Identification and Estimation of Causal Effects of Treatment Regimes with Duration Outcomes*

Johan Vikström†, Per Johansson‡, Ingeborg Waernbaum§

September 25, 2012

Abstract

This paper considers identification and estimation of average treatment effects on discrete survival time under unconfoundedness in a setting where the event under study and the treatments can occur at any point in time. In this setting we propose conditions under which the survival time under a specific treatment regime can be identified. Also, we introduce a novel weighting estimator, that is shown to be asymptotically unbiased. The estimator re-weights the outcomes of all that follow the treatment regime of interest until they diverge from this treatment regime.

Keywords: Treatment effects; dynamic treatment assignment; dynamic selection; program evaluation; matching
JEL classification: C14, C4

*We are grateful for helpful suggestions from seminar participants at IFAU Uppsala. Vikström acknowledges financial support from the Jan Wallander and Tom Hedelius Foundation.

†IFAU Uppsala and UCLS, johan.vikstrom@ifau.uu.se

‡IFAU Uppsala, Department of Economics, Uppsala University and IZA, per.johansson@ifau.uu.se

§Department of Statistics, Umeå University and IFAU Uppsala, ingeborg.waernbaum@stat.umu.se

1 Introduction

We consider identification and estimation of the effect of a treatment regime, i.e. a sequence of treatments given in an initial state on the time in the very same state under unconfoundedness. Exits out of the initial state and treatments are allowed to occur at any point in time. This set-up typically arises in evaluations of active labor market policy (ALMP) programs were e.g. training often is assigned at any elapsed unemployment duration and the outcome is time in unemployment. Another important example is medical treatments given after some waiting time and were the outcome is survival time.

The setting under study introduces problems due to both dynamic treatment assignment and dynamic selection. The former due to the fact that treatment is not offered once and the latter since the individuals eventually leave the initial state. The dynamic treatment assignment implies that we no longer can rely on methods from the static evaluation literature (see e.g. Heckman et al., 1999 and Imbens, 2004 for overviews). The dynamic selection means that we are not able to condition on the entire treatment path as in e.g. Lechner and Miquel (2010), since the actual start of the treatment is unobserved if the individual leaves the initial state before receiving treatment.[1]

Aiming to address the problems associated with analyzing survival time when the time to treatment is not fixed Fredriksson and Johansson (2008) introduced a matching estimator that for a given starting time t of the treatment contrasts the outcomes for the treated at t with the outcomes for the not-yet treated at t. The key insight is that the outcomes of the not-yet treated are censored when they become treated. The identification is based on a single unconfoundedness assumption. This estimator is also discussed in de Luna and Johansson (2010). A similar approach is introduced by Crepon et al. (2009), which instead states the identifying assumptions in the form of an unconfoundedness and a separate no-anticipation assumption. The latter is done by relating the results to Abbring and van den Berg (2003)'s Timing-of-Events approach. Another approach is to consider the effect of treatment now versus waiting for treatment as in Sianesi (2004).

In this paper we contribute to the literature in a number of ways. We consider a general setting were treatments are allowed to start and stop at any point in time. We focus on average effects of a main treatment regime on the survival rate throughout a specific interval compared to the same survival rate under a reference treatment. The setting with a single time to treatment considered in Fredriksson and Johansson (2008) and Crepon et al. (2009) is a special case of this setting. For our general setting we explicitly show that for unanticipated treatments the average effect on the survival rate some time after the start of treatment is identified under sequential unconfoundedness among survivors. This essentially is a longitudinal version of the regular static unconfoundedness assumption. This also clarifies identification for the setting considered in Fredriksson and Johansson (2008) and Crepon et al. (2009).

Our main contribution, however, concerns estimation. We introduce a weighted Kaplan-Meier (1958) type of estimator, that is shown to be asymptotically unbiased. It allows

[1] In the statistical literature, Robins and colleagues consider a similar setting, see e.g. Robins et al. (2000) and Hernan et al. (2001).

estimation of a range of interesting average effects in a dynamic setting. The estimator re-weights the outcomes of all that follow the treatment regime of interest until they diverge from this treatment regime. As illustration consider the effect of treatment given at t compared to never treatment. At t the re-weighting assures that the outcomes of the not-yet treated are re-weighted in order to mimic the distribution of the confounders in the population of treated at t. In the subsequent periods the weights are adjusted in order to control for fact that without weighting individuals with characteristics that makes them less likely to enter treatment will be over represented in the sample of not-yet treated. The weights are based on propensity scores in each period. We show that the weighted estimator provides asymptotically unbiased estimates. The standard errors are obtained by bootstrapping. We explore the small sample properties of the estimators using an extended Monte Carlo simulation.

We also contrast our estimator to the estimator proposed by Fredriksson and Johansson (2008). We conclude that their estimator in general is biased. The intuition behind this result is that their estimator is based on a pooled sample of matched controls, consisting of controls with different values of the confounders. Although under the underlying assumptions censoring into treatment is random conditional on the confounders, the censoring is not random in the pooled sample of controls, and this introduces bias.

In section 2, we present the theoretical framework. Section 3 discusses the treatment effect of interest, which resemblance the average treatment effect on the treated often considered in the static matching literature. In section 4 we introduce our weighted estimator, and the small sample properties of our estimator are studied in a Monte Carlo simulation presented in section 5. Section 6 illustrates our estimator using data on work practice for unemployed individuals in Sweden. Section 7 concludes.

2 Model

We consider identification and estimation of the average effect of a treatment where the outcome is a transition from an initial state to a destination state. In the following we consider a generalization of the potential outcome framework (Neyman, 1990; Rubin, 1974) for longitudinal data. Throughout the paper we assume a random sample of N individuals $i = 1, \ldots, N$, although the index i is suppressed for the random variables defined below. We assume discrete time points, $t = 0, 1, \ldots$ and a binary treatment D_t with realized values $d_t \in \mathcal{D}_t$ where $\mathcal{D}_t$ is the set $\{0, 1\}$. We denote by $\overline{D}_t = \{D_1, \ldots, D_t\}$ the sequence of treatments given at the different time points, commonly referred to as a treatment regime, see e.g., Hernan et al. (2001). For each individual there are 2^t possible realizations of a regime $\overline{D}_t$. Throughout this paper we will make the usual assumption that there is no effect of the treatment before it starts.[2].

For sake of presentation we use this directly in our notation. For each time we can then consider a binary potential outcome $Y_t^{\overline{d}_t}$, an indicator of a transtion in period t if the treatment regime had been $\overline{d}_t \in \mathcal{D}_t$. Since for each t a potential outcome $Y_t^{\overline{d}_t}$ is defined, we

[2] Abbring and van den Berg (2003) call this the no-anticipation assumption, which is also discussed by e.g. Abbring and Heckman (2008)

have that $\overline{Y}_t^{\overline{d}_t}$ is the sequence of potential outcomes $\overline{Y}_t^{\overline{d}_t} = \{\overline{Y}_1^{\overline{d}_1}, \ldots, \overline{Y}_t^{\overline{d}_t}\}$. The observed outcome, Y_t, corresponds to the individuals actual treatment regime

$$Y_t = \sum_{\overline{d}_t \in \mathcal{D}_t} I(\overline{D}_t = \overline{d}_t) Y_t^{\overline{d}_t} \tag{1}$$

where $I()$ is an indicator function. In addition we also denote by X a vector of baseline covariates observed for all individuals.

The population of interest is the individuals that are in an initial state at time origin. Transitions, i.e. the variable taking the value $Y_t^{\overline{d}_t} = 1$ as well as the start of treatment could occur at $t = 1, \ldots T$. This situation occurs, for instance, for a medical treatment given after some waiting time, $t > 1$, were the outcome is survival time, or for active labor market training given after some time in unemployment, were the outcome is time to employment. We assume that treatment is assigned at the beginning of the discrete time period, so that treated responses are observed in all periods. We consider non-censored survival times in the initial state.

Our parameter of interest is the average affect of a treatment regime $\overline{d}_t$ on the probability to survive from a starting point t' to a end point t compared to survival throughout the same time interval under a reference treatment regime for the population following treatment regime $\overline{d}$ and survive up until t'

$$ATET_{t,t'}(\overline{d}_t, \overline{d}_t^*) = \tag{2}$$

$$\Pr(\overline{Y}_t^{\overline{d}_t} = 0 | \overline{D}_{t'} = \overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}} = 0) - \Pr(\overline{Y}_t^{\overline{d}_t^*} = 0 | \overline{D}_{t'} = \overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}} = 0).$$

The results of the paper concerns identification and estimation of this parameter of interest. It resemblances the average treatment effect on the treated often considered in the static matching literature. One differences is that this average effect is taken over the population of treated at t' that survive up until t', and not over the full population of treated at t'.

