Summer 2010

Identifying the Effect of a Welfare-to-Work Program Using Capacity Constraint: a New York City Quasi-Experinmet

John Ifcher
Santa Clara University, jifcher@scu.edu

Identifying the Effect of a Welfare-To-Work Program Using Program Capacity Constraints:

A New York City Quasi-Experiment

by John Ifcher *

Right running heading: Identifying Program Effects Using Capacity Constraints

February 2009

Abstract

In 1999 general assistance recipients in New York City were required to participate in a job training and outplacement assistance program. Initially, recipients were enrolled in 'waves' due to capacity constraints. The program's impact is identified using a quasi-experiment in which selectees are compared to concomitantly eligible non-selectees. Selectees are 15 percentage points more likely to start a job and 10 percentage points more likely to exit welfare than are non-selectees. This methodology is important since random-assignment experiments can be costly and difficult to implement. Further, experiments are not impervious to criticism; this procedure addresses three of five known shortcomings.

* Santa Clara University, Department of Economics, 500 El Camino Real, Lucas 216B, Santa Clara, CA, 95053, 408-554-5579 (phone), 408-554-2331 (fax), jifcher@scu.edu.

INTRODUCTION

Over the past two decades U.S. welfare programs have been transformed. A central objective of the reforms has been to help recipients move from welfare to work. In response, a variety of 'welfare-to-work' programs have been implemented. They typically include one or more of the following components: unpaid work experience, classroom instruction, on-the-job training, financial incentives, and job search assistance.

A substantial literature has developed to identify the resulting changes in welfare use, employment, well-being, and family structure [recent reviews include Blank, 2002; Grogger and Karoly, 2005; Moffitt, 2003]. The results largely support the following two conclusions: First, in the short run, at least, welfare-to-work programs had the intended effect, reducing welfare use and increasing employment. Second, the most effective welfare-to-work programs are those that include mandatory work requirements with an emphasis on job placement [Grogger and Karoly, 2005].

These findings are largely derived from well-designed random-assignment experiments. While such experiments are certainly an excellent method for identifying the effect of a welfare-to-work program, they are not impervious to criticism. The following five shortcomings have been identified and are summarized in Grogger and Karoly [2005]: First, since experimental programs are often implemented by above-average managers, it is unclear whether successful pilot programs can be expanded without losses in effectiveness. Second, experiments miss some general equilibrium effects. For example, whereas full-scale implementation of a welfare-to-work program might crowd-out other job seekers or even subsequent program participants, a pilot program might not. Third, the 'message' of the pilot program may cross-over from the

program group to the control group. Fourth, experiments that only include current welfare recipients do not capture entry effects. That is, a welfare-to-work program may change the attractiveness of receiving welfare and therefore not only effect the exit rate, but also the entry rate. Finally, random-assignment experiments can be costly and hard to implement.

The first four shortcomings are noteworthy since each could cause experimental estimates to be biased. For example, the first presumably introduces a positive bias and the third a negative bias. Thus, available estimates, which are largely based on experimental evidence, are potentially biased with the direction and magnitude of the overall bias unknown. Moreover, no new random-assignment experiments have been conducted since the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA).[1]

A few non-experimental and quasi-experimental studies have explored the effect of welfare-to-work programs. Two of these studies have challenged the claim that the most effective welfare-to-work programs are those that emphasize rapid job placement [Hotz et al, 2006; Dyke et al, 2006]. These authors find that more intensive training, which emphasizes human capital development, generates larger positive effects. These effects, however, take longer to emerge than do the effects of programs that emphasize job placement.[2]

The paucity of non-experimental and quasi-experimental studies is largely due to two factors. First, there has been a shortage of clever strategies (in the absence of random-assignment) to identify the effects of welfare-to-work programs using administrative data. Second, welfare-to-work programs vary along numerous dimensions,

making it difficult to parameterize the programs in a manner that is useful for identification. Thus, data from nationally representative surveys has not been successfully used to estimate the effect of welfare-to-work programs.[3]

In this paper, I demonstrate a quasi-experimental identification strategy to estimate the effect of a welfare-to-work program. To do so, I take advantage of a quirk in the program's administration. Specifically, when the program was initiated, all eligible General Assistance (GA) recipients could not participate concurrently due to capacity constraints.[4] Rather, GA recipients were selected for the program in 'waves.' The wave enrollment process creates the opportunity to identify the effect of the program by comparing 'selectees,' recipients who were selected on a given date to 'non-selectees,' recipients who were eligible, but not selected, on that date. The results indicate that the program increased the likelihood that a recipient started a job by 15 percentage points and increased the likelihood that a recipient exited welfare by 10 percentage points.

This quasi-experiment is similar to a random-assignment experiment in that it enables one to estimate the effect of a welfare-to-work program. It also ameliorates three of the five previously discussed shortcomings of random-assignment experiments. First, the program was not a small-scale, pilot program implemented by above-average managers. Rather, it was the principal welfare-to-work program for all welfare recipients in New York City (NYC). Second, this analysis should capture general equilibrium effects in the labor market given: (1) the size of program, over 10,000 GA recipients were enrolled during the program' first year, and (2) the number of welfare recipients who started a job, over 100,000 recipients of Family Assistance (FA) and GA combined

reported finding a job during the program's first two years. Finally, this program was not a costly, or difficult to implement, random-assignment experiment.

The next section of this paper presents an overview of welfare reform in NYC. The third section describes the quasi-experimental identification strategy. The fourth section discusses the empirical implementation. The fifth section presents the results. The sixth section describes a robustness check that was performed. The final section discusses the findings and presents a brief fiscal cost benefit analysis.

WELFARE REFORM IN NYC

In 1994, Rudolph Giuliani, newly elected mayor of NYC, made reforming the City's welfare programs a priority. At the time, NYC had over one million welfare recipients, including almost 300,000 GA recipients. The City was spending approximately three billion dollars annually on welfare programs, including one billion dollars on GA.

In 1995, NYC initiated the Work, Accountability, and You (NYCWAY) program. A central tenet of NYCWAY was that able-bodied welfare recipients were required to work in exchange for their benefits.[5] Almost all recipients fulfilled this requirement by participating in a Work Experience Program (WEP) assignment 21 hours per week.[6] The majority of WEP participants worked outdoors removing litter, weeds, and graffiti from parks, vacant lots, streets, and highways. Over the next six years, the number of welfare recipients in NYC declined by more than 50% and the number of GA recipients declined by 70%.

In 1999, NYC created the Employment Services and Placement Program (ESPP), a job training and outplacement assistance program. With the implementation of the

4

ESPP, welfare recipients were required to participate in a WEP assignment three days a week and in the ESPP two days a week. This increased, from 21 to 35, the number of hours per week that recipients were required to spend in structured activities.

