IFAU
Institute for Evaluation of Labour Market and Education Policy

Evaluation of sequences of treatments with application to active labor market policies

Johan Vikström

WORKING PAPER 2015:5

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

Evaluation of sequences of treatments with application to active labor market policies[a]

by

Johan Vikström[b]

16[th] March, 2015

Abstract
This paper proposes a new framework for analyzing the effects of sequences of treatments with duration outcomes. Applications include sequences of active labor market policies assigned at specific unemployment durations and sequences of medical treatments. We consider evaluation under unconfoundedness and propose conditions under which the survival time under a specific treatment regime can be identified. We introduce inverse probability weighting estimators for various average effects. The finite sample properties of the estimators are investigated in a simulation study. The new estimator is applied to Swedish data on participants in training, in a work practice program and in subsidized employment. One result is that enrolling an unemployed person twice in the same program or in two different programs one after the other leads to longer unemployment spells compared to only participating in a single program once.

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

[a]I am grateful for helpful suggestions from Gerard J. van den Berg, Per Johansson, Martin Huber, Helena Holmlund, Oskar Nordström Skans, Ingeborg Waernbaum and seminar participants at IFAU Uppsala and the 5th joint IZA/IFAU conference.
[b]IFAU-Uppsala, and UCLS, E-mail: johan.vikstrom@ifau.uu.se

Table of contents

1 Introduction

Active labor market policies include a wide range of programs, such as training, practice programs, subsidized employment and job search assistance, and often the unemployed are enrolled in multiple programs during a single unemployment spell. For instance, in Sweden, almost 24% of the program participants participate in more than one program during one unemployment spell.[1] Unemployed workers also frequently participate in the same program more than once. Thus, from a policy perspective, evaluations of combinations of programs and sequences of the same program are of key interest, as these offer more comprehensive evidence than evaluations that only study the effects of the first program.

The effects of sequences of policies have been studied by, e.g., Lechner (2008, 2009), Lechner and Miquel (2010) and Lechner and Wiehler (2013).[2] These papers develop a framework for causal analysis of sequences of treatments, and propose and implement matching and inverse probability weighting (IPW) estimators for various average effects. One key aspect is that the causal comparison is between actually taking an entire sequence of treatments compared to taking another sequence (e.g., no treatment). The outcome of interest is the difference between two potential outcomes at a single point in time. This evaluation approach is very useful in a number of settings.

This paper also considers the identification and estimation of the effects of sequences of treatments, but in a different setting than Lechner (2008, 2009) and Lechner and Miquel (2010). We consider evaluation when the outcome under study is the time in an initial state and when the treatments are only offered while the individuals are in the initial state. For active labor market policy (ALMP) programs, this corresponds to an evaluation of the effects of programs for the unemployed on time in unemployment. One sequence of treatments could, for instance, be to enroll unemployed individuals in training after six months of unemployment and those still unemployed after twelve months in subsidized employment. Another example is multiple episodes of the same program. Other applications might concern comparisons between short and long programs and sequences of

[1] This number is conditional on participating at least once in a program during the time period 1999-2006.

[2] There is also a parallel epidemiological literature, see e.g. Robins (1986, 1998) and Hernan et al. (2001).

medical treatments.

The fact that the outcome of interest is the time in unemployment and the fact that treatments are only given to those who are still unemployed have several implications. First, since the treatments are only given to unemployed individuals, we argue that it is useful to compare treatment regimes in terms of protocols stipulating when each treatment should start. In other words, the effect of interest is the effect on time in unemployment of being enrolled in and following a specific sequence until employment is found, and not the effect of staying unemployed until the entire treatment sequence is finished. One advantage is that such treatment protocols are highly policy relevant since they could be manipulated by policymakers.

Another advantage is that it captures the full effect of the sequence of treatments. As an illustration, consider the comparison between two periods of training and no treatment in both periods. The two periods of training could affect the two-period survival rate in unemployment in several ways. For instance, the first period of training might affect the job finding rate in both the first and the second period, so that some unemployed individuals leave unemployment after the training in the first period. The framework proposed in this paper captures such first period effects as well as any effects in the second period, because we study the effects of a treatment sequence that is followed until employment is found. As a comparison, note that any analyses using only those who do not find employment before the end of the entire sequence by definition disregard the effects in the first period.

Moreover, in a duration setting, the identification of the causal effect is complicated by the fact that the potential outcome at time t can only be observed for individuals conditional on that (s)he has survived up to time $t-1$. This dynamic selection problem implies, for instance, that the difference in hazard rates cannot be identified without model assumptions since conditioning jointly on the outcome and a counterfactual treatment at time $t-1$ is not possible (van den Berg, 2001). For that reason, this paper focuses on average effects on the survival rate.

The dynamic selection problem also affects the causal comparison. In a duration setting, any analyses that compare entire sequences of treatments run into problems with dynamic selection. The reason is that conditioning the sample on individuals completing an entire sequence implies conditioning on survival under a specific series of treatments, i.e. conditioning on future outcomes. Thus, comparing two entire sequences implies comparing two different groups of survivals and following the reasoning of van den Berg (2001), this creates a difficult dynamic selection problem. The problem is particularly severe if yoy compares very long treatment sequences. This is another important reason why this paper argues that one useful causal comparison is between treatment regimes in terms of protocols stipulating when each treatment should start, because as shown below this deals with the dynamic selection problem.

Besides proposing a framework for analyzing the effects of sequences of treatments, henceforth referred to as a treatment regime, when the outcome is a duration, this paper contributes in several other ways. We study both treatment regimes defined by sequences of the same treatment and sequences defined by multiple types of programs. We study the estimation and identification of the difference in survival under two treatment regimes, and show that this parameter is identified under an assumption of unconfoundedness among survivors conditional on a set of observed covariates.

We also propose IPW estimators for various average effects. Initially, results are provided under selection on time-invariant individual characteristics and later on these results are extended to allow for selection on time-varying characteristics. The relevant assumption is that conditional on survival up until a certain time period, treatment assignment in that time period is random conditional on the possibly time-varying covariates. Standard errors could be obtained by bootstrapping. The finite sample properties of the estimator are investigated in simulations for a simple two-period setting by showing results for the estimator's bias, variance and size. Here, a naive estimation approach not adjusting for any covariates results is compared to the weighting estimator. All this partly builds on the work by Vikström (2014), which considers the effects of a single treatment on the survival rate in an initial state, and also studies IPW estimation.

As an illustration of the estimator, re-consider the comparison of two periods of training (treatment regime) and no treatment (comparison regime). In the first period, under unconfoundedness, the re-employment rate under the treatment (comparison) regime is obtained by weighting the outcome responses of the training participants (non-treated) in the first period in order to mimic the distribution of the confounders among the population of interest. In the second period, we need to capture the re-employment rates for those still unemployed under each regime, since the re-employment rate in the second period together with the re-employment rate in the first period give the two-period survival rate. One estimation problem is that only a subset of the training participants who are not re-employed in the first period participate in training in the second period, and this creates selection. However, under sequential unconfoundedness among survivors, we can correct for this selection by adding the second part of the weights. Similar weighting occurs for the comparison regime and the same reasoning applies to an arbitrary number of time periods. This yields a series of unique weights for each time period that gives the re-employment rate in each period under each treatment regime, and together these re-employment rates provide the survival rates under each regime.

