IFAU
Institute for Evaluation of Labour
Market and Education Policy

IPW estimation and related estimators for evaluation of active labor market policies in a dynamic setting

Johan Vikström

WORKING PAPER 2014:1

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

IFAU also provides funding for research projects within its areas of interest. The deadline for applications is October 1 each year. Since the researchers at IFAU are mainly economists, researchers from other disciplines are encouraged to apply for funding.

IFAU is run by a Director-General. The institute has a scientific council, consisting of a chairman, the Director-General and five other members. Among other things, the scientific council proposes a decision for the allocation of research grants. A reference group including representatives for employer organizations and trade unions, as well as the ministries and authorities concerned is also connected to the institute.

Postal address: P.O. Box 513, 751 20 Uppsala
Visiting address: Kyrkogårdsgatan 6, Uppsala
Phone: +46 18 471 70 70
Fax: +46 18 471 70 71
ifau@ifau.uu.se
www.ifau.se

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

ISSN 1651-1166

IPW estimation and related estimators for evaluation of active labor market policies in a dynamic setting[a]

by

Johan Vikström[b]

10th January, 2014

Abstract
This paper considers treatment evaluation in a discrete time setting in which treatment could start at any point in time. A typical application is an active labor market policy program which could start after any elapsed unemployment duration. It is shown that various average effects on survival time are identified under unconfoundedness and no-anticipation and inverse probability weighting (IPW) estimators are provided for these effects. The estimators are applied to a Swedish work practice program. The IPW estimator is compared with related estimators. One conclusion is that the matching estimator proposed by Fredriksson and Johansson (2008)[c] overlooks a selective censoring problem.[d]

Keywords: Treatment effects; dynamic treatment assignment; program evaluation; work practice
JEL-codes: C14, C4

[a]Gerard J. van den Berg, Gregory Jolivet, Per Johansson, Bas van der Klauuw, Ingeborg Waernbaum and seminar participants at IFAU Uppsala, the 4^{th} IZA/IFAU conference and SOFI are acknowledged, for valuable comments. Financial support from the Jan Wallander and Tom Hedelius Foundation is acknowledged.
[b]IFAU Uppsala, and UCLS, e-post: johan.vikstrom@ifau.uu.se
[c]Fredriksson P. and P. Johansson (2008), "Dynamic Treatment Assignment: The Consequences for Evaluations Using Observational Data", *Journal of Business & Economic Statistics*, 26:4, 435–445.
[d]STATA software to compute the estimators in this paper is available on ifau.se.

Table of contens

1 Introduction

A common feature of many active labor market policy (ALMP) programs is that the program could start after many different unemployment durations. By now it is rather well known that this dynamic nature of the treatment assignment introduces several methodological issues, which affect the selection of a proper control group. The main issue is that currently non-treated individuals might become treated later on. The implication of this is that unconfoundedness-based methods that use static treatment status, defined as enrollment into treatment before exit from unemployment, are no longer valid (see discussions in e.g. Sianesi 2004, Fredriksson and Johansson 2008, Crépon et al. 2009). The reason is that static treatment status depends on survival time (i.e., the outcome), since the probability of treatment enrollment by construction increases with the time in unemployment, and this confounds any analysis solely based on static treatment indicators.

As a response, several papers explicitly address the dynamic nature of the identification and estimation problem. Sianesi (2004) develops an ingenious way to transform the dynamic problem into a static problem by focusing on the effect of treatment now vs. continuing to wait for treatment. Important applications of this approach includes e.g. Sianesi (2008), Fitzenberger et al. (2008) and Biewen et al. (2013). Several other papers focus on the average effect of treatment after some elapsed duration compared with never receiving treatment, and this is also the average effect considered in this paper. For instance, both Fredriksson and Johansson (2008) and Crépon et al. (2009) utilize the outcomes of the not-yet treated to obtain the counterfactual outcome under never treatment. One difference is that Fredriksson and Johansson (2008) impose a single sequential unconfoundedness assumption, whereas Crépon et al. (2009) discuss identification under separate unconfoundedness and no-anticipation assumptions. In a related paper, Kastoryano and van der Klauuw (2011) compare different evaluation approaches in a dynamic setting. Other influential studies are Lechner (1999) and Gerfin and Lechner (2002).

This paper contributes to this literature in several ways. It considers identification and estimation of the effect of a treatment given in an initial state on survival time in the very same state under unconfoundedness. The key feature is that exits out of the initial state and the start of treatment are allowed to occur at any point in time. Besides ALMP programs, an important example of this setting is, for instance, a medical treatment implemented at various times after the onset of the disease.

The main contribution of this paper is to provide an inverse probability weighting (IPW) estimator for the average treatment effect on the treated for treatment in a certain period against no treatment now nor thereafter. One advantage of the IPW approach is that once the scores forming the weights are estimated no additional functional form assumptions are needed. The finite sample properties of the IPW estimators and related estimators are explored in a Monte Carlo simulation.

As an illustration of the estimator consider the average effect on the treated at t. The survival rate under treatment is obtained directly from those actually treated at t. The counterfactual exit rate under no treatment at t is estimated by weighting the outcomes of the not-yet treated at t in order to mimic the distribution of the confounders in the population of treated at t. In subsequent periods some of the not-yet treated at t become treated, and this creates selective censoring in the group of not-yet treated. However, under unconfoundedness the weights at $t+1$ correct for this selective censoring, so that the IPW estimator gives the desired exit rate at $t+1$. A series of unique weights for each time period provide exit rates in all subsequent periods and together these exit rates constitute the counterfactual survival rate. The weights are also adjusted to handle standard right-censoring. An estimator for the average effect averaged over all pre-treatment durations is also suggested.

Another important contribution is a re-examination of the properties of the estimator proposed by Fredriksson and Johansson (2008) and further discussed by de Luna and Johansson (2010). Like the IPW estimators proposed in this paper this estimator also uses the not-yet treated as control group. The conclusion is that Fredriksson and Johansson (2008) overlooked a selective censoring problem. Briefly consider the intuition behind this result. For a given time to treatment t the first step of the Fredriksson and Johansson (2008) estimator is one-to-one matching of treated and non-treated at t. In the second step, treated and matched controls are pooled into two separate groups and the average effect is obtained by contrasting the survival rates in the two groups. Importantly, in this step any matched control starting treatment after t is considered right-censored when (s)he starts treatment. This right-censoring of the matched controls, however, introduces non-random censoring. The intuition is that although treatment assignments are random within each matched pair the right-censoring due to subsequent treatment assignment is not random within the pooled sample of matched controls, and this confounds the analysis unless treatment assignment does not depend on the observed characteristics. This is also confirmed by extensive Monte Carlo simulations. The IPW estimator is also compared to the blocking estimator in Crépon et al. (2009).

This paper also explicitly discusses identification under selection on observables. The discussion builds upon the work by Fredriksson and Johansson (2008) and Crépon et al. (2009). The latter paper discusses identification under unconfoundedness and no-anticipation. This paper aims to clarify and highlight the results in Crépon et al. (2009) by providing explicit step by step identification results, which formally explain the exact role of the unconfoundedness assumption and the no-anticipation assumption. The identification part of the paper also includes a detailed discussion of several potential control groups. This collects and discusses results from several previous papers including Sianesi (2004), Fredriksson and Johansson (2008) and Crépon et al. (2009) in an unified framework, aiming to clarify how a proper control group could be constructed.

Besides the above discussed papers on evaluation under selection on observables with a single treatment, this paper is related to several other strands of the literature. Robins (1986), Lechner (2008, 2009) and Lechner and Miquel (2010) also discuss dynamic settings, but focus on the effects of sequences of treatments. Evaluation of sequences of treatments is also considered in the companion paper Vikström et al. (2013), which provides IPW estimators of average effects of sequences of treatments, allowing individuals to enter and exit from treatment multiple times.

This paper is also related to the Timing-of-Events (ToE) approach by Abbring and van den Berg (2003), which also consider evaluation in a dynamic setting, in which exits and treatments are allowed to occur at any point in time. The main difference compared with this paper is that the ToE approach allows the selection into treatment to be based on both observed and unobserved heterogeneity. This is achieved at the expense of imposing the mixed proportional hazard structure, whereas the IPW approach in this paper requires no parametric assumptions. The ToE approach and methods based on selection on observables, such as the IPW approach in this paper, are therefore complementary approaches, applicable under different treatment assignment processes.

The IPW estimators are illustrated using data from a Swedish work practice program. Data for the period 2003-2006 are used and the result is that the program increases the employment rate 15 months after enrollment in the program with 6-12 percentage points compared with no treatment.

The rest of the paper is organized as follows. Section 2 presents the evaluation framework and section 2 presents identification results and the discussion of the selection of a proper control group. Section 4 introduces the IPW estimator and section 5 re-examines the properties of the Fredriksson and Johansson (2008) estimator. Section 6 gives the simulation results for the weighted estimator and related estimators, and section 7 reports the results from the application. Section 8 concludes.

2 Evaluation framework

This paper considers average effects of a treatment given in an initial state on survival time in the very same state. Each individual enters the single treatment at most once. Transitions as well as the start of treatment could occur at any point in discrete time. Time to treatment start is denoted by S with realized values $s \in [1,\infty)$. Let $Y_t(s)$ be an indicator of a transition in period t if treated at s. The potential outcome if never treated is denoted by $Y_t(0)$ and the observed outcome in period t is Y_t. Denote by $\overline{Y}_t(s)$ the sequence of potential outcomes $\overline{Y}_t(s) = \{Y_1(s),\ldots,Y_t(s)\}$, and $\overline{Y}_t$ is the sequence of observed outcomes $\overline{Y}_t = \{Y_1,\ldots,Y_t\}$. Throughout the paper assume a sample of N individuals $i = 1,\ldots,N$.