3 Identification

We consider identification of $ATET_{t,t'}(\overline{d}_t, \overline{d}_t^*)$ if we have data on the selection to treatment, such that it is reasonable to assume that the probability distribution of the potential durations is independent of assignment to treatment when conditioning on observed covariates. In a dynamic environment one may consider several different unconfoundedness assumptions, see e.g. the discussion in Ridder and Vikström (2011). In this paper we assume that unconfoundedness holds among survivors

Assumption 1 (Sequential unconfoundedness among survivors) *For all t, $\overline{d}_{t-1}$ and all $\overline{d}_s^*$, $s \geq t$ with the first $t-1$ components equal to $\overline{d}_{t-1}$*

$$D_t \perp \overline{Y}_s^{\overline{d}_s^*} \quad t = s, s+1, \ldots | X, \overline{D}_{t-1} = \overline{d}_{t-1}, \overline{Y}_{t-1}^{\overline{d}_{t-1}} = 0.$$

where X are the observed covariates. That is conditional on survival up until t and a specific dynamic treatment regime $\overline{d}_{t-1}$ treatment assignment in period t is random conditional on the observed covariates. Assumption 1 holds in all situations where decisions are made sequentially based on the survivor experience up to a certain time point. For instance, if case workers assign unemployed individuals to ALMP programs based on time in unemployment and a set of observed covariates. It is easy to show that it also holds in the more restrictive setting where entire treatment regimes are randomly assigned conditional on covariates at at time 0.

We also make an overlap condition

Assumption 2 (Overlap) *For all t and $\overline{d}_{t-1}$*

$$\Pr(D_t = 1 | X, \overline{D}_{t-1} = \overline{d}_{t-1}, \overline{Y}_{t-1}^{\overline{d}_{t-1}} = 0) = p_t(X,\overline{d}_{t-1}) < 1.$$

Initially, we consider cases when $\overline{d}_{t'-1} = \overline{d}^*_{t'-1}$, i.e. when the two dynamic treatment regimes of interest are identical up to the point from where we consider survival. Naturally, in most applications t' will be the time point where $\overline{d}$ and $\overline{d}^*$ diverge for the first time. The key identification problems are that the treatment may change at any point in time and that the actual treatment regime is unobserved after the individual left the initial state. As already discussed in the introduction this has important implications for the choice of control group. In a dynamic setting where treatment is only imposed once Fredriksson and Johansson (2008) and Crepon et al. (2009) show that those not-yet treated at t could be used as control group to estimate the exit rate rates under no-treatment at t for those treated before t. Our setting is more general as we allow treatment to start and end at any point in time. In this setting we show that the exit rates under treatment regime $\overline{d}_t$ could be identified using those that follow the treatment regime $\overline{d}$ up until t but not necessarily after t as control group. This resemblance the argument of using all not-yet treated as control group made by Fredriksson and Johansson (2008) and Crepon et al. (2009).

Our main identification result is summarized in Theorem 1

Theorem 1 (Identification of ATET) *Suppose that Assumption 1 and 2 hold. If $\overline{d}_{t'} = \overline{d}^*_{t'}$ then for all $t > t'$ we have the following point identification result*

$$ATET_{t,t'}(\overline{d}_t, \overline{d}^*_t) = \tag{3}$$

$$\mathbb{E}_{\mathbf{X}|\overline{D}_{t'}=\overline{d}_{t'},\overline{Y}_{t'-1}=0} \prod_{k=t'}^{t} \Pr(Y_k = 0 | X, \overline{D}_k = \overline{d}_k, \overline{Y}_{k-1} = 0) -$$

$$\mathbb{E}_{\mathbf{X}|\overline{D}_{t'}=\overline{d}_{t'},\overline{Y}_{t'-1}=0} \prod_{k=t'}^{t} \Pr(Y_k = 0 | X, \overline{D}_k = \overline{d}^*_k, \overline{Y}_{k-1} = 0).$$

Proof Here consider the case when $t' = t - 1$. Results for for arbitrary t and t' are reported in the appendix. For $t' = t - 1$

$$ATET_{t,t'}(\overline{d}_t, \overline{d}^*_t) =$$

$$\Pr(Y_t^{\overline{d}_t}=0, Y_{t'}^{\overline{d}_{t'}}=0|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0) - \Pr(Y_t^{\overline{d}_t^*}=0, Y_{t'}^{\overline{d}_{t'}^*}=0|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0)$$

Under assumption 1 (UC) and using the observational rule 1 we have

$$\begin{aligned}
\Pr(Y_t^{\overline{d}_t} &= 0, Y_{t'}^{\overline{d}_{t'}}=0|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0) = \\
\mathbb{E}_{\mathbf{X}|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0} \Pr(Y_t^{\overline{d}_t} &= 0, Y_{t'}^{\overline{d}_{t'}}=0|X, \overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0) = \\
\mathbb{E}_{\mathbf{X}|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0} \Pr(Y_t^{\overline{d}_t} &= 0|X, \overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'}^{\overline{d}_{t'}}=0) \Pr(Y_{t'}^{\overline{d}_{t'}}=0|X, \overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0) \\
&= (\text{UC at } t) = \\
\mathbb{E}_{\mathbf{X}|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0} \Pr(Y_t^{\overline{d}_t} &= 0|X, \overline{D}_{t}=\overline{d}_{t}, \overline{Y}_{t'}^{\overline{d}_{t'}}=0) \Pr(Y_{t'}^{\overline{d}_{t'}}=0|X, \overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0) \\
&= (\text{obs. rule}) = \\
\mathbb{E}_{\mathbf{X}|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}=0} \Pr(Y_t &= 0|X, \overline{D}_{t}=\overline{d}_{t}, \overline{Y}_{t'}=0) \Pr(Y_{t'}=0|X, \overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}=0) \\
&= \mathbb{E}_{\mathbf{X}|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}=0} \prod_{k=t'}^{t} \Pr(Y_k=0|X, \overline{D}_k=\overline{d}_k, \overline{Y}_{k-1}=0,
\end{aligned}$$

and in addition since $\overline{d}_{t'-1}=\overline{d}^*_{t'-1}$ we have

$$\begin{aligned}
\Pr(Y_t^{\overline{d}_t^*} &= 0, Y_{t'}^{\overline{d}_{t'}^*}=0|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0) = \\
\mathbb{E}_{\mathbf{X}|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0} \Pr(Y_t^{\overline{d}_t^*} &= 0, Y_{t'}^{\overline{d}_{t'}^*}=0|X, \overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0) = \\
\mathbb{E}_{\mathbf{X}|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0} \Pr(Y_t^{\overline{d}_t^*} &= 0, Y_{t'}^{\overline{d}_{t'}^*}=0|X, D_{t'}=d_{t'}, \overline{D}_{t'-1}=\overline{d}^*_{t'-1}, \overline{Y}_{t'-1}^{\overline{d}^*_{t'-1}}=0) = \\
&= (\text{UC at } t^*) = \\
\mathbb{E}_{\mathbf{X}|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0} \Pr(Y_t^{\overline{d}_t^*} &= 0, Y_{t'}^{\overline{d}_{t'}^*}=0|X, \overline{D}_{t'}=\overline{d}^*_{t'}, \overline{Y}_{t'-1}^{\overline{d}^*_{t'-1}}=0) = \\
\mathbb{E}_{\mathbf{X}|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0} \Pr(Y_t^{\overline{d}_t^*} &= 0|X, \overline{D}_{t'}=\overline{d}^*_{t'}, \overline{Y}_{t'}^{\overline{d}^*_{t'}}=0) \Pr(Y_{t'}^{\overline{d}^*_{t'}}=0|X, \overline{D}_{t'}=\overline{d}^*_{t'}, \overline{Y}_{t'-1}^{\overline{d}^*_{t'-1}}=0) \\
&= (\text{UC at } t) = \\
\mathbb{E}_{\mathbf{X}|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0} \Pr(Y_t^{\overline{d}_t^*} &= 0|X, \overline{D}_{t}=\overline{d}^*_{t}, \overline{Y}_{t'}^{\overline{d}^*_{t'}}=0) \Pr(Y_{t'}^{\overline{d}^*_{t'}}=0|X, \overline{D}_{t'}=\overline{d}^*_{t'}, \overline{Y}_{t'-1}^{\overline{d}^*_{t'-1}}=0) \\
&= (\text{obs. rule}) = \\
\mathbb{E}_{\mathbf{X}|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}=0} \Pr(Y_t &= 0|X, \overline{D}_{t}=\overline{d}^*_{t}, \overline{Y}_{t'}=0) \Pr(Y_{t'}=0|X, \overline{D}_{t'}=\overline{d}^*_{t'}, \overline{Y}_{t'-1}=0) \\
&= \mathbb{E}_{\mathbf{X}|\overline{D}_{t'}=\overline{d}_{t'}, \overline{Y}_{t'-1}=0} \prod_{k=t'}^{t} \Pr(Y_k=0|X, \overline{D}_k=\overline{d}^*_k, \overline{Y}_{k-1}=0)
\end{aligned}$$

$\square$.