Eleven private contractors were hired to provide ESPP services. All had a history of providing social services in NYC. The contractors were paid on a performance basis. They received a substantial payment for each recipient they placed in a job; the average payment was approximately $3,000. The contractors focused on developing participants' 'soft skills' including résumé writing and interview techniques, and helped participants arrange job interviews. ESPP contractors were required to attempt to place each participant for at least six weeks.

Prior to implementation, there were over 10,000 GA recipients who were eligible for the ESPP. To be eligible, a GA recipient had to be participating in a WEP assignment and job-ready.[7] Since all eligible GA recipients could not be accommodated simultaneously, recipients were enrolled in waves. Selectees were informed by mail of their status, instructed to report to the proper location at a prescribed date and time (the program start date was typically two weeks after the selection date), and advised that they would be sanctioned if they failed to comply with the program's requirements. New waves were formed every two weeks until each eligible GA recipient was selected or had become ineligible.

Recipients were selected for each wave centrally by computer programmers. The selection process did not include intake interviews or objective assessments. The intention was to generate a random sample of selectees for each wave stratified by borough. In the fourth section, I report that the selection process did not approximate a

random one and discuss the attendant adjustments that are made to the estimates presented in this paper.

QUASI-EXPERIMENTAL IDENTIFICATION STRATEGY

Generally, to identify the effect of a program one compares the outcomes of treated and untreated individuals. In the case of the ESPP, recipients who remained eligible should have ultimately been treated. Thus, recipients who remained untreated must have become ineligible prior to being selected for the ESPP and do not make a valid comparison group.[8]

An Alternative Approach to Identify the Effect of the ESPP

It is possible, however, to estimate the effect of the ESPP taking advantage of the wave enrollment process. To do so, I compare all selectees from a given wave to all 'non-selectees,' recipients who were eligible but not selected, for the same wave. By performing this comparison for each of the first 17 waves, and combining all selectees into one group, the program group, and all non-selectees into another group, the control group (see Figure 1), I can estimate the Program Effect (PE).[9] Specifically, the PE is defined as,

$$E[Y_i^M (P_i = 1)] - E[Y_i^M (P_i = 0)] \qquad (1)$$

where $Y_i^M (P_i)$ is an indicator function which equals one if individual i started a job within M months of his or her inclusion in the program or control group and zero otherwise, and is a function of whether recipient i was a member of the program group, P_i=1, or the control group, P_i=0. <<Figure 1 here>>

All selectees are placed in the program group, and considered treated, regardless of whether or not they participated in the ESPP. This even includes recipients who failed to attend the ESPP orientation. Hence the PE is an 'Intent to Treat' (IT) effect and should not suffer from self-selection bias.[10] Members of the program group remain in the program group for the remainder of the study; that is, there is no program group attrition.

Further, this identification strategy does not suffer from control group attrition. Each member of the control group remains in the control group for the remainder of the study. For example, control group members who subsequently were selected, and or exited welfare, remain in the control group.

On the other hand, the identification strategy does suffer from control group contamination. Over 70% of control group members were selected for the ESPP subsequent to their inclusion therein. Consequently, the PE is negatively biased and a conservative estimate of the program's true effect.

Adjusting for Control Group Contamination

To adjust for control group contamination, a set of additional, restricted control groups is created in which recipients are removed from the control group if they were selected within a given number of waves of their inclusion in the control group. For example, a control group with a one-wave restriction on being selected is created by removing members of the original control group if they were selected in the subsequent wave. So, a non-selectee from the first wave, who was a selectee for the second wave, would be excluded from this restricted control group.

It is important to note that the original control group includes two distinct cohorts: (1) non-selectees who remained eligible and were ultimately selected, and (2) non-

selectees who became ineligible prior to being selected. This distinction is important because all members of the original control group who are removed to form the restricted control groups must have remained eligible (since they were selected). That is, they must have been in the first cohort. For the restricted control group to be a valid comparison group for the program group, the ratio of members of cohort one to members of cohort two must be the same in the restricted control group as in the original control group. Accordingly, the weight placed on members of the first cohort in the restricted control group is increased. Specifically, the weight is the reciprocal of the probability that members of cohort one were not selected for the subsequent wave (see Figure 2).

<<Figure 2 here>>

I utilize the above procedure recursively to create control groups with up to a thirty-wave restriction on being selected. Imposing longer restrictions produces very large weights that could cause unstable estimates. For example, the maximum weight is 286 when a thirty-two-wave restriction on being selected is imposed. In this paper, I use the initial, unrestricted control group, denoted *control group: no restriction,* and control groups that restrict being selected for an even number of waves. Thus, the restricted control groups have a one-month, two-month, three-month, …, or fifteen-month restriction on being selected (assuming that each month contains two-waves, or twenty-eight days). They are denoted *control group: 1M restriction*, *control group: 2M restriction*, *control group: 3M restriction*, …., and *control group: 15M restriction*, respectively.

EMPIRICAL IMPLEMENTATION

With NYC's permission, the data for this study were extracted from an administrative database. The case history and available demographic characteristics were compiled for each eligible GA recipient. It should be noted that NYC only collected basic demographic information, such as race, gender, and date of birth, for each welfare recipient. No additional information is available.

The program and control groups were formed as described in the prior section. Of all selectees and non-selectees (n = 64,800), just over half are men and about 90 percent are nonwhite. Their average age is 47 and they are likely to live in the Bronx, Brooklyn, or Manhattan (see Table 1). Comparing members of the program group (n = 6,782) and control group (n = 58,018), one observes that the average age, years on welfare, and gender and racial distributions appear similar. A difference of means test, however, reveals that only the gender distribution is not significantly different. The distribution of borough of residence is disparate as well. This is presumably the result of stratifying the selection by borough. <<Table 1 here>>

Selection Process Was Not Random

Given the results of the difference of means tests discussed above, it is unlikely that the selection process approximated a random one. To confirm this impression, I test whether recipients' demographic characteristics significantly impact the probability of being selected. Specifically, the following probit equation is estimated,

$$P[P_i^{wb} = 1] = F(\alpha^{wb} + \sum_{c=1}^{C} \lambda_c^{wb} x_{ic}^{wb} + \varepsilon_i^{wb}) \qquad (2)$$

where P_i^{wb} is a program group dummy that equals one if individual i was a selectee and zero if individual i was a non-selectee in wave w and borough b; and x_{ic}^{wb} is a series of C

demographic characteristics for individual i in wave w and borough b. To decrease the chance of generating spurious results, equation (2) is only estimated for the three largest waves in each of the four largest boroughs; the number of selectees per borough varied widely across waves.