In order to illustrate the type of questions that could be evaluated within our framework, we apply our estimator to Swedish population register data on a work practice program, a training program and a subsidized employment scheme. We study both the effects of different sequences of work practice episodes and the effects of different combinations of the three programs. We analyze the impact of the timing of the first program as well as the spacing between the first and the second program in the analyzed sequence. In the analysis of sequences of work practice episodes, we examine the effects of specific combinations of the first entry time and the second entry time with no work practice as comparison regime. We find that early first-time enrollment between 91 and 270 days leads to a significant increase in re-employment rates. For later enrollments, we find no significant effects. We also find that enrolling an unemployed individual in the program twice has negative effects on the re-employment rates compared to only participating once. This holds regardless of whether the second work practice episode starts shortly after or several months after the first program episode. Our detailed analysis of different sequences of program episodes shows that this is because a second program episode shortly after the first is associated with both greater negative locking-in effects and greater positive post-program effects, and on average these two effects cancel each other out.

The analysis of different combinations of training, work practice and subsidized employment also produces a number of interesting results. Both the training program and the subsidized employment scheme are associated with substantially greater locking-in effects than the work practice program, which is explained by longer program durations. For the training program this is counteracted by a greater post-program effect, so that the total effect of training and work practice is about the same. For subsidized employment the timing is important. Enrollment before 225 days is mainly associated with large locking-in effects and small post-program effects, while enrollment after 270 days could lead to a reduction in total unemployment. From the analysis of sequences of programs, we conclude that in most cases a second program episode does not reduce the total unemployment more than a single program episode. This holds both when the first program consists of work practice and when it consists of training. It also holds for most timings of the first program and most spacings between the two program episodes. One exception is that for some pre-treatment durations, work practice straight after training is beneficial. One reason could be that the work practice episode allows the training participant to gain experience quickly in the occupation for which (s)he has been trained for.

The paper proceeds as follows. Section 2 introduces the theoretical framework and assumptions. Section 3 describes the IPW estimator. Section 4 presents a simulation study highlighting the properties of the proposed estimator while Section 5 presents the application. Section 6 concludes.

2 The causal framework

We consider the identification and estimation of average effects on survival time in an initial state. We assume discrete time points, $t = 1, 2, \ldots$.[3] In each time period there are M mutually exclusive treatments and no treatment. Treatment in time period t is denoted by D_t with realized values $d_t \in \mathscr{D}_t$ where $\mathscr{D}_t$ is the set $\{0, M\}$. We use the notation $\overline{d}_t$ for a particular sequence of treatments. The random sequence consisting of treatment regimes at all time points, $t = 1, 2, \ldots$, is denoted $\overline{D}$. Subsequently, the notation $\overline{A}_t$ is used

[3] Defining causal effects in continuous time introduces several technical issues that would distract from the conceptual issues (Gill and Robins, 2001).

to denote a sequence $\overline{A}_t = \{A_1, \ldots, A_t\}$.

For each time period we consider a binary potential outcome $Y_t^{\overline{d}}$, an indicator of a transition in period t if the treatment regime had been $\overline{d} \in \overline{\mathscr{D}}$, where $\overline{\mathscr{D}}$ is the domain of possible realizations of the full treatment regime $\overline{D}$. Since for each t a potential outcome $Y_t^{\overline{d}_t}$ is defined, we have that $\overline{Y}_t^{\overline{d}}$ is the sequence of potential outcomes $\overline{Y}_t^{\overline{d}} = \{Y_1^{\overline{d}}, \ldots, Y_t^{\overline{d}}\}$ under treatment regime d. We write $\overline{Y}_t^{\overline{d}} = 0$ for the event of survival up to and including t under regime d. For instance, if Y is an indicator of a transition out of unemployment this is the event of remaining unemployed for at least t time periods. Similar notations are used for other variables.

Throughout the paper we assume that the potential outcome $Y_t^{\overline{d}} = Y_t^{\overline{d}_t}$, that is the potential outcome at time t, is the same for the same realized sequence of treatments up until t regardless of what the future assignments are at times $t+1, t+, \ldots$. This means that future treatments do not affect current outcomes, since the outcome in period t is the same for all treatment regimes with the first t components equal to $\overline{d}_t$ regardless of treatments after t. Abbring and van den Berg (2003) call this the no-anticipation assumption, and it is also discussed by, e.g., Abbring and Heckman (2008). It holds if individuals are unaware of future treatments or if they do not alter their behavior as a response to knowledge of future treatments.

The observed outcome, Y_t, corresponds to the individuals' actual treatment regime

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

where $I()$ is an indicator function. For clarification, note that in survival time settings it is common to assume that $Y_t^{\overline{d}_t} = 1$ is an event leading to an absorbing state. This means that for individuals where $Y_{t'}^{\overline{d}_{t'}} = 1$ we have that for all $t > t'$, $Y_t^{\overline{d}_t}$ is no longer observed. As an illustration, if an individual leaves unemployment at t, the observed data do not contain any information on transitions out of unemployment after t.

Finally, X is the baseline covariates observed for all individuals and measured at time origo. Initially, they are assumed to be time-invariant. Later on we will allow for time-varying covariates.

2.1 Causal effects

In this survival time setting, there are several interesting average effects of sequences of treatments. Initially, consider a two-period case and the comparison of two periods with the first treatment (11) and two periods of no treatment (00). The main points of this Section can be illustrated in this setting. We discuss the extension to the general case at the end of this Section. One alternative is to study the average effect of actually taking a full sequence of treatments, here called conditional average treatment effects. In the second period, one conditional average effect is

$$\text{CATET}_2(11,00) = \tag{2}$$

$$E(Y_2^{11}|D_2 = 1, D_1 = 1, Y_1^1 = 0) - E(Y_2^{00}|D_2 = 1, D_1 = 1, Y_1^1 = 0)$$

i.e., the average effect on the outcome in the second period for those actually treated in both periods. It is important to note that those actually treated in both periods consist of those treated in the first period ($D_1 = 1$) who remain in the initial state ($Y_1 = 0$) and who are subsequently treated in the second period ($D_2 = 1$). This is an interesting and valid treatment effect as it compares two potential outcomes for the same population. However, one problem with the CATET is that it essentially ignores the effect of the treatment in the first period. The reason is that it contrasts potential outcomes in the second period for a group that remain in the initial state until the second period, so that any differential survival in the first period is ignored. Another important issue concerns the identification and estimation of CATET. This is because CATET is a conditional transition rate for a group that survives under treatment in the first period. It is well known that the identification of such conditional transition rates is difficult (see e.g. van den Berg, 2001). Let us discuss this in detail under the assumption that both treatment in the first period and treatment in the second period are randomly assigned. Through randomization we have from the observed outcomes in the second period

$$E(Y_2|D_2 = 1, D_1 = 1, Y_1 = 0) = E(Y_2^{11}|D_2 = 1, D_1 = 1, Y_1^1 = 0) \tag{3}$$

$$E(Y_2|D_2=0,D_1=0,Y_1=0)=E(Y_2^{00}|D_2=0,D_1=0,Y_1^0=0) \tag{4}$$

where the left-hand sides of (3) and (4) involve only observed variables. From (3) we obtain the second period outcomes for those who are actually treated in both time periods. From (4) we obtain the outcome under two periods of no treatment for those surviving the first period without treatment, but without further assumptions this does not identify the second term on the right-hand side of (2). The reason is that the second term of (2) is for survivors in the first period under treatment whereas the identified average from (4) is for survivors under treatment in the first period. Note that both these issues arise since the treatments are given in a survival time setting. In a survival time setting outcomes as well as treatments are unobserved in the second period for those with a transition in the first period. The insights and estimators in this paper are thus only applicable to survival settings, such as ALMP programs for unemployed individuals and many medical treatments.