Next, if $\overline{d}_{t'-1} \neq \overline{d}^*_{t'-1}$. Then using the same reasoning as above we have

$$\Pr(Y_t^{\overline{d}^*_t} = 0, Y_{t'}^{\overline{d}^*_{t'}} = 0 | \overline{D}_{t'} = \overline{d}^*_{t'}, \overline{Y}_{t'-1}^{\overline{d}^*_{t'-1}} = 0) = \tag{4}$$

$$\Pr(Y_t^{\overline{d}^*_t} = 0, Y_{t'}^{\overline{d}^*_{t'}} = 0 | D_{t'} = d^*_{t'}, \overline{D}_{t'-1} = \overline{d}^*_{t'-1}, \overline{Y}_{t'-1}^{\overline{d}^*_{t'-1}} = 0) =$$

$$\mathbb{E}_{\mathbf{X}|\overline{D}_{t'}=\overline{d}^*_{t'}, \overline{Y}_{t'-1}=0} \prod_{k=t'}^{t} \Pr(Y_k = 0 | X, \overline{D}_k = \overline{d}^*_k, \overline{Y}_{k-1} = 0).$$

However, if $\overline{d}_{t'-1} \neq \overline{d}^*_{t'-1}$ then because of the dynamic selection up until $t'-1$ the population with $\overline{D}_{t'} = \overline{d}_{t'-1}, \overline{Y}_{t'-1}^{\overline{d}_{t'-1}} = 0$ will, in general, differ from the population with $\overline{D}_{t'} = \overline{d}^*_{t'-1}, \overline{Y}_{t'-1}^{\overline{d}^*_{t'-1}} = 0$, so that the observed outcomes for the latter population is not enough to point identify the outcomes of the former population. That is if $\overline{d}_{t'} \neq \overline{d}^*_{t'}$ then $ATET_{t,t'}(\overline{d}_t, \overline{d}^*_t)$, in general, is not point identified.

4 Weighted estimation

In this section we consider estimation of $ATET_{t,t'}(\overline{d}_t, \overline{d}^*_t)$ for cases when $\overline{d}_{t'-1} = \overline{d}^*_{t'-1}$. We introduce a weighted Kaplan-Meier (KM) type of estimator, which estimates the outcomes under treatment regime $\overline{d}_t$ ($\overline{d}^*_t$) using those that follow the treatment regime $\overline{d}$ ($\overline{d}^*$) up until t but not necessarily after t. This follows from the same reasoning as in the section on identification. We have

$$\widehat{ATET}_{t,t'}(\overline{d}_t, \overline{d}^*_t) = \tag{5}$$

$$\prod_{k=t'}^{t} \left[1 - \frac{\sum_{i \in \overline{D}_{t'-1}=\overline{d}_{t'-1}, \overline{Y}_{t'-1}=0} w_{k,t'}^{\overline{d}_k}(\overline{d}_{t'}) Y_{k,i} \mathbf{1}(\overline{Y}_{k-1,i} = 0) \mathbf{1}(\overline{D}_{k,i} = \overline{d}_k)}{\sum_{i \in \overline{D}_{t'-1}=\overline{d}_{t'-1}, \overline{Y}_{t'-1}=0} w_{k,t'}^{\overline{d}_k}(\overline{d}_{t'}) \mathbf{1}(\overline{Y}_{k-1,i} = 0) \mathbf{1}(\overline{D}_{k,i} = \overline{d}_k)} \right] -$$

$$\prod_{k=t'}^{t} \left[1 - \frac{\sum_{i \in \overline{D}_{t'-1}=\overline{d}_{t'-1}, \overline{Y}_{t'-1}=0} w_{k,t'}^{\overline{d}^*_k}(\overline{d}_{t'}) Y_{k,i} \mathbf{1}(\overline{Y}_{k-1,i} = 0) \mathbf{1}(\overline{D}_{k,i} = \overline{d}^*_k)}{\sum_{i \in \overline{D}_{t'-1}=\overline{d}_{t'-1}, \overline{Y}_{t'-1}=0} w_{k,t'}^{\overline{d}^*_k}(\overline{d}_{t'}) \mathbf{1}(\overline{Y}_{k-1,i} = 0) \mathbf{1}(\overline{D}_{k,i} = \overline{d}^*_k)} \right]$$

where

$$w_{k,t'}^{\overline{d}_k}(\overline{d}_{t'}) = \frac{1}{\prod_{m=t'+1}^{k} p_{d_m}(X, \overline{d}_{m-1})} \tag{6}$$

and

$$w_{k,t'}^{\overline{d}^*_k}(\overline{d}_{t'}) = \frac{p_{d_{t'}}(X, \overline{d}_{t'-1})}{p_{d^*_{t'}}(X, \overline{d}^*_{t'-1}) \prod_{m=t'+1}^{k} p_{d^*_m}(X, \overline{d}^*_{m-1})} \tag{7}$$

and

$$p_{d_m}(X, \overline{d}_{m-1}) = \Pr(D_m = d_m | X, \overline{D}_{m-1} = \overline{d}_{m-1}, \overline{Y}_{m-1}^{\overline{d}_{m-1}} = 0). \tag{8}$$

In the appendix we show that this provides an asymptotically unbiased estimator of the $ATET$.

Several things are important to note about the estimator given by (5). First, it is a weighted KM estimator, since the numerator re-weights the observed exits in a certain period and the denominator re-weights the individuals at risk shortly before this time period. Second, as already mentioned individuals are used in the estimation as long as they follow the treatment regime of interest. At the point of divergence they are regarded as censored. Naturally, individuals are followed the entire time in the initial state if they exit while being on the treatment regime of interest. Third, individuals with a certain set of covariates are, in general, given different weight in different periods. This follows since the weights depends on the entire censoring up until the specific period, k. Fourth, the weights depend on the inverse probability of remaining on the treatment regime of interest conditional on observed covariates.

Fifth, briefly consider the intuition behind the weights. Take the case when $\overline{d}_t$ is a sequence of 0-s until t' and the remaining values are 1, and when $\overline{d}_t^*$ is a sequence of 0-s. In this case the $ATET_{t,t'}(\overline{d}_t, \overline{d}_t^*)$ reflects the effect on the average survival rate when comparing no treatment up until t' and always treated thereafter with never treated. Then, at t' we have $w_{k,t'}^{\overline{d}_k}(\overline{d}_{t'}) = 1$ and

$$w_{k,t'}^{\overline{d}_k^*}(\overline{d}_{t'}) = \frac{p_{d_{t'}}(X,\overline{d}_{t'-1})}{p_{d_{t'}^*}(X,\overline{d}_{t'-1}^*)} = \frac{p_{d_{t'}}(X,\overline{d}_{t'-1})}{1 - p_{d_{t'}}(X,\overline{d}_{t'-1})} \tag{9}$$

where the second equality follows since $\overline{d}_{t'-1} = \overline{d}_{t'-1}^*$, and since $\overline{d}_{t'} = 1$ and $\overline{d}_{t'}^* = 0$. Here the outcomes of the individuals following treatment regime $\overline{d}_{t'}^*$ are re-weighted in order to mimic the distribution of X in the population taking treatment regime $\overline{d}_{t'}$. It is also exactly the same weights encountered in the static matching literature. For the static case we have from e.g. Wooldridge (2010) the following estimator of the average treatment effect on the treated

$$ATET = \mathbb{E}\left\{\frac{(D - p(X))\,Y}{\rho(1 - p(X))}\right\} = \mathbb{E}\left\{\frac{1}{\rho}DY\right\} - \mathbb{E}\left\{\frac{p(X)}{\rho(1 - p(X))}(1 - D)Y\right\}, \tag{10}$$

where D and Y are treatment and outcome in the static setting, and $\rho = \Pr(D = 1)$, i.e. the unconditional probability of entering treatment. Note that also here the outcomes of the non-treated is not re-weighted and the outcomes of the non-treated are weighted with the probability of entering treatment divided by 1 minus this probability. The only difference is the presence of ρ. However, note that in our dynamic case with a KM type of estimator this is captured by the expressions in the denominators.[3]

