1318459
	Grade12	0.1902404	0.0457895	0.1902577	0.0644864
Time	Year98	0.3847283	0.0496395	0.3847448	0.0972272

Hausman Test Statistic = 0.98, does not reject the MLE estimates.

Table 4 Mean Group Impacts and Equilibrium Impacts

Variable Category	Parameter	Group Impacts		Equilibrium Impacts
		Job-Seeker Group	Job-holder Group	
WorkFirst Participation	LJSS1	0.0845	0.0066*	0.037
	LJSS2	0.0215	0.004*	0.003
	LJSS3	0.008	0.002*	0.007
	Ltotal_AS	-0.012	-0.0175	-0.027
	Ltotal_PS	0.026	-0.0188*	0.024
Employment History	lunemploycount	-0.020	-0.02*	-0.0019
	lemploycount	0.029	0.019	0.044
Earnings history	Pre_earn	0.00001	0.00001	0.00001
Family	num_adlt	-0.028	-0.0267	-0.051
	num_chld	0	0.006	0.0056
	Age_youngest	0.002	0.001	0.0038
	Married	-0.069	0.009*	-0.021
Race	Whites	-0.0289	-0.0306	-0.039
	Blacks	0.0465	0.0057*	-0.025
	Hispanics	0.0329	0.004*	0.016
Language and Education	English	0.1055	-0.0815	0.005
	grade12	0.0484	0.0292	0.054
Geographic Information	Region1	0.0365	0.0185	0.060
	Region2	0.0233	0.033*	0.0244
	Region3	0.0439	0.025*	0.0491
Local economy	Unemployrate	0	-0.003	-0.003
Time	year98	0.0831	0.0616	0.0899
	year99	0.054	-0.01	0.0130
	quarter1	0.069	-0.067	-0.084
	quarter2	0	0.0327	-0.022

* Insignificant variables are considered as having coefficient zero in calculating the unconditional equilibrium effects, and its long run impacts on job seekers (job holders) are considered as zero.

Table 5 With or Without Individual Heterogeneity *

		Without Individual Heterogeneity		With Individual Heterogeneity	
Probability of being employed		Parameter	Chi-Square	Parameter	Standard Error
	LJSS1_0	0.3228	91.40	0.35	0.176
	LJSS2_0	0.1053	5.19	0.20	0.24
Probability of staying employed	LJSS1_1	0.044	1.128	.	.
	LJSS2_1	-0.016	0.066	.	.
	LJSS1_1-LJSS1_0	.	.	0.14	0.25
	LJSS2_1-LJSS2_0	.	.	-0.09	0.29

* LJSS1_0 and LJSS2_0 are the impacts of the first and the second JSS on the probability of being employed respectively.
LJSS1_1 and LJSS2_1 are the impacts of the first and the second JSS on the probability of staying employed respectively.