For these reasons, we focus on average effects on the survival rate and the comparison of the probability to survive from a starting point to time t compared to survival throughout the same time interval under a reference treatment regime. In the two-period case to a comparison of two periods of the first treatment with no treatment we have

$$\text{SATE}(11,00)= \tag{5}$$

$$\Pr(Y_2^{11}=0,Y_1^1=0)-\Pr(Y_2^0=0,Y_1^0=0),$$

i.e. the average effect of treatment in the first period and for the first-period survivals also treatment in the second period compared to no treatment in a similar way. This takes into account any differential survival due to treatment in the first period as well as any differential survival in the second period, because for those who have already exited in the first period the joint probability of survival in both periods is by definition zero. The effect should thus be interpreted as the effect of a sequence given to those still in the initial state, and not the effect of actually surviving and taking an entire sequence of treatments, that is, a comparison of treatment in the first period and then for survivors also treatment in the second period. The effects are, as shown below, also identified under a single unconfoundedness assumption. You could also study average effects for specific subpopulations, such as the average effect on those treated in the first period.

In the general case we have the contrast between the two treatment regimes $\overline{d}_t$ and $\overline{d}_t^*$ that could be any sequences of the M treatments:

$$\mathrm{SATE}(\overline{d}_t, \overline{d}_t^*) = \Pr(\overline{Y}_t^{\overline{d}_t} = 0) - \Pr(\overline{Y}_t^{\overline{d}_t^*} = 0) \tag{6}$$

The results of the paper concern the identification and estimation of this parameter of interest. Contrasting regimes with respect to a starting point, $t > 1$ could also be considered for a population that follows the same treatment sequence in the time points before the new starting point of interest.

2.2 Identification

We consider the identification of $\mathrm{SATE}(\overline{d}_t, \overline{d}_t^*)$ if we have data on the selection to treatment, and as such 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 longitudinal settings, you may impose several different unconfoundedness assumptions, see e.g. the discussion in Lechner and Miquel (2010). In this paper, we assume that unconfoundedness holds among survivors

A.1 For all t, $\overline{d}_{t-1}$ and all $\overline{d}_s^*$, $s \geq t$ with the first $t-1$ components equal to $\overline{d}_{t-1}$

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

That is, for individual surviving up until t under regime $\overline{d}_{t-1}$ treatment assignment in period t is random conditional on the observed covariates. Assumption $A.1$ holds in all situations in which decisions are made sequentially based on the survivor experience up to a certain point in time, for instance, if case workers assign unemployed individuals to ALMP programs based on time in unemployment and a set of observed covariates.

Let us compare assumption $A.1$ to the weak dynamic conditional independence assumption (W-DCIA) in Lechner and Miquel (2010), which implies that conditional on treatments up until t and observed covariates the potential outcomes are independent of treatments in period t. Importantly, the observed covariates might include observable outcomes in previous periods. Since $A.1$ is conditional on survival up until t (the outcome) and treatments up until t one similarity between the W-DCIA and $A.1$ is that both assumptions are about independence conditional on treatments and outcomes in previous periods. However, in settings with survival outcomes the two assumptions have different practical implications. In Lechner and Miquel (2010) the sample consists of individuals who remain in the initial state so that they experience the full sequence of the treatments. In the first period this means that treatment assignments should be unconfounded in a sample that is constructed conditional on future treatments, and this might lead to violations of the W-DCIA. Note that this concern only applies to survival settings, such as evaluations of the effect of ALMP programs. The causal framework in this paper and assumption $A.1$ do not imply any conditioning on future outcomes.[4]

We also make an overlap assumption

A.2 For all t, $\overline{d}_{t-1}$ and d_t

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

Our main identification result is summarized in Theorem 1.

Theorem 1 (Identification of ATE) *Suppose that* $A.1$ *and* $A.2$ *hold then*

$$\mathrm{SATE}(\overline{d}_t, \overline{d}_t^*) =$$

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

Proof See appendix.

The identification follows from the fact that the probability of the observed survival of the individuals conditional on the covariates and the treatment regime can be used to estimate the corresponding probability for the potential outcomes under the same regime.

[4]Lechner and Miquel (2010) also allow the covariates to include time-varying confounders. Note that in Section 3.3 the results are extended to allow for selection on time-varying covariates.

3 Estimation

In this Section, we consider the estimation of SATE$(\overline{d}_t, \overline{d}_t^*)$. We introduce a IPW estimator, which estimates the outcomes under treatment regime $\overline{d}_t$ (or $\overline{d}_t^*$) using those who follow the treatment regime $\overline{d}$ (or $\overline{d}^*$) up until t but not necessarily after t. This follows from the same reasoning as in the section on identification. We have

$$\widehat{\text{SATE}}(\overline{d}_t, \overline{d}_t^*) = \prod_{k=1}^{t}\left[1 - \frac{\sum_{i=1}^{N} \hat{w}_{\overline{d}_{k,i}} Y_{k,i} \mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(\overline{D}_{k,i}=\overline{d}_k)}{\sum_{i=1}^{N} \hat{w}_{\overline{d}_{k,i}} \mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(\overline{D}_{k,i}=\overline{d}_k)}\right] - \tag{7}$$

$$\prod_{k=1}^{t}\left[1 - \frac{\sum_{i=1}^{N} \hat{w}_{\overline{d}^*_{k,i}} Y_{k,i} \mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(\overline{D}_{k,i}=\overline{d}^*_k)}{\sum_{i=1}^{N} \hat{w}_{\overline{d}^*_{k,i}} \mathbf{1}(\overline{Y}_{k-1,i}=0)\mathbf{1}(\overline{D}_{k,i}=\overline{d}^*_k)}\right]$$

with the estimated weights

$$\hat{w}_{\overline{d}_{k,i}} = \frac{1}{\hat{p}_{d_1}(X_i)\prod_{m=2}^{k}\hat{p}_{d_m}(X_i,\overline{d}_{m-1})},\ \hat{w}_{\overline{d}^*_k} = \frac{1}{\hat{p}_{d^*_1}(X_i)\prod_{m=2}^{k}\hat{p}_{d^*_m}(X_i,\overline{d}^*_{m-1})}.$$

Here, $p_{d_t}(X,\overline{d}_{t-1})$ are propensity scores, i.e. the probability to obtain treatment d_t in period t given survival up until t under treatment regime $\overline{d}_{t-1}$ and covariates X. We have

$$p_{d_t}(X,\overline{d}_{t-1}) = Pr(D_t = d_t | X, \overline{D}_{t-1} = \overline{d}_{t-1}, \overline{Y}.$$

In practice, the true propensity scores, $p_{d_t}(X,\overline{d}_{t-1})$, are replaced by estimated propensity scores, $\hat{p}_{d_t}(X,\overline{d}_{t-1})$.

In the appendix we show that these weights allow us to the estimate the SATE$(\overline{d}_t, \overline{d}_t^*)$. It could be considered as a weighted Kaplan-Meier estimator, since the numerator weights the observed exits in a certain period and the denominator weights the risk set in the same period. This produces the discrete transition rates in each period, and taken together they constitute the survival rates under the two treatment regimes. Individuals are used in the estimation as long as they follow the treatment regime of interest, and at the point of divergence they are regarded as censored. The weights account for this selective censoring. Note that the weights are normalized by construction.

