Journal of Econometrics 145 (2008) 98–108

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Evaluating the effectiveness of Washington state repeated job search services on the employment rate of prime-age female welfare recipients[☆]

Cheng Hsiao[a,b,*], Yan Shen[c], Boqing Wang[d], Greg Weeks[e]

[a] Department of Economics, University of Southern California, CA 90089-0253, United States
[b] Department of Economics and Finance, City University of Hong Kong, Hong Kong
[c] China Center for Economic Research, Peking University, Beijing, 100871, PR China
[d] Department of Social and Health Services, Research and Data Analysis, PO Box 45204, Olympia, WA 98504-5204, United States
[e] Employment and Security Department, Labor Market and Economic Analysis Branch, PO Box 9046, Olympia, WA 98504-9046, United States

ARTICLE INFO

Article history:
Available online 21 June 2008

Keywords:
Transition probability
Repeated job search services
Conditional independence
Conditional maximum likelihood estimator

ABSTRACT

This paper uses an unbalanced panel dataset to evaluate how repeated job search services (JSS) 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 into account issues of sample attrition, sample refreshment and duration dependence. We also generalize Honoré and Kyriazidou's [Honoré, B.E., Kyriazidou, E., 2000. Panel data discrete choice models with lagged dependent variables. Econometrica 68 (4), 839–874] conditional maximum likelihood estimator to allow for the presence of individual-specific effects. A limited information test is suggested to test for selection issues in non-experimental data. The specification tests indicate that the (conditional on the set of the confounding variables considered) assumptions of no selection due to unobservables and/or no unobserved individual-specific effects are not violated. Our findings indicate that the first job search service does have positive and significant impacts on the employment rate. However, providing repeated JSS to the same client has no significant impact. Further, we find that there are significant experience-enhancing effects. These findings suggest that providing one job search services training to individuals may have a lasting impact on raising their employment rates.

Executive summary

This paper uses an unbalanced panel dataset to evaluate the effects of repeated job search services (JSS) on the employment rates of the prime-age female welfare recipients in the state of Washington. The JSS are the main services provided by the WorkFirst program under the Federal Temporary Assistance for Needy Family (TANF) program in the state of Washington. Since the average annual expenditure per recipient of TANF in 1998 was $12,363 and the total annual expenditure was $830 million to the State of Washington, policy makers are particularly interested in finding out how effective the JSS program is and whether repeatedly participating in JSS could fundamentally change the labor market outcomes.

We propose a transitional probability model of finding employment or staying on employment as a means to take into account issues arising from sample attrition, sample refreshment and duration dependence. Being state dependent, employment or non-employment, a transition probability model accommodates dynamics in a simple format. We estimate our transitional probability model under the assumptions of (1) conditional independence (CIA) (namely, participation in JSS can be considered exogenous conditional on the confounding variables) and (2) no individual-specific effects. We also generalize Honoré and Kyriazidou's (2000) conditional maximum likelihood estimator to allow for the presence of individual-specific effects. A limited information test is suggested to test for selection issues in non-experimental data. The specification tests indicate that the assumptions of no selection due to unobservables and/or no unobserved individual-specific effects are not violated.

☆ This research was supported in part by Employment Security Department, State of Washington, but does not necessarily represent an official position of the Employment Security Department. We thank M. Kaiser, Geert Ridder, R. Sickles, two referees and participants in the *Use of Econometrics in Informing Public Policy Makers* Conference for very helpful comments and suggestions. A similar conditional maximum likelihood estimator for a state-dependent coefficient and individual effects logit model was also found in [D'Addio and Honore, 2002. Duration dependence and time-varying variables in discrete time duration models. Working Paper No. 2002-13. Department of Economics, University of Aarhus] after the completion of the preliminary version of this paper.

* Corresponding author at: Department of Economics, University of Southern California, CA 90089-0253, United States.
E-mail address: chsiao@rcf.usc.edu (C. Hsiao).

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.05.011

We find that only the first Job Search Services had positive and statistically significant impacts on employment rates. The probabilities of employment are increased by about 3.6% for the non-employed and by 3.4% for the employed. Repeating JSS does not appear to raise the probability of employment. We also find that each additional quarter of non-employment reduces the probability of employment by 2.9% and each additional quarter of employment raises the probability of employment by 2.3%. The "experience-enhancing" effect together with the finding that the first JSS does raise the probability of employment appears to provide empirical support for requiring TANF recipients to engage in employment-related activities, and a focus on short-term less expensive job search activities can be beneficial. However, repeating JSS does not appear to yield any additional benefit. It appears that for those who have already taken one JSS, perhaps other training programs, such as long-term human capital augment activities could be more beneficial rather than prodding them to repeatedly taking JSS.

1. Introduction

This paper uses an unbalanced panel dataset to evaluate the effects of repeated job search services (JSS) on the employment rates of the prime-age female welfare recipients in the state of Washington. The JSS are the main services provided by the WorkFirst program under the Federal Temporary Assistance for Needy Family (TANF) program in the state of Washington. The criterion for an individual to be on TANF is the income level. The TANF program was established in 1996 by the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) to replace the Aid to Families with Dependent Children (AFDC) program. The TANF requires all adults who receive cash welfare assistance to work or to engage in employment-related activities or face sanctions. The WorkFirst program was initiated in 1997 using block grants from TANF. TANF recipients are required to participate in the main activities of JSS that may last up to 12 continuous weeks and include classroom instructions on how to find jobs; pre-employment training; high-wage/high-demand training, etc. A TANF recipient is also allowed to participate in the JSS repeatedly with the approval of case managers. However, since the average annual expenditure per recipient of TANF in 1998 was $12,363 and the total annual expenditure was $830 million to the State of Washington, policy makers are particularly interested in finding out how effective the JSS program is and whether repeatedly participating in JSS could fundamentally change the labor market outcomes.

A central issue for using non-experimental data for program evaluation is the non-randomness of program participants and non-participants (e.g. see the survey of Friedlander et al. (1997)). The measurements may be subject to various selection bias (e.g., Angrist et al. (1996), Heckman and Robb (1985), and Rosenbaum and Rubin (1983)). With panel data outcomes of the same individual before and after the treatment could be observed. Moreover, panel data provides information on duration of employment and unemployment spell which is considered important but is often not available in cross-sectional data (e.g., Heckman et al. (1999)).

Contrary to the early studies using panel data (e.g., Bassi (1984), Ashenfelter and Card (1985), Heckman and Hotz (1989), Hotz et al. (2006) and Hotz et al. (2002)), our dataset is not a balanced panel data. A significant feature of our data is that clients entered and left the TANF program at different time periods. There is also the issue of right censoring because the data period ends at the last quarter of 2000. Table 1 shows the frequency of new entrants of our sample at each quarter from the second quarter of 1998 to the last quarter of 2000. For instance, only about 3.4% clients have been observed over the complete sample period. In this paper, we propose a transitional probability model of finding employment or staying on employment as a means to take into account issues arising from sample attrition, sample refreshment and duration dependence. Being state dependent, employment or non-employment, a transition probability model accommodates dynamics in a simple format. Under the assumption that conditional on certain regressors, the transition probability from one state at time period $t-1$ to another state at time period t is homogeneous, the observed sample provides some information on the unknown parameters characterizing the transition probability as long as the time series observations for a given individual exceed two and there are sufficient mixtures of samples in all four possible states. There is no need to be concerned about when an individual entered and exited the sample. Moreover, from the estimated transitional probabilities, it is possible to trace out an individual's path of employment over time and duration of employment.

Table 1
Frequency distribution of new entrants in each quarter

Quarter	Freq.	Percent	Cum.
1998.II	1 224	3.40	3.40
1998.III	2 233	6.21	9.61
1998.IV	2 920	8.12	17.73
1999.I	3 050	8.48	26.21
1999.II	3 420	9.51	35.71
1999.III	3 935	10.94	46.65
1999.IV	3 793	10.54	57.19
2000.I	4 070	11.31	68.51
2000.II	4 065	11.30	79.81
2000.III	3 755	10.44	90.25
2000.IV	3 509	9.75	100.00
Total	35 974	100.00	

We estimate our transitional probability model under the assumptions of (1) conditional independence (CIA) (namely, participation in JSS can be considered exogenous conditional on the confounding variables) and (2) no individual-specific effects. Diagnostic tests are then proposed to examine the validity of these assumptions.