The coefficients on age, and years continuously on welfare, provide strong evidence that the selection process was not random. Six of twelve coefficients on age, and seven of twelve coefficients on years continuously on welfare, are positive and significantly different than zero (see Table 2). Thus, the probability of being selected increased with age and welfare tenure.[11] For the other demographic characteristics there was no discernable selection pattern. <<Table 2 here>>

One thing is certain though; eligible recipients were selected solely using information that was stored in the administrative database. The selection process was centralized and conducted by computer programmers. Individual caseworkers were not involved in any manner. No intake interviews or objective assessments were conducted prior to selection. In other words, the selection process was conducted without human discretion. Such a selection process, even if it did not approximate a random one, should not disturb the necessary assumption that there was no systematic selection on unobserved characteristics. Consequently, by including covariates in the analysis, one should be able to adjust for the observed differences.

Adjusting for observed characteristics

Since recipients who were older, and who had longer welfare spells, should have been less likely to start a job and or exit welfare, the PE as defined in equation (1) is potentially negatively biased. To adjust for the observed differences, a program dummy

and a series of observed characteristics are regressed on an outcome dummy. Specifically, the following equation is estimated,

$$y_i^M = \alpha^M + \beta^M P_i + \sum_{c=1}^{C} \lambda_c^M x_{ic} + \sum_{j=1}^{5} \delta_j^M B_{ij} + \sum_{k=1}^{17} \gamma_k^M W_{ik} + \sum_{j=1}^{5}\sum_{k=1}^{17} \eta_{jk}^M (B_{ij} * W_{ik}) + \varepsilon_i^M \quad (3)$$

where y_i^M is an outcome dummy that equals one if individual i started a job within M months of being a selectee or non-selectee, and zero otherwise; P_i is a program group dummy that equals one if individual i was in the program group and zero if individual i was in the control group; x_{ic} is a series of C demographic characteristics for individual i at the time that he or she became a selectee or non-selectee; B_{ij} is a borough dummy that equals one if individual i resides in borough j and zero otherwise; and W_{ik} is a wave dummy that equals one if individual i was a selectee or non-selectee in wave k and zero otherwise. Equation (3) is estimated using OLS for values of M between 0.5 and 26.[12]

RESULTS

Estimating equation (3), one finds that the PE first increases in M, and then after peaking, decreases in M, as control group contamination increases (see Figure 3). The peak PE is positive and significantly different than zero using each control group. The magnitude of the peak PE increases – as well as the number of months required to reach it – as the length of the restriction on being selected increases. For example, the peak PE is 0.075 (t = 19.8, p = 0) when M = 4, 0.131 (t = 16.2, p = 0) when M = 8, and 0.152 (t = 12.7, p = 0) when M = 11.5 using *control groups: no restriction, 6M, and 12M restriction*, respectively. <<Figure 3 here>>

Further, the PEs using *control group: 1M restriction* are everywhere below the PEs using *control group: 2M restriction* when $M > 2$. The former do not diverge from the latter until the former's one-month restriction on being selected expires. This pattern is easy to observe in Figure 4a, which presents the PEs using the first four restricted control groups only. Figure 4a also reveals that the pattern recurs when one compares the PEs using *control groups: 2M* and *3M restriction,* and *control groups: 3M* and *4M restriction*. In fact, this pattern is present using each control group between *control groups: 1M* and *12M restriction*. After that, it stops. The peak PEs, and the PEs in the vicinity of the peak, are quite similar using *control groups: 12M, 13M, 14M,* and *15M restriction* (see Figure 4b). That is, further correction for control group contamination (beyond a twelve-month restriction on being selected) does not significantly increase the peak PE. Moreover, the peak PE using *control group: 14M restriction* is slightly larger than is the peak PE using *control group: 15M restriction* (0.158 versus 0.155). Thus, the ESPP appears to increase the likelihood of starting a job by approximately 15 percentage points. <<Figure 4 here>>

This employment effect appears to persist. Comparing the post-peak PEs using *control groups: 12M, 13M, 14M,* and *15M restriction*, one observes that the longer the restriction, the more months it takes for a consistent post-peak decline to begin. For each of these control groups, a consistent post-peak decline begins only after the restriction on being selected has ended. For example, the PEs using *control group: 12M restriction* begin to consistently decrease in M when $M > 12$. This pattern is easy to observe in Figure 4b, which presents the PEs using the four most restricted control groups only. Thus, the PEs using *control group: 15M restriction* provides a good estimate of the impact of the ESPP for at least fifteen months post selection.

The employment effect is presumably not the result of the underlying economic conditions, since including borough and wave dummies as well as the interaction term should control for any macroeconomic shocks that may have occurred during the study period. Further, prematurely terminating the study period on September 11th, 2001 does not materially affect the findings; after September 11th, the unemployment rate increased by over two percentage points in NYC.

Estimating equation (3) without covariates, one finds that the PEs decline. For example, the PEs decrease by an average of 0.003 and 0.008 using *control groups: 6M and 12M restriction*, respectively (see Figure 5). These declines are not statistically significant. Further, they were expected. Recall that the probability of being selected for the ESPP was greater for recipients who were older and had longer welfare spells. Such recipients were assumed to be less likely to have started a job. Additional support for this assumption can be found in the coefficients on age, which are negative and significantly different than zero for all values of *M* (see Table 3). <<Figure 5 here>> <<Table 3 here>>

Starting a job is not the sole outcome by which the ESPP should be evaluated. For the ESPP to be considered a true success, it should have also increased the probability that recipients exited welfare after starting a job.[13] Specifically, equation (3) is estimated with y_i^M redefined as an outcome dummy that equals one if individual *i* started a job and exited welfare for a six-month period within *M* months of becoming a selectee or non-selectee, and zero otherwise.[14]

The resulting PEs are similar in many respects to the previously estimated PEs. Again,

- the PE first increases in *M*, and then peaks (after peaking, however, it does not substantially decrease in *M*);

- the peak PE is positive and significantly different than zero using each control group;

- the magnitude of the peak PE increases as the length of the restriction increases; and

- the PEs estimated using less restricted control groups are everywhere below the PEs estimated using more restricted control groups after the formers' restriction on being selected expires.

The peak PE is 0.052 (t = 10.4, p = 0) when *M* = 18.5, 0.091 (t = 9.9, p = 0) when *M* =

17, and 0.111 (t = 9.6, p = 0) when *M* = 17 using *control groups: no restriction, 6M,* and

12M, respectively (see Figure 6). Thus, it appears that the ESPP increases the likelihood

that recipients start a job and exit welfare by approximately 11 percentage points.