Let us discuss the intuition behind the weights. Take the case in which $\overline{d}_t$ is a sequence of ones and $\overline{d}_t^*$ is a sequence of zeros, i.e. a comparison of always being treated against never being treated. Then, for $t = 1$ we have

$$w_{d_1} = \frac{1}{p_{d_1}(X)} = \frac{1}{\Pr(D_1 = 1|X)}, \; w_{d_1^*} = \frac{1}{p_{d_1^*}(X)} = \frac{1}{1 - p_{d_1}(X)} = \frac{1}{1 - \Pr(D_1 = 1|X)},$$

i.e. the standard IPW average treatment effect weights from the static evaluation literature. Naturally, in the first period, no selective censoring has occurred so the only purpose of the weights is to re-weight the outcomes of the individuals on $\overline{d}_t$ and $\overline{d}_t^*$ in order to mimic the distribution of the covariates in the full population. Next, for $t = 2$ we have

$$w_{\overline{d}_2} = \frac{1}{\Pr(D_1 = 1|X)\Pr(D_2 = 1|X, D_1 = 1, Y_1 = 0)},$$

and

$$w_{\overline{d}_2^*} = \frac{1}{[1 - \Pr(D_1 = 1|X)]\Pr(D_2 = 0|X, D_1 = 0, Y_1 = 0)}$$

Here, the weights serve two purposes. Besides re-weighting in order to mimic the distribution of the covariates in the full population as for $t = 1$, the weights also correct for the selective censoring due to treatment assignment in period 2. Specifically, the weights depend on the inverse probability of remaining on the treatment regime of interest conditional on the observed covariates, so that individuals with covariates such that they are more likely to diverge from the regime of interest are given a greater weight. Finally, note that the weights given to individuals change over time, since the weights depend on the entire censoring up until a specific period.

One way to obtain standard errors is to use bootstrapping. Since the selection probabilities, $p_{d_t}(X, \overline{d}_{t-1})$, and the weights are re-estimated in each bootstrap replication, this accounts for variation in both the estimation of weights and the outcome equation. If the selection probabilities are estimated using parametric models, such as logit or probit, the joint estimation problem could be expressed in a GMM framework. The finite sample properties of the estimator are explored in a Monte Carlo simulation presented in Section 4.

3.1 Trimming

It is well known that IPW estimation might be sensitive to extreme values of the propensity scores, since then single observations might receive too large weight (see e.g., Frölich 2004, Busso et al. 2009, Huber et al. 2013). One way to overcome this problem is trimming, i.e. removing observations with extreme values of the weights. There are several different trimming approaches. For instance, Crump et al. (2009) conclude that for many applications a reasonable rule of thumb is to remove observations with a propensity score outside the range [0.1, 0.9]. More recently, Huber et al. (2013) proposes a three-step approach for the average treatment effect on the treated. For average treatment effects their approach implies setting the weights to zero for all treated (controls) whose share of the sum of all weights in the treatment (control) group is greater than t% (e.g., 4%). Thereafter, normalize the weights again and finally discard observations whose propensity score is smaller (greater) than the maximum (minimum) of the minimum (maximum) scores among the treated and controls. In this paper, the weights are a function of several propensity scores and the weights given to a certain individual change with the survival time. We, therefore, consider a slightly modified version of the trimming approach in Huber et al. (2013). Firstly, using the t% rule, obtain the cut-off values $\overline{w}_{\overline{d}_k}$ and $\overline{w}_{\overline{d}_k^*}$ for all k. Then, compute the weights in both treatment regimes for all k and only use individuals whose weights are below $\overline{w}_{\overline{d}_k}$ and $\overline{w}_{\overline{d}_k^*}$ in all time periods. This assures that extreme values are discarded and that the same type of individuals are discarded in both treatment arms. This approach is similar to the one used by Lechner (2009).

3.2 Censoring

The results in the previous Sections hold if there is no regular right censoring of the survival time in the initial state. In many applications, this is a restrictive assumption since many ongoing spells are censored, for example, due to dropout from the study. Here, we allow censoring that occurs at any point in time. Let C_t be a censoring indicator for period t and we denote $\overline{C}_t = \{C_1, \ldots, C_t\}$. We assume that censoring occurs at the beginning of the discrete time period, i.e., before treatment assignment.

We consider the case when the censoring is allowed to depend on a set of observed covariates (as before denoted by X). We call this independent censoring among survivors conditional on observed covariates:

A.3 For all t, $\overline{d}_{t-1}$ and all $\overline{d}^*_s$, $s \geq t$ with the first $t-1$ components equal to $\overline{d}_{t-1}$

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

Note that this assumption resembles the sequential unconfoundedness assumption among survivors. The only difference is that that the former concerns treatment assignment and assumption $A.3$ concerns censoring. In applications on the effects of ALMP programs censoring might, for instance, occur if the employment office loses contact with some unemployed persons or if there are calendar time restrictions. Assumption $A.3$ implies that all such censoring should be random conditional on the observed covariates. For completeness, we also need to assume that the censoring rate in all time periods is less than one for all $\overline{d}_t, t$ (assumption A.4)

$$\Pr(C_t = 1 | X, \overline{D}_{t-1} = \overline{d}_{t-1}, \overline{Y}_{t-1} = 0, \overline{C}_{t-1} = 0) = c_t(X, \overline{d}_{t-1}) < 1.$$

In the appendix we show that $\text{SATE}(\overline{d}_t, \overline{d}^*_t)$ is identified under assumptions $A.1 - A.4$. Concerning estimation, the censoring introduces an additional dimension as it implies selective divergence from the treatment regime of interest and due to censoring. This means that we also have to weight with the inverse probability of being censored in each time period. We have the following weighted estimator:

$$\widehat{\text{SATE}}(\overline{d}_t, \overline{d}^*_t) = \tag{8}$$

$$\prod_{k=t'}^{t} \left[1 - \frac{\sum_{C_1=0} \hat{w}^C_{\overline{d}_k}(d_1) Y_{k,i} \mathbf{1}(\overline{Y}_{k-1,i} = 0) \mathbf{1}(\overline{D}_{k,i} = \overline{d}_k) \mathbf{1}(\overline{C}_{k,i} = 0)}{\sum_{C_1=0} \hat{w}^C_{\overline{d}_k}(d_1) \mathbf{1}(\overline{Y}_{k-1,i} = 0) \mathbf{1}(\overline{D}_{k,i} = \overline{d}_k) \mathbf{1}(\overline{C}_{k,i} = 0)} \right] -$$
$$\prod_{k=t'}^{t} \left[1 - \frac{\sum_{C_1=0} \hat{w}^C_{\overline{d}^*_k}(d_1) Y_{k,i} \mathbf{1}(\overline{Y}_{k-1,i} = 0) \mathbf{1}(\overline{D}_{k,i} = \overline{d}^*_k) \mathbf{1}(\overline{C}_{k,i} = 0)}{\sum_{C_1=0} \hat{w}^C_{\overline{d}^*_k}(d_1) \mathbf{1}(\overline{Y}_{k-1,i} = 0) \mathbf{1}(\overline{D}_{k,i} = \overline{d}^*_k) \mathbf{1}(\overline{C}_{k,i} = 0)} \right]$$

where

$$\hat{w}^{C}_{\overline{d}_k} = \frac{1}{\hat{p}_{d_1}(X)\prod_{m=2}^{k}\hat{p}_{d_m}(X,\overline{d}_{m-1})\prod_{m=2}^{k}\hat{c}_m(X,\overline{d}_{m-1})}$$

and

$$\hat{w}^{C}_{\overline{d}^*_k} = \frac{1}{\hat{p}_{d^*_1}(X)\prod_{m=2}^{k}\hat{p}_{d^*_m}(X,\overline{d}^*_{m-1})\prod_{m=2}^{k}\hat{c}_m(X,\overline{d}^*_{m-1})}.$$

In the appendix we show that this estimator solves the selection problem.