Our findings show that only the first job search services have positive and statistically significant effects on the employment rate regardless of whether one is initially employed or non-employed. However, providing repeated JSS to the non-employed clients or to those who are already employed has no statistically significant impacts on the probability of finding employment or staying employed. Furthermore, the probability of employment is also influenced by the duration in employment or non-employment, 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 checking procedures are proposed and conducted in Section 4. Conclusions are given in Section 5. Detailed descriptions of our data are presented in Appendix.

2. The model

Most empirical investigations evaluating government training programs using experimental and non-experimental data do not consider the effects of the repeated job trainings. Considering sequential treatments from an intertemporal optimization framework is very complicated. Most literature follow the lead of Robins (1986) and Gill and Robins (2001) by treating sequential treatments as some form of sequential randomization (e.g. Lechner and Miquel (2001), Lechner (2004)). Under the assumption of some form of weak or strong dynamic conditional independence assumption (e.g. Lechner (2004)), the participations of the treatment

Table 2
Frequency distribution of number of JSS taken

Number of JSS	Freq.	Percent	Cum.
0	8,856	24.62	24.62
1	13,446	37.38	61.99
2	8,968	24.93	86.92
3	3,075	8.55	95.47
4	1,058	2.94	98.41
5	353	0.98	99.39
6	148	0.41	99.81
7	45	0.13	99.93
8	18	0.05	99.98
9	6	0.02	100.00
10	1	0.00	100.00
Total	35,974	100.00	

are essentially treated as exogenous. Unfortunately, the demand on the data to measure the effects of different treatment sequences is huge and complicated. For instance, if there are two periods, there are four possible states for the two-period sequences, (0, 0), (1, 0), (0, 1) and (1, 1), where 1 indicates receiving treatment in that period, and 0 not. If there are n periods, then there will be 2^n possible states for the n-period sequences. Matching estimates will have to be computed for each possible state to control the confounding effects of observables that vary across individuals and over time. Therefore, for the n-period data, $\sum_{t=1}^{n} 2^t$ matching estimates will have to be computed. Such a huge number of measurements might fail to convey a clear picture to policy makers. Therefore, in this paper we propose separating the timing effects and treatment effects.

2.1. A transitional probability model for the outcomes

In this section we propose a transitional probability framework to take into 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 at time t and the value 0 otherwise, $t = t_i, t_i + 1, \ldots, T_i$, where t_i and T_i denote the first period and the last period that client i is observed. We assume that y_{it} depends on the previous state, $y_{i,t-1}$, previous JSS treatments, and strictly exogenous socio-demographic variables, $\mathbf{x}_{it}$. Conditional on $s = y_{i,t-1} = 0$ or 1, the potential state of employment is given by

$$y_{it}^{1s*} = \alpha_i^{1s} + \mathbf{x}_{it}'\boldsymbol{\beta}^{1s} + \mathbf{D}_{it}'\boldsymbol{\gamma}^{1s} + \varepsilon_{it}^{1s}, \quad s = 0, 1, \tag{1}$$

and the potential state of unemployment is given by

$$y_{it}^{0s*} = \alpha_i^{0s} + \mathbf{x}_{it}'\boldsymbol{\beta}^{0s} + \mathbf{D}_{it}'\boldsymbol{\gamma}^{0s} + \varepsilon_{it}^{0s}, \quad s = 0, 1, \tag{2}$$

where $\mathbf{D}_{it}$ is a 4×1 vector that takes the value $\mathbf{D}_{it}' = (0, 0, 0, 0)$, $\mathbf{D}_{it}' = (1, 0, 0, 0)$, $\mathbf{D}_{it}' = (1, 1, 0, 0)$, $\mathbf{D}_{it}' = (1, 1, 1, 0)$, and $\mathbf{D}_{it}' = (1, 1, 1, 1)$ if the ith individual at time t received no JSS, 1 JSS, 2 JSS, 3 JSS and 4 or more JSS before time period t, respectively.[1] We group 4 or more JSS into 4 JSS dummy because over 98% of clients took not more than 4 JSS treatments (Table 2). Let

$$\begin{aligned} y_{it}^{s*} &= y_{it}^{1s*} - y_{it}^{0s*} \\ &= \alpha_i^{s} + \mathbf{x}_{it}'\boldsymbol{\beta}^{s} + \mathbf{D}_{it}'\boldsymbol{\gamma}^{s} + \varepsilon_{it}^{s}, \quad s = 0, 1, \end{aligned} \tag{3}$$

where $\alpha_i^s = \alpha_i^{1s} - \alpha_i^{0s}$, $\boldsymbol{\beta}^s = \boldsymbol{\beta}^{1s} - \boldsymbol{\beta}^{0s}$, $\boldsymbol{\gamma}^s = \boldsymbol{\gamma}^{1s} - \boldsymbol{\gamma}^{0s}$. Combining y_{it}^{1*} and y_{it}^{0*} into one equation yields

$$\begin{aligned} y_{it}^{*} &= \alpha_i^{0}\left(1 + \delta_i y_{i,t-1}\right) + \mathbf{x}_{it}'\left(\boldsymbol{\beta}^{0} + \mathbf{b}y_{i,t-1}\right) \\ &\quad + \mathbf{D}_{it}'\left(\boldsymbol{\gamma}^{0} + \mathbf{g}y_{i,t-1}\right) + \varepsilon_{it}, \end{aligned} \tag{4}$$

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

$$y_{it} = \begin{cases} 1 & \text{if } y_{it}^* > 0, \\ 0 & \text{if } y_{it}^* \le 0. \end{cases} \tag{5}$$

We assume that

A1 The error ε_{it} follows an independent type I extreme value distribution.

A2 Conditional on $y_{i,t-1}$ and $\mathbf{x}_i' = \left(\mathbf{x}_{i,t_i}', \ldots, \mathbf{x}_{i,T_i}'\right)$, the distribution of ε_{it} and D_{it} are independent (Conditional Independence Assumption (CIA) (Rosenbaum and Rubin, 1983) or Ignorable Treatment Assignment Assumption (Heckman and Robb, 1985; Holland, 1986)),[2]

$$\varepsilon_{it} \perp D_{it} | y_{i,t-1}, \mathbf{x}_i. \tag{6}$$

A3 Conditional on $y_{i,t-1}$ and $\mathbf{x}_i$, $\alpha_i^{1s} = \alpha_i^{0s} = c$ so there are no unobserved individual-specific effects.

Under A1–A3,

$$\begin{aligned} &\Pr\left(y_{it} = 1 | \mathbf{x}_{it}, y_{i,t-1}\right) \\ &= \frac{\exp\left(c(1 + \delta y_{i,t-1}) + \mathbf{x}_{it}'(\boldsymbol{\beta}^0 + \mathbf{b}y_{i,t-1}) + \mathbf{D}_{it}'\left(\boldsymbol{\gamma}^0 + \mathbf{g}y_{i,t-1}\right)\right)}{1 + \exp\left(c(1 + \delta y_{i,t-1}) + \mathbf{x}_{it}'(\boldsymbol{\beta}^0 + \mathbf{b}y_{i,t-1}) + \mathbf{D}_{it}'\left(\boldsymbol{\gamma}^0 + \mathbf{g}y_{i,t-1}\right)\right)} \\ &= F_{it}, \quad \text{for } t = t_i + 1, \ldots, T_i, i = 1, \ldots, N. \end{aligned} \tag{7}$$

However, the lagged value of y before t_i is unobserved. We follow Heckman (1981) to approximate the initial state by

A4

$$\Pr\left(y_{it_i} = 1 | \mathbf{x}_i, D_{it_i}\right) = \frac{\exp\{\bar{c} + \bar{\mathbf{x}}_i'\mathbf{b}^* + D_{it_i}'\boldsymbol{\gamma}^*\}}{1 + \exp\{\bar{c} + \bar{\mathbf{x}}_i'\mathbf{b}^* + D_{it_i}'\boldsymbol{\gamma}^*\}} = P_{it_i},$$

where $\bar{\mathbf{x}}_i = \frac{1}{(T_i - t_i + 1)}\sum_{t_i}^{T_i}\mathbf{x}_{it}$, $\bar{c}$, $\mathbf{b}^*$ and $\boldsymbol{\gamma}^*$ are parameters to be estimated.