The first parameter of interest is the average treatment effect of treatment at s' on the probability of surviving to time point t compared with survival throughout the same interval if treated at $s'' > s'$ for the population starting treatment at s'

$$\mathrm{ATET}_t(s',s'') = \tag{1}$$

$$\Pr(\overline{Y}_t(s') = 0 | S = s', \overline{Y}_{s'-1}(s') = 0) - \Pr(\overline{Y}_t(s'') = 0 | S = s', \overline{Y}_{s'-1}(s') = 0).$$

It resembles the average treatment effect on the treated often considered in the static matching literature, but one difference is that this average effect is taken over the population of treated at s' that survives up until s' and not over the full population of treated at s'. An effect of particular interest is the average effect on the treated of treatment at s against no treatment now nor thereafter (i.e., $s' = s$ and $s'' = 0$). The following short-hand notation is used for this average effect

$$\text{ATET}_t(s) = \Pr(\overline{Y}_t(s) = 0 | S = s, \overline{Y}_{s-1}(s) = 0) - \Pr(\overline{Y}_t(0) = 0 | S = s, \overline{Y}_{s-1}(s) = 0). \quad (2)$$

In this paper, the interest lies in cases with selection on observables, such that conditional on a set of observed covariates, treatment assignment is independent of the potential outcomes. In the dynamic setting considered here, we have

$$S \perp Y_t(s) \quad \forall t, s \quad | X. \quad \text{(U.1)}$$

Besides this unconfoundedness (conditional independence) assumption, another key assumption is no-anticipation, which implies that any future treatment does not affect current outcomes. Formally,

$$\Pr(Y_t(s') = 1) = \Pr(Y_t(s'') = 1) \quad , \forall t < \min(s', s''). \quad \text{(N.A.)}$$

The assumption is fulfilled if individuals are unaware of future treatments or if they do not alter their behavior as a response to knowledge of future treatments. The importance of this assumption for evaluations in dynamic settings was highlighted by Abbring and van den Berg (2003), and subsequently discussed by e.g. Abbring and Heckman (2008).

3 Identification and the choice of control group

Let us consider identification of $\text{ATET}_t(s)$ under assumptions U.1 and N.A. in detail. First and trivially, the survival function under treatment at s is directly identified by the outcomes of those actually treated at s

$$\Pr\left(\overline{Y}_t(s)=0|S=s,\overline{Y}_{s-1}(s)=0\right)=\Pr(\overline{Y}_t=0|S=s,\overline{Y}_{s-1}=0). \tag{3}$$

The main issue instead is how to select a proper control group in order to identify the counterfactual outcome $\Pr(\overline{Y}_t(0)=0|S=s,\overline{Y}_{s-1}(s)=0)$. One key problem is that the start of treatment could occur at any point in time, so that individuals not treated at t might be treated at $t+1$, $t+2$, ... Another problem is that the start of treatment is unobserved if the individual leaves the initial state before receiving treatment. This has several implications for the choice of control group.

In general there are three potential control groups to treated at s. The first is to use individuals who are non-treated at s but possibly treated later, i.e. individuals still waiting for treatment. This control group is used by e.g. Sianesi (2004, 2008). Clearly, under assumption U.1 and for a given X those still waiting for treatment at s are comparable to those treated at s. However, using individuals still waiting for treatment as control group leads to a special kind of treatment effect. Let us consider this in detail, and for sake of presentation take the case with $s=1$. Then, from the observed outcomes of those still waiting for treatment at $s=1$, we have

$$\begin{aligned}
&\Pr(\overline{Y}_t=0|X,S>1)=\sum_{k=2}^{\infty}\Pr(S=k|X,S>1)\Pr(\overline{Y}_t(k)=0|X,S=k)= \\
\text{(N.A.)}\quad &\sum_{k=t+1}^{\infty}\Pr(S=k|X,S>1)\Pr(\overline{Y}_t(0)=0|X,S=k)+ \\
&\sum_{k=2}^{t}\Pr(S=k|X,S>1)\Pr(Y_t(k)=...=Y_k(k)=Y_{k-1}(0)...=Y_1(0)=0|X,S=k)= \\
\text{(U.1)}\quad &\Pr(S>t|X,S>1)\Pr(\overline{Y}_t(0)=0|X,S=1)+ \\
&\sum_{k=2}^{t}\Pr(S=k|X,S>1)\Pr(Y_t(k)=...=Y_k(k)=Y_{k-1}(0)...=Y_1(0)=0|X,S=1).
\end{aligned}$$

Thus, for all $t>1$ the estimated outcome is a weighted average of the survival rate under no treatment for the treated at s and survival rates that partly depend on treatment responses. Even if the obtained counterfactual outcome is well defined, it might be difficult to interpret if a large fraction of those still waiting for treatment at s is treated later on. For that reason, control groups enabling us to estimate the $\text{ATET}_t(s)$ are considered, since such analyses complement analysis based on those still waiting as control group.

The second potential control group is some subset of "never-treated" individuals, that is individuals who are observed to exit the initial state before becoming treated. Note that this group includes those who actually never would have received treatment, but also individuals who would have been enrolled in treatment shortly after their exit from the initial state. Define T as the observed time in the initial state. Then, formally, the "never-treated" group consists of individuals with $s \leq T < S$. This means that the "never-treated" are endogenously selected partly on the outcome itself. Specifically, individuals with a specific S are included in the control group if they exit early, but not if they exit after S.

Instead, another choice is to successively use all not-yet treated at t to estimate the exit rate under no-treatment at t for those treated at s. This idea is discussed by e.g. Fredriksson and Johansson (2008), Crépon et al. (2009) and Kastoryano and Klaauw (2011). This paper follows Crépon et al. (2009) and discusses identification under unconfoundedness and an explicit no-anticipation assumption. Here, the contribution is to provide step by step identification results, with the aim to clarify the exact role of the two identifying assumptions. Initially, note that

$$\Pr(\overline{Y}_t(0) = 0|S = s, \overline{Y}_{s-1}(s) = 0) = E_{X|S=s,\overline{Y}_{s-1}=0} \Pr(\overline{Y}_t(0) = 0|X, S = s, \overline{Y}_{s-1}(s) = 0). \tag{4}$$

Next, for a given X and using assumptions U.1 and N.A.

$$
\begin{aligned}
& \Pr(\overline{Y}_t(0)=0|X,S=s,\overline{Y}_{s-1}(s)=0) = && (5)\\
\text{(N.A.)}\quad & \Pr(\overline{Y}_t(0)=0|X,S=s,\overline{Y}_{s-1}(0)=0) = \\
& \prod_{m=s}^{t} \Pr(Y_m(0)=0|X,S=s,\overline{Y}_{m-1}(0)=0) = \\
\text{(U.1)}\quad & \prod_{m=s}^{t} \Pr(Y_m(0)=0|X,S>m,\overline{Y}_{m-1}(0)=0) = \\
& \prod_{m=s}^{t} \sum_{k=m+1}^{\infty} \Pr(S=k|X,S>m,\overline{Y}_{m-1}(0)=0)\Pr(Y_m(0)=0|X,S=k,\overline{Y}_{m-1}(0)=0) = \\
\text{(N.A.)}\quad & \prod_{m=s}^{t} \sum_{k=m+1}^{\infty} \Pr(S=k|X,S>m,\overline{Y}_{m-1}=0)\Pr(Y_m(k)=0|X,S=k,\overline{Y}_{m-1}(k)=0) = \\
& \prod_{m=s}^{t} \sum_{k=m+1}^{\infty} \Pr(S=k|X,S>m,\overline{Y}_{m-1}=0)\Pr(Y_m=0|X,S=k,\overline{Y}_{m-1}=0) = \\
& \prod_{m=s}^{t} \Pr(Y_m=0|X,S>m,\overline{Y}_{m-1}=0).
\end{aligned}
$$

Note that in each period only not-yet treated individuals are used, so that the control group successively changes as some previously non-treated individuals start treatment, but this is not a problem for identification since this successive treatment assignment process is assumed to be random for a given X. In addition, note the independent use of the selection on observables assumption and the no-anticipation assumption. The selection on observables assumption relates to the allocation of treatment across individuals, and assures that the treated and the not-yet treated have similar potential outcomes. The no-anticipation assumption concerns the relationship between different potential outcomes for a given individual, and assures that the outcomes of the not-yet treated at t could be used to mimic the outcomes under never treatment if even if some of the not-yet treated at t become treated at $t+1,\ldots.$

From the results in (3), (4) and (5), we obtain

$$
\text{ATET}_t(s) = \Pr(\overline{Y}_t=0|S=s,\overline{Y}_{s-1}=0) - E_{X|S=s,\overline{Y}_{s-1}=0} \prod_{m=s}^{t} \Pr(Y_m=0|X,S>m,\overline{Y}_{m-1}=0).
$$

Identification of $\text{ATET}_t(s',s'')$ follows using similar reasoning.

4 Weighted estimation

This section proposes IPW estimators for the average treatment effect on the treated. In appendix A.2 it is shown that an asymptotically unbiased estimator of $\mathrm{ATET}_t(s)$ is:

$$\widehat{\mathrm{ATET}}_t(s) = \tag{6}$$

$$\prod_{k=s}^{t}\left[1-\frac{\sum_{i\in \overline{Y}_{k-1,i}=0,S_i=s} Y_{k,i}}{\sum_{i\in \overline{Y}_{k-1,i}=0,S_i=s} 1}\right]-\prod_{k=s}^{t}\left[1-\frac{\sum_{i\in \overline{Y}_{k-1,i}=0,S_i>k} w_{k,i}(s)Y_{k,i}}{\sum_{i\in \overline{Y}_{k-1,i}=0,S_i>k} w_{k,i}(s)}\right]$$

with

$$w_{k,i}(s) = \frac{p_s(X_i)}{1-p_s(X_i)}\frac{1}{\prod_{m=s+1}^{k} 1-p_m(X_i)},\ p_t(X_i) = \Pr(S=t|X_i, S\geq t, \overline{Y}_{t-1}=0).$$

In practice, estimated propensity scores are used instead of the true propensity scores. See Hirano et al. (2003) for a discussion of the implications of using estimated scores instead of the true scores. One way to obtain standard errors is bootstrapping. The estimator for $\mathrm{ATET}_t(s', s'')$ is given in appendix A.3.