Next, using the same example we have at $t' + 1$

$$w_{k,t'}^{\overline{d}_k^*}(\overline{d}_{t'+1}) = \frac{p_{d_{t'}}(X,\overline{d}_{t'-1})}{p_{d_{t'}^*}(X,\overline{d}_{t'-1}^*)p_{d_{t'+1}^*}(X,\overline{d}_{t'}^*)}, \tag{11}$$

so that the re-weighting serves two purposes. Besides re-weighting in order to mimic the distribution of X among the population taking treatment regime $\overline{d}_{t'}$ the weights also correct

[3] For instance, at t' for our example the denominator in the first part of the estimator is $\sum D_{t',i}$.

for the selective censoring due to treatment assignment in period $t' + 1$. More specifically, in period $t' + 1$ a fraction of all not-yet treated, i.e. on $\overline{d}^*$, becomes treated, i.e. diverge from $\overline{d}^*$. This outflow is selective since the probability of diverging depends on the observed covariates. Our estimator corrects for this selective drop-out by giving individuals with X characteristics that makes likely to diverge from $\overline{d}^*$ are given larger weight.

Standard errors are obtained by bootstrapping. In practice the true propensity scores are replaced by estimated propensity scores. The small sample properties of the estimator is explored in a Monte Carlo simulation presented in section 5.

5 Simulation results

In this section we investigate the properties of the estimator introduced in section 4. In this simulation we assume that if treated the individual remains treated in all subsequent periods. One reason for this is that we wish to compare our estimator to other proposed estimators for this special case. We assume that transitions out of the initial state are given by

$$\Pr(Y_{i,t} = 1|\overline{Y}_{i,t-1} = 0) = [1 + \exp(-(a_0 + a_1 X_i + a_2 V_i^Y))]^{-1} \tag{12}$$

and that transitions into treatment are given by

$$\Pr(D_{i,t} = 1|\overline{D}_{i,t-1} = 0) = [1 + \exp(-(b_0 + b_1 X_i + b_2 V_i^D))]^{-1}, \tag{13}$$

where X is assumed to be observed by the econometrician, and V^D and V^Y are unobserved. All three are taken as independently uniformly distributed on [-0.5,0.5]. In all simulations a_0 = -3.0, and b_0 is -4.5 or -2.0. This corresponds to cases with low and moderately high rate of inflow into treatment, respectively. For the other parameters we consider three cases. In a setting with neither observed nor unobserved heterogeneity we take $a_1 = b_1 = a_2 = b_1 = 0$, the setting with only observed heterogeneity has $a_1 = b_1 = 1$, and in the most extended setting with unobserved heterogeneity we in addition set $a_2 = b_1 = 1$. Note that the latter setting allows the treatment and the outcome to be correlated through X, but not through the unobserved effects, so that the sequential unconfoundedness assumption is fulfilled.

In the sample design we generate samples of sizes 20,000, 12,000 and 8,000. The number of replications is 300. When we implement our weighted KM estimator we estimate the treatment propensities using logit regression models and the standard errors are calculated using bootstrap (99 replications).

As comparison we also run simulations for the estimator proposed by Fredriksson and Johansson (2008). They propose a two-step matching estimator. In the first step a matched sample of treated and controls is constructed using one-to-one matching. In the second step this matched sample construct estimates of the survival rates under treatment and no-treatment using two separate un-weighted KM estimators. In practice we implement this estimator using 1-nearest neighbor propensity score matching, where the propensity score is estimated using logit.

Tables 1 and 2 also report simulation results for the Fredriksson and Johansson (2008) estimator. Based on the results for bias and size we confirm that this estimator performs well if there are no heterogeneity in the transition rate out of the initial state and in the transition rate into treatment. However, in our second specification that includes correlated observed heterogeneity the FJ estimator is severely biased. In our specification when individuals with a high transition rate out the initial state also have a high probability to enter treatment the FJ estimator, is as theoretically, predicted negatively biased. The bias is even more severe when unobserved heterogeneity is included in our third specification.

6 Application to Swedish ALMP

In this section we study the effects of a work practice program governed by the Swedish public employment service (PES). The aim of the program, directed toward long term unemployed, is to provide the unemployed individual with practical experience and to maintain or strengthen their productivity. The work practice could take place at both private and public employers. The maximum duration is 6 months.

The population is taken from the population register Händel administrated by the PES. As one requirement for payment from the unemployment insurance is that one is registered at the PES this implies that we basically have full coverage of all unemployed job seekers in Sweden. The register contatins information on the time when an individual (i) became unemployed, (ii) entered into a labor market program, and (iii) the time of exit from unemployment. We also have information on the reason for the exit (employment, education, social assistance, sickness or disability insurance programs or unknown reason). Händel also includes a number of personal characteristics recorded at the beginning of the unemployment spell and information on eligibility for unemployment insurance (UI).[4] To this data we have matched information on marital status, household characteristics (e.g. number of children), labor income and income from various insurance programs (e.g. sickness and disability insurance programs) schemes from the population register LOUISE.

We sample all unemployed individuals in ages 25-55 at the time of entry into unemployment who has a spell of unemployment longer than 6 months in the period January 1, 1999 to December 31, 2003. We exclude unemployment spells starting within 180 days from the last spell. For this population we construct detailed information on previous unemployment episodes which is used as control variables. We aggregate the daily spell data to monthly intervals.[5]We focus on individuals that enter work practice between 6 and 27 months of un-

[4]Unemployment benefits can be paid in two different ways, a fixed basic compensation or an income-related amount based on previous earnings. To receive any compensation, the unemployed person must be at least 20 and fulfill: i) the basic conditions, and ii) the work condition. The basic conditions are a set of rules for the unemployed. For instance, they state that he or she should be partially or completely unemployed and prepared to accept suitable job offers. The work condition specifies that the unemployed person must have been employed for approximately 6 out of the last 12 months preceding unemployment. If these requirements are met, the unemployed person is qualified for the fixed basic compensation. To be eligible for the higher income-related compensation you also need to have been a member of an unemployment insurance fund for at least twelve months preceding the first day of unemployment.

[5]We mainly do this for presentation reasons. Remeber that we estimate specific average effects for each possible start of treatment. Another reason is measurement error in the daily information. Finally, no-anticipation implying that individuals do not know or cannot predict the exact timing of their assignment to work practice is more likely to hold for monthly data then for daily data.

employment. An unemployment spells that ended for other reasons than employment are treated as a right-censored durations.[6]

Table 3 provides descriptive statistics for our sample of work practice spells and all other unemployment spells used in the analysis. In total we have more than 500,000 unemployment spells and 8.4% (46,495 individuals) of the spells concerns participation in work practice. 36.3% (16,878 individuals) of those entering the program starts within 9 months of unemployment. Only 3.2 percent (1,488 individuals) of the participants starts in months 22-27. This is, thus, a large and extensive program, with a high degree of dynamic treatment assignment. 62.7% of all observations are uncensored spells and among these spells the mean (median) unemployment duration is 14.3 (10) months.

In the estimation we take use of estimated propensity scores. We estimate logit regression models in which we include gender, age, age squared, number of unemployment days in the last 5 years, level of education (5 categories), indicator for UI entitlement, region of residence (22 regions), indicator for at least one child in the household, marital status, country of origin (3 categories) labor income, social benefits and unemployment insurance benefits and calendar year. We include incomes and benefits both one and two years before the start of the unemployment. Table 4 presents descriptive statistics on a subset of these covariates for the treated and for the individuals that leave unemployment before enrollment into work practice. From this table we can see that males, immigrants and individuals eligible for UI are overrepresented among the participants. Furthermore the participants have more extensive unemployment history. Individuals living in the Stockholm and Goteborg metropolitan statistical area (MSA) and individuals with a university degree are less likely work practice participants. All in all the two populations of treated and non-treated are rather similar. If anything, the program participants are negatively selected population, so that they in absence of treatment would have had longer unemployment spells compared to other unemployed.