Under A1–A4, the likelihood function of $\left(y_{it_i}, y_{i,(t_i+1)}, \ldots, y_{iT_i}\right)$, $i = 1, \ldots, N$ is equal to

$$L = \prod_{i=1}^{N}\prod_{t=t_i}^{T_i} F_{it}^{y_{it}}\left(1 - F_{it}\right)^{1 - y_{it}} P_{it_i}^{y_{it_i}}\left(1 - P_{it_i}\right)^{(1 - y_{it_i})}. \tag{8}$$

Because all the parameters vary depending on the outcome in the previous period, one can obtain a full set of estimates by obtaining first one subset, then the other in two separate maximizations. That is, the maximum likelihood estimator (MLE) can be obtained by maximizing two separate likelihood functions of the binary response models, one conditional on $y_{i,t-1} = 1$, the other on $y_{i,t-1} = 0$.

[1] If the ith individual received only one training before $t = 2$, $\mathbf{D}_{i2}' = (1, 0, 0, 0)$, but $\mathbf{D}_{i3}' = (1, 1, 0, 0)$ if he received another training before $t = 3$. If an individual's treatment status remains unchanged between time t and $t - 1$, then $D_{it} = D_{i,t-1}$. We assume no decaying effect on JSS for the transition probability since the span of our data is only about two years. It may be treated as a first order approximation. Of course, we may generalize the specification of (3) to introduce the decaying effect. However, this may introduce considerable multicollinearity in the data. We hope that the inclusion of length of employment and unemployment could capture part of the time effects raised by Card and Sullivan (1988) and Heckman and Smith (1999) since the length of employment or unemployment changes with most recent changes in employment or unemployment status.

[2] As pointed out by a referee, under A2 one can also estimate the treatment effect by matching. The advantages of matching are that no functional form or distribution assumptions need to be made. The disadvantages are that (1) one only estimates the treatment effect, (2) the estimated treatment effects are sensitive to the way matching adjustments are computed. The advantages of parametric approach are that (1) in addition to the treatment effects, we can also estimate the effects of other socio-demographic variables and (2) efficient estimation methods can be implemented. The disadvantages are that both functional form and distribution assumption need to be imposed, and if these assumptions are not valid, the inference is biased.

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 non-employed and employed, respectively? Second, what are the unconditional impacts of JSS and other characteristics on the employment rate if an individual is randomly drawn from welfare recipients (i.e., regardless of an individual's employment status)?

Let $d_{it}^{m} = 1$ if the ith individual received exactly m JSS before time period t and after t and 0 otherwise. We define the treatment effect of the mth JSS for individual i conditional on $d_{it}^{m-1} = 1, y_{i,t-1} = s, s = 0, 1$ as changes of probability of employment,

$$\Delta_{it}^{sm}\left(\mathbf{x}, d_{t}^{m} = 1\right) = \Pr\left(y_{it} = 1 | y_{i,t-1} = s, \mathbf{x}, d_{it}^{m} = 1\right) - \Pr\left(y_{it} = 1 | y_{i,t-1} = s, \mathbf{x}, d_{it}^{m-1} = 1\right).$$

The average treatment effect (ATE) of the mth JSS of TANF recipients conditional on last period's employment $s, s = 0, 1$, is defined as

$$\text{ATE}\left(\Delta^{ms}\right) = \int \left[\Pr\left(y_{it} = 1 | y_{i,t-1}, \mathbf{x}, d_{it}^{m} = 1\right) - \Pr\left(y_{it} = 1 | y_{i,t-1}, \mathbf{x}, d_{it}^{m-1} = 1\right)\right] f(\mathbf{x})\, \mathrm{d}\mathbf{x}, \quad (9)$$

where $f(\mathbf{x})$ denotes the probability density function of $\mathbf{x}$ of TANF recipients. The average treatment effect of the treated (TT) of the mth JSS for those employed, $y_{i,t-1} = 1$, or non-employed, $y_{i,t-1} = 0$ conditional on $d_{it}^{m-1} = 1$, is defined as

$$\text{TT}\left(\Delta^{ms}\right) = \int \left[\Pr\left(y_{it} = 1 | y_{i,t-1}, \mathbf{x}, d_{it}^{m} = 1\right) - \Pr\left(y_{it} = 1 | y_{i,t-1}, \mathbf{x}, d_{it}^{m-1} = 1\right)\right] f\left(\mathbf{x} | d_{t}^{m-1} = 1\right) \mathrm{d}\mathbf{x}. \quad (10)$$

The ATE of the mth JSS is the mean impact of the mth JSS if a random TANF client is assigned m JSS given he has already received $(m-1)$ JSS. The TT of the mth JSS is the mean impact of the mth JSS if the same selection rule for assigning treatment applies to the future. If $\mathbf{x}_{it}$ is randomly drawn, TT (Δ^{ms}) or ATE (Δ^{ms}) can be approximated by taking the sample average of the predicted probabilities of those who have $d_{it}^{m-1} = d, d = 0, 1$. We can also evaluate the impact of one-unit change of x_{ij} on $P_{ist}, s = 0, 1$, by $\partial P_{ist}/\partial x_{ij}$ if x_{ij} is continuous, where $P_{ist} = \Pr\left(y_{it} = 1 | y_{i,t-1} = s, \mathbf{x}\right)$. The population impact of a one-unit change of x_j is $\int \frac{\partial P_{ist}}{\partial x_{ij}} \mathrm{d}F(x_i)$. Assuming that our sample is a random sample, this impact can be approximated by $\frac{1}{N}\sum_i \frac{\partial P_{ist}}{\partial x_{ij}}$.

The transitional probability framework also allows one to trace out an individual's dynamic path of P_{it} from its initial state, following the rule of

$$\begin{bmatrix} P_{it} \\ 1 - P_{it} \end{bmatrix} = \begin{bmatrix} P_{i1,t-1} & (1 - P_{i1,t-1}) \\ P_{i0,t-1} & (1 - P_{i0,t-1}) \end{bmatrix} \begin{bmatrix} P_{i,t-1} \\ 1 - P_{i,t-1} \end{bmatrix},$$

where P_{it} denotes the marginal or unconditional probability of employment. Since the transition probability depends on $\mathbf{x}$, an individual's dynamic path depends on her past, current and future $\mathbf{x}$. We also evaluate the impact of $\mathbf{x}$ on the long-run marginal probability of employment regardless of an individual's employment status by letting $\pi_i^1 = F_i^0/[1 - F_i^1 + F_i^0]$, where F_i^s denotes the probability of employment given the previous state is s, where $s = 0, 1$, and are evaluated at the time series mean of $\mathbf{x}_{it}, \bar{\mathbf{x}}_i$.

3. Findings

In this section we provide estimates of model (7) using panel data from Washington State female TANF recipients between the age of 25–35 from the second quarter of 1998 to the third quarter of 2000.[3] Following the suggestion of the literature (e.g., Ashenfelter and Card (1985), Friedlander et al. (1997) and Hotz et al. (2006)) we consider the following types of socio-demographic factors as possible explanatory variables that determine an individual's current employment status: (i) previous participation in the WorkFirst program such as JSS, alternative services (AS) (for clients who could not participate in JSS directly due to problems like drug abuses and family violence), and previous participation in post-employment services (PS) (for clients who have got at least part time jobs); (ii) duration dependence such as number of quarters employed or non-employed; (iii) earnings 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) education dummy such as grade 12 and above dummies; (vii) local economy such as local non-employment rate; and (viii) geographic and time dummies. A full description of these variables are presented in Table 3 and the descriptive statistics are provided in Table 4.[4]

Tables 5 and 6 (columns (1)) provide the estimates of model (7) with all JSS variables and socio-demographic variables included. For those who are non-employed we note that the first job search services have significant impacts on the probability of employment (at 5% level), while the rest of the job search services have insignificant impacts; and second, the longer an individual stays non-employed, the less likely she will be employed (the estimated coefficient for duration of non-employment is −0.16). These two results put together make a strong case for the state to provide job search services to non-employed individuals quickly to get them out of non-employment. The negative coefficient for number of alternative services previously taken may be viewed as capturing the attributes of a client that are deemed not employable since people who take alternative services are people who have problems of substance abuse or domestic violence.

For those who are employed Table 6 also indicates that only the first JSS has statistically significant positive impacts on the employment rate. Further, some variables that are significant for the non-employed turn out to be insignificant for the employed. Age of the youngest child is no longer statistically significant for the probability of staying employed once she is employed, neither is marital status, nor is non-employment duration relevant to the probability of staying on the job. On the other hand, the longer an individual is employed, the higher the probability that she will stay employed the next period. This seems to suggest that once a client is employed, what matters is not non-employment history, but her employment history. Furthermore, previous post-employment services appear beneficial for staying employed.