<<Figure 6 here>>

ROBUSTNESS CHECK

One might be concerned that 'starting a job' is better observed for members of the

program group than for members of the control group.[15] This is not likely since NYC set

the ambitious goal of moving 100,000 recipients from welfare to work in 2000. Thus, the

City had a strong incentive to accurately record each job that was found by a recipient.

Nevertheless, to rule out this possibility, I estimate the effect of the ESPP on recipients'

welfare use. The ESPP should have increased welfare exits, if it increased employment.

Specifically, equation (3) is estimated with y_i^M redefined as an outcome dummy that

equals one if individual *i* has exited welfare for a six-month period within *M* months of

becoming a selectee or non-selectee and zero otherwise.[16]

14

The resulting PEs are again similar in many respects to the PEs estimated using the initial dependent variable (see Figure 7). Again,

- the PE first increases in M, and then after peaking, decreases in M, as control group contamination increases;
- the peak PE is positive and significantly different than zero using each control group;
- the magnitude of the peak PE increases as the length of the restriction on being selected increases; and
- the PEs estimated using less restricted control groups are everywhere below the PEs estimated using more restricted control groups after the formers' restriction on being selected expires.

The peak PE is 0.040 (t = 7.2, p = 0) when $M = 8$, 0.076 (t = 6.4, p = 0) when $M = 13$, 0.094 (t = 5.4, p = 0) when $M = 13.5$, and 0.103 (t = 4.8, p = 0) when $M = 13.5$ using *control groups: no restriction, 6M, 12M*, and *15M*, respectively. Thus, it appears that the ESPP increases the likelihood that recipients exit welfare by approximately 10 percentage points. <<Figure 7 here>>

DISCUSSION

The primary contribution of this paper is the development of a new quasi-experimental approach to identify the effect of a welfare-to-work program. This is important for the following four reasons. First, random-assignment experiments are often costly and difficult to implement. Second, random-assignment experiments are not impervious to criticism. This quasi-experiment addresses three of five known shortcomings. Third, this identification strategy can be used to estimate the effect of other government programs that face capacity constraints. For example, the Section 8 program is perpetually oversubscribed as are many high performing public schools (in districts

with school choice).[17] Fourth, this identification strategy can be utilized ex post. In contrast, random-assignment experiments need to be designed ex ante.

The quasi-experiment developed in this paper is not without its own shortcomings. Most notably, control group contamination is an inescapable side effect of the identification strategy. Even after adjusting for control group contamination to the extent possible, one is only able to estimate short-term effects. In the long-run, the PE's negative bias returns after the restriction on being selected expires. This limitation, only being able to estimate short-term effects, is not unique to this study. Most research regarding welfare reform only makes uses a few years of follow-up data [Friedlander et al, 1997].

Another unique aspect of this study is that it estimates the impact of a welfare-to-work program on GA recipients; GA is virtually unstudied to date. Yet, GA programs serve a large population of economically vulnerable individuals who should be of interest to researchers. The ESPP appears to have had a persistent positive impact on GA recipients in NYC. It increased the probability that they exited welfare and started a job by 10 and 15 percentage points, respectively.[18]

Previous studies that estimated the effect of similar welfare-to-work programs – those with mandatory work requirements and an emphasis on job placement – on FA recipients found that welfare exits and employment increased by an average of 6 and 9 percentage points, respectively [Grogger and Karoly, 2005]. The ESPP's effect on GA recipients is somewhat larger than the above reported averages. Whether this difference is due to the distinct population (GA versus FA recipients), the unique identification strategy, or another factor is unclear.

This research also contributes to the literature by studying the effect of a welfare-to-work program in NYC. Neither random-assignment experiments nor non-experimental studies have focused on a welfare-to-work program there. Yet, 23 percent of all GA recipients, and 7 percent of all FA recipients, lived in NYC in 1996. The only study that focuses on the impact of welfare reform in NYC is Chernick and Reimers [2004]. This study found that welfare use declined and earnings increased for 'at risk' individuals after PWRORA.[19] The authors acknowledged, however, that the observed changes could have been caused by welfare reform, the robust economy, or another factor.

Finally, one should consider whether the benefits of the ESPP outweighed the costs. Preferably, one would conduct a comprehensive Cost-Benefit Analysis (CBA) as was done in Orr et al [1996]. Unfortunately, such a CBA is not possible due to missing data. Specifically, recipients' earnings are not available. Increased earnings are the primary benefit of a training program for an enrollee.

It is possible, however, to conduct a rough fiscal CBA. The primary cost for the government were the fees paid to the ESPP contractors. The average fee was $3,000 per placement. Some placed selectees, however, would have found a job without the help of an ESPP contractor. Thus, we need to determine the cost of a 'new' placement, one that would not have occurred without the ESPP. To do so, consider that approximately 35 percent of selectees started a job during the study period and that the ESPP's employment effect was approximately 15 percentage points, thus NYC had to pay for 35 placements to generate 15 new placements. Consequently, the cost of a new placement was approximately $7,000 (= $3,000 × (35 ÷ 15)).

The primary benefit for the government was a reduction in GA benefits. In early 2000, a typical GA recipient received $350 per month, or $4,200 annually. Thus, the average selectee had to remain off welfare for 1.66 years (= $7,000 ÷ 4,200) for the primary fiscal benefit to outweigh the primary fiscal cost. There is not enough post selection data, unfortunately, to determine whether this is the case.[20] However, at the end of the study period, over two-thirds of the selectees who started a job were still off welfare. These recipients had already been off welfare for a year, on average. Projecting into the future, if one assumes that one-third of the selectees who started a job, and who were still off welfare at the end of the study, returned to welfare in each of the three subsequent years, then the primary fiscal benefits would surpass the primary fiscal costs. Further, there are numerous additional fiscal benefits, for example, decreased expenditures on Food Stamps or Medicaid, and few additional fiscal costs; data regarding these additional fiscal benefits are not available. Thus, it appears likely that the ESPP would pass a fiscal CBA.

NOTES

[1] Prior to passage of the PRWORA in 1996, states could receive a 'waiver,' permission from the federal government, to implement pilot welfare-to-work programs for a portion

of their Aid to Families with Dependent Children caseload. As a condition of receiving the waiver, the state had to evaluate the pilot program using a random-assignment experiment.

[2] Also using a quasi-experimental approach, Autor and Houseman [2005] have demonstrated that, unlike direct-hire jobs, temporary help jobs do not change the likelihood that the recipients are employed one to two years post placement.

[3] This is generally true for other components of welfare reform as well. One exception is time limits. Using a non-experimental identification strategy, Grogger [2003] demonstrates that time limits decrease welfare use.