3.3 Selection on time-varying covariates

So far, we have considered estimation and inference 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 the assignment to treatment when conditioning on *time-invariant* covariates. We now allow for selection on *time-variant* observed characteristics. In order to rule out effects of D_t on X the covariates determining treatment assignment in t should be measured before assignments are made. For that reason we use the notation X_{t^-} for the observed covariates at t, where t^- indicates that X is measured at least slightly before t. We consider identification and estimation under a generalization of the sequential unconfoundedness among survivors assumption **A.5** For all t, $\overline{d}_{t-1}$ and all $\overline{d}^*_s$, $s \geq t$ with the first $t-1$ components equal to $\overline{d}_{t-1}$

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

The only difference compared to assumption $A.1$ is that treatment assignments in a certain period are allowed to depend on the covariate values in that time period instead of on the covariates measured at the start of the spell. Note that X_t could include the covariate values in the previous time period (e.g X_{t-1}). Concerning evaluations of ALMP programs, some unemployed persons might get divorced and/or experience a deterioration in health during an unemployment spell, and such time-varying characteristics might on impact program enrollment.

In the appendix we show that SATE$(\overline{d}_t, \overline{d}^*_t)$ is identified under assumption $A.5$ and an overlap condition similar to the one reported in assumption $A.2$. Here let us illustrate using a two-period illustration. Note that a sufficiently large sample provides information on the outcomes under each treatment regime for each combination of X_1 and X_2. We denote these averages as $Y(\overline{d}_1,X_1)$ and $Y(\overline{d}_2,X_1,X_2)$. Under assumption $A.5$ these averages are informative about the average potential outcomes under each treatment regime. The observed data also gives the distribution of X_1 in the full population, $f_1(X_1)$, and the distribution of X_2 conditional on X_1 among the survivors under each treatment regime, $f(X_2|X_1,\overline{d}_1)$. It is then possible to use the probability distributions $f_1(X_1)$ and $f(X_2|X_1,\overline{d}_1)$ to align the average outcomes $Y(\overline{d}_1,X_1)$ and $Y(\overline{d}_2,X_1,X_2)$ in order to obtain the desired average potential outcomes under each treatment regime.

In terms of estimation, the estimator given by (7) still holds but now with the following weights

$$\hat{w}_{\overline{d}_{k,i}} = \frac{1}{\hat{p}_{d_1}(X_{i1})\prod_{m=2}^{k}\hat{p}_{d_m}(X_{it},\overline{d}_{m-1})},\ \hat{w}_{\overline{d}^*_k} = \frac{1}{\hat{p}_{d^*_1}(X_{i1})\prod_{m=2}^{k}\hat{p}_{d^*_m}(X_{it},\overline{d}^*_{m-1})}.$$

where

$$p_{d_t}(X_t,\overline{d}_{t-1}) = \Pr(D_t = d_t|X_t,\overline{D}_{t-1} = \overline{d}_{t-1},\overline{Y}_{t-1} = 0). \tag{9}$$

In the appendix we explore the properties of this estimator. The weights have the same structure as for time-invariant confounders. The only difference is that the propensity score at each t depends on the covariate values at t. Essentially, with only time-invariant covariates, the probability of remaining in a specific regime depends on the duration time and these time-invariant characteristics. In cases with time-variant characteristics, this probability depends instead on the covariate values in each time period. Otherwise the reasoning behind the weights is the same.

4 Simulations

We investigate the small-sample properties of our estimator for a two-period setting with a single treatment. The units may be either treated or non treated in the first period and units that survive the first period may be either treated or non treated in the second period. We focus on ATE$(1,0)$, ATE$(10,00)$ and ATE$(11,00)$ and generate the potential outcomes of all sequences of treatments, $D_1 = 0$, $D_1 = 1$, $\overline{D}_2 = 0,0$, $\overline{D}_2 = 1,0$ and $\overline{D}_2 = 1,1$. Let us use the notation $\theta_{Y_1^0} = \Pr(Y_1(0) = 1|X_1, X_2, V_Y)$, $\theta_{Y_1^0} = \Pr(Y_1(1) = 1|X_1, X_2, V_Y)$, $\theta_{Y_2^{00}} = \Pr(Y_2(0,0) = 1|X_1, X_2, V_Y)$, $\theta_{Y_2^{10}} = \Pr(Y_2(1,0) = 1|X_1, X_2, V_Y)$ and $\theta_{Y_2^{11}} = \Pr(Y_2(1,1) = 1|X_1, X_2, V_Y)$. Additionally, $\theta_{D_1} = \Pr(D_1 = 1|X_1, X_2, V_D)$ and $\theta_{D_2} = \Pr(D_2 = 1|X_1, X_2, V_D)$.

We consider three data generating processes: The first is our baseline model with similar selection into treatment in both periods. Model 2 allows for time-varying selection into treatment in the sense that one covariate determines treatment assignments in the first period and the other covariate determines assignments in the second period. Finally, model 3 allows for an interaction between the two covariates. This generates more variation in the actual selection probabilities, and allows us to test to what extent the IPW approach is robust to selection probabilities closer to zero and one. Specifically, we have:

DGP 1: Baseline model

$$\begin{aligned}
\theta_{Y_1^0} = \theta_{Y_2^{00}} = &\quad f(-1.5 + X_1 + X_2 + V_Y) \\
\theta_{Y_1^1} = \theta_{Y_2^{11}} = \theta_{Y_2^{10}} = &\quad f(-1.5 + \gamma + X_1 + X_2 + V_Y) \\
\theta_{D_1} = \theta_{D_2} = &\quad f(X_1 + X_2 + V_D) \\
\text{with} &\quad f(h) = [1 + \exp(-h)]^{-1},
\end{aligned}$$

DGP 2: Time-varying selection effect of covariates

$$\begin{aligned}
\theta_{Y_1^0} = \theta_{Y_2^{00}} = &\quad f(-1.5 + X_1 + X_2 + V_Y) \\
\theta_{Y_1^1} = \theta_{Y_2^{11}} = \theta_{Y_2^{10}} = &\quad f(-1.5 + \gamma + X_1 + X_2 + V_Y) \\
\theta_{D_1} = &\quad f(X_1 + V_D) \\
\theta_{D_2} = &\quad f(X_2 + V_D)
\end{aligned}$$

DGP 3: Interaction between covariates

$$\begin{aligned}
\theta_{Y_1^0} = \theta_{Y_2^{00}} = &\quad f(-1.5 + X_1 + X_2 + X_1X_2 + V_Y) \\
\theta_{Y_1^1} = \theta_{Y_2^{11}} = \theta_{Y_2^{10}} = &\quad f(-1.5 + \gamma + X_1 + X_2 + X_1X_2 + V_Y) \\
\theta_{D_1} = \theta_{D_2} = &\quad f(X_1 + X_2 + V_D)
\end{aligned}$$

We use four covariates (X_1, X_2, V_Y, V_D), where X_1, X_2 are assumed to be observed by the data analyst and V_D, V_Y are unobserved. All four are independent random variables drawn from a uniform distribution on the interval [-1,1]. The fact that X_1, X_2 affect both treatment assignment and the outcome causes the selection we have to deal with. Note that V_D and V_Y are independently distributed and that implies that the sequential unconfoundedness assumption holds in all three cases even though all models include unobserved effects. We either set $\gamma = 0$ or $\gamma = 1$; that is, we generate data with and without a treatment effect. Note that model 3 allows for a fairly strong interaction, which creates selection probabilities very close to zero and one. For model 3 the average minimum denominator is 0.029 in the first period and 0.0022 in the second period, while for model 1 these values are 0.159 and 0.030. For that reason we use the trimming rule described in Section 3.1. We either set t to 4% or to 100% for the untrimmed case. For models 1 and 2 the trimming is of minor importance, but for model 3 the largest average weight in the second period is reduced from 10.6% to 4% when the number of observations is 1200.