As only the first JSS appears to have significant impacts for the non-employed and the employed, we re-estimate the model by dropping all other JSS dummies for both employed and non-employed and present the results in columns (2) of Tables 4 and 5,

[3] The WorkFirst program formally started in the fourth quarter of 1997, but the data started in the second quarter of 1998 as there were very few observations in the fourth quarter of 1997 and the first quarter of 1998. However, this shall not cause measurement problems for individuals in this sample, since their information had been recorded through their participations in AFDC (Aids for Families with Dependent Children). For some participants the record can be traced back to the middle of 1980s. Consequently, the measurements of variables like previous total unemployed quarters, previous total employed quarters are not affected by the actual starting point of the data.

[4] We define non-employed as those who had zero quarterly earnings reported. Information on the data may be found in Lerch and Mayfield (1999) or the Appendix.

Table 3
Variable definitions

Variable category	Variable name	Definitions
WorkFirst participation	First JSS	Indicator for whether the first Job Search Services (JSS) has been taken before period t.
	Second JSS	Indicator for whether the second JSS has been taken before period t.
	Third JSS	Indicator for whether the third JSS has been taken before period t.
	Four or more JSS	Indicator for whether the individual has taken four or more JSS before period t.
	Previous total AS	Total number of Alternative Services (AS) taken before period t.
	Previous total PS	Total number of Post-employment Services (PS) taken before period t.
Employment history	Previous non-employment quarters	Total unemployed quarters before period t.
	Previous employment quarters	Total employed quarters before period t.
Earnings history	Pre-WorkFirst earnings	Earnings before the quarter that client entered the WorkFirst, measured in dollars.
	Previous wage rate	Wage rate of previous job.
Family	Number of adults	Number of adults in the assistance unit.
	Number of kids	Number of children in the assistance unit.
	Age of the youngest kid	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[a]	Whites	Race indicator. 1 indicates client is white.
	Blacks	Race indicator. 1 indicates client is black.
	Hispanics	Race indicator. 1 indicates client is Hispanics.
Education	grade 12	Binary indicator with 1 indicates client's highest grade higher than 12.
Geographic	Region 1–Region 5	Location indicator. Region i indicates client is from Region i, $i = 1, \ldots, 5$
Local economy	Unemployment rate	The unemployment rate of the county that the client was located at time t.

[a] Asian and Native American are the omitted category.

respectively. The estimates that drop the insignificant JSS dummies look very close to those with the full set of JSS dummies used as regressors. We therefore focus our discussion of the impacts of one-unit change of a variable based on the model with only 1 JSS used as a conditional variable.

The mean marginal impacts of one-unit change in explanatory variables to the probability of being employed for the non-employed and for the employed are reported in columns (1) and (2) of Table 7, respectively. This table shows that the first JSS increases the probability of being employed by 3.6% for those who were non-employed in the previous period, and increases the probability of staying employed by 3.4%.

Column (3) of Table 7 presents the impacts of the repeated JSS participations as well as other characteristics regardless of an individual's initial state of employment evaluated at $\bar{\mathbf{x}}_i$. It shows that in the long run, the first job search service increases the probability of being employed by 4.7%.[5] It also implies that the longer one stays non-employed, 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.[6] This information together with the finding that unconditionally, JSS have positive effects on the probability of employment may shed additional light on the cost and benefit analysis between the education-first or the employment-first strategy.

Table 4
Descriptive statistics

Variable	Mean	Min	Max
First JSS	0.630	0	1
Second JSS	0.219	0	1
Third JSS	0.073	0	1
Four or more JSS	0.025	0	1
Number of kids	2.407	0	12
Age of the youngest kid	5.064	0	18
Previous total AS	0.645	0	9
Previous total PS	0.127	0	5
Pre-WorkFirst earnings	321.535	0	20 110
Previous non-employment quarters	1.463	0	11
Previous employment quarters	0.871	0	10
Number of adults	1.183	0	4
Married	0.182	0	1
Whites	0.669	0	1
Hispanics	0.115	0	1
Grade 12	0.139	0	1
Region 1	0.141	0	1
Region 2	0.141	0	1
Region 3	0.077	0	1
Region 4	0.209	0	1
Region 5	0.219	0	1
Unemployment rate	5.602	2.566	15.871
Previous wage rate	570.241	0	15 734

4. Diagnostic checking

The inference reported in Section 3 requires the validity of the conditional independence assumption (CIA, A3). As pointed out by a referee, CIA implies that there are no unobserved individual-specific effects that could be correlated with $\mathbf{w}_{it}$, $\mathbf{w}'_{it} = (\mathbf{x}'_{it}, \mathbf{D}'_{it})$. If CIA assumption does not hold, then our maximum likelihood estimates are biased. In this section we propose specification tests to check whether it is appropriate to impose this assumption. We start by first relaxing the restrictive nature of the conditional independence by assuming a weaker version of conditional independence, namely, conditional on the individual-specific effects and $\mathbf{x}$, $(y^{s*}_{it} \perp \mathbf{D}_{it} | \alpha^s_i, \mathbf{x}_{it})$. We propose a conditional maximum likelihood estimator to allow for the presence of unobserved individual-specific effects conditional on $\mathbf{D}_{it}$ being predetermined, then suggest a Hausman (1978) type specification test to test for the presence of individual-specific effects. We then propose a limited information framework to test for the strong version of the conditional independence assumption, namely, $(y^{s*}_{it} \perp \mathbf{D}_{it} | \mathbf{x}_{it})$ by simultaneously testing for the presence of individual-specific effects and endogeneity of participation in JSS. We propose to test these assumptions sequentially conditional on other assumptions being valid to allow for more efficient use of sample observations or to 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 have more power relative to particular alternative hypotheses 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

5 Hotz et al. (2006) finds a 4.3% gain for the San Diego county and 13.6% gain for the Riverside county for the California GAIN program.

6 It should be noted that staying employed does not necessarily mean staying on the same job.

Table 5
MLE estimations for initially non-employed clients

	(1)	(2)	(3)
First JSS	0.203***	0.189***	0.244***
	(0.038)	(0.036)	(0.049)
Second JSS	−0.074		−0.086
	(0.051)		(0.067)
Third JSS	0.101		0.094
	(0.084)		(0.108)
Four or More JSS	0.188		0.193
	(0.127)		(0.127)
Number of kids	−0.018	−0.018	−0.017
	(0.014)	(0.014)	(0.014)
Age of the youngest kid	0.018***	0.018***	0.019***
	(0.004)	(0.004)	(0.004)
Previous total AS	−0.091***	−0.096***	−0.090***
	(0.021)	(0.020)	(0.021)
Previous total PS	0.130**	0.130**	0.130**
	(0.062)	(0.062)	(0.062)
Pre-WorkFirst Earnings	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)
Previous non-employment quarters	−0.160***	−0.152***	−0.131***
	(0.018)	(0.014)	(0.029)
Previous employment quarters	0.129***	0.134***	0.120
	(0.035)	(0.034)	(0.074)
Number of adults	−0.167***	−0.167***	−0.168***
	(0.047)	(0.047)	(0.047)
Married	−0.225***	−0.224***	−0.225***
	(0.051)	(0.051)	(0.051)
Whites	−0.166***	−0.166***	−0.163***
	(0.039)	(0.039)	(0.039)
Hispanics	0.076	0.074	0.075
	(0.055)	(0.055)	(0.056)
Grade 12	0.190***	0.190***	0.189***
	(0.046)	(0.046)	(0.046)
Region 1	0.219***	0.217***	0.227***
	(0.056)	(0.056)	(0.057)
Region 2	0.178***	0.179***	0.174***
	(0.063)	(0.063)	(0.064)
Region 3	0.295***	0.294***	0.296***
	(0.068)	(0.068)	(0.072)
Region 4	−0.049	−0.046	−0.048
	(0.059)	(0.059)	(0.062)
Region 5	−0.019	−0.019	−0.016
	(0.050)	(0.050)	(0.053)
Unemployment rate	−0.022**	−0.022**	−0.020
	(0.010)	(0.010)	(0.016)
Previous wage rate	0.0001	0.0001	−0.00005
	(0.0001)	(0.0001)	(0.00003)
Constant	−0.597***	−0.608***	−0.689***
	(0.105)	(0.104)	(0.140)

Standard errors in parentheses. Column (1) The base model. Column (2) The model without insignificant Job Search Indicators. Column (3) The estimated coefficients are from the exogeneity test models, the instruments and their interactions with the explanatory variables are not presented but available from the authors. The omitted instruments and the interactions are: number of cases handled by each case manager (mgrcasenum), number of clients participating in each CSO (csonum), and the interactions of mgrcasenum and with previous non-employment quarters, previous employment quarters, previous wage rate, the four JSS indicators, and interaction of csonum with employment quarters, previous wage rate, the four JSS indicators. The likelihood ratio test statistic is 15.07, with a p value of 0.58.