In order for our estimator to provide consistent estimates the sequential unconfoundedness among survivors needs to hold. The implication of this is that conditional on our large set of covariates treatment is randomly assigned among individuals still unemployed at each unemployment duration. that individuals that treated early does not have shorter/longer durations in the absence of the program than those treated later or never treated. Case workers in Sweden have large influence over enrollment into different programs why self selection to the PES program is less important (see e.g. Eriksson, 1997; Carling and Richardsson, 2001). Moreover, Eriksson and Lagerström (2006, 2012) show that previous unemployment episodes status negatively affects the chances of becoming employed. Time in unemployment should therefore by important for treatment assignment. Also, note that we condition on detailed

[6]Unemployment benefits can be paid in two different ways, a fixed basic compensation or an income-related amount based on previous earnings. To receive any compensation, the unemployed person must be at least 20 and fulfill: i) the basic conditions, and ii) the work condition. The basic conditions are a set of rules for the unemployed. For instance, they state that he or she should be partially or completely unemployed and prepared to accept suitable job offers. The work condition specifies that the unemployed person must have been employed for approximately 6 out of the last 12 months preceding unemployment. If these requirements are met, the unemployed person is qualified for the fixed basic compensation. To be eligible for the higher income-related compensation you also need to have been a member of an unemployment insurance fund for at least twelve months preceding the first day of unemployment.

information on previous unemployment spells. Besides sequential unconfoundedness among survivors treatment needs to be unanticpated, i.e. unemployed should not be able to predict the exact timing of their assignment to work practice. Since there are several unpredictable events leading to program enrollment, such as caseworkers discretionary power, we believe that this assumption is fulfilled.

Treatment assignment probabilities each waiting time using logit regression models. As illustration we present effect estimates using our weighted estimator and estimates using the pre-matched but un-weighted estimator by Fredriksson and Johansson (2008). It is implemented using a one-to-one nearest neighborhood matching strategy. We match on the same estimated propensity score as was used for our weighting estimator.

Figure 2 displays the results for a selection of waiting or enrollment times. They are presented using the difference in fraction re-employed as outcome for each of the 24 months after program start. Initially, consider the results for enrollment after 7 months of unemployment. For this enrollment time we find that for the first 2-3 months the exit rates are lower among program participants. This is in line with results from training and employment subsidy programs (see e.g. Forslund et al., 2004; Forslund and Vikström, 2011). This locking-in effect most likely is due to lower search effort during the program. After this initial period participant gradually catch-up and about 10 months after the enrollment to the program the fraction employed is higher. This effect on employment rate is maintained during the entire follow-up horizon (up to 25 month after entering into the program). We find similar patterns for waiting times up to 13 months. For longer waiting times we also find locking-in effects, but participants never fully catch-up with those who did not enroll at the given time. The problem with the evaluation of enrolling at longer time periods is that few individuals start their work practice at a given month why the precision of our estimates is low. Considering the comparison with the Fredriksson and Johansson (2008) estimator we can see that there are quite small differences for waiting times up to 16 months. For longer waiting times to treatment there are some differences.

7 Conclusions

In this paper, we have implemented a weighted KM estimator for the average effects on survival time of dynamic treatment regimes under sequential unconfoundedness. The identification and estimation problems arise since the dynamic setting considered in this paper is plagued by both dynamic treatment assignment as well as dynamic selection. The former due to the fact that treatment is allowed to start and stop at any point in time. The latter since the outcome of interest is the duration in an initial state.

References

Abbring, J. and Heckman, J. (2008). Dynamic policy analysis. In Matyas, L. and Sevestre, P., editors, *The Econometrics of Panel Data*, volume 46, pages 795–863.

Abbring, J. H. and Van Den Berg, G. J. (2003). The nonparametric identification of treatment effects in duration models. *Econometrica*, 71(5):1491–1517.

Carling, K. and Richardson, K. (2001). The relative efficiency of labor market programs: Swedish experience from the 1990's. IFAU Working Paper 2001:2.

Crepon, B., Ferracci, M., Jolivet, G., and van den Berg, G. J. (2009). Active labor market policy effects in a dynamic setting. *Journal of the European Economic Association*, 7(2-3):595–605.

de Luna, X. and Johansson, P. (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(7):2122 – 2137.

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

Eriksson, S. and Lagerströ, J. (2006). Competition between employed and unemployed job applicants: Swedish evidence. *Scandinavian Journal of Economics*, 108:373–396.

Eriksson, S. and Lagerströ, J. (2012). Detecting discrimination in the hiring process: Evidence from an internet-based search channel. *Empirical Economics*, 43:537–563.

Forslund, A., Johansson, P., and Lindqvist, L. (2004). Employment subsidies - a fast lane from unemployment to work? IFAU Working paper 2004:18.

Forslund, A. and Vikström, J. (2011). Arbetsmarknadspolitikens effekter p sysselsttning och arbetslshet en versikt. IFAU Report 2011:7.

Fredriksson, P. and Johansson, P. (2008). Dynamic treatment assignment. *Journal of Business & Economic Statistics*, 26(4):435–445.

Heckman, J. J., Lalonde, R. J., and Smith, J. A. (1999). The economics and econometrics of active labor market programs. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3 of *Handbook of Labor Economics*, chapter 31, pages 1865–2097. Elsevier.

Hernan, M. A., Robins, J., and Brumback, B. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association*, 96:440–448.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*, 86(1):4–29.

Kaplan, M. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481.

Lechner, M. and Miquel, R. (2010). Identification of the effects of dynamic treatments by sequential conditional independence assumptions. *Empirical Economics*, 39(1):111–137.

Neyman, J. (1990). On the application of probability theory to agricultural experiments, essay on principles. Translated by D.M. Dabrowska and edited by T.P. Speed, *Statistical Science*, **5**, 465-472.

Ridder, G. and Vikström, J. (2011). Bounds on Treatment Effects on Transitions. IFAU Working paper 2011:6.

Robins, J., Hernan, M., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11:550–560.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, 66:688–701.

Sianesi, B. (2004). An evaluation of the swedish system of active labor market programs in the 1990s. *The Review of Economics and Statistics*, 86(1):133–155.

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

Tables and Figures

Figure 1: Transition rate into program by time in unemployment

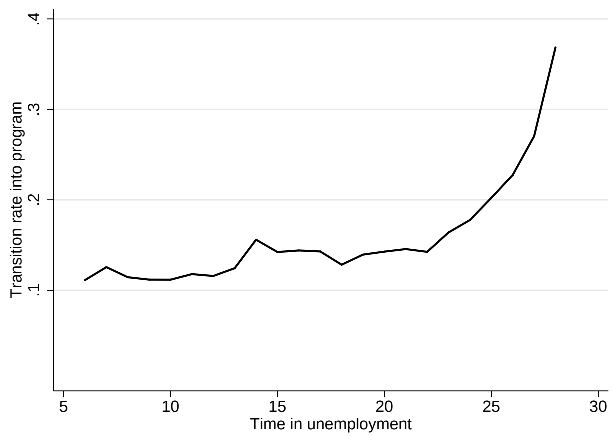
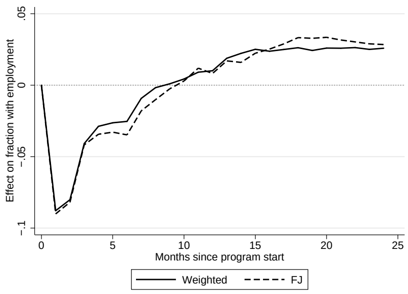
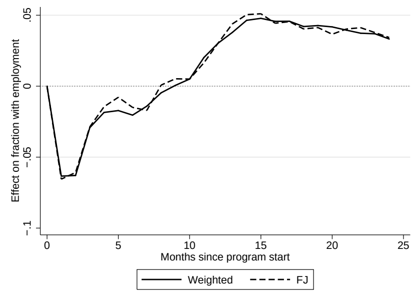


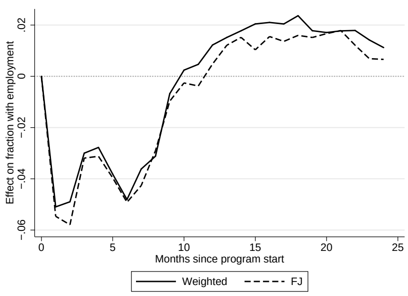
Figure 2: Effect of work practice on fraction reemployed. By time to program start