[4] GA is cash assistance for financially needy individuals who are not covered by federally funded income maintenance programs.

[5] Other components of NYCWAY included enhanced detection of welfare fraud, increased substance abuse treatment, and diversion to federally funded disability insurance where appropriate.

[6] Participants received no compensation other than their welfare benefits and a nominal stipend for carfare and lunch.

[7] GA recipients were deemed job-ready if they did not have physical or emotional barriers to employment. Private contractors were hired to evaluation recipients' job-readiness. Over 90 percent of GA recipients who were already participating in a WEP assignment were found to be job-ready.

[8] Becoming ineligible was not exogenous. Recipients could become ineligible by exiting welfare or failing to comply with various program requirements.

[9] After 17 waves, 6,782 recipients had been selected, 953 remained eligible, and 3,791 had become ineligible.

[10] In contrast, one could estimate the effect of participating in the ESPP, a 'Treatment on the Treated' (TT) effect. To do so, one could multiply the IT effect by the inverse of the fraction of selectees who received treatment – that is, participated in the ESPP [Orr, 1996, p. 107-108]. For the TT estimate to be unbiased the following assumption needs to hold: The 'threat' of treatment cannot have an effect that is similar to the effect of the treatment itself; otherwise, the TT estimate will be positively biased. This assumption is unlikely to hold since the treatment is mandatory and costly for participants [Black et al, 2003]. IT and TT effects are further discussed in Katz, Kling, and Liebman [2001].

[11] The computer programmers who conducted the selection process mistakenly believed that sorting the list of eligible recipients by borough would cause each resulting borough list to be randomly ordered. Thus they simply selected recipients from the top of these lists.

[12] This approach enables one to estimate a very general, non-parametric hazard rate. Corrected standard errors are calculated by clustering the observations by individual. This is necessary since some individuals appear in the dataset repeatedly. $M = 26$ is the maximum number of months for which there is post selection data for each recipient. Again, each month is assumed to have 28 days. So there are 26 M's in two years.

[13] It would have been interesting to study whether recipients remained employed. However, NYC did not collect job retention data, and New York State was unwilling to provide the unemployment insurance wage records for study participants. Thus, this

measure is used. It provides some information regarding whether recipients remained employed, since recipients who lose their job presumably would return to welfare.

[14] This analysis is limited to M ≤ 18, since there are only two years of follow-up data and six months of follow-up data are needed to determine whether a recipient exited welfare.

[15] This concern arises since ESPP contractors were paid for each job placement. Thus the contractors had a strong incentive to make sure that each placement was recorded.

[16] Again, this analysis is limited to M ≤ 18, since there are only two years of follow-up data and six months of follow-up data are needed to determine whether a recipient exited welfare.

[17] In a future study, this identification strategy will be used to estimate the ESPP's effect on GA recipients' homeless shelter use.

[18] It is also likely that the ESPP increased earnings, since most selectees were placed in direct hire jobs and such jobs have been shown to increase earnings and quarters of employment [Autor and Houseman, 2005].

[19] At risk households were defined as those headed by a mother with a minor child as well as those headed by an uneducated, non-elderly, childless adult.

[20] Previous studies have suffered from the same limitation and have projected the benefits of short-term training programs into the future as well [Friedlander, 1997].

REFERENCES

Autor, David, and Susan Houseman. 2005. Do Temporary Help Jobs Improve Labor Market Outcomes for Low-Skill Workers? Evidence from Random Assignments. Upjohn Institute Staff Working Paper No. 05-124, Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Black, Dan A., Jeffrey A. Smith, Mark C. Berger, and Brett J. Noel. 2003. Is the Threat of Reemployment Services More effective than the Services Themselves? Evidence from Random Assignment in the UI System. *The American Economic Review*, 93(4): 1313-1327.

Blank, Rebecca M. 2002. Evaluating Welfare Reform in the United States. *Journal of Economic Literature,* 40(4): 1105-1166.

Chernick, Howard, and Cordelia Reimers. 2004. The Decline in Welfare Receipt in New York City: Push vs. Pull. *Eastern Economic Journal*, 30(2): 3-29.

Dyke, Andrew, Carolyn J. Heinrich, Peter R. Mueser, Kenneth R. Troske, and Kyung-Seong Jeon. 2006. The Effects of Welfare-to-Work Program Activities on Labor Market Outcomes. *Journal of Labor Economics*, 24(3): 567-607.

Friedlander, David H., David H. Greenberg, and Philip K. Robins. 1997. Evaluating Government Training Programs for the Economically Disadvantaged. *Journal of Economic Literature*, 35(4): 1809-1855.

Grogger, Jeffrey. 2003. The Effect of Time Limits, the EITC, and Other Policy Changes on Welfare Use, Work, and Income Among Female-Headed Families. *The Review of Economics and Statistics,* 85(2): 394-408.

Grogger, Jeffrey, and Lynn A. Karoly. 2005. *Welfare Reform: Effects of a Decade of Change*. Cambridge, MA: Harvard University Press.

Hotz, V. Joseph, Guido W. Imbens, and Jacob A. Klerman. 2006. Evaluating the Differential Effects of Alternative Welfare-to-Work Training Components: A Reanalysis of the California GAIN Program. *Journal of Labor Economics,* 24(3): 521-566.

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman. 2001. Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment. *The Quarterly Journal of Economics*, 116(2): 607-654.

Moffitt, Robert A. 2003. The Temporary Assistance for Needy Families Program, in *Means-Tested Transfer Programs in the United States,* edited by Robert A. Moffitt. Chicago, IL: The University of Chicago Press, 291-363.

Orr, Larry L., Harold S. Bloom, Stephen H. Bell, Fred Doolittle, Winston Lin, and George Cave. 1996. *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study.* Washington, DC: The Urban Institute Press.

Table 1

Demographic Characteristics of Members of the Program and Control Group

Demographic characteristic	All recipients (1)	Program group (2)	Control group (3)
Observations	64,800	6,782	58,018
Male	0.541 (0.002)	0.539 (0.006)	0.541 (0.002)
Age	47.324 (0.034)	**48.199** (0.101)	**47.221** (0.036)
Years continuously on welfare	3.379 (0.012)	**3.716** (0.037)	**3.340** (0.013)
Race			
Asian	0.012 (0.000)	**0.009** (0.001)	**0.012** (0.000)
Black	0.492 (0.002)	**0.478** (0.006)	**0.493** (0.002)
Hispanic	0.352 (0.002)	**0.371** (0.006)	**0.349** (0.002)
White	0.095 (0.001)	**0.086** (0.003)	**0.096** (0.001)
Not reported	0.042 (0.001)	**0.050** (0.003)	**0.042** (0.001)
Borough of residence			
Bronx	0.389 (0.002)	**0.308** (0.006)	**0.399** (0.002)
Brooklyn	0.309 (0.002)	**0.265** (0.005)	**0.314** (0.002)
Manhattan	0.180 (0.002)	**0.289** (0.006)	**0.168** (0.002)
Queens	0.111 (0.001)	**0.124** (0.004)	**0.110** (0.001)
Staten Island	0.007 (0.000)	**0.014** (0.001)	**0.006** (0.000)

Standard errors are given in parenthesis.
Bold means are significantly different for the program and control group.