Naturally, if the interest lies in the average effect on the treated at s the actually observed outcomes of those treated at s, could be used to estimate the survival rate under treatment. Thus, in the first part of the estimator the exit rate in each period under treatment is obtained as the unweighted fraction leaving the initial state among those treated at s.

The counterfactual outcome under no-treatment is obtained using untreated survivors at s, i.e. those not-yet treated at s. Under assumption U.1, the treated at s and the not-yet treated at s are comparable if we adjust for the fact that due to the assignment process the distribution of X differs between the two populations. Thus, the counterfactual exit rate at s is obtained by weighting the not-yet treated at risk at s and the exits among this group. Note that the weights at s essentially follow from the IPW estimators of average effects on the treated in the static evaluation literature (see e.g. Wooldridge, 2010), since in this period the only purpose of the weights is to adjust for covariate differences between the treated and not-yet treated.

At $s+1$, i.e. in the second period after the start of treatment, a fraction of the not-yet treated at s that survives up until $s+1$ starts treatment. This creates selective censoring in the group of not-yet treated. However, under assumption U.1, assignments at $s+1$ only depend on observed covariates, so that the selective censoring could be taken into account by weighting the outcomes of the not-yet treated at $s+1$, and this is the purpose of the second part of the weights. The implication is that individuals still not-yet treated with covariates such that they have a high probability to start treatment are given larger weight, and this corrects for the selective censoring due to treatment assignment at $s+1$. Again, the exit rate is obtained as the weighted exits divided by the weighted risk set. Similar weighting occurs at $s+2$, but then the weights take the selective censoring at both $s+1$ and $s+2$ into account, and so on.

Overlap. In practice, an overlap condition is required. For a specific $\text{ATET}_t(s)$ we have:

$$p_k(X) < 1 \quad \forall s \leq k \leq t.$$

Concerning applications, if some individuals have characteristics that make them very likely or very unlikely to enter treatment common support restrictions may be important. Note that in this dynamic setting common support needs to be imposed for all time periods between s and t. Two approaches are minima and maxima comparisons, and trimming (see e.g. the discussion in Caliendo and Kopeinig, 2008).

Right-censoring. The estimators above ignore regular right-censoring, which is common in many applications due to, for instance, drop-out from the study or a limited follow-up period. Standard right-censoring creates another source of selective censoring, but unlike treatment assignment, this selective censoring affects both the treated and the not-yet treated. Formally, let C be the censoring time, and consider estimation if the censoring is independent conditional on covariates

$$C \perp Y_t(s) \quad \forall t,s \quad | X. \tag{C.1}$$

Under this assumption, one can use the observed covariates for individuals with censored and uncensored durations in order to correct for the right-censoring using a similar IPW approach as when correcting for censoring due to treatment assignment. The exact expressions for the weights are reported in appendix A.3.

Aggregation over pre-treatment durations. The $\text{ATET}_t(s)$ provides estimates for each separate pre-treatment duration. From a policy perspective, one might also be interested in the overall effect, that is the average effect on the treated averaged over all pre-treatment durations. Specifically, the overall effect on the probability of surviving t' time periods after the start of the treatment could be estimated as:

$$\widehat{\text{ATET}}_{t'} = \sum_s \widehat{P}(s) \widehat{\text{ATET}}_{s+t'}(s),$$

i.e. as an average over relevant pre-treatment durations. Here, $\widehat{P}(s) = \frac{n_s}{\sum_s n_s}$, were n_s is the number of treated at s, so that $\widehat{P}(s)$ is the fraction in the sample of treated starting treatment at s.

5 A related estimator

The weighted estimators introduced in the previous section are based on using the not-yet treated as control group. Estimation using not-yet treated as control group has previously been proposed by e.g. Fredriksson and Johansson (2008) (FJ henceforth). This section re-examines the properties of the FJ estimator. FJ propose a two-step matching estimator. In the first step, treated at s are matched to untreated survivors at s using one-to-one matching. In the second step, the samples of matched treated and matched controls are used to construct unweighted estimates of the survival rates under treatment and no-treatment. In this second step any matched control starting treatment after t is considered right-censored when (s)he starts treatment. Formally, let $j(i)$ be the index for the selected match for individual i. Then, the FJ estimator is:

$$\widehat{\text{ATET}}_t^{FJ}(s) = \prod_{k=s}^{t} \left[1 - \frac{\sum_{i \in \overline{Y}_{k-1,i}=0, S_i=s} Y_k}{\sum_{i \in \overline{Y}_{k-1,i}=0, S_i=s} 1} \right] \tag{7}$$

$$- \prod_{k=s}^{t} \left[1 - \frac{\sum_{i \in S_i=s, \overline{Y}_{s-1,i}=0} Y_{k,j(i)} \mathbf{1}(\overline{Y}_{k-1,j(i)} = 0) \mathbf{1}(S_{j(i)} > k)}{\sum_{i \in S_i=s, \overline{Y}_{s-1,i}=0} \mathbf{1}(\overline{Y}_{k-1,j(i)} = 0) \mathbf{1}(S_{j(i)} > k)} \right].$$

From this we have that the second step is standard Kaplan-Meier (1958) estimates of the survivor functions, so that the FJ estimator is easy-to-use in practice.[1]

Note that the estimator of the survival rate under treatment is identical to the weighted estimator presented in this paper. Instead the differences lies in the estimation of the survival rate under no-treatment. The most important difference is that in the FJ estimator all not-yet treated are given equal weight in the first period as well as in all subsequent time periods. That is the FJ estimator does not correct for the selective censoring due to treatment assignment in periods $s+1$, $s+2$,.... Intuitively, even if the censoring due to treatment in the control group is random conditional on X the censoring is not random in the pooled control group consisting of individuals with different values of the covariates, and this might introduce substantial bias.

Let us consider this in detail. For sake of presentation take the case when X takes a finite number of values, so that with a large sample of untreated individuals exact matches are available for each treated. Then, from appendix A.4

$$p\lim_{N\to\infty} \widehat{\text{ATET}}_t^{FJ}(s) = \Pr(\overline{Y}_t(s)=0|S=s,\overline{Y}_{s-1}(s)=0) - \tag{8}$$

$$\prod_{k=s}^{t}\left[1-\frac{\mathbb{E}_{X|S=s,\overline{Y}_{s-1}=0}\Pr(Y_k(0)=1,\overline{Y}_{k-1}(0)=0|X,S=s,\overline{Y}_{s-1}=0)\prod_{m=s+1}^{k}1-p_m(X)}{\mathbb{E}_{X|S=s,\overline{Y}_{s-1}=0}\Pr(Y_{k-1}(0)=0|X,S=s,\overline{Y}_{s-1}=0)\prod_{m=s+1}^{k}1-p_m(X)}\right].$$

From this expression we have that the problem with FJ estimator is the presence of $\prod_{k=s+1}^{t}1-p_k(X)$ in both the denominator and the numerator of the second part of the expression. The consequence is that without re-weighting individuals with X characteristics that makes them less likely to enter treatment will be overrepresented among the not-yet treated, and this confounds the comparison of the treated and the not-yet treated if these X characteristics also affects the outcome.

Equation (8) also shows that the Fredriksson and Johansson (2008) estimator is asymptotically unbiased in some special cases. This holds if, for instance, no one becomes

[1] Concerning the details in FJ, from equation (9) in FJ we have their estimator, $\widehat{\Delta}(w,s)$, for the average effect on the survival rate. Even though FJ never explicitly define the effect of interest, $\Delta(w,s)$, we implicitly have from their text that their $\Delta(w,s)$ is the average effect on the treated, $\text{ATET}_{s+w}(s)$, considered in this paper. This interpretation is also supported by the application in FJ as well as the reformulation of the FJ estimator in de Luna and Johansson (2010).

treated between s and t ($\prod_{k=s+1}^{t} 1 - p_k(X) = 1$), or if treatment assignment does not depend on X ($p_m(X) = p_m$). That is in cases without dynamic treatment assignment and/or under random treatment assignment, i.e. cases in which one could argue that the matching approach is redundant. Equation (8) also provides some intuition about how the selective censoring affects the FJ estimator. In general, if individuals with X characteristics that makes them likely to enter treatment also on average have higher exit rates the survival rate under no-treatment will be underestimated, since then the exit rates among the remaining not-yet treated will be too low. In addition, a high treatment rate at $s+1,...$, in general, implies more extensive selective censoring.

Besides problems due to unaccounted selective censoring caused by the treatment assignment among the not-yet treated the Fredriksson and Johansson (2008) estimator is problematic for two other reasons. First, any standard right-censoring that depends on observed covariates introduces additional selective censoring, affecting the survivor functions under both treatment and no treatment if both treated and matched controls are subject to right-censoring. Second, the aggregated estimator in Fredriksson and Johansson (2008) for the effect on the survival rate aggregated over all pretreatment durations introduce additional issues, since the matched controls are pooled over all pretreatment durations without correcting for the fact that the censoring rate due to treatment assignment among the not-yet treated might differ across pre-treatment durations.[2]

6 Monte Carlo simulation

This section examines the finite sample properties of the weighted estimator introduced in this paper, and compares it with related estimators. Data are generated using a logistic model for the hazard rate out of the initial state

$$\Pr(Y_t = 1 | \overline{Y}_{t-1} = 0, X, V_Y) = [1 + \exp(-(3.0 + b_Y X + c_Y V_Y))]^{-1}$$

[2] In their preferred aggregation scheme, Fredriksson and Johansson (2008) pool matched treated and controls over pre-treatment durations, then aggregated hazard rates are calculated, and finally aggregated survival functions are constructed using the aggregated hazard rates. Then, if the treatment rates vary across time periods, the censoring due to treatment assignment among the not-yet treated differs across periods, so that matched controls at s' and s'' are censored at different rates and this introduces additional selective censoring.

and for the hazard rate into treatment

$$\Pr(S=t|S\geq t,X,V_S)=[1+\exp(-(a_S+b_SX+c_SV_S))]^{-1},$$

where X is assumed to be observed by the econometrician, and V_S and V_Y are unobserved. All three are independently uniformly distributed on the interval [-1,1]. Three baseline settings are considered: no heterogeneity ($b_Y=b_S=c_Y=c_S=0$), observed heterogeneity ($b_Y=b_S>0, c_Y=c_S=0$) and a full heterogeneity setting ($b_Y=b_S>0, c_Y=c_S=1$). I set a_S equal to -2.0 or -3.0, i.e. either a high or a low treatment rate.[3] Samples of size 10,000 are generated, and the number of replications is 10,000. The propensity scores in the weighted estimator are estimated with a correct logistic model specification. The standard errors are calculated using bootstrap (99 replications). I consider $\text{ATET}_t(1)$, i.e. the effect of treatment in the first period, but very similar results are obtained for other enrollment times. Common support is imposed using a standard minima and maxima comparison. That is individuals whose propensity score is smaller (larger) than the maximum (minimum) of the minimum (maximum) scores among the treated and the non-treated survivors, respectively. This is imposed for all of the ten first time periods.