We generate samples of sizes 300, 1200 and 4800. This allows us to immediately asses whether the estimators are $\sqrt{N}$-convergent, since then the standard error should decrease with 50% when the sample size increases by a factor of four. The number of replications is 10000. The propensity scores are estimated under correct model specification. We obtain standard errors using bootstrapping (99 replications).

Table 1 reports the simulation results for ATE$(11,00)$. Panel A reports the results for the three models using a sample size of 1200 and three different levels of trimming. For comparison, we also present simulation results from a naive model, where we make no adjustments for the observed covariates. The bias of this naive estimator is presented in columns 1 and 6. As expected, the naive estimator is severely biased, and this holds for all models. From columns 2 and 6 we see that our weighted estimator reduces this bias to virtually zero and is more than 100 times smaller than the bias of the naive estimator.[5] For model 1 and 2 the results for the size of the test with a nominal size of 5% also show that our estimator has roughly the correct size. In all cases, the size is very close to 5%. For the model with an interaction effect (model 3), the untrimmed IPW estimator does

[5]Note that for presentation reasons all bias estimates have been multiplied by 100.

Table 1: Simulation results - ATE$(11,00)$

Panel A: No treatment effect (N=1200, $\gamma = 0$)

	Period 1					Period 2				
	bias naive [1]	bias [2]	size [3]	rmse [4]	se [5]	bias naive [6]	bias [7]	size [8]	rmse [9]	se [10]
Untrimmed										
DGP 1	-8.4	0.010	0.051	0.025	0.025	-15.9	-0.060	0.057	0.038	0.038
DGP 2	-4.5	-0.055	0.050	0.024	0.024	-8.9	-0.108	0.051	0.034	0.034
DGP 3	-21.9	-0.548	0.090	0.039	0.038	-32.2	-1.045	0.087	0.055	0.054
Trimming 4%										
DGP 1	-8.4	-0.015	0.051	0.025	0.025	-15.9	-0.066	0.051	0.038	0.037
DGP 2	-4.5	0.029	0.052	0.024	0.024	-8.9	0.045	0.054	0.034	0.034
DGP 3	-21.9	0.067	0.047	0.028	0.028	-32.2	0.135	0.048	0.044	0.044
Trimming 2%										
DGP 1	-8.4	-0.033	0.049	0.026	0.026	-15.9	-0.036	0.049	0.038	0.038
DGP 2	-4.5	-0.007	0.051	0.024	0.024	-8.9	-0.050	0.049	0.033	0.033
DGP 3	-21.9	0.118	0.045	0.029	0.029	-32.2	0.243	0.047	0.044	0.044

Panel B: Treatment effect (N=1200, $\gamma = 1$, Trimming 2%)

	Period 1					Period 2				
	bias naive [1]	bias [2]	size [3]	rmse [4]	se [5]	bias naive [6]	bias [7]	size [8]	rmse [9]	se [10]
DGP 1	-17.4	0.045	0.049	0.029	0.029	-22.1	0.125	0.049	0.041	0.041
DGP 2	-12.8	0.061	0.052	0.026	0.026	-14.7	-0.005	0.051	0.034	0.034
DGP 3	-28.4	0.069	0.046	0.034	0.034	-33.2	0.205	0.054	0.048	0.048

Panel C: Sample size (DGP 1, $\gamma = 0$)

	Period 1					Period 2				
	bias naive [1]	bias [2]	size [3]	rmse [4]	se [5]	bias naive [6]	bias [7]	size [8]	rmse [9]	se [10]
N=300	-8.4	0.019	0.046	0.052	0.052	-15.9	-0.035	0.050	0.077	0.077
N=1200	-8.4	0.010	0.051	0.025	0.025	-15.9	-0.060	0.057	0.038	0.038
N=4800	-8.4	-0.022	0.054	0.012	0.012	-15.9	-0.016	0.054	0.019	0.019

Note: Bias naive is bias from a naive model with no adjustment for covariates. Both bias naive and bias of our weighted estimator are multiplied by 100. Size is for 5% level tests. The bootstrap standard errors is based on bootstrap (99 replications). The results are based on 10,000 replications, and using a logistic simulation model.

not have correct size, as expected. The bias is also larger for the second time period. This is because the propensity scores enters multiplicativelly, and this creates more extreme weights. For this model the bias is also substantially greater than for models 1 and 2. However, the trimmed IPW estimator reduces this bias to virtually zero and also has roughly correct size. This holds for both t=2% and t=4%.

Panel B reports similar estimates for models with a substantial treatment effect. Concerning trimming, we set t=2% in all cases. These results are very similar to the results in Panel A for no treatment effect. Panel C of *Table 1* shows that our estimator performs well for all three sample sizes, and that the standard errors and mean squared error decrease

Table 2: Simulation results - ATE$(10,00)$

Panel A: No treatment effect (N=1200, $\gamma=0$)

	Period 1					Period 2				
	bias naive	bias	size	rmse	se	bias naive	bias	size	rmse	se
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Untrimmed										
DGP 1	-8.4	-0.038	0.051	0.025	0.025	-11.1	-0.020	0.055	0.038	0.038
DGP 2	-4.4	-0.027	0.052	0.024	0.024	-6.1	-0.044	0.052	0.035	0.035
DGP 3	-21.9	-0.461	0.085	0.039	0.039	-24.4	-0.610	0.076	0.052	0.052
Trimming 4%										
DGP 1	-8.4	0.016	0.046	0.025	0.025	-11.1	0.000	0.053	0.038	0.038
DGP 2	-4.4	-0.011	0.050	0.024	0.024	-6.1	-0.066	0.052	0.035	0.035
DGP 3	-21.9	0.039	0.051	0.026	0.026	-24.4	0.087	0.052	0.039	0.039
Trimming 2%										
DGP 1	-8.4	-0.023	0.049	0.024	0.024	-11.1	-0.012	0.048	0.037	0.037
DGP 2	-4.4	-0.043	0.054	0.024	0.024	-6.1	-0.030	0.053	0.035	0.035
DGP 3	-21.9	0.053	0.044	0.025	0.025	-24.4	0.102	0.046	0.038	0.038

Panel B: Treatment effect (N=1200, $\gamma=1$, Trimming 2%)

	Period 1					Period 2				
	bias naive	bias	size	rmse	se	bias naive	bias	size	rmse	se
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
DGP 1	-17.4	-0.023	0.051	0.028	0.028	-8.6	0.027	0.050	0.037	0.037
DGP 2	-12.8	0.008	0.050	0.026	0.026	-2.9	0.012	0.051	0.034	0.034
DGP 3	-28.3	0.081	0.082	0.030	0.030	-20.1	0.123	0.048	0.040	0.040

Panel C: Sample size (DGP 1, $\gamma=0$)

	Period 1					Period 2				
	bias naive	bias	size	rmse	se	bias naive	bias	size	rmse	se
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
N=300	-8.2	0.080	0.048	0.050	0.050	-11.0	0.115	0.052	0.076	0.076
N=1200	-8.4	-0.038	0.051	0.025	0.025	-11.1	-0.020	0.055	0.038	0.038
N=4800	-8.3	0.000	0.054	0.013	0.013	-11.1	-0.033	0.053	0.019	0.019

Note: Bias naive is bias from a naive model with no adjustment for covariates. Both bias naive and bias of our weighted estimator are multiplied by 100. Size is for 5% level tests. The bootstrap standard errors is based on bootstrap (99 replications). The results are based on 10,000 replications, and using a logistic simulation model.

quite rapidly as the sample size increases from 300 to 1200 and up to 4800. The size of the standard errors decreases by roughly 50 percent when the sample size increases by a factor of four. Finally, *Table 2* reports simulation results for ATE$(10,00)$, which all are very similar to the results for ATE$(11,00)$.