** Significant at 5%.

*** Significant at 1%.

endogeneity of $\mathbf{D}_{it}$ and the presence of unobserved individual-specific effects.[7]

4.1. 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 and slope coefficients assuming $\mathbf{D}_{it}$ is predetermined. We note that under our weaker version of CIA, when (A3) does not hold, both α_i^1 and α_i^0 (or $\alpha_i^0\delta_i$ and α_i^0) appear in (4). If α_i^1 and α_i^0 are treated as randomly distributed, one can obtain the MLE provided their conditional distributions given $\mathbf{w}$ 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 the need for multiple integrations of α_i^1 and α_i^0 over the $(T_i - t_i)$ period. On the other hand, if α_i^1 and α_i^0 are treated as fixed, there is no need to specify their distributions conditional on $\mathbf{w}_i$, *a priori*. Therefore, we focus on fixed effect models for (7). However, there are not enough time series observations to obtain consistent estimates of α_i^1 and α_i^0, hence $\boldsymbol{\beta}$ or $\boldsymbol{\gamma}$ (e.g. see Hsiao (2003)) no matter how large N is.

A consistent estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}', \mathbf{b}', \mathbf{g}')'$ can be derived if one can transform (7) into a model without incidental parameters

[7] In other words, we distinguish the source of correlations as those due to the presence of individual-specific effects or correlations between $\mathbf{D}_{it}$ and the random error term, ε_{it}.

Table 6
Maximum likelihood estimates for initially employed clients

	(1)	(2)	(3)
First JSS	0.180***	0.165***	0.112
	(0.044)	(0.042)	(0.096)
Second JSS	−0.066		−0.119
	(0.069)		(0.152)
Third JSS	−0.060		−0.049
	(0.120)		(0.259)
Four or More JSS	−0.123		−0.387
	(0.191)		(0.413)
Number of kids	−0.008	−0.008	−0.008
	(0.017)	(0.017)	(0.018)
Age of the youngest kid	0.009	0.009	0.009
	(0.006)	(0.006)	(0.006)
Previous total AS	−0.043	−0.039	−0.041
	(0.035)	(0.035)	(0.035)
Previous total PS	0.099**	0.101**	0.095**
	(0.040)	(0.040)	(0.041)
Pre-WorkFirst earnings	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)
Previous non-employment quarters	−0.038	−0.056	0.094
	(0.042)	(0.041)	(0.089)
Previous employment quarters	0.051**	0.034*	0.097**
	(0.021)	(0.018)	(0.043)
Number of adults	−0.164***	−0.165***	−0.162***
	(0.060)	(0.060)	(0.060)
Married	0.078	0.078	0.076
	(0.063)	(0.063)	(0.064)
Whites	−0.116**	−0.111**	−0.113**
	(0.048)	(0.048)	(0.048)
Hispanics	0.092	0.094	0.101
	(0.067)	(0.067)	(0.067)
Grade 12	0.086	0.086	0.081
	(0.058)	(0.058)	(0.058)
Region 1	0.105	0.105	0.121*
	(0.068)	(0.068)	(0.069)
Region 2	0.055	0.059	0.056
	(0.077)	(0.077)	(0.078)
Region 3	−0.074	−0.076	−0.052
	(0.082)	(0.082)	(0.088)
Region 4	0.002	0.001	0.040
	(0.077)	(0.077)	(0.081)
Region 5	−0.113*	−0.113*	−0.119*
	(0.065)	(0.065)	(0.070)
Unemployment rate	−0.047***	−0.047***	−0.067***
	(0.012)	(0.012)	(0.018)
Previous wage rate	0.0004***	0.0004***	0.0004***
	(0.00002)	(0.00002)	(0.00004)
Constant	0.766***	0.786***	0.757***
	(0.132)	(0.131)	(0.185)

Standard errors in parentheses. Column (1) The base model. Column (2) The model without insignificant Job Search Indicators. Column (3) The estimated coefficients are from the exogeneity test models, the instruments and their interactions with the explanatory variables are not presented but available from the authors. They are available from the authors upon request. The omitted instruments and the interactions are: number of cases handled by each case manager (mgrcasenum), number of clients participating in each CSO (csonum), and the interactions of mgrcasenum and with previous non-employment quarters, previous employment quarters, previous wage rate, the four JSS indicators, and interaction of csonum with employment quarters, previous wage rate, the four JSS indicators. The likelihood ratio test statistic is 20.91, with a p value of 0.23.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

α_i^1 and α_i^0, $i = 1, \ldots, N$. The logit form allows such a transformation for observations satisfying certain conditions. For instance, consider the case that $T_i - t_i = 3$ and consider two events

$$A = \{y_{i0}, y_{i1} = 0, y_{i2} = 1, y_{i3}\},$$
$$B = \{y_{i0}, y_{i1} = 1, y_{i2} = 0, y_{i3}\},$$

where t_i is normalized to be 0. Under the assumptions that $\mathbf{w}_{i2} = \mathbf{w}_{i3}$ and $y_{i0} = y_{i3}$,

$$P\left(A|A \cup B, \mathbf{w}_{it}, \alpha_i^0, \delta_i\right) = \frac{1}{1 + \exp\left[(\mathbf{w}_{i1} - \mathbf{w}_{i2})\,\boldsymbol{\theta}_0 + (\mathbf{w}_{i1}y_{i0} - \mathbf{w}_{i3}y_{i3})\,\boldsymbol{\theta}_1\right]}, \tag{11}$$

where $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$, $\boldsymbol{\theta}_1 = (\mathbf{b}', \mathbf{g}')'$. The conditional probability no longer depends on α_i^0 and δ_i. Therefore, we propose to estimate $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ by maximizing the objective function

$$\sum_{i=1}^{N} \sum_{t_i \le t \le T_i - 3} 1(y_{i,t+1} + y_{i,t+2} = 1) \cdot 1(y_{it} = y_{i,t+3}) \times 1(\mathbf{D}_{i,t+2} = \mathbf{D}_{i,t+3}) \cdot K\left(\frac{\mathbf{x}_{i,t+2} - \mathbf{x}_{i,t+3}}{\sigma_n}\right) \times \ln\left(\frac{\exp[(\mathbf{w}_{i,t+1} - \mathbf{w}_{i,t+2})\tilde{\boldsymbol{\theta}}_0 + (\mathbf{w}_{i,t+1}y_{it} - \mathbf{w}_{i,t+3}y_{i,t+3})\tilde{\boldsymbol{\theta}}_1]^{y_{i,t+1}}}{1 + \exp[(\mathbf{w}_{i,t+1} - \mathbf{w}_{i,t+2})\tilde{\boldsymbol{\theta}}_0 + (\mathbf{w}_{i,t+1}y_{it} - \mathbf{w}_{i,t+3}y_{i,t+3})\tilde{\boldsymbol{\theta}}_1]}\right) \tag{12}$$

with respect to $\tilde{\boldsymbol{\theta}}_0$ and $\tilde{\boldsymbol{\theta}}_1$ over the parameter space, where $\mathbf{1}\,(A)$ denotes the indicator function, $K(\frac{\mathbf{x}_{i,t+2} - \mathbf{x}_{i,t+3}}{\sigma_n})$ denotes a kernel density function that gives more weight to those observations

Table 7
Mean group impacts and equilibrium impacts

Parameter	Group impacts		Random individual
	The non-employed group (1)	The employed group (2)	(3)
First JSS	0.036	0.034	0.047
Second JSS	0.000	0.000	0.000
Third JSS	0.000	0.000	0.000
Four or More JSS	0.00001	0.000	0.000
Previous total AS	−0.016	−0.008	−0.016
Previous total PS	0.023	0.018	0.028
Pre-WorkFirst earnings	0.00001	0.00001	0.00002
Previous non-employment quarters	−0.029	−0.0007	−0.016
Previous employment quarters	0.023	0.009	0.023
Age of the youngest kid	0.003	0.002	0.0035
Number of adults	−0.033	−0.037	−0.040
Number of kids	−0.018	−0.0016	−0.003
Married	−0.039	0.014	−0.212
Whites	−0.030	−0.018	−0.035
Hispanics	0.076	0.018	0.010
Grade 12	0.035	0.016	0.035
Region 1	0.042	0.027	0.041
Region 2	0.025	0.010	0.029
Region 3	0.034	−0.013	0.031
Region 4	−0.037	0.0003	0.0002
Region 5	−0.0034	−0.0213	−0.013
Unemployment rate	−0.004	−0.008	−0.008
Previous wage rate	0.0001	0.00008	0.00005

whose $\mathbf{x}_{i,t+2}$ are closer to $\mathbf{x}_{i,t+3}$, and σ_n is a bandwidth that shrinks toward 0 as n increases.