Besides the IPW estimator introduced in this paper, the properties of two related estimators are also explored. The first is the FJ estimator discussed in section 5. As in FJ 1-nearest neighbor propensity score matching is applied, in which the scores are estimated using logistic regression models. The second related estimator is the blocking estimator proposed by Crépon et al. (2009) (CFJV henceforth). In this approach propensity scores are estimated for each time to treatment, s. The scores could be estimated in several different ways. CFJV uses a proportional hazard model with a piecewise constant baseline function and unobserved heterogeneity. Another way, utilized here, is to estimate separate logit models for each s, using only the survivors at each s. Note that data are generated using logistic models, so this implies using the correct functional form. Then, the treated at s and the non-treated at s are divided into blocks based on the predicted scores. For

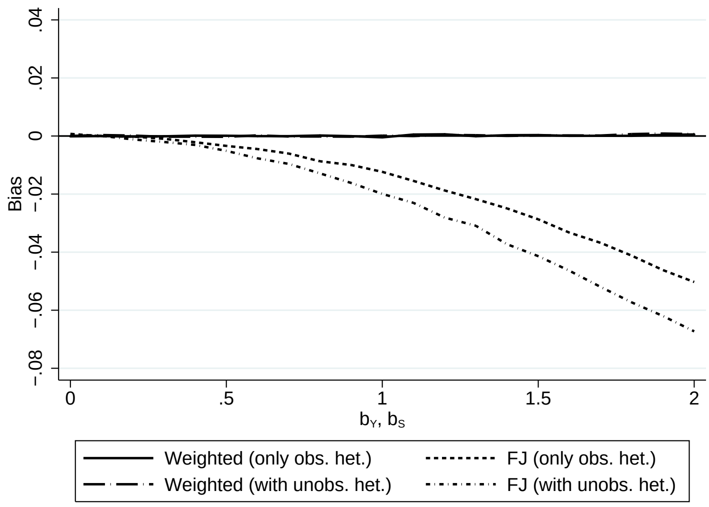
[3] In the case with $b_Y=b_S=c_Y=c_S=1$ this implies that about 14% respectively 6% start treatment in each period.

block b, the average effect on the survival function is

$$\prod_{k=s}^{t}\left[1-\frac{\sum_{i\in J_b} Y_{k,i}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i=s)}{\sum_{i\in J_b}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i=s)}\right]-\prod_{k=s}^{t}\left[1-\frac{\sum_{i\in J_b} Y_{k,i}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i>k)}{\sum_{i\in J_b}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i>k)}\right],$$

where J_b denotes the set of indices for all individuals in block b. Note that the hazard rate in each period is the fraction leaving the initial state among the treatment group and the control group, of which the latter only consists of not-yet treated individuals. The overall average effect on the treated is obtained by averaging over all blocks using the distribution of the score function in the treatment group. 10 blocks are used and the standard errors are obtained using bootstrap (99 replications). In both the FJ and the CFJV approach common support is imposed using the same minima and maxima comparison as for the IPW estimator. The only difference is that common support is only imposed in the first period, since the subsequent scores play no role in these two approaches.

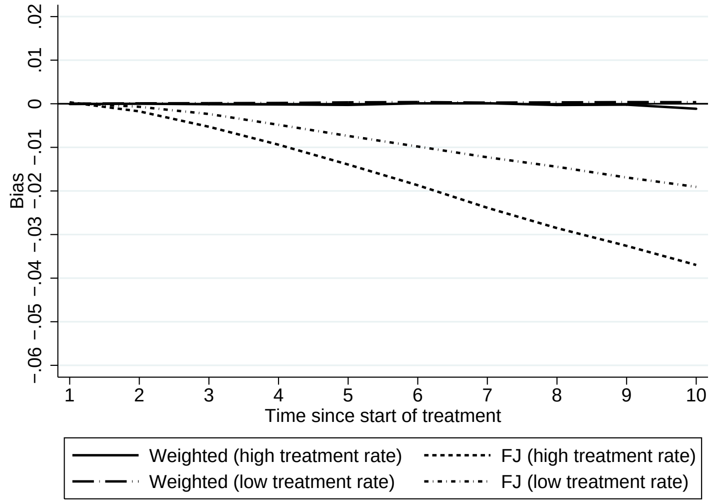
Figure 1: Bias results for the IPW estimator and the FJ estimator. By observed and unobserved covariate impact



Note: Bias for $\text{ATET}_{10}(1)$, i.e. the average effect of treatment in the first period on survival 10 periods. b_Y and b_S measures the impact of the observed covariate in the exit rate and treatment rate equation, respectively. Data generating processes for the logistic simulation models described in Section 5. IPW estimator for the average effect on the treated and FJ is the Fredriksson and Johansson (2008) estimator. Samples of sizes 10,000 and results are based on 10,000 replications.

Initially, *Figure 1* compares the bias of the weighted estimator and the FJ estimator for a selection of values of b_Y and b_S, with and without unobserved heterogeneity. Specifically the bias is for survival ten periods ($\text{ATET}_{10}(1)$), and note that low (high) values of

Figure 2: Bias results for the IPW estimator and the FJ estimator. By time since the start of the treatment and treatment rate



Note: Bias for $\text{ATET}_t(1)$, where t is time since the start of the treatment. Data generating processes for the logistic simulation models described in Section 5. IPW estimator for the average effect on the treated and FJ is the Fredriksson and Johansson (2008) estimator. Samples of sizes 10,000 and results are based on 10,000 replications.

b_Y, b_S correspond to limited (extensive) observed heterogeneity. The results in the figure show that the bias of the weighted estimator is small in all cases. In the no heterogeneity setting (with $b_Y = b_S = 0$) the bias of the FJ estimator is small, but with observed heterogeneity in the model the FJ estimator is, as theoretically expected, biased. The bias is increasing in b_Y,b_S, so that more pronounced observed heterogeneity leads to larger bias. The figure also shows that the bias of the FJ estimator is reinforced by uncorrelated unobserved heterogeneity. Next, *Figure 2* shows how the bias varies by time since the start of the treatment for low and high treatment rate in a setting with $b_Y = b_S = c_Y = c_S = 1$. It is directly apparent that the bias of the FJ estimator increases with time since the start of the treatment and with the treatment rate. Naturally, with a higher treatment rate the selective drop-out in each period among the not-yet treated is larger, and the total selective drop-out increases with time since the start of the treatment. This confirms the theoretical conclusions in section 5.

All these properties are confirmed by the full simulation results reported in *Table 1*. Results in this table for tests with a nominal size of 5% show that the IPW estimator also has the correct size. *Table 1* also shows that for our three baseline settings the bias of the Crépon et al. (2009) blocking estimator is small and of correct size. Besides the three

Table 1: Monte Carlo simulation. Comparison between IPW and related estimators

t	IPW			FJ			CFJV		
	size [1]	bias [2]	se [3]	size [4]	bias [5]	se [6]	size [7]	bias [8]	se [9]
No heterogeneity									
1	.058	-.0063	.0093	.062	.012	.013	.06	-.0095	.0094
2	.056	.01	.013	.06	.0067	.018	.056	-.03	.013
3	.056	.0024	.015	.061	-.0063	.022	.052	-.013	.015
4	.054	-.0049	.017	.064	.017	.025	.052	-.016	.017
5	.053	.0068	.018	.065	-.02	.027	.051	-.011	.018
6	.052	.0037	.019	.066	-.035	.03	.05	-.0082	.019
8	.05	-.01	.021	.068	-.041	.033	.051	-.0018	.021
10	.05	-.014	.022	.068	-.026	.037	.05	.019	.022
Observed heterogeneity									
1	.054	-.013	.0096	.067	.033	.013	.057	.006	.0097
2	.05	-.014	.013	.062	.0066	.019	.054	-.00016	.013
3	.052	-.011	.016	.065	-.073	.022	.05	-.0049	.016
4	.054	-.0012	.017	.064	-.2	.025	.053	-.016	.018
5	.051	.014	.019	.067	-.37	.028	.056	-.015	.019
6	.052	-.00065	.02	.069	-.5	.03	.053	-.0099	.02
8	.052	.015	.022	.079	-.85	.034	.053	.014	.022
10	.052	.011	.023	.09	-1.2	.037	.052	.0081	.023
Unobserved and observed heterogeneity									
1	.055	.0062	.01	.073	-.0017	.014	.057	-.015	.01
2	.052	-.00094	.014	.071	-.11	.019	.051	-.12	.014
3	.05	-.0088	.016	.074	-.28	.022	.054	-.3	.016
4	.054	-.01	.017	.081	-.49	.025	.057	-.51	.017
5	.053	.0078	.018	.087	-.69	.027	.07	-.74	.018
6	.054	.011	.019	.093	-.93	.029	.082	-1	.019
8	.054	.034	.02	.1	-1.4	.032	.11	-1.5	.02
10	.051	.037	.021	.12	-1.7	.036	.15	-1.9	.021
Time-varying selection effect of covariates									
1	.051	.017	.01	.072	-.0063	.014	.053	-.0023	.01
2	.053	.028	.014	.071	-.097	.019	.054	-.12	.014
3	.052	.044	.016	.078	-.27	.023	.054	-.3	.016
4	.051	.046	.017	.079	-.46	.025	.063	-.5	.017
5	.054	.053	.018	.088	-.71	.027	.071	-.73	.018
6	.053	.033	.019	.095	-.96	.029	.082	-.97	.019
8	.054	.014	.02	.11	-1.4	.032	.11	-1.4	.02
10	.053	.011	.021	.12	-1.8	.035	.14	-1.8	.021