Table 2

Coefficients from Estimating Equation (2) for Select Borough-Wave Pairs

Demographic characteristic	Bronx 3rd wave	Bronx 4th wave	Bronx 12th wave	Brooklyn 3rd wave	Brooklyn 11th wave	Brooklyn 16th wave	Manhattan 1st wave	Manhattan 2nd wave	Manhattan 3rd wave	Queens 4th wave	Queens 5th wave	Queens 6th wave
Male	-0.0434	0.0592	0.0086	*0.1885*	-0.1316	**0.3983**	*0.1398*	-0.0197	0.0175	0.1497	-0.1053	-0.0159
	(0.0632)	(0.0675)	(0.0732)	(0.0755)	(0.0904)	(0.1151)	(0.0700)	(0.0716)	(0.1047)	(0.1195)	(0.1216)	(0.1348)
Age	*0.0077*	0.0016	**0.0177**	**0.0119**	0.0031	0.0083	0.0070	*0.0089*	0.0109	0.0070	**0.0206**	*0.0182*
	(0.0037)	(0.0039)	(0.0044)	(0.0046)	(0.0054)	(0.0071)	(0.0044)	(0.0043)	(0.0062)	(0.0069)	(0.0076)	(0.0080)
Years continuously on welfare	**0.0800**	**0.0674**	-0.0199	**0.0438**	**-0.0573**	**0.0447**	**0.1091**	**0.0789**	-0.0099	**0.0935**	0.0123	-0.0418
	(0.0095)	(0.0103)	(0.0123)	(0.0122)	(0.0163)	(0.0170)	(0.0110)	(0.0125)	(0.0201)	(0.0193)	(0.0214)	(0.0265)
Race												
Asian	-0.0227	-	-	-	-0.1606	**-1.5415**	-0.5790	-0.6611	0.5272	0.1396	-0.3055	0.2656
	(0.4157)				(0.5724)	(0.5917)	(0.4508)	(0.4702)	(0.5006)	(0.4093)	(0.4715)	(0.4792)
Black	-0.0717	-0.0478	0.1155	-0.0862	0.2813	-0.4173	*-0.2989*	-0.0398	-0.2199	-0.0563	0.0506	0.2926
	(0.1324)	(0.1456)	(0.1931)	(0.1551)	(0.2430)	(0.2618)	(0.1325)	(0.1728)	(0.2677)	(0.2828)	(0.3126)	(0.3627)
Hispanic	0.0867	-0.0490	0.2344	-0.0503	0.3712	-0.3528	-0.2489	-0.0200	0.1130	0.0994	0.1166	0.5131
	(0.1271)	(0.1420)	(0.1906)	(0.1634)	(0.2516)	(0.2753)	(0.1326)	(0.1734)	(0.2674)	(0.2976)	(0.3280)	(0.3796)
White	0.2063	-0.1216	-0.0126	-0.3179	0.1867	*-0.6867*	-0.1420	-0.2260	-0.4293	-0.1750	-0.0310	0.0423
	(0.1827)	(0.2156)	(0.2555)	(0.1894)	(0.2652)	(0.2920)	(0.1824)	(0.2211)	(0.3510)	(0.3029)	(0.3296)	(0.3826)

Standard errors are given in parenthesis.
Bold coefficients are significantly different than zero at $p < 0.01$
Bold Italicized coefficients are significantly different than zero at $p < 0.05$

Table 3

Coefficients from Estimating Equation (3) using *Control Group: 6M restriction*

Demographic characteristic	Eight months (sixteen waves) post selection ($M = 8$)			Two years post selection ($M = 26$)		
	(1)	(2)	(3)	(4)	(5)	(6)
PE	**0.1310** (0.0078)	**0.1348** (0.0077)	**0.1309** (0.0081)	**0.075** (0.01)	**0.0823** (0.0125)	**0.0823** (0.0129)
Male		-0.0092 (0.0110)	-0.0066 (0.0112)		-0.0182 (0.0208)	-0.0147 (0.0208)
Age		**-0.0016** (0.0006)	**-0.0017** (0.0006)		**-0.0042** (0.0011)	**-0.0044** (0.0011)
Years continuously on welfare		-0.0020 (0.0016)	-0.0024 (0.0017)		0.0015 (0.0034)	-0.0002 (0.0034)
Race						
Asian		0.0835 (0.0673)	0.0690 (0.0661)		0.0566 (0.1083)	0.0393 (0.1057)
Black		0.0096 (0.0220)	0.0007 (0.0218)		0.0199 (0.0460)	0.0116 (0.0467)
Hispanic		-0.0051 (0.0222)	-0.0042 (0.0220)		-0.0516 (0.0460)	-0.0431 (0.0471)
White		0.0055 (0.0263)	-0.0165 (0.0266)		-0.0199 (0.0540)	-0.0514 (0.0547)
Borough of residence						
Bronx			-0.0538 (0.1184)			-0.0764 (0.1650)
Brooklyn			0.0159 (0.1190)			0.0417 (0.1651)
Manhattan			-0.0090 (0.1220)			0.0238 (0.1692)
Queens			0.0456 (0.1320)			0.0233 (0.1807)
Wave dummy	No	No	Yes	No	No	Yes
Interaction term	No	No	Yes	No	No	Yes