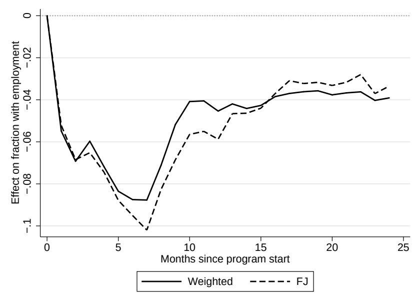
(a) 7 months waiting time



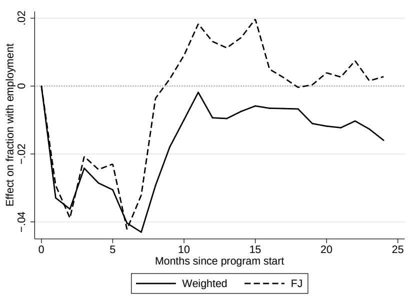
(b) 10 months waiting time



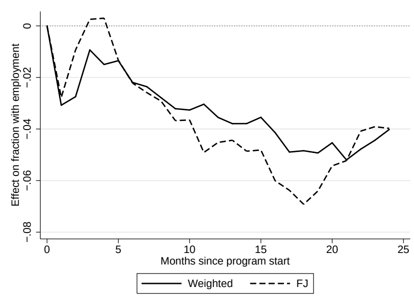
(c) 13 months waiting time



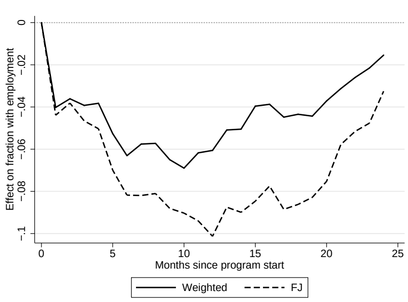
(d) 16 months waiting time



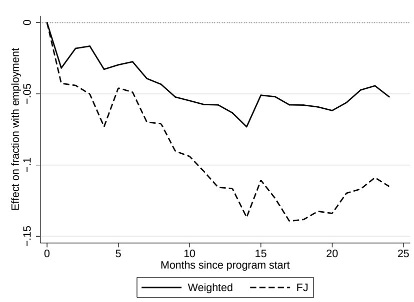
(e) 21 months waiting time



(f) 23 months waiting time



(g) 25 months waiting time



(h) 27 months waiting time

Table 1: Monte Carlo results for the weighted estimator under different model specifications

		Weighted			FJ	
	Size	Bias	MSE	Size	Bias	MSE
No heterogeneity $[a_1 = b_1 = a_2 = b_2 = 0]$, low inflow $\beta_0 = -4.5$						
t = 1	0.090	-0.99	0.25	0.047	-1.36	0.41
t = 2	0.053	-0.94	0.37	0.050	-0.63	0.78
t = 3	0.057	-2.46	0.57	0.057	0.37	1.14
t = 5	0.063	-1.90	0.89	0.053	-2.16	1.65
t = 7	0.070	-0.78	1.11	0.053	-6.18	2.07
t = 10	0.070	1.14	1.18	0.040	-5.33	2.00
No heterogeneity $[a_1 = b_1 = a_2 = b_2 = 0]$, large inflow: $\beta_0 = -2.0$						
t = 1	0.057	-0.36	0.02	0.070	0.00	0.04
t = 2	0.047	-0.41	0.03	0.063	-0.39	0.07
t = 3	0.047	-0.23	0.05	0.063	-0.31	0.10
t = 5	0.063	-0.45	0.08	0.090	-0.72	0.19
t = 7	0.073	-0.32	0.13	0.073	-1.40	0.32
t = 10	0.057	-0.07	0.14	0.087	-1.22	0.38
Observed heterogeneity $[a_1 = b_1 = 1, a_2 = b_2 = 0]$, low inflow: $\beta_0 = -4.5$						
t = 1	0.040	0.21	0.18	0.077	0.76	0.55
t = 2	0.023	-0.94	0.34	0.060	1.06	0.87
t = 3	0.047	-0.63	0.46	0.053	0.22	1.10
t = 5	0.037	0.67	0.56	0.057	-0.25	1.67
t = 7	0.063	-0.05	0.71	0.063	-2.81	2.29
t = 10	0.050	-0.80	0.67	0.050	-3.99	2.27
Observed heterogeneity $[a_1 = b_1 = 1, a_2 = b_2 = 0]$, large inflow: $\beta_0 = -2.0$						
t = 1	0.043	-0.47	0.02	0.060	-0.05	0.04
t = 2	0.040	-0.28	0.04	0.093	-0.35	0.10
t = 3	0.037	-0.13	0.05	0.103	-1.18	0.15
t = 5	0.053	-0.03	0.08	0.097	-3.85	0.22
t = 7	0.043	-0.13	0.11	0.113	-10.76	0.44
t = 10	0.057	0.00	0.12	0.117	-11.87	0.47
Observed and unobserved heterogeneity $[a_1 = b_1 = a_2 = b_2 = 1]$, low inflow: $\beta_0 = -4.5$						
t = 1	0.070	-1.53	0.20	0.037	-1.17	0.35
t = 2	0.050	-1.43	0.29	0.053	-2.67	0.63
t = 3	0.037	-0.73	0.35	0.050	-1.79	0.79
t = 5	0.030	-0.85	0.35	0.073	-1.40	0.94
t = 7	0.050	0.18	0.32	0.060	-4.15	0.74
t = 10	0.067	-0.20	0.31	0.047	-4.35	0.67
Observed and unobserved heterogeneity $[a_1 = b_1 = a_2 = b_2 = 1]$, large inflow: $\beta_0 = -2.0$						
t = 1	0.047	-0.14	0.03	0.077	-0.05	0.07
t = 2	0.047	-0.08	0.05	0.090	-1.85	0.14
t = 3	0.080	-0.68	0.08	0.127	-4.38	0.20
t = 5	0.053	-0.84	0.09	0.170	-9.76	0.32
t = 7	0.053	-0.77	0.10	0.230	-16.69	0.56
t = 10	0.080	-0.64	0.10	0.260	-19.03	0.63

Note: Results for a logistic simulation model (a_0 =-3.0 in all specifications). The standard errors for the weighted estimator is based on bootstrap (99 replications). Size is for 5% level tests for the treatment parameter which enters the model with a true coefficient equal to zero. The table also reports the mean bias (Bias) and mean squared error (MSE). The design uses 20,000 observations and the results are based on 300 replications.

Table 2: Monte Carlo results for the weighted estimator under different sample sizes

		Weighted			FJ	
	Size	Bias	MSE	Size	Bias	MSE
Sample size 20,000						
t = 1	0.047	-0.14	0.03	0.077	-0.05	0.07
t = 2	0.047	-0.08	0.05	0.090	-1.85	0.14
t = 3	0.080	-0.68	0.08	0.127	-4.38	0.20
t = 5	0.053	-0.84	0.09	0.170	-9.76	0.32
t = 7	0.037	-0.82	0.09	0.200	-13.69	0.44
t = 10	0.080	-0.64	0.10	0.260	-19.03	0.63
Sample size 12,000						
t = 1	0.050	0.36	0.03	0.063	0.20	0.07
t = 2	0.060	0.33	0.07	0.067	0.07	0.14
t = 3	0.057	0.45	0.10	0.053	-0.83	0.22
t = 5	0.057	0.33	0.15	0.073	-1.88	0.35
t = 7	0.050	0.18	0.17	0.093	-5.53	0.52
t = 10	0.033	-0.80	0.20	0.127	-10.57	0.74
Sample size 8,000						
t = 1	0.067	-0.38	0.06	0.023	0.00	0.09
t = 2	0.067	-0.72	0.10	0.067	-0.05	0.21
t = 3	0.037	-0.13	0.14	0.053	-0.23	0.32
t = 5	0.047	-0.20	0.23	0.060	-2.56	0.48
t = 7	0.060	0.11	0.29	0.083	-6.87	0.69
t = 10	0.057	-2.25	0.36	0.100	-13.67	1.09

Note: Results for a logistic simulation model with a_0 =-3.0, a_0 =-2.0 and $a_1 = b_1 = a_2 = b_2 = 1$ in all specifications. The standard errors for the weighted estimator is based on bootstrap (99 replications). Size is for 5% level tests for the treatment parameter which enters the model with a true coefficient equal to zero. The table also reports the mean bias (Bias) and mean squared error (MSE). The results are based on 300 replications.