Note: Data generating processes for the logistics simulation models described in Section 6. IPW estimates with bootstraped standard errors (99 replications). FJ is the Fredriksson and Johansson (2008) matching estimator implemented using 1-nearest neighbor propensity score matching. CFJV is the Crepon et al. 2009) blocking estimator applied using 10 blocks and bootstraped standard errors (99 replications). Bias has been multiplied by 100. Size is for 5% level tests. The results are based on 10,000 replications.

baseline models, a model with two covariates (both uniform on [-1,1]) with time-varying selection is considered:

$$\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, X_1, X_2, V_Y) = [1+\exp(-(3.0+X_1+X_2+V_Y))]^{-1}$$
$$\Pr(S = t|S \geq t, X_1, X_2, V_S) = [1+\exp(-(2.0+X_1+V_S))]^{-1}, \quad t = 1$$
$$\Pr(S = t|S \geq t, X_1, X_2, V_S) = [1+\exp(-(2.0+X_2+V_S))]^{-1}, \quad t > 1,$$

i.e. one covariate only affects selection into treatment in the first period and the other covariate affects selection in all subsequent periods. The results from this model, reported in the fourth panel of *Table 1*, show that the bias of the weighted estimator is small, but that both the FJ estimator and the blocking estimator are severely biased. The intuition behind the results for the blocking estimator is that the blocking is only based on the selection mechanism at $t = 1$ and not on the subsequent selection into treatment, and this leads to bias in the current setting with a time-varying impact of the covariates.

7 Application

This section illustrates the estimators discussed in this paper using data from a Swedish work practice program. Participants are enrolled for a maximum of six months, but often for a shorter period, and the work practice could take place at both private and public firms. The aim of the program is to provide the unemployed with practical experience in a certain profession, and thereby maintain and/or strengthen their professional competence. The program could start at any point in time and the main outcome of interest is time in unemployment, offering a setting in which the IPW estimators proposed in this paper are applicable.

Register data from the Swedish employment offices are used in the analysis. The data cover all registered unemployed persons and contain day-by-day information on the unemployment status, entries into and exits from active labor market programs, as well as the reason for the unemployment spell to end (as a rule, this is re-employment, but some times it is a transition into education or other insurance schemes). I sample all unemployment spells that start between January 1, 2003 and December 31, 2006. The daily data are aggregated into 30 days intervals. Concerning re-employment, it is required employment (full-time, part-time or subsidized) to be retained for at least 30 days. Unemployment spells that terminate for other reasons than re-employment are considered to be right-censored durations until re-employment, and this censoring is assumed to be random conditional on observed covariates. The spells are also right-censored if the unemployed participates in any another program before entering the work practice program. The analysis is restricted to everyone in ages 25-55 at the time of entry into unemployment.

The employment office data also include a number of personal characteristics recorded at the beginning of the unemployment spell and information on UI eligibility. They are also used to construct information on previous employment history (e.g. the number of previous unemployment spells). Yearly population register data are also used to construct information on previous labor income and income from various insurance schemes. *Ta-*

Table 2: Sample statistics for work practice participants and non-treated

	Untreated	Work practice after					
		All	4 months	5 months	6 months	7 months	8 months
# observations	321146	14432	2109	1887	1535	1304	1348
Mean survival time	9.2	21.4	12.9	14.1	15.5	17.2	18.9
Censored (%)	45.6	22.1	20.2	20.5	18.4	17.8	19.1
Female (%)	51.6	52.2	54.8	55.8	53.0	53.0	54.7
Age	36.7	37.0	35.8	36.0	36.3	36.3	36.7
High school education (%)	43.4	43.1	45.4	46.5	45.1	43.8	41.2
University education (%)	35.9	36.8	36.0	35.9	36.7	37.8	40.5
Foreign born (%)	27.9	32.3	31.2	29.5	27.0	26.8	29.0
Children in household (%)	48.3	50.6	51.4	50.1	51.0	50.8	51.3
Married (%)	35.5	39.0	37.0	36.8	36.1	36.9	38.9
Stockholm MSA (%)	21.3	11.7	10.4	10.7	9.6	10.1	9.3
Gothenburg MSA (%)	17.4	12.0	11.1	9.6	12.6	10.3	10.5
Skane MSA (%)	14.5	13.6	12.6	12.7	11.3	11.9	13.5
North (%)	13.4	19.9	26.6	23.9	23.3	21.6	21.4
South (%)	11.1	12.7	13.2	13.4	12.9	12.0	14.3
UI eligiable (%)	85.4	87.7	86.1	86.9	89.7	89.0	88.9
Unemployment record (days)	259.5	273.8	279.4	276.2	277.9	272.1	275.6
UI year -1 (%)	32.6	29.1	32.9	33.6	33.3	29.8	31.2
UI year -2 (%)	31.4	28.6	29.7	31.4	31.9	29.8	29.5
Labor income year -1	118792	111355	102904	96960	104844	115708	110078
Labor income year -2	117655	109175	105266	98305	104150	109498	107259
Social benefits -1 (%)	12.8	12.9	12.9	12.2	11.9	12.0	11.1
Social benefits -2 (%)	11.9	11.6	11.2	11.1	10.2	9.8	9.7

Note: Time periods is in months. All variables recorded at the start of the unemployment spell. Previous unemployment is in days of unemployment during 5 years before the start of the unemployment spell. Labor income in SEK. Unemployment insurance benefits and social benefits indicators for non-zero insurance earnings.

ble 2 provides sample statistics on a subset of the covariates used in the analysis.[4] The statistics show that there are more males, university educated, foreign born, married and parents, and less big city residents (Stockholm and Gothenburg) among the participants. The treated further have more extensive unemployment record and slightly lower previous labor income. The table also reports statistics by enrollment time, showing that females, non-university educated, Swedish born and big city residents are overrepresented among participants enrolling relatively late in work practice.

In order to apply the weighted estimator, two main assumptions have to be fulfilled;

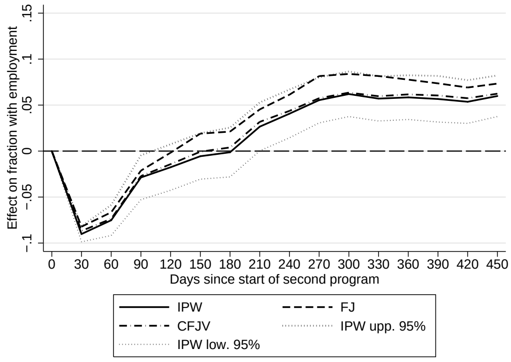
[4]The exact covariates used in the analysis are gender, age, age squared, indicator for at least one child in the household, marital status, country of origin (3 categories), level of education (5 categories), region of residence (22 regions), inflow year dummies, indicator for UI entitlement, number of unemployment days in the last 5 years, and labor income, social assistance and unemployment insurance benefits one and two years before start of the unemployment spell.

unconfoundedness and no-anticipation. No-anticipation means that the unemployed should not be able to predict the exact timing of their enrollment in work practice. Several recent studies find that unemployed react to information about future treatments (see e.g. Black et al., 2003; Crépon et al. 2013). However, in most cases the individual is informed about the work practice program shortly before the start of the program, and this prohibits any substantial anticipation effects. There are also several unpredictable events leading to program enrollment. For instance, case workers have large influence over enrollment decisions (see e.g. Eriksson, 1997; Carling and Richardson, 2001) and substantial discretionary power in their daily work, and this makes it difficult for the unemployed to predict future treatments. The analysis includes a large number of background characteristics, including extensive controls for previous income and unemployment. Because of the large set of covariates and the substantial case worker discretion, the unconfoundedness assumption should also be fulfilled.

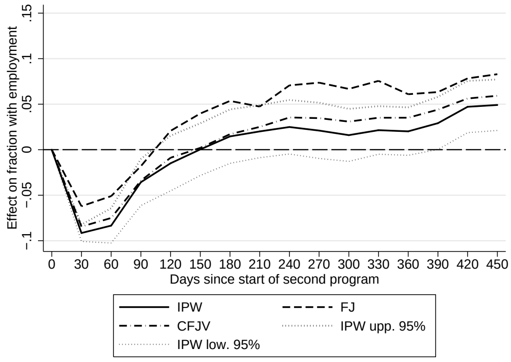
Concerning estimation details, the treatment propensity scores and censoring probabilities are estimated using logistic regression models, and standard errors are obtained using bootstrap (99 replications). For each pre-treatment duration common support is imposed over the 15 months after the start of the treatment using the maxima and minima comparison described in Section 6. On average this excludes about 10% of the treated and 15% of the non-treated at each pre-treatment duration. The IPW estimator is also compared with the two-step matching estimator in FJ and the blocking estimator in CFJV. The FJ estimator is implemented using 1-nearest neighbor propensity score matching. The CFJV estimator is applied using 20 blocks and standard errors are obtained using bootstrap (99 replications). Common support is imposed in both cases.