5 Application to Swedish ALMP

This Section studies the effects of a work practice program, a training program and subsidized employment programs governed by the Swedish public employment service (PES). These are the three main types of programs for unemployed workers in Sweden. Initially, we study the effects of different sequences of work practice episodes. Specifically, we explore treatment regimes defined by a specific combination of the first and the second entry time. For instance, one treatment regime is to enroll those who are still unemployed after five months in work practice and then those who are still unemployed after seven months in work practice once more. By studying a wide range of different treatment regimes, we aim to answer a number of questions: What is the overall effect of the work practice program? Is early enrollment better than late enrollment? What are the effects of being assigned work practice a second time and does this effect depend on the time that has passed since the first episode?

Thereafter we explore different combinations of work practice, training and subsidized employment programs. We compare different sequences, such as first training and then work practice, first work practice then training and so on. We also explore how much the timing of the first program matters and the effect of the spacing between two program episodes. This allows us to identify the situations in which it is most beneficial to enroll an unemployed individual in more than one program during the same unemployment spell.

5.1 The programs

The aim of the *work practice* program is to provide long-term unemployed individuals with practical experience and employer contacts in order to maintain and strengthen their productivity. The participants should perform regular tasks at regular firms, even though they are not employed by them. The work practice can take place at both private and public employers. The duration of the program duration does not normally exceed six months, and is in fact usually much shorter. Participants receive a grant for their participation in this and the other programs. Those who are entitled to unemployment insurance (UI) benefits receive a grant equal to their UI benefits.

The main purpose of the *training* program is to improve the skills of the unemployed and thereby enhance their chances of obtaining a job. The contents of the courses should be directed towards the upgrading of skills or the acquisition of skills that are in short supply or that are expected to be in short supply. These could be computer skills, technical skills, manufacturing skills, and skills in services and medical health care.

The *subsidized employment* scheme includes several different programs that all subsidize the labor costs for a longer period of time. This includes both hiring subsidies targeted at public and private employers as well start-up subsidies. The hiring subsidy programs consist of the general employment subsidy that subsidizes 50% of the total wage costs for a maximum of 6 months and the extended employment subsidy at a subsidy level of 75% for the first 6 months and then 25% for a further 18 months.[6] The start-up subsidies are aimed at unemployed individuals who are considered to be well placed to engage in business. The participants receive a grant equal to their UI benefits while participating and setting up their own business (maximum 6 months).

5.2 Data and estimation details

The population is taken from the register Händel administrated by the PES, which includes all job seekers in Sweden. The register contains daily information on the time when an individual (i) became unemployed, (ii) entered into a labor market program and (iii) exited from unemployment. It also includes information on the reason for the exit (employment, education, social assistance, disability or sickness insurance programs and lost contact), and personal characteristics recorded at the beginning of the unemployment spell. To this data we merge information on marital status, household characteristics (e.g. number of children), labor income and income from various insurance schemes (e.g. sickness and disability) from the population register LOUISE. We also use the unemployment records to construct detailed information on past labor market history (e.g. number and length of previous spells).

We sample all unemployed individuals in Händel between January 1, 1999 to December 31, 2006 who were between 25 and 55 at the time of entry into unemployment. The study ends in April 8, 2011. A spell of unemployment is defined as an uninterrupted period of time when an unemployed individual is registered at the employment office as either full-time unemployed or participating in a program. The spell ends when the unemployed individual finds employment for a minimum period of 30 days. Spells with exits for other reasons than employment, such as lost contact, sickness or end of study,

[6] In 1998 the general employment subsidy was called the individual employment subsidy, but the rules were the same.

Table 3: Number of individuals with one and several program episodes

First program	No second program		Second Work practice		Second training		Second subsidized job	
	#	%	#	%	#	%	#	%
Work Practice	46747	66	12268	17	3409	5	8073	11
Training	28426	76	2950	8	3199	9	3043	8
Subsidized job	8729	80	877	8	327	3	952	9

are censored. We assume that the censoring is ignorable conditional on the covariates. We ignore participation in any other program, so that the counterfactual state to the work practice is regular service at the employment office, including other types of programs. We aggregate the daily spell data to 45-days intervals.

Our sample consists of 1,032,668 unemployment spells. In total, 16.6% of the spells concern participation in any of the three programs, and in 43% of these spells, the first program is work practice. The same number for training is 43.6% and for subsidized employment 13.4%. Moreover, 61.5% of all spells are uncensored. *Table 3* presents the number and the fraction of unemployed individuals who participate once or twice in the three programs. The first row presents statistics for the unemployed whose the first program is work practice. It shows that 66% of them only participated in work practice, while 17% participated in work practice a second time. Work practice is thus an extensive program with many successive entries and exits in and out of the program. We also see that of these work practice participants, 5% and 11% later on enroll in training and subsidized employment. For training participants, 22% took part in another program within the same unemployment spell. The same number for subsidized employment is 20%. All this suggests that traditional evaluations of labor market programs, which usually only focus on the effects of the first program in an unemployment spell, may miss interesting aspects by ignoring those who participate in a program more than once, and this motivates the analysis of sequences of treatments.

We use logit regression models to estimate the propensity scores.[7] In the conditioning set, we include gender, age, number of unemployment days in the last five years, level

[7] For regimes defined by a single work practice episode the propensity scores are estimated separately in each time period for the censoring due to a first program episode in the control group and jointly for all time periods for the censoring in the treatment arm due to a second program episode. The latter joint logit model includes time period dummies. Similar joint estimations are performed when the regime of interest is defined by first and second entry times.

of education (3 categories), indicator for UI entitlement, region of residence (6 regions), an indicator for at least one child in the household, marital status, foreign born, labor income, social assistance, unemployment benefits and calendar year (for inflow). We include incomes and benefits for both one and two years before the start of the spell of unemployment.

Table 4 presents descriptive statistics on a subset of these covariates for the program participants and the non-participants. From this table we can see that males, immigrants and individuals eligible for UI are overrepresented among those participating in work practice. We also see that participants in training and subsidized employment have on average a lower level of education and are more often males and living in the northern parts of Sweden.

In order for our estimator to be consistent, the sequential unconfoundedness among survivors and the no-anticipation assumptions need to hold.[8] The no-anticipation assumption implies that there should be no causal effect of future treatments. It holds if the unemployed are unaware of future assignments into work practice, or if they do not react to such information. Since there are several unpredictable events leading to program enrollment, such as discretionary power of caseworkers, we believe that this assumption is fulfilled. The sequential unconfoundedness assumption is fulfilled if conditional on a set of covariates and a given treatment path treatment in the next period is randomly assigned among individuals who are still unemployed. There are several reasons why we think this is a plausible assumption.

Firstly, we condition on a large set of covariates, including detailed information on past unemployment history and individual background characteristics. We also control for regional labor market conditions using a set of regional indicators. Several studies have examined the importance of including different types of control variables when evaluating training programs. Besides using basic socioeconomic variables Heckman et al. (1998) and Heckman and Smith (1999) stress that it is important to control for previous unemployment, lagged earnings and local labor market characteristics. Using German

[8] Apart from these two main assumptions, the censoring needs to be unconfounded. Similar censoring assumptions are routinely imposed when using duration models.