Table 3: Sample statistics for time in unemployment and time until work practice

Regardless of treatment	
# spells	553,510
% spells with work practice	8.4
% uncensored	62.7
average unemployment duration	14.3 (11.4)
median unemployment duration	10
Concerning spells with work practice	
# spells	46,247
% not censored	71.4
average unemployment duration	26.7 (17.9)
median unemployment duration	21
average time to program start	13.9 (8.6)
median time to program start	11
% program start in	
6-9 months	36.3
10-12 months	17.5
13-15 months	14.5
16-18 months	9.0
19-21 months	5.9
22-24 months	4.1
25-27 months	3.2

Notes: The time unit is month. Standard deviations in parentheses.

Table 4: Sample statistics treated and non-treated in our sample

Variable	No work practice	Work practice
Female (%)	51.1	49.6
Age	37.2 (8.2)	38.1 (8.4)
Foreign born (%)	30.9	33.1 ()
Married (%)	38.7	39.7 ()
Children in household (%)	49.2	49.5 ()
Eligible for UI (%)	82.5	89.0 ()
High School education (%)	44.5	46.6
University education (%)	32.2	29.7
Stockholm MSA (%)	19.6	10.3
Goteborg MSA (%)	18.3	12.8
Skåne (%)	14.5	14.4
North (%)	13.9	20.8
South (%)	11.1	13.5
Previous unemployment	326 (390)	416 (452)
Labor income year -1	112,900 (131,000)	101,300 (110,500)
Labor income year -2	107,500 (129,800)	95,800 (107,600)
Social benefits year -1 (%)	15.4	15.8
Social benefits year -2 (%)	15.9	16.7
UI benefits year -1 (%)	30.5	36.2
UI benefits year -2 (%)	31.2	37.0

Note: Previous unemployment is in days of unemployment during 5 years before the start of the unemployment spell. Labor income is in SEK. Standard deviations in parentheses.

Appendix: Proofs

In this appendix the population with $\overline{D}_{t-1}=\overline{d}_{t-1},\overline{Y}_{t-1}^{\overline{d}_{t-1}}=0$ will be denoted W_t.

Properties of (5)

First, consider the first part of the estimator when $k=t'$. Since $w_{t',t'}^{\overline{d}_{t'}}(\overline{d}_{t'})=1$ and using the observational rule in (1) we, trivially, have

$$\mathbb{E}\left[w_{t',t'}^{\overline{d}_{t'}}(\overline{d}_{t'})Y_{t',i}\mathbf{1}(\overline{D}_{t',i}=\overline{d}_{t'})|X,W_{t'-1}\right]=\Pr\left(\overline{Y}_{t'}^{\overline{d}_{t'}}=1,\overline{D}_{t'}=\overline{d}_{t'}|X,W_{t'-1}\right), \tag{A.1}$$

and

$$\mathbb{E}\left[w_{t',t'}^{\overline{d}_{t'}}(\overline{d}_{t'})\mathbf{1}(\overline{D}_{t',i}=\overline{d}_{t'})|X,W_{t'-1}\right]=\Pr\left(\overline{D}_{t'}=\overline{d}_{t'}|X,W_{t'-1}\right). \tag{A.2}$$

Second, consider the second part of the estimator when $k=t'$. If assumption 1 holds, using (1) and since $\overline{d}_{t'-1}=\overline{d}^*_{t'-1}$

$$\mathbb{E}\left[w_{t',t'}^{\overline{d}^*_{t'}}(\overline{d}_{t'})Y_{t',i}\mathbf{1}(\overline{D}_{t',i}=\overline{d}^*_{t'})|X,W_{t'-1}\right]=\mathbb{E}\left[\frac{p_{d_{t'}}(X,\overline{d}_{t'-1})Y_{t'}^{\overline{d}^*_{t'}}\mathbf{1}(\overline{D}_{t'}=\overline{d}^*_{t'})}{p_{d^*_{t'}}(X,\overline{d}^*_{t'-1})}|X,W_{t'-1}\right]= \tag{A.3}$$

$$\mathbb{E}\left[\overline{Y}_{t'}^{\overline{d}^*_{t'}}|X,\overline{D}_{t'}=\overline{d}^*_{t'},\overline{Y}_{t'-1}^{\overline{d}^*_{t'-1}}=0\right]\frac{p_{d_{t'}}(X,\overline{d}_{t'-1})p_{d^*_{t'}}(X,\overline{d}^*_{t'-1})}{p_{d^*_{t'}}(X,\overline{d}^*_{t'-1})}=$$

$$\mathbb{E}\left[\overline{Y}_{t'}^{\overline{d}^*_{t'}}|X,\overline{D}_{t'}=\overline{d}^*_{t'},\overline{Y}_{t'-1}^{\overline{d}^*_{t'-1}}=0\right]p_{d_{t'}}(X,\overline{d}_{t'-1}).$$

Using UC at t' and that $\overline{d}_{t'-1}=\overline{d}^*_{t'-1}$

$$\mathbb{E}\left[\overline{Y}_{t'}^{\overline{d}^*_{t'}}|X,\overline{D}_{t'}=\overline{d}_{t'},\overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0\right]p_{d_{t'}}(X,\overline{d}_{t'-1})=\Pr\left(\overline{Y}_{t'}^{\overline{d}^*_{t'}}=1,\overline{D}_{t'}=\overline{d}_{t'}|X,W_{t'-1}\right). \tag{A.4}$$

Using similar reasoning as above

$$\mathbb{E}\left[w_{t',t'}^{\overline{d}^*_{t'}}(\overline{d}_{t'})\mathbf{1}(\overline{D}_{t',i}=\overline{d}^*_{t'})X,W_{t'-1}\right]=\Pr\left(\overline{D}_{t'}=\overline{d}_{t'}|X,W_{t'-1}\right). \tag{A.5}$$

Third, consider the first part of the estimator when $k=t'+1$. If assumption 1 holds and using (1)

$$\mathbb{E}\left[w_{t'+1,t'}^{\overline{d}_{t'+1}}(\overline{d}_{t'})Y_{t'+1,i}\mathbf{1}\left(\overline{Y}_{t',i}=0\right)\mathbf{1}(\overline{D}_{t'+1,i}=\overline{d}_{t'+1})|X,W_{t'-1}\right]= \tag{A.6}$$

$$\mathbb{E}\left[\frac{1}{p_{d_{t'+1}}(X,\overline{d}_{t'})}Y_{t'+1}^{\overline{d}_{t'+1}}\mathbf{1}(D_{t'+1}=d_{t'+1})\mathbf{1}(Y_{t'}^{\overline{d}_{t'}}=0)\mathbf{1}(D_{t'}=d_{t'})|X,W_{t'-1}\right]=$$

$$\frac{\mathbb{E}\left[Y_{t'+1}^{\overline{d}_{t'+1}}|X,\overline{D}_{t'+1}=\overline{d}_{t'+1},\overline{Y}_{t'}^{\overline{d}_{t'}}=0\right]p_{d_{t'+1}}(X,\overline{d}_{t'})\Pr\left(\overline{Y}_{t'}^{\overline{d}_{t'}}=0|X,\overline{D}_{t'}=\overline{d}_{t'},\overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0\right)p_{d_{t'}}(X,\overline{d}_{t'-1})}{p_{d_{t'+1}}(X,\overline{d}_{t'})}=$$

$$\mathbb{E}\left[Y_{t'+1}^{\overline{d}_{t'+1}}|X,\overline{D}_{t'+1}=\overline{d}_{t'+1},\overline{Y}_{t'}^{\overline{d}_{t'}}=0\right]\Pr\left(\overline{Y}_{t'}^{\overline{d}_{t'}}=0|X,\overline{D}_{t'}=\overline{d}_{t'},\overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0)p_{d_{t'}}(X,\overline{d}_{t'-1}\right)$$

$$=(\text{UC at } t'+1)=$$

$$\mathbb{E}\left[Y_{t'+1}^{\overline{d}_{t'+1}}|X,\overline{D}_{t'}=\overline{d}_{t'},\overline{Y}_{t'}^{\overline{d}_{t'}}=0\right]\Pr\left(\overline{Y}_{t'}^{\overline{d}_{t'}}=0|X,\overline{D}_{t'}=\overline{d}_{t'},\overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0\right)p_{d_{t'}}(X,\overline{d}_{t'-1})=$$

$$\Pr\left(Y_{t'+1}^{\overline{d}_{t'+1}}=1,\overline{Y}_{t'}^{\overline{d}_{t'}}=0,\overline{D}_{t'}=\overline{d}_{t'}|X,W_{t'-1}\right),$$

and by similiar reasoning

$$\mathbb{E}\left[w_{t'+1,t'}^{\overline{d}_{t'+1}}(\overline{d}_{t'})\mathbf{1}(\overline{Y}_{t',i}=0)\mathbf{1}(\overline{D}_{t'+1,i}=\overline{d}_{t'+1})|X,W_{t'-1}\right]=\Pr\left(\overline{Y}_{t'}^{\overline{d}_{t'}}=0,\overline{D}_{t'}=\overline{d}_{t'}|X,W_{t'-1}\right). \tag{A.7}$$