Initially, consider the results for $\text{ATET}_t(s)$ in *Table 3* and illustrated for a selection of enrollment time in *Figure 3*. Note that the results are for the effect on the fraction re-employed instead of the effect on the survival rate. In all cases, there are substantial locking-in effects with lower employment rates during the first months after enrollment. In the first month after assignment the reemployment is about 10 percentage points lower in the treatment group. After this period participants catch up and after about 4-6 months the re-employment rate among the treated is the same as for the controls. 15 months

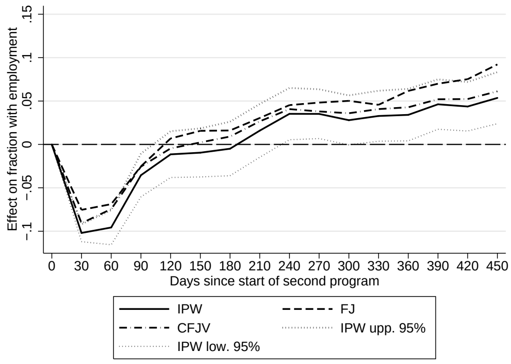
Figure 3: Effect of work practice on fraction reemployed. By pre-treatment duration



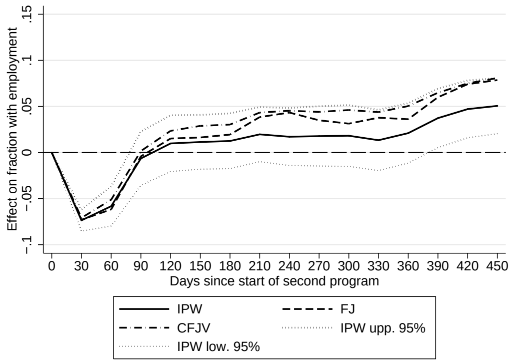
(a) Treatment after 5 months



(b) Treatment after 6 months



(c) Treatment after 7 months



(d) Treatment after 8 months

Note: IPW estimates with bootstraped standard errors (99 replications). FJ is the Fredriksson and Johansson (2008) matching estimator and CFJV the Crépon et al. (2009) blocking estimator.

after enrollment the employment rate on average is about 4-6 percentage points higher among the participants. The size of the effect varies with enrollment time, but the more extensive results in *Table 3* indicate no clear pattern that early enrollment is better than late enrollment.[5]

Table 3 and *Figure 3* also present the results using the FJ and CFJV estimators. First, consider treatment after 6 months. Shortly after the start of the treatment, the difference between all estimators are small. For follow-up times beyond 4-5 months the difference between the IPW and the FJ estimator is large and in some case even significant. This

[5] Vikström et al. 2013 also studies the effects of the work practice program, but focuses on effects sequences of work practice episodes. In a related paper, Forslund et al. (2013) using the Fredriksson and Johansson (2008) estimator, finds that work practice leads to an increased employment rate and conclude that the time period studied in this paper seems to be associated with particularly positive effects on the re-employment rate.

Table 3: Effects of work practice on fraction reemployed by time to program start. Comparison between IPW and related estimators

	4 months			5 months			6 months		
Time	IPW	FJ	CFJV	IPW	FJ	CFJV	IPW	FJ	CFJV
1	-.0903	-.0836	-.088	-.0902	-.0822	-.0868	-.0915	-.0621	-.0845
	(.00388)	(.0079)	(.00413)	(.00435)	(.00839)	(.00461)	(.00474)	(.00867)	(.00432)
2	-.0729	-.0731	-.0711	-.0752	-.0666	-.0737	-.0834	-.051	-.075
	(.00843)	(.0122)	(.00766)	(.00846)	(.0126)	(.00843)	(.00975)	(.0139)	(.00882)
3	-.0292	-.0265	-.0268	-.0287	-.0213	-.0272	-.0357	-.0181	-.034
	(.0118)	(.0145)	(.0109)	(.0123)	(.0153)	(.0114)	(.0131)	(.0166)	(.0107)
4	-.0293	-.0232	-.0241	-.0176	-.00174	-.0142	-.0149	.0207	-.00884
	(.0122)	(.0157)	(.0113)	(.0128)	(.0164)	(.0124)	(.0155)	(.0178)	(.0123)
5	-.0167	-.0212	-.00861	-.00559	.019	-.00066	.000322	.0397	.00203
	(.0124)	(.0164)	(.0118)	(.0127)	(.0171)	(.013)	(.0146)	(.0186)	(.0115)
7	.0127	.0161	.0233	.0266	.0453	.0318	.0201	.0474	.025
	(.013)	(.0169)	(.0103)	(.0134)	(.0179)	(.0115)	(.0147)	(.0197)	(.0111)
10	.0289	.0537	.0451	.062	.0839	.0635	.016	.0668	.0308
	(.0127)	(.017)	(.00954)	(.0125)	(.0182)	(.011)	(.0147)	(.0202)	(.0119)
12	.0288	.0404	.0438	.0584	.0777	.0616	.0202	.0609	.0351
	(.0114)	(.0171)	(.00983)	(.0123)	(.0183)	(.0107)	(.0134)	(.0205)	(.0122)
15	.0486	.0642	.0575	.0599	.0734	.0624	.0491	.0829	.0592
	(.0102)	(.0171)	(.00934)	(.0114)	(.0184)	(.0113)	(.0143)	(.0206)	(.0125)

	7 months			8 months			9 months		
Time	IPW	FJ	CFJV	IPW	FJ	CFJV	IPW	FJ	CFJV
1	-.102	-.0754	-.0908	-.0736	-.0729	-.0709	-.0682	-.0705	-.0666
	(.00512)	(.00995)	(.00482)	(.00598)	(.00963)	(.00499)	(.00639)	(.0105)	(.00505)
2	-.0957	-.0688	-.0745	-.0582	-.0617	-.0511	-.0622	-.0491	-.0573
	(.0102)	(.015)	(.0104)	(.011)	(.0143)	(.0105)	(.0123)	(.0156)	(.00987)
3	-.0355	-.0253	-.0256	-.00642	-.00411	.00163	-.0163	-.0073	-.0217
	(.0127)	(.0179)	(.0125)	(.0149)	(.0172)	(.0123)	(.0158)	(.0191)	(.0131)
4	-.0114	.00689	-.00457	.00988	.0152	.0235	-.00731	.00761	-.0107
	(.0136)	(.0193)	(.0124)	(.0155)	(.0187)	(.0126)	(.0182)	(.0211)	(.0145)
5	-.0095	.0157	.00231	.0114	.0163	.0289	-.00937	.016	-.00854
	(.0143)	(.0202)	(.0136)	(.015)	(.0196)	(.0136)	(.0203)	(.0224)	(.0162)
7	.0159	.0305	.0266	.0197	.0384	.0433	-.0177	.0157	-.0065
	(.0157)	(.0214)	(.0148)	(.0151)	(.0206)	(.0148)	(.02)	(.0243)	(.0167)
10	.0279	.0502	.0359	.0182	.0314	.0461	.014	.0536	.0138
	(.0146)	(.0219)	(.0139)	(.017)	(.0215)	(.0149)	(.022)	(.0256)	(.0165)
12	.0341	.0617	.0428	.021	.0361	.0504	.048	.0739	.0518
	(.0153)	(.022)	(.0132)	(.0166)	(.0217)	(.0151)	(.0213)	(.0263)	(.0151)
15	.0536	.0922	.0612	.0506	.0786	.0808	.0528	.0948	.0719
	(.0152)	(.0219)	(.0126)	(.0154)	(.0216)	(.0151)	(.0205)	(.0265)	(.0142)

Note: Time is in number of months since program start. IPW estimates with bootstraped standard errors (99 replications). FJ is the Fredriksson and Johansson (2008) matching estimator implemented using 1-nearest neighbor propensity score matching. CFJV is the Crepon et al. 2009) blocking estimator applied using 20 blocks and bootstraped standard errors (99 replications). Standard errors in parenthesis.

follows the theoretical properties, since the selective drop-out which will bias the FJ estimator, in general, increases with time since the start of the treatment. Second, for other enrollment times (e.g. enrollment after 7 months) the difference between the IPW and the FJ estimator is smaller. The results also indicate some differences between the IPW and the CFJV estimator. For some enrollment times (e.g. 5 months) the difference is small and for other enrollment times (e.g. 8 months) the difference is larger.

8 Conclusions

This paper has re-considered treatment evaluation under unconfoundedness in a dynamic treatment assignment setting in which treatment could start at any point in time. The outcome of interest is survival time and together with the dynamic treatment assignment this introduces well-known methodological issues. Building upon previous results, it has been shown that a range of average effects and average effects on the treated, including the effect of starting treatment in a certain time period compared with never receiving treatment, are identified under unconfoundedness and no-anticipation.

This paper has also introduced IPW estimators for average effects on the treated, in situations with and without standard right-censoring. These estimators include a series of unique weights for each time period and use the not-yet treated in each period to estimate the counterfactual survival rate. The IPW estimator has been compared with some other proposed estimators of effects on survival time in a setting with dynamic treatment assignment. One conclusion is that the two-step matching estimator suggested by Fredriksson and Johansson (2008) ignores a selective censoring problem that confounds the analysis unless treatment assignment does not depend on the observed characteristics. The conclusion is based on both simulations and theoretical results for the asymptotic properties of the estimator.

An analysis of data from a Swedish ALMP program illustrates the usefulness of the identification results and the IPW estimators. Since the ALMP program could start at any elapsed unemployment duration and the outcome of interest is time in unemployment this offers a key application of the estimators introduced in this paper. The result is that participation in the program leads to significantly increased employment rates compared with never taking treatment.

References

Abbring J.H. and G.J. van den Berg (2003), “The non-parametric identification of treatment effects in duration models”, *Econometrica*, 71, 1491–1517.

Abbring J.H. and J.J. Heckman (2008) “Dynamic Policy Analysis”, In Mátyás L. and P. Sevestre (Eds.), *The Econometrics of Panel Data*, Chap. 24, Berlin Heidelberg: Springer Verlag, 796–863.

Black D., J. Smith, M. Berger and B. Noel (2003) “Is the Threat of Reemployment Services More Effective than the Services Themselves? Evidence from Random Assignment in the UI System”, *American Economic Review*, 93:4, 1313–1327.

Biewen M., B. Fitzenberger, A. Osikominu and M. Paul (2013), “The Effectiveness of Public Sponsored Training Revisited: The Importance of Data and Methodological Choices”, forthcoming in Journal of Labor Economics.

Carling K. and K. Richardson, K. (2001), “The Relative Efficiency of Labor Market Programs: Swedish Experience From the 1990”, *Labour Economics*, 26:4, 335–354.

Caliendo M. and S. Kopeinig (2008), “Some Practical Guidance for the Implementation of Propensity Score Matching”, *Journal of Economic Surveys*, 22:1, 31–72.

Crépon, B., M. Ferracci, G. Jolivet and G.J. van den Berg (2009), “Active Labor Market Policy Effects in a Dynamic Setting”, *Journal of the European Economic Association*, 7, 595–605.