Theorem 1. *Let* $\mathbf{q}_{it} = [\mathbf{w}_{i,t+1} - \mathbf{w}_{i,t+2}\ \mathbf{w}_{i,t+1}y_{it} - \mathbf{w}_{i,t+2}y_{i,t+3}]$, $\boldsymbol{\psi} = (\boldsymbol{\theta}_0', \boldsymbol{\theta}_1')$,

$$h_{it}(\boldsymbol{\psi}) = 1\left(y_{i,t+1} + y_{i,t+2} = 1\right) \cdot 1(y_{it} = y_{i,t+3}) \times 1\left(\mathbf{D}_{i,t+2} = \mathbf{D}_{i,t+3}\right) \times \ln\left(\frac{\exp(\mathbf{q}_{it}\boldsymbol{\psi})^{y_{i,t+1}}}{1 + \exp(\mathbf{q}_{it}\boldsymbol{\psi})}\right). \quad (13)$$

Suppose that

C1 *The error term* ε_{it} *is independently, identically distributed and is independent of* $\mathbf{D}_{it}, \mathbf{x}_{it}, \alpha_i^s$.

C2 $\{(\mathbf{y}_i, \mathbf{w}_i), i = 1, \ldots, N\}$ *is a random sample of N observations from a distribution that satisfies* (6)*, where* $\mathbf{y}_i = (y_{it_i}, y_{i,t_i+1}, \ldots, y_{i,T_i})$, $\mathbf{w}_i = (w_{it_i}, w_{i,t_i+1}, \ldots, w_{i,T_i})$.

C3 *The true values of the parameters of interest,* $\boldsymbol{\psi}_0$*, are in the parameter space* Ψ*, which is a compact subset of the Euclidean K-space* (R^K)*, where* $K = k + q + 1$.

C4 (i) *The random vector* $\mathbf{x}_{i,t+2} - \mathbf{x}_{i,t+3}$ *conditional on* $\mathbf{D}_{i,t+2} = \mathbf{D}_{i,t+3}$ *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) $\Pr\left(\mathbf{D}_{i,t+2} = \mathbf{D}_{i,t+3}\right) > 0$.

C5 $E[\|\mathbf{w}_{i,t+2} - \mathbf{w}_{i,t+3}\|^2|\mathbf{A}]$ *and* $E[\|\mathbf{w}_{i,t+1}y_{it} - \mathbf{w}_{i,t+3}y_{i,t+3}\|^2|\mathbf{A}]$ *are bounded on their supports, where* $\mathbf{A} = [(\mathbf{x}_{i,t+2} - \mathbf{x}_{i,t+3}), \mathbf{D}_{i,t+2} = \mathbf{D}_{i,t+3}]$ *for assumptions* (C4)–(C6).

C6 *The function* $E(h(\boldsymbol{\psi})|\mathbf{A})$ *is continuous in a neighborhood of zero for all* $\psi \in \Psi$.

C7 *The functions* $E[(\mathbf{w}_{i,t+1}y_{it} - \mathbf{w}_{i,t+2}y_{i,t+3})(\mathbf{w}_{t+1}y_{it} - \mathbf{w}_{i2}y_{i,t+3})'|\mathbf{A}]$ *and* $E[(\mathbf{w}_{i,t+1} - \mathbf{w}_{i,t+2})(\mathbf{w}_{i,t+1} - \mathbf{w}_{i,t+2})'|\mathbf{A}]$ *and have full column rank in the neighborhood of zero.*

C8 $K : R^k \to R$ *is a function of bounded variation that satisfies:* (i) $\sup_{v \in R}|K(v)| < \infty$, (ii) $\int |K(v)|\,\mathrm{d}v < \infty$, *and* (iii) $\int K(v)\,\mathrm{d}v = 1$.

C9 σ_n *is a sequence of positive numbers that satisfies:* $\sigma_n \to 0$ *as* $n \to \infty$*, but* σ_n *converges to 0 at a slower rate than* $n \to \infty$ *(e.g.* σ_n *is of order* $n^{-1/(2h+1)}$ *for a finite positive integer h).*

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

$$\max_{\boldsymbol{\psi} \in \Psi} \sum_{i=1}^{N} \sum_{t_i \le t \le T_i - 3} K\left(\frac{\mathbf{x}_{i,t+2} - \mathbf{x}_{i,t+3}}{\sigma_n}\right) h_{it}(\boldsymbol{\psi}), \quad (14)$$

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

Assumptions (C3)–(C6) are the regularity conditions required for the objective function to converge to a non-stochastic limit which is uniquely maximized at $\boldsymbol{\psi}_0$ by a law of large numbers. Assumption (C7) is required for the identification of $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. Assumptions (C8) and (C9) 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. Similarly 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 $\mathbf{x}_{i,t+2} = \mathbf{x}_{i,t+3}$, we also need to condition on $\mathbf{D}_{i,t+2} = \mathbf{D}_{i,t+3}$. The proof of consistency and asymptotic normality follows straightforwardly those in Honoré and Kyriazidou (2000).

Whether individual-specific effects need to be controlled is critical for the reliability of the MLE. 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: less than 10% of the observations used in MLE satisfy the key conditions that $y_{i,t+1} + y_{i,t+2} = 1$, $y_{it} = y_{i,t+3}$ and $\mathbf{D}_{i,t+2} = \mathbf{D}_{i,t+3}$ for the consistency of the conditional MLE.[8] Furthermore, very few clients have taken more than two JSS. As the conditional MLE fails to converge with all JSS included, we

[8] As pointed out by a referee that in a heterogeneous treatment effects world (e.g., Imbens and Angrist (1994)), the mean impacts of treatment on the treated for those less than 10% of the sample that meets the required conditions could be different from that of the treated sample as a whole. The heterogeneous treatment effects are not taken up in our parametric specifications which assumes homogeneous marginal effects across individuals.

include only one JSS in the conditional logit estimation as this is our maintained hypothesis anyway.

The conditional MLE yields estimates with larger variances. The Hausman statistic for misspecification is 0.27 for the job-seeker group and 0.98 for the job-holder group, which are not significant at a chi-square distribution with one degree of freedom (3.84 at 5% level). These results appear to suggest that there is no evidence of the presence of no unobserved individual heterogeneity conditioning on observed client characteristics, hence do not appear to contradict using the MLE model to evaluate the effectiveness of repeated JSS.

4.2. A limited information test for conditional independence assumption

One of the critical issues in obtaining accurate estimates of the treatment effects is to control the impact of selection process. Our estimates are obtained under the assumption that the selection is exogenously determined. If this assumption is violated, our inference is biased. Wooldridge (2003) has proposed a likelihood ratio test of conditional independence by specifying a complete model for $(y_{it}, \mathbf{D}'_{it})$. However, a complete specification of sequential treatments from an intertemporal optimization framework is very complicated. Moreover, a full information approach could be more sensitive to specification errors. Therefore, in this subsection, we propose to test for CIA from a limited information framework.

Let the latent response function,

$$\mathbf{D}^*_{it} = \mathbf{d}\left(\mathbf{z}_{it}, \mathbf{x}_{it}\right) + \mathbf{v}_{it} \tag{15}$$

determine the outcome $\mathbf{D}_{it}$, where $\mathbf{z}_{it}$ denotes those variables that affect participation, but do not affect the current outcomes of y_{it}, and $\mathbf{v}_{it}$ is the error term that is orthogonal to $\mathbf{z}_{it}$ and $\mathbf{v}_{it}$. The $\mathbf{z}_{it}$ variables can be case-manager designation numbers, number of clients attending the same Community Service Office (CSO), etc. because case managers assigned WorkFirst participants to job search services. Participants were assigned to second and higher job search services when the case manager thought it would be helpful for them to go into job search. The way case managers are assigned clients varies from office to office. In some offices, it is a simple rotation as new clients come in, they are assigned to the next case manager who is up in the queue. Other offices assign them by the letter of the clients' last name — one case manager may get A through G and another H through M. In no office, according to a WorkFirst manager, was there a case where one manager was assigned all the easy cases and another the more difficult cases.