Standard errors are given in parenthesis.
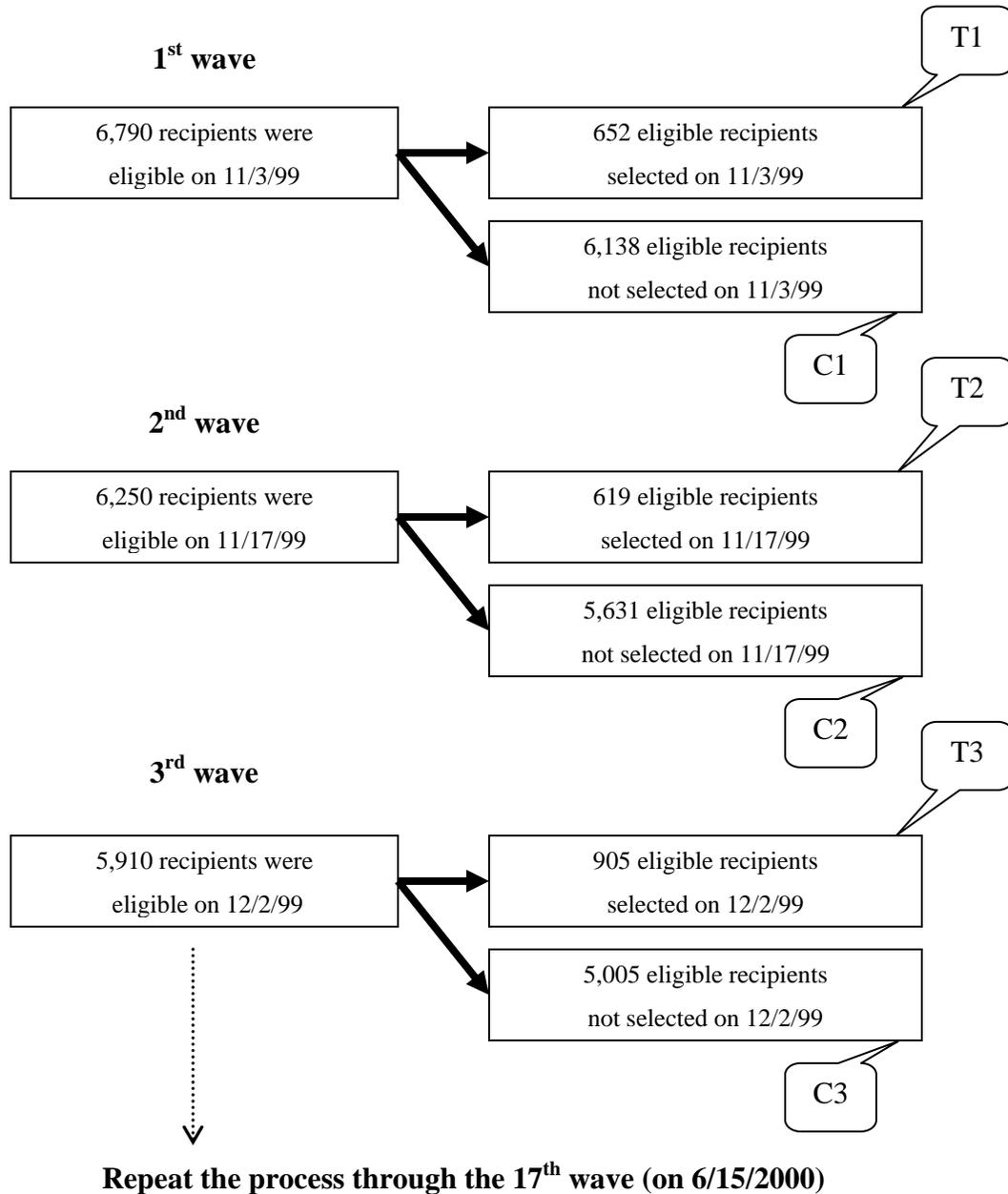**Bold** coefficients are significantly different than zero at $p < 0.01$

Figure 1

The Formation of the Program and Control Groups

The program group:
All selectees from the first 17 waves. Specifically it is the union of T1, T2, T3, ….., T17.

The control group:
All non-selectees from the first 17 waves. Specifically it is the union of C1, C2, C3, ….., C17[1].

1st wave

| 6,790 recipients were eligible on 11/3/99 |

→ | 652 eligible recipients selected on 11/3/99 | T1

→ | 6,138 eligible recipients not selected on 11/3/99 | C1

2nd wave

| 6,250 recipients were eligible on 11/17/99 |

→ | 619 eligible recipients selected on 11/17/99 | T2

→ | 5,631 eligible recipients not selected on 11/17/99 | C2

3rd wave

| 5,910 recipients were eligible on 12/2/99 |

→ | 905 eligible recipients selected on 12/2/99 | T3

→ | 5,005 eligible recipients not selected on 12/2/99 | C3

Repeat the process through the 17th wave (on 6/15/2000)

---

[1] Note that many recipients were non-selectees in multiple waves. Consequently, many recipients are members of the control group multiple times.

Figure 2

The Formation of the Program Group and the Control Group with a One-Wave Restriction on Being Selected

The treatment group:
All selectees from the first 17 waves. Specifically it is the union of T1, T2, T3, ….., T17.

The control Group with a one-wave restriction on being selected
All recipients who were:

1. non-selectees during any of the first 17 waves and
2. not selected in the wave subsequent to the one in which they were a non-selectee.

Specifically it is a weighted union of C1a, C1b, C2a, C2b, C3a, C3b, ….., C17a, C17b. The weight placed on each member of each a-series cohort is equal to one. The weight placed on each member of each b-series cohort is equal to the reciprocal of the probability of not being selected in the subsequent wave, conditional on being eligible to be selected in that wave, for example, for members of C1b the weight is the reciprocal of (5,248/5,860), or (1/0.896) = 1.116.