Crépon, B., M. Ferracci, G. Jolivet and G.J. van den Berg (2013), “Dynamic Treatment Evaluation Using Data on Information Shocks”, mimeo.

de Luna X. and P. Johansson (2010), “Non-parametric Inference for the Effect of a Treatment on Survival Times with Application in the Health and Social Sciences”, *Journal of Statistical Planning and Inference*, 140, 2122–2137.

Gerfin M. and M. Lechner (2002), “A Microeconometric Evaluation of the Active Labour Market Policy in Switzerland”, *Economic Journal*, 112, 854–893.

Hirano K., G. imbens and G. Ridder (2003), “ Efficient estimation of average treatment effects using the estimated propensity score”, *Econometrica*, 71, 1161–1189.

Eriksson, M. (1997), “To choose or not to choose: Choice and choice set models, Umeå Economic Studies 443, Department of Economics”, Umeåa University.

Fitzenberg B., A. Osikominu and R. Völter (2008), “Get Training or Wait? Long-Run Employment Effects of Training Programs for theUnemployed in West Germany”, *Annales d’Économie et de Statistique,* 91/92, 321-355.

Fredriksson P. and P. Johansson (2008), “Dynamic Treatment Assignment: The Consequences for Evaluations Using Observational Data”, *Journal of Business & Economic Statistics*, 26:4, 435–445.

Forslund A., L. Liljeberg, L. von Trott zu Solz (2013), “Job practice: an evaluation and comparison with vocational labour market training programmes”, IFAU Working Paper 2013:6.

Kaplan E. and P. Meier (1958), “Nonparametric Estimation From Incomplete Observations”, *Journal of American Statistical Association*, 53, 457–481.

Kastoryano S. and B. van der Klauuw (2011), “Dynamic Evaluation of Job Search Assistance”, IZA DP No.5424.

Lechner M.(1999), “Earnings and employment effects of continuous off-the-job training in East Germany after unification”, *Journal of Business and Economic Statistics,* 17, 74-90.

Lechner M. (2008), “Matching Estimation of Dynamic Treatment Models: Some Practical Issues”, In. Millimet D., Smith J. and Vytlacil E. (Eds.), *Advances in Econometrics 21, Modelling and Evaluating Treatment Effects in Econometrics*, Emerald Group Publishing Limited, 289–333.

Lechner M. (2009), “Sequential Causal Models for the Evaluation of Labor Market Programs”, *Journal of Business & Economic Statistics*, 27:1, 71–83.

Lechner M. and R. Miquel (2010), “Identification of the Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions”, *Empirical Economics*, 39, 111–137.

Robins J. (1986), “A New Approach to Causal Inference in Mortality Studies With Sustained Exposure Periods: Application to Control of the Healthy Worker Survivor Effect”, *Mathematical Modelling*, 7, 1293–1512.

Sianesi B. (2004), “An Evaluation of the Swedish System of Active Labour Market Programmes in the 1990s”, *Review of Economics and Statistics*, 86, 133–155.

Sianesi B. (2008), “Differential effects of active labour market programs for the unemployed”, *Labour Economics*, 15(3), 370–399.

Vikström J., P. Johansson and I. Waernbaum (2013), “Identification and Estimation of Causal Effects of Treatment Regimes with Duration Outcomes”, mimeo IFAU Uppsala.

Wooldridge J. M. (2010). *Econometric Analysis of Cross Section and Panel Data (2nd ed.)*, MIT Press.

Appendix

A.1 Properties of $\widehat{\mathrm{ATET}}_t(s)$

For the first part of the estimator

$$p\lim_{N\to\infty}\prod_{k=s}^{t}\left[1-\frac{\sum_{i\in\overline{Y}_{k-1,i}=0,S_i=s}Y_k}{\sum_{i\in\overline{Y}_{k-1,i}=0,S_i=s}1}\right]=p\lim_{N\to\infty}\prod_{k=s}^{t}\left[1-\frac{\sum_i Y_{k,i}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i=s)}{\sum_i \mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i=s)}\right]= \tag{A.1}$$

$$\prod_{k=s}^{t}\left[1-\frac{p\lim_{N\to\infty}\frac{1}{N}\sum_i Y_{k,i}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i=s)}{p\lim_{N\to\infty}\frac{1}{N}\sum_i \mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i=s)}\right]=$$

$$\prod_{k=s}^{t}\left[1-\frac{\Pr(Y_k(s)=1,\overline{Y}_{k-1}(s)=0,S=s)}{\Pr(\overline{Y}_{k-1}(s)=0,S=s)}\right]=\Pr(\overline{Y}_t(s)=0|S=s,\overline{Y}_{s-1}(s)=0).$$

For the second part of the estimator

$$p\lim_{N\to\infty}\prod_{k=s}^{t}\left[1-\frac{\sum_{i\in\overline{Y}_{k-1,i}=0,S_i>k}w_{k,i}(s)Y_k}{\sum_{i\in\overline{Y}_{k-1,i}=0,S_i>k}w_{k,i}(s)}\right]= \tag{A.2}$$

$$p\lim_{N\to\infty}\prod_{k=s}^{t}\left[1-\frac{\sum_i w_k(s)Y_{k,i}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i>s)}{\sum_i w_k(s)\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i>s)}\right]=$$

$$\prod_{k=s}^{t}\left[1-\frac{p\lim_{N\to\infty}\frac{1}{N}\sum_i w_k(s)Y_{k,i}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i>s)}{p\lim_{N\to\infty}\frac{1}{N}\sum_i w_k(s)\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i>s)}\right]=$$

$$\prod_{k=s}^{t}\left[1-\frac{\mathbb{E}[w_k(s)Y_k\mathbf{1}(\overline{Y}_{k-1}=0)\mathbf{1}(S>k)]}{\mathbb{E}[w_k(s)\mathbf{1}(\overline{Y}_{k-1}=0)\mathbf{1}(S>k)]}\right]=\prod_{k=s}^{t}\left[1-\frac{\mathbb{E}_X\mathbb{E}[w_k(s)Y_k\mathbf{1}(\overline{Y}_{k-1}=0)\mathbf{1}(S>k)|X]}{\mathbb{E}_X\mathbb{E}[w_k(s)\mathbf{1}(\overline{Y}_{k-1}=0)\mathbf{1}(S>k)|X]}\right].$$

Next, for presentation reasons, consider $s=1,k=2$:

$$\mathbb{E}[w_k(s)Y_k\mathbf{1}(\overline{Y}_{k-1}(s)=0)\mathbf{1}(S>k)|X]=\mathbb{E}_X[w_2(1)Y_2\mathbf{1}(Y_1=0)\mathbf{1}(S>2)|X]=$$
$$\frac{p_1(X)}{1-p_1(X)}\frac{\mathbb{E}_X[Y_2\mathbf{1}(Y_1=0)\mathbf{1}(S>2)|X]}{1-p_2(X)}\overset{N.A}{=}$$
$$\frac{p_1(X)}{1-p_1(X)}\frac{\mathbb{E}_X[Y_2(0)\mathbf{1}(Y_1(0)=0)\mathbf{1}(S>2)|X]}{1-p_2(X)}$$
$$\frac{p_1(X)\Pr(Y_2(0)=1|X,Y_1(0)=0,S>2)(1-p_2(X))\Pr(Y_1(0)=0|X,S>1)(1-p_1(X))}{(1-p_1(X))(1-p_2(X))}=$$
$$p_1(X)\Pr(Y_2(0)=1|X,Y_1(0)=0,S>2)\Pr(Y_1(0)=0|X,S>1)\overset{U.1}{=}$$
$$p_1(X)\Pr(Y_2(0)=1|X,Y_1(0)=0,S=1)\Pr(Y_1(0)=0|X,S=1)=$$
$$\Pr(Y_2(0)=1,Y_1(0)=0,S=1|X),$$

and by similar reasoning (using N.A. and U.1)

$$\mathbb{E}[w_k(s)Y_k\mathbf{1}(\overline{Y}_{k-1}(s)=0)\mathbf{1}(S>k)|X] = \Pr(Y_k(0)=1,\overline{Y}_{k-1}(0)=0,S=s|X), \tag{A.3}$$

also

$$\mathbb{E}[w_k(s)\mathbf{1}(\overline{Y}_{k-1}(s)=0)\mathbf{1}(S>k)|X] = \Pr(\overline{Y}_{k-1}(0)=0,S=s|X). \tag{A.4}$$

Using (A.2)-(A.4)

$$p\lim_{N\to\infty}\prod_{k=s}^{t}\left[1-\frac{\sum_i w_k(s)Y_{k,i}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i>s)}{\sum_i w_k(s)\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i>s)}\right]= \tag{A.5}$$

$$\prod_{k=s}^{t}\left[1-\frac{\mathbb{E}_X\Pr(Y_k(0)=1,\overline{Y}_{k-1}(0)=0,S=s|X)}{\mathbb{E}_X\Pr(\overline{Y}_{k-1}(0)=0,S=s|X)}\right]=$$

$$\prod_{k=s}^{t}\left[1-\frac{\Pr(Y_k(0)=1,\overline{Y}_{k-1}(0)=0,S=s)}{\Pr(\overline{Y}_{k-1}(0)=0,S=s)}\right]=\Pr(\overline{Y}_t(0)=0|S=s,\overline{Y}_{s-1}(s)=0).$$

Then, using (A.1) and (A.5) $p\lim_{N\to\infty}\widehat{\mathrm{ATET}_t}(s)=\mathrm{ATET}_t(s)$.