If $\boldsymbol{\varepsilon}_{it}$ and $\mathbf{v}_{it}$ are orthogonal, namely, CIA holds, then the marginal density of ε_{it} is identical to the conditional density of ε_{it} conditional on $\mathbf{D}_{it}, f\left(\varepsilon_{it}\right) = f\left(\varepsilon_{it} | \mathbf{D}_{it}\right)$. It follows from (4) that

$$H_0 : E\left(y_{it} | y_{i,t-1}, \mathbf{D}_{it}, \mathbf{x}_{it}, \mathbf{z}_{it}, \boldsymbol{\alpha}_i\right) = E\left(y_{it} | y_{i,t-1}, \mathbf{D}_{it}, \mathbf{x}_{it}, \boldsymbol{\alpha}_i\right), \tag{16}$$

or if there is no presence of individual-specific effects,

$$H_0^* : E\left(y_{it} | y_{i,t-1}, \mathbf{D}_{it}, \mathbf{x}_{it}, \mathbf{z}_{it}\right) = E\left(y_{it} | y_{i,t-1}, \mathbf{D}_{it}, \mathbf{x}_{it}\right), \tag{17}$$

where $\boldsymbol{\alpha}_i = \left(\boldsymbol{\alpha}_i^1, \boldsymbol{\alpha}_i^0\right) = \left(\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^0\right)$. If $\boldsymbol{\varepsilon}_{it}$ and $\mathbf{v}_{it}$ are correlated, then CIA does not hold, then

$$E\left(y^*_{it} | y_{i,t-1}, \mathbf{D}_{it}, \mathbf{x}_{it}, \boldsymbol{\alpha}_i\right) = \alpha_i^0\left(1 + \delta_i y_{i,t-1}\right) + \mathbf{x}'_{it}\left(\boldsymbol{\beta}^0 + \mathbf{b} y_{i,t-1}\right) + \mathbf{D}'_{it}\left(\boldsymbol{\gamma}^0 + \mathbf{g} y_{i,t-1}\right) + E\left(\varepsilon_{it} | \mathbf{D}_{it}, \boldsymbol{\alpha}_i\right), \tag{18}$$

where the selection factor[9]

$$E\left(\varepsilon_{it} | \mathbf{D}_{it}, \boldsymbol{\alpha}_i\right) = \lambda\left(\mathbf{x}_{it}, \mathbf{z}_{it}, \boldsymbol{\alpha}_i\right) \tag{19}$$

is a function of $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$. In other words,

$$H_1 : E\left(y_{it} | y_{i,t-1}, \mathbf{D}_{it}, \mathbf{x}_{it}, \mathbf{z}_{it}, \boldsymbol{\alpha}_i\right) \neq E\left(y_{it} | y_{i,t-1}, \mathbf{D}_{it}, \mathbf{x}_{it}, \boldsymbol{\alpha}_i\right), \tag{20}$$

or if there is no presence of individual-specific effects,

$$H_1^* : E\left(y_{it} | y_{i,t-1}, \mathbf{D}_{it}, \mathbf{x}_{it}, \mathbf{z}_{it}\right) \neq E\left(y_{it} | y_{i,t-1}, \mathbf{D}_{it}, \mathbf{x}_{it}\right). \tag{21}$$

We approximate $E\left(y_{it} | y_{i,t-1}, \mathbf{D}_{it}, \mathbf{x}_{it}, \mathbf{z}_{it}, \boldsymbol{\alpha}_i\right)$ by[10]

$$F\left[\boldsymbol{\alpha}_i^*\left(1 + \delta_i\right) y_{i,t-1} + \mathbf{x}'_{it}\left(\boldsymbol{\beta}^0 + \mathbf{b} y_{i,t-1}\right) + \mathbf{D}'_{it}\left(\boldsymbol{\gamma}^0 + \mathbf{g} y_{i,t-1}\right) + \mathbf{p}\left(\mathbf{z}_{it}, \mathbf{x}_{it}, y_{i,t-1}\right)\right], \tag{22}$$

where $\mathbf{p}\left(\mathbf{z}_{it}, \mathbf{x}_{it}, y_{i,t-1}\right)$ could be a polynomial or series function of $\left(\mathbf{z}_{it}, \mathbf{x}_{it}, y_{i,t-1}\right)$. For instance, if we approximate $\mathbf{p}\left(\cdot\right)$ by

$$\mathbf{p}\left(\mathbf{z}_{it}, \mathbf{x}_{it}, y_{i,t-1}\right) = \mathbf{a}'_1 \mathbf{z}_{it} + \mathbf{a}'_2\left(\mathbf{z}_{it}\mathbf{z}'_{it}\right) + \mathbf{a}'_3\left(\mathbf{z}_{it}\mathbf{x}'_{it}\right) + \mathbf{a}'_4\left(\mathbf{z}_{it} y_{i,t-1}\right), \tag{23}$$

then a test of H_0 or H_0^* against H_1 or H_1^* is equivalent to the test

$$\tilde{H}_0 : \mathbf{a}_1 = \mathbf{0}, \mathbf{a}_2 = \mathbf{0}, \mathbf{a}_3 = \mathbf{0}, \text{ and } \mathbf{a}_4 = \mathbf{0}.$$

As is generally recognized, the choices of management practices and institutional structure can affect program effectiveness, we consider the following variables for the specification of $\mathbf{p}$: (1) The number of cases each case manager handles because case managers were exogenously assigned by the Washington TANF offices. Further, if a case manager handles more cases than the other for the sample period, on average, the clients from those handling more clients would receive less attention, hence their probabilities of participation could be affected. The number of cases occurring in each community service office could also be another variable for $\mathbf{p}$, as more clients getting services from the same community office could affect their probabilities of participations in JSS. We also consider interactions of these two variables with other explanatory variables. Because of severe multicollinearity in the expanded explanatory variables, we only include interaction terms of those variables with those that are individual specific and time varying. This gives 17 additional variables in $\mathbf{p}$. Tables 5 and 6, column 3, report the MLE estimates of $\boldsymbol{\theta}$ with the $\mathbf{p}$ variables included as additional explanatory variables assuming no individual-specific effects. We note that the coefficients of JSS variables hardly changed with the addition of $\mathbf{p}$. The likelihood ratio test of H_0^* vs. H_1^* is 15.07 for the non-employed group and 20.91 for the employed group, which are not significant at a chi-square distribution with 17 degrees of freedom (the critical value is 27.59 at the 5% significance level and it is 24.76 at the 10% significance level).

We also test H_0 vs. H_1 using the conditional MLE approach to allow for the presence of individual-specific effects. Due to multicollinearity there are 7 additional variables in $\mathbf{p}$ under the conditional MLE. The likelihood ratio statistic is 8.02 for the non-employed and 10.63 for the employed, both are insignificant at a chi-square distribution with 7 degrees of freedom (the critical value is 15.50 at the 5% significance level and it is 13.36 at the 10% significance level).

9 For example, see Amemiya (1984), Eq. (10.7)) or Robinson (1988) for the case when the outcome is binary.

10 We suppose that $\lambda(\mathbf{x}_{it}, \mathbf{z}_{it}, \boldsymbol{\alpha}_i) \simeq \tilde{\alpha}_i(1+\tilde{\delta}_i y_{i,t-1}) + p\left(\mathbf{x}_{it}, \mathbf{z}_{it}, y_{i,t-1}\right)$, and merge the impact of $\tilde{\alpha}_i(1+\tilde{\delta}_i y_{i,t-1})$ with $\alpha_i^0(1+\delta_i y_{i,t-1})$ in (22).

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. Using 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 on employment rate. We found that only the first Job Search Services had positive and statistically significant impacts on employment rates. The probabilities of employment are increased by about 3.6% for the non-employed and by 3.4% for the employed. Repeating JSS does not appear to raise the probability of employment.