Fourth, consider the second part of the estimator when for $k=t'+1$. If assumption 1 holds, using (1) and since $\overline{d}_{t'-1}=\overline{d}^*_{t'-1}$

$$\mathbb{E}\left[w_{t'+1,t'}^{\overline{d}^*_k}(\overline{d}_{t'})Y_{t'+1,i}\mathbf{1}(\overline{Y}_{t',i}=0)\mathbf{1}(\overline{D}_{t'+1,i}=\overline{d}^*_{t'+1})|X,W_{t'-1}\right]= \tag{A.8}$$

$$\mathbb{E}\left[\frac{p_{d_{t'}}(X,\overline{d}_{t'-1})Y_{t'+1}^{\overline{d}^*_{t'+1}}\mathbf{1}(D_{t'+1}=d^*_{t'+1})\mathbf{1}(Y_{t'}^{\overline{d}^*_{t'}}=0)\mathbf{1}(D_{t'}=d^*_{t'})}{p_{d^*_{t'}}(X,\overline{d}^*_{t'-1})p_{d^*_{t'+1}}(X,\overline{d}^*_{t'})}|X,W_{t'-1}\right]=$$

$$=\text{outcomes in period } t'=$$

$$\frac{p_{d_{t'}}(X,\overline{d}_{t'-1})\mathbb{E}\left[Y_{t'+1}^{\overline{d}^*_{t'+1}}\mathbf{1}(D_{t'+1}=d^*_{t'+1})|X,\overline{D}_{t'+1}=\overline{d}^*_{t'+1},\overline{Y}_{t'}^{\overline{d}^*_{t'}}=0\right]\Pr\left(\overline{Y}_{t'}^{\overline{d}^*_{t'}}=0|X,\overline{D}_{t'}=\overline{d}^*_{t'},\overline{Y}_{t'-1}^{\overline{d}^*_{t'-1}}=0\right)}{p_{d^*_{t'+1}}(X,\overline{d}^*_{t'})}=$$

$$=\text{outcomes in period } t'=$$

$$\mathbb{E}\left[Y_{t'+1}^{\overline{d}^*_{t'+1}}|X,\overline{D}_{t'+1}=\overline{d}^*_{t'+1},\overline{Y}_{t'}^{\overline{d}^*_{t'}}=0\right]\Pr\left(\overline{Y}_{t'}^{\overline{d}^*_{t'}}=0|X,\overline{D}_{t'}=\overline{d}^*_{t'},\overline{Y}_{t'-1}^{\overline{d}^*_{t'-1}}=0\right)p_{d_{t'}}(X,\overline{d}_{t'-1})$$

$$=(\text{UC at } t'+1)=$$

$$\mathbb{E}\left[Y_{t'+1}^{\overline{d}^*_{t'+1}}|X,\overline{D}_{t'}=\overline{d}^*_{t'},\overline{Y}_{t'}^{\overline{d}^*_{t'}}=0\right]\Pr\left(\overline{Y}_{t'}^{\overline{d}^*_{t'}}=0|X,\overline{D}_{t'}=\overline{d}^*_{t'},\overline{Y}_{t'-1}^{\overline{d}^*_{t'-1}}=0\right)p_{d_{t'}}(X,\overline{d}_{t'-1})$$

$$=(\text{UC at } t')=$$

$$\mathbb{E}\left[Y_{t'+1}^{\overline{d}^*_{t'+1}}|X,\overline{D}_{t'}=\overline{d}_{t'},\overline{Y}_{t'}^{\overline{d}^*_{t'}}=0\right]\Pr\left(\overline{Y}_{t'}^{\overline{d}^*_{t'}}=0|X,\overline{D}_{t'}=\overline{d}_{t'},\overline{Y}_{t'-1}^{\overline{d}^*_{t'-1}}=0\right)p_{d_{t'}}(X,\overline{d}_{t'-1})$$

$$\Pr\left(Y_{t'+1}^{\overline{d}^*_{t'+1}}=1,\overline{Y}_{t'}^{\overline{d}^*_{t'}}=0,\overline{D}_{t'}=\overline{d}_{t'}|X,W_{t'-1}\right),$$

and by similiar reasoning

$$\mathbb{E}\left[w_{t'+1,t'}^{\overline{d}^*_k}(\overline{d}_{t'})\mathbf{1}(\overline{Y}_{t',i}=0)\mathbf{1}(\overline{D}_{t'+1,i}=\overline{d}^*_{t'+1})|X,W_{t'-1}\right]=\Pr\left(\overline{Y}_{t'}^{\overline{d}^*_{t'}}=0,\overline{D}_{t'}=\overline{d}_{t'}|X,W_{t'-1}\right). \tag{A.9}$$

Then taking the probability limit of (5)

$$p\lim_{N\to\infty}\widehat{ATET}_{t,t'}(\overline{d}_t,\overline{d}_t^*) = \tag{A.10}$$

$$\prod_{k=t'}^{t}\left[1-\frac{\mathbb{E}\left[w_{k,t'}^{\overline{d}_k}(\overline{d}_{t'})Y_{k,i}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(\overline{D}_{k,i}=\overline{d}_k)\right]}{\mathbb{E}\left[w_{k,t'}^{\overline{d}_k}(\overline{d}_{t'})\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(\overline{D}_{k,i}=\overline{d}_k)\right]}\right]-$$

$$\prod_{k=t'}^{t}\left[1-\frac{\mathbb{E}\left[w_{k,t'}^{\overline{d}_k^*}(\overline{d}_{t'})Y_{k,i}\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(\overline{D}_{k,i}=\overline{d}_k^*)\right]}{\mathbb{E}\left[w_{k,t'}^{\overline{d}_k^*}(\overline{d}_{t'})\mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(\overline{D}_{k,i}=\overline{d}_k^*)\right]}\right].$$

Using that the equations (A.1)-(A.9) the unconditional on X are the same too and similiar reasoning for $k=t'+1,...,t$ we have

$$p\lim_{N\to\infty}\widehat{ATET}_{t,t'}(\overline{d}_t,\overline{d}_t^*) =$$

$$\prod_{k=t'}^{t}\left[1-\frac{\Pr\left(Y_k^{\overline{d}_k}=1,\overline{Y}_{k-1}^{\overline{d}_{k-1}}=0,\overline{D}_{t'}=\overline{d}_{t'}|\mathbf{X},\pi_{t'-1}\right)}{\Pr\left(\overline{Y}_{k-1}^{\overline{d}_{k-1}}=0,\overline{D}_{t'}=\overline{d}_{t'}|\mathbf{X},\pi_{t'-1}\right)}\right]-$$

$$\prod_{k=t'}^{t}\left[1-\frac{\Pr\left(Y_k^{\overline{d}_k^*}=1,\overline{Y}_{k-1}^{\overline{d}_{k-1}^*}=0,\overline{D}_{t'}=\overline{d}_{t'}|\mathbf{X},\pi_{t'-1}\right)}{\Pr\left(\overline{Y}_{k-1}^{\overline{d}_{t'}^*}=0,\overline{D}_{t'}=\overline{d}_{t'}|\mathbf{X},\pi_{t'-1}\right)}\right]$$

$$\prod_{k=t'}^{t}\left[1-\Pr\left(Y_k^{\overline{d}_k}=1,|\mathbf{X},\overline{Y}_{k-1}^{\overline{d}_{k-1}}=0,\overline{D}_{t'}=\overline{d}_{t'}\right)\right]-$$

$$\prod_{k=t'}^{t}\left[1-\Pr\left(Y_k^{\overline{d}_k^*}=1,|\mathbf{X},\overline{Y}_{k-1}^{\overline{d}_{k-1}^*}=0,\overline{D}_{t'}=\overline{d}_{t'}\right)\right]=$$

$$\Pr(\overline{Y}_t^{\overline{d}_t}=0|\overline{D}_{t'}=\overline{d}_{t'},\overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0)-\Pr(\overline{Y}_t^{\overline{d}_t^*}=0|\overline{D}_{t'}=\overline{d}_{t'},\overline{Y}_{t'-1}^{\overline{d}_{t'-1}}=0)=ATET_{t,t'}(\overline{d}_t,\overline{d}_t^*).$$