A.2 Estimator of $\mathrm{ATET}_t(s^i,s^j)$

$$\widehat{\mathrm{ATET}}_t(s',s'')=\prod_{k=s'}^{t}\left[1-\frac{\sum_i Y_{k,i}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i=s')}{\sum_i\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i=s')}\right]-$$

$$\prod_{k=s''}^{t}\left[1-\frac{\sum_i w_{k,i}(s',s'')Y_{k,i}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i=s'')}{\sum_i w_{k,i}(s',s'')\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i=s'')}\right]*$$

$$\prod_{k=s'}^{s''-1}\left[1-\frac{\sum_i w_{k,i}(s',s'')Y_{k,i}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i>k)}{\sum_i w_{k,i}(s',s'')\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i>k)}\right]$$

with

$$w_{k,i}(s',s'')=\frac{p_{s'}(X_i)}{1-p_{s'}(X_i)}\frac{1}{\prod_{m=s'+1}^{s''-1}1-p_m(X_i)}\frac{1}{p_{s''}(X_i)}.$$

A.3 Estimator of $\mathrm{ATET}_t(s)$ with right censoring

$$\widehat{\mathrm{ATET}}_t(s)=\prod_{k=s}^{t}\left[1-\frac{\sum_i w^C_{k(s),i}Y_{k,i}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i=s)\mathbf{1}(C_i>s)}{\sum_i w^C_{k(s),i}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i=s)\mathbf{1}(C_i>s)}\right]$$

$$-\prod_{k=s}^{t}\left[1-\frac{\sum_i w^C_{k(0),i}(s)Y_k\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i>k)\mathbf{1}(C_i>k)}{\sum_i w^C_{k(0),i}(s)\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(S_i>k)\mathbf{1}(C_i>k)}\right]$$

with

$$w^C_{k(s),i}(s) = \frac{1}{\prod_{m=s+1}^{k}[1-c_m(X_i)]}$$

and

$$w^C_{k(0),i}(s) = \frac{p_s(X_i)}{1-p_s(X_i)} \frac{1}{\prod_{m=s+1}^{k}[1-p_m(X_i)][1-c_m(X_i)]}$$

and

$$c_t(X_i) = \Pr(C_i = t|X_i, S_i \geq t, \overline{Y}_{t-1,i} = 0).$$

A.4 Properties of $\widehat{\text{ATET}}_t^{FJ}(s)$

In the remainder of the appendix the following notation is used

$$\begin{aligned} h_t^0(0,X) &= \Pr(Y_t(0) = 1|X, \overline{Y}_{t-1}(0) = 0, S > t). \\ h_t^0(s,X) &= \Pr(Y_t(0) = 1|X, \overline{Y}_{t-1}(0) = 0, S = s). \end{aligned}$$

Define $N_s = \sum_{i \in S_i = s, \overline{Y}_{s-1,i}=0} 1$. First, from section A.2

$$p \lim_{N_s \to \infty} \prod_{k=s}^{t} \left[1 - \frac{\sum_i Y_{k,i} \mathbf{1}(\overline{Y}_{k-1,i} = 0)\mathbf{1}(S_i = s)}{\sum_i \mathbf{1}(\overline{Y}_{k-1,i} = 0)\mathbf{1}(S_i = s)}\right] = \Pr(\overline{Y}_t(s) = 0|S = s, \overline{Y}_{s-1}(s) = 0). \tag{A.6}$$

Second,

$$p \lim_{N_s \to \infty} \prod_{k=s}^{t} \left[1 - \frac{\sum_{i \in S_i = s, \overline{Y}_{s-1,i}=0} Y_{k,j(i)} \mathbf{1}(\overline{Y}_{k-1,j(i)} = 0)\mathbf{1}(S_{j(i)} > k)}{\sum_{i \in S_i = s, \overline{Y}_{s-1,i}=0} \mathbf{1}(\overline{Y}_{k-1,j(i)} = 0)\mathbf{1}(S_{j(i)} > k)}\right] =$$

$$\prod_{k=s}^{t} \left[1 - \frac{p \lim_{N \to \infty} \frac{1}{N_s} \sum_{i \in S_i = s, \overline{Y}_{s-1,i}=0} Y_{k,j(i)} \mathbf{1}(\overline{Y}_{k-1,j(i)} = 0)\mathbf{1}(S_{j(i)} > k)}{p \lim_{N \to \infty} \frac{1}{N_s} \sum_{i \in S_i = s, \overline{Y}_{s-1,i}=0} \mathbf{1}(\overline{Y}_{k-1,j(i)} = 0)\mathbf{1}(S_{j(i)} > k)}\right],$$

If X is finite and since the controls are selected from untreated survivors at s we can express this as

$$\prod_{k=s}^{t} \left[1 - \frac{p \lim_{N \to \infty} \frac{1}{N_s} \sum_{i \in S_i = s, \overline{Y}_{s-1,i}=0} Y_{k,j(i)} \mathbf{1}(\overline{Y}_{k-1,j(i)} = 0)\mathbf{1}(S_{j(i)} > k)}{p \lim_{N \to \infty} \frac{1}{N_s} \sum_{i \in S_i = s, \overline{Y}_{s-1,i}=0} \mathbf{1}(\overline{Y}_{k-1,j(i)} = 0)\mathbf{1}(S_{j(i)} > k)}\right] =$$

$$\prod_{k=s}^{t} \left[1 - \frac{\mathbb{E}_{X|S=s,\overline{Y}_{s-1}=0} \mathbb{E}[Y_k \mathbf{1}(\overline{Y}_{k-1} = 0)\mathbf{1}(S > k)|X, S > s, \overline{Y}_{s-1} = 0]}{\mathbb{E}_{X|S=s,\overline{Y}_{s-1}=0} \mathbb{E}[\mathbf{1}(\overline{Y}_{k-1} = 0)\mathbf{1}(S > k)|X, S > s, \overline{Y}_{s-1} = 0]}\right].$$

Next, under assumptions (N.A.) and (U.1)

$$
\begin{aligned}
&\mathbb{E}[Y_k \mathbf{1}(\overline{Y}_{k-1}=0)\mathbf{1}(S>k)|X,S>s,\overline{Y}_{s-1}=0] = && \text{(A.7)}\\
\text{(N.A.)}\quad &\mathbb{E}[Y_k(0)\mathbf{1}(\overline{Y}_{k-1}(0)=0)\mathbf{1}(S>k)|X,S>s,\overline{Y}_{s-1}(0)=0] = \\
&h_k^0(0,X)[1-p_k(X)]\prod_{m=s}^{k-1}[1-h_m^0(0,X)][1-p_m(X)] = \\
\text{(U.1)}\quad &h_k^s(0,X)[1-p_k(X)]\prod_{m=s}^{k-1}[1-h_m^s(0,X)][1-p_m(X)] = \\
&\Pr(Y_k(0)=1,\overline{Y}_{k-1}(0)=0|X,S=s,\overline{Y}_{s-1}=0)\prod_{m=s+1}^{k}1-p_m(X).
\end{aligned}
$$

By similar reasoning

$$
\mathbb{E}[\mathbf{1}(\overline{Y}_{k-1}=0)\mathbf{1}(S>k)|X,S>s,\overline{Y}_{s-1}=0] = \Pr(Y_{k-1}(0)=0|X,S=s,\overline{Y}_{s-1}=0)\prod_{m=s+1}^{k}1-p_m(X). \tag{A.8}
$$

Third, taking $p\lim$ of (7) and using (A.6) as well as (A.7)-(A.8)

$$
p\lim_{N\to\infty}\widehat{\text{ATET}}_t^{FJ}(s) = \Pr(\overline{Y}_t(s)=0|S=s,\overline{Y}_{s-1}(s)=0)-
$$

$$
\prod_{k=s}^{t}\left[1-\frac{\mathbb{E}_{X|S=s,\overline{Y}_{s-1}=0}\Pr(Y_k(0)=1,\overline{Y}_{k-1}(0)=0|X,S=s,\overline{Y}_{s-1}=0)\prod_{m=s+1}^{k}1-p_m(X)}{\mathbb{E}_{X|S=s,\overline{Y}_{s-1}=0}\Pr(Y_{k-1}(0)=0|X,S=s,\overline{Y}_{s-1}=0)\prod_{m=s+1}^{k}1-p_m(X)}\right].
$$

Publication series published by IFAU – latest issues

Rapporter/Reports

2013:20 Hensvik Lena and Oskar Nordström Skans ”Kontakter och ungdomars arbetsmarknadsinträde”

2013:21 Dahlberg Matz, Eva Mörk and Katarina Thorén ”Jobbtorg Stockholm – resultat från en enkätundersökning”

2013:22 Sibbmark Kristina ”Arbetsmarknadspolitisk översikt 2012”

2013:23 Hedlin Maria and Magnus Åberg ”Vara med i gänget?” – Yrkessocialisation och genus i två gymnasieprogram”

2013:24 Alam Moudud, Kenneth Carling and Ola Nääs ”Har kommunala sommarjobb under gymnasieåren en positiv effekt på arbetskarriären senare i livet?”

2013:25 Lundin Andreas and Tomas Hemmingsson ”Prediktorer för arbetslöshet och förtidspension”

2013:26 Egebark Johan and Niklas Kaunitz ”Sänkta arbetsgivaravgifter för unga”

2014:1 Assadi Anahita ”En profilfråga: Hur använder arbetsförmedlare bedömningsstödet?”

Working papers

2013:20 Wondratschek Verena, Karin Edmark and Markus Frölich ”The short- and long-term effects of school choice on student outcomes – evidence from a school choice reform in Sweden”

2013:21 Edmark Karin and Roger Gordon ”Taxes and the choice of organizational form by entrepreneurs in Sweden”

2013:22 Golsteyn Bart H.H., Hans Grönqvist and Lena Lindahl ”Time preferences and lifetime outcomes”

2013:23 Hensvik Lena and Oskar Nordström Skans “Networks and youth labor market entry”

2013:24 Alam Moudud, Kenneth Carling and Ola Nääs “The effect of summer jobs on post-schooling incomes”

2013:25 Lundin Andreas and Tomas Hemmingsson “Adolescent predictors of unemployment and disability pension across the life course – a longitudinal study of selection in 49 321 Swedish men”

2013:26 van den Berg Gerard J., Arne Uhlendorff and Joachim Wolff “Sanctions for young welfare recipients”

2013:27 Egebark Johan and Niklas Kaunitz “Do payroll tax cuts raise youth employment?”

2014:1 Vikström Johan “IPW estimation and related estimators for evaluation of active labor market policies in a dynamic setting”

Dissertation series

2013:1 Vikman Ulrika “Benefits or work? Social programs and labor supply”