The TANF has three unique features compared with previous welfare programs. First, there are the universal requirements for recipients to work.[11] All states require adults who receive cash welfare assistance to work or to engage in employment-related activities. Second, State governments usually provide cash assistance and other assistance such as child care, transportation assistance to the disadvantaged to make the low-paying work financially attractive to welfare recipients. Third, a 60-month time limit is imposed on each welfare recipient.[12] The principal reasoning behind these appears to be that having a low-paying job now is better than waiting for a high-paying job in the future. We find that each additional quarter of non-employment reduces the probability of employment by 2.9% and each additional quarter of employment raises the probability of employment by 2.3%. The "experience-enhancing" effect together with the finding that the first JSS does raise the probability of employment appears to provide empirical support for requiring TANF recipients to engage in employment-related activities, and a focus on short-term less expensive job search activities can be beneficial. However, repeating JSS does not appear to yield any additional benefit. It appears that for those who have already taken one JSS, perhaps other training programs, such as long-term human capital augment activities could be more beneficial rather than prodding them to repeatedly taking JSS.

Our analysis is based on those who were new entrants to TANF. It will be worthwhile to expand the data to include later years when some recipients approach their 60-month time limits to investigate the effects of 60-month limit on employment outcome. The expanded data may also allow us to investigate more thoroughly the selection issues and the presence of individual effects. The validity of our inference depends on the validity of our assumption, notably the conditional independence assumption and no unobserved individual-specific effects conditional on observables for welfare recipients (or more precisely, TANF recipients). In this paper we have also suggested a conditional maximum likelihood estimator to allow for the presence of individual-specific effects and a limited information procedure to test for the conditional independence assumption. However, given the constraints of our data, the power of these tests could be dubious.

Finally, as pointed out by a referee we are "using State Unemployment Insurance (UI) earnings data matched to program records from Washington State. In many state UI systems there is no way to distinguish between a person who actually earns zero in a quarter and a person who is working in the uncovered sector or in another state". These systematic and random measurement errors could seriously bias our estimation results and deserve further study.

[11] Rules for exemption can be found in http://www.spdp.org/tanf/work/sorksumm.htm.

[12] As pointed out by a referee, the presence of the life-time 60-month time limit may be a reason to exercise caution in the interpretation of the estimated long-run steady-state effects.

Appendix. The WorkFirst in the State of Washington: A description of provided services

The data came from the administrative records of the WorkFirst program in Washington state. In order to evaluate the program, the state developed a data file that was an extract from the legacy data systems used to operate the program. These legacy systems, used to manage the WorkFirst program included the Jobs Automated Systems (JAS), and Automated Client Evaluation System (ACES). Additionally, the data for this study came from the Unemployment Insurance data file that includes the covered employment, earnings and hours for all Washington employees covered by the UI system. This earnings data was matched by Social Security Numbers to the clients identified in the WorkFirst data to determine program outcomes.

WorkFirst is a mandatory program under TANF (Temporary Assistance for Needy Families) in the State of Washington. The WorkFirst program serves three groups of people: parents and children aged sixteen or older who receive cash assistance under the temporary assistance for needy families (TANF), general assistance for pregnant women (GA-S) or state family assistance (SFA) programs; parents who no longer receive cash assistance and need some continuing support to remain self-sufficient; and low income parents who support their family without applying for or relying on cash assistance. Its mandatory nature comes from the fact that if one refuses to take activities required by WorkFirst case managers, he/she will be sanctioned. Sanction means reducing a certain percent of cash receipt that one can receive from the TANF program.

When clients first enter WorkFirst, they will work together with case managers to develop the Individual Responsibility Plan (IRP).

The following services can be introduced to clients based on their background:

1. Job Search Services. When clients first enter WorkFirst, the initial focus is to assist them in finding employment. Therefore they will be first introduced to job search. Periods of job search may last up to twelve continuous weeks. The first four weeks are coded as Job Search Initial (JI). They may also be directed to Job Search Workshop (JW). Job Search Services may include one or more of the following forms: (1) classroom instruction that helps in finding job openings, complete applications, practice interviews and apply other skills and abilities with a job search specialist or a group of fellow job seekers; (2) pre-employment training; (3) high-wage/high-demand training.

   By the end of the first four weeks, a job search specialist will determine whether one should continue in job search. Job search will end when (1) clients find a job and work 20 hours or more at an unsubsidized job; or (2) clients become exempt from WorkFirst requirements; or (3) client's situation changes and he is temporarily deferred from continuing with job search; or (4) Job search specialists have determined that clients need additional skills and/or experience to find a job; or (5) clients have not found a job at the end of the job search period. Participants were assigned to second and higher order job search services when the case manager thought it would be helpful for them to go into job search.
2. Alternative services: For those who are unable to find employment because of problems with substance abuse, domestic violence etc. were referred to alternative services.
3. Post-employment services: Employed clients on the caseload have access to mentors, job-specific education, career planning and other services intended to help them stay employed and find higher-paying jobs. Employed clients who leave the caseload are eligible for post-employment services for up to one year after exiting TANF.

References

Angrist, Imbens, Rubin, 1996. Idenfication of causal effects using instrumental variables. Journal of the American Statistical Association 462–465.

Amemiya, T., 1984. Advanced Econometrics. Harvard University Press.

Ashenfelter, O.C., Card, C., 1985. Using the longitudinal structure on earnings to estimate the effect of training programs. The Review of Economics and Statistics 67, 648–660.

Bassi, L.J., 1984. Estimating the effect of training programs with non-random selection. The Review of Economic and Statistics 66 (1), 36–43.

Card, Sullivan, 1988. Measuring the effect of subsidized training programs on movements in and out of employment. Econometrica 56 (3), 497–530.

D'Addio, A., Honore, B., 2002. Duration dependence and time-varying variables in discrete time duration models. Working Paper No. 2002-13. Department of Economics, University of Aarhus.

Friedlander, D., Greenberg, D., et al., 1997. Evaluating government training programs for the economically disadvantaged. Journal of Economic Literature 35 (4), 1809–1855.

Gill, R.D., Robins, J.M., 2001. Causal inference for complex longitudinal data: The continuous case. Annals of Statistics 29 (6).

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 6, 1251–1271.

Heckman, J.J., 1981. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In: Manski, C.F., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA, pp. 179–195.

Heckman, J.J., Hotz, V.J., 1989. Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. Journal of American Statistical Association 84, 862–874.

Heckman, J.J., Smith, J.A., 1999. The pre-program earnings dip and the determinants of participation in a social program: Implications for simple program evaluation strategies. NBER Working Papers 6983. National Bureau of Economic Research, Inc.

Heckman, J., Robb, R., 1985. Alternative methods for evaluating the impact of interventions: An overview. Journal of Econometrics 30, 239–267.

Heckman, J., LaLonde, R.J., Smith, J., 1999. The economics and econometrics of active labor market programs. In: Ashenfelter, O., Card, David (Eds.), Handbook of Labor Economics, vol. III.

Holland, P., 1986. Statistics and causal inference. Journal of the American Statistical Association 81, 945–960.

Honoré, B.E., Kyriazidou, E., 2000. Panel data discrete choice models with lagged dependent variables. Econometrica 68 (4), 839–874.

Hotz, J., Imbens, G., Klerman, J., 2006. Evaluating the differential effects of alternative welfare-to-work training components: A re-analysis of the California GAIN program. Journal of Labor Economics 24 (3), 521–566.

Hotz, J., Mullin, C., Scholz, J.K., 2002. Welfare, employment and income: Evidence on the effects of benefit reductions from California. American Economic Review 92 (2), 380–384.

Hsiao, C., 2003. Analysis of Panel Data, second ed. Cambridge University Press, Cambridge.

Imbens, G., Angrist, J., 1994. Identification and estimation of local average treatment effects. Econometrica 62 (2), 467–476.

Lechner, M., 2004. Sequential matching estimation of dynamic causal models. Discussion paper no. 2004-06, Department of Economics, University of St. Gallen.

Lerch, S., Mayfield, J., 1999. Welfare and Employment Outcomes of the WorkFirst Program. Washington State Institute for Public Policy.

Lechner, M., Miquel, R., 2001. A potential outcome approach to dynamic programme evaluation: Nonparametric identification. Memio. University of St. Gallen.

Robins, J.M., 1986. A new approach to causal inference in mortality studies with a sustained exposure period—Application to control of the healthy worker survivor effect. Mathematical Modelling 7, 1393–1512.

Robinson, P.M., 1988. Root-N-consistent semiparametric regression. Econometrica 56, 931–954.

Rosenbaum, P.R., Rubin, D.B., 1983. The central role of propensity score in observational studies for causal effects. Biometrika 70, 41–55.

Wooldridge, J., 2003. Econometric Analysis of Cross Section and Panel Data. MIT Press.