Figure 3

Estimated effect of ESPP on starting a job using all control groups
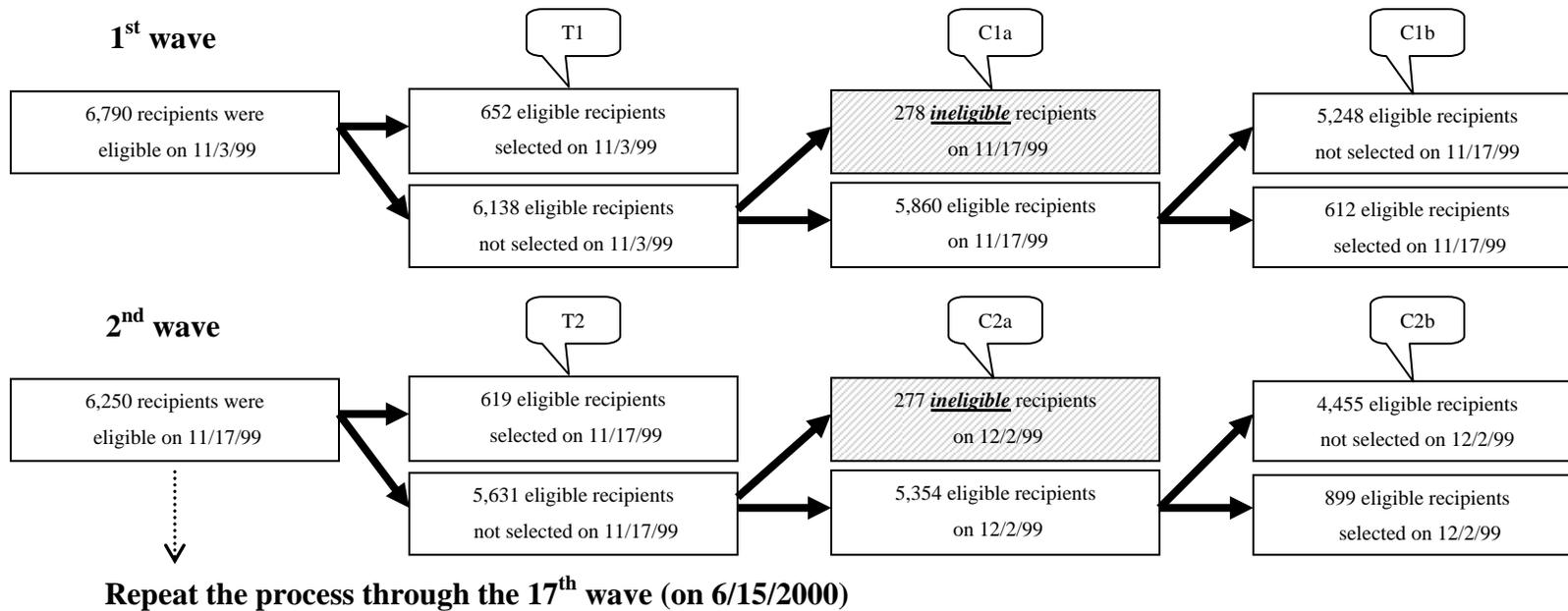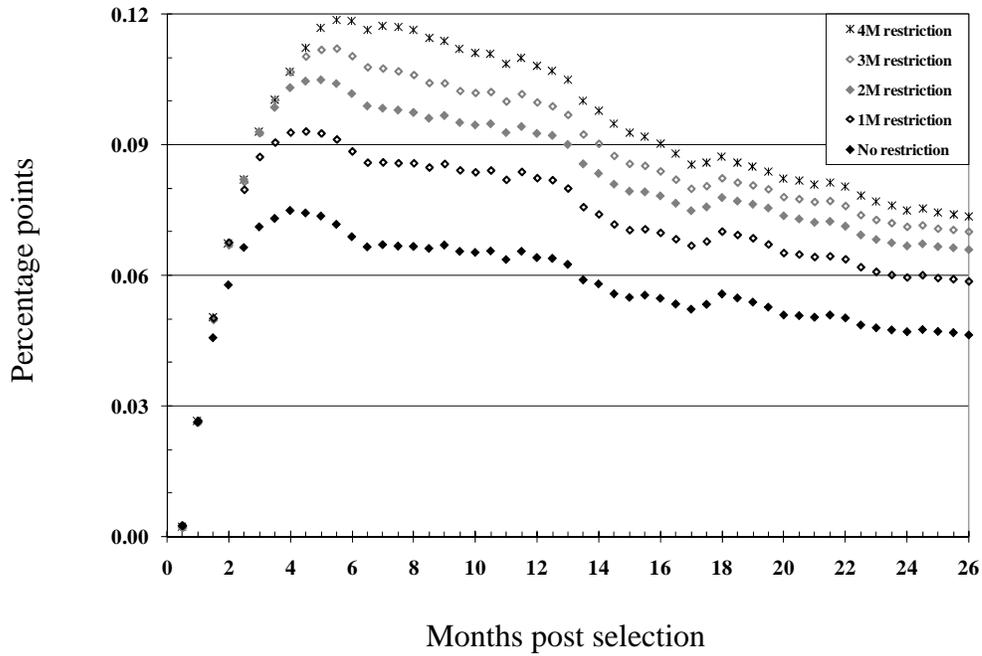
Figure 4

a) Estimated effect of ESPP on starting a job using four least restricted control groups

b) Estimated effect of ESPP on starting a job using four most restricted control groups
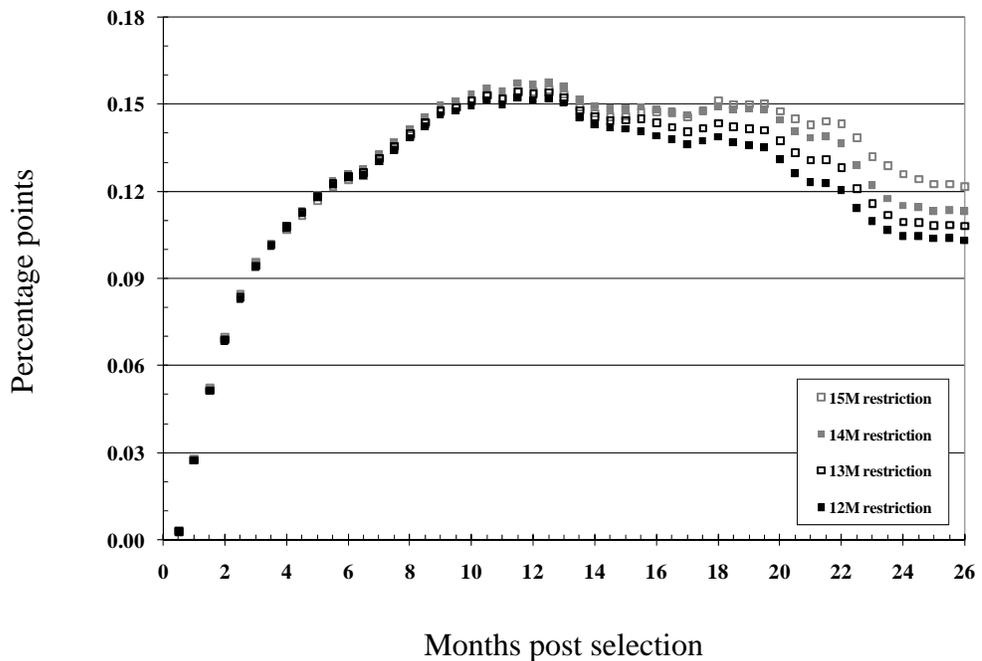
Figure 5

Estimated effect of ESPP with various covariates included using *control groups: 6M* and *12M restriction*

Legend:
- All covariates (6M restriction)
- Age, race, time on welfare, and gender included (6M restriction)
- No covariates (6M restriction)
- All covariates (12M restriction)
- Age, race, time on welfare, and gender included (12M restriction)
- No covariates (12M restriction)

Y-axis: Percentage points
X-axis: Months post selection

Figure 6

Estimated effect of ESPP on starting a job and exiting welfare using all control groups

Figure 7

Estimated effect of ESPP on exiting welfare using all control groups