Table 4: Sample statistics for program participants and non-participants. First program in the unemployment spell

Variable	Control	Work practice	Training	Subsidized employment
# obs.	990,493	88,071	69,576	12,071
Female (%)	51.9	52.7	46.3	43.2
Age	36.9	37.8	37.2	40.3
Married (%)	36.2	39.0	36.7	40.3
Children in household (%)	47.8	50.1	50.7	46.1
High School education (%)	45.3	47.4	56.1	46.1
University education (%)	30.9	27.2	23.0	25.5
Ages 35-44 (%)	32.5	33.7	34.8	34.7
Ages 44- (%)	20.0	23.0	19.8	33.7
Eligible for UI (%)	82.0	86.0	89.2	90.7
Stockholm MSA (%)	19.9	10.2	12.7	12.6
Goteborg MSA (%)	17.9	12.2	15.2	19.1
Skåne (%)	14.4	13.9	11.7	16.5
North (%)	14.1	22.4	18.0	16.6
South (%)	11.2	13.7	13.4	11.1
Previous unemployment	330.8	419.7	380.9	496.5
Labor income year -1	105,600	87,700	117,500	88,900
Labor income year -2	102,100	84,900	108,900	96,900
Social benefits year -1 (%)	14.9	17.4	13.7	19.5
Social benefits year -2 (%)	15.7	18.9	15.1	20.6
UI benefits year -1 (%)	33.6	37.3	39.7	40.7
UI benefits year -2 (%)	34.0	38.1	38.3	41.2
Inflow year 2000 (%)	11.6	13.9	17.4	10.6
Inflow year 2001 (%)	11.9	13.2	13.0	13.9
Inflow year 2002 (%)	12.8	11.9	13.1	14.8
Inflow year 2003 (%)	13.6	12.6	8.6	15.5
Inflow year 2004 (%)	13.8	15.3	9.0	14.9
Inflow year 2005 (%)	12.9	11.3	9.4	10.6
Inflow year 2006 (%)	11.8	7.4	9.3	8.8

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

data Lechner and Wunsch (2013) obtain similar results, and conclude that controlling for socioeconomic variables, regional dummies and short-run labor market history removes most of the bias. Note that we control for all these variables. However, in a recent study using German data, Biewen et al. (2014) find that conditioning on employment history makes little difference if you control for the elapsed unemployment duration before treatment.

Second, case-workers in Sweden have a large influence over enrollment in different programs, so that self selection is less important (see, e.g., Eriksson, 1997; Carling and Richardsson, 2001). Third, note that the unconfoundedness assumption is for survivors and not for the whole population of unemployed persons, and it is not farfetched to assume that treatment assignment conditional on survival is less selective.

5.3 Results

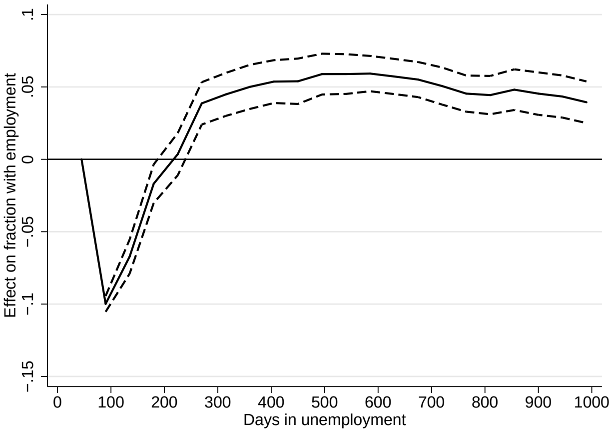
Sequences of work practice episodes

We start with the results for sequences of work practice episodes. Initially, we focus on the effects of the first work practice program during the unemployment spell. That is, we compare starting work practice after a certain number of months and not enrolling a second time with never enrolling in the program. *Figure 1*a displays results for enrollment after 136-180 days. For this enrollment time we find that for roughly the first 100 days the exit rates are lower among the program participants. This is the familiar locking-in effect found in most studies of training and employment subsidy programs which is caused by lower individual search effort during the actual program. After this initial period, participants gradually catch-up and about 270 days after enrollment in the program the fraction re-employed is significantly higher among the treated.

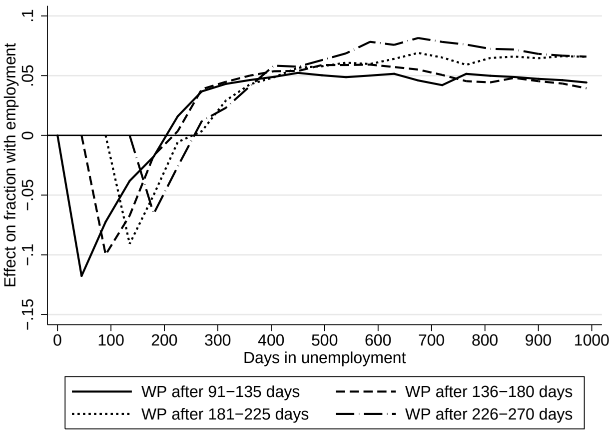
In order to evaluate the relative importance of the negative locking-in effect and the positive post-program effect, *Table 5* presents the effects on the cumulated re-employment rates during the first 36 months of unemployment. The cumulated effects have been scaled in such a way that they could be interpreted as the effect on the average truncated unemployment duration in months. From the first row of column 1 we have that work practice after 136-180 days decreases average time in unemployment by about 1.0 months or 8.6%.

*Figure 1*b and *Figure 1*c present results for other enrollment times. For enrollment times up to 360 days we find similar results as for enrollment after 136-180 days, including substantial locking-in effects, gradual catching up and eventually higher re-employment rates among the treated. For enrollment times beyond 360 days, we also find a similar pattern, but the re-employment rate never becomes significantly higher among the treated compared to the never treated. The fact that early enrollment is relatively better than late

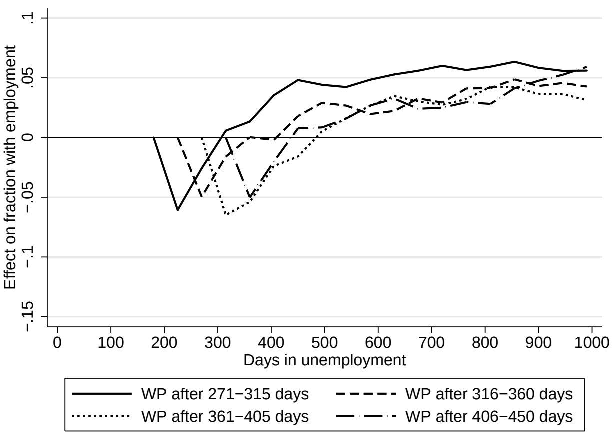
Figure 1: Estimates of the effect of the first work practice episode on fraction re-employed



(a) Treatment after 136-180 days (with standard errors)



(b) Treatment after 91-270 days



(c) Treatment after 271-450 days

Table 5: Work practice regimes and cumulative employment rates

		Time between first and second program		
Time to first	No second	45-135 days	136-225 days	226-315 days
	(1)	(2)	(3)	(4)
91-135 days	-0.86	0.42	0.01	0.61
	(0.22)	(0.65)	(0.54)	(0.51)
136-180 days	-1.00	-0.05	0.14	-0.09
	(0.22)	(0.96)	(0.66)	(0.64)
181-225 days	-1.20	-0.65	-0.31	-0.08
	(0.29)	(0.96)	(0.73)	(0.79)
271-315 days	-1.45	-1.03	-0.59	
	(0.27)	(1.03)	(0.77)	
226-270 days	-1.00	0.83	0.55	
	(0.26)	(1.06)	(1.40)	
316-360 days	-0.56	-0.17		
	(0.31)	(1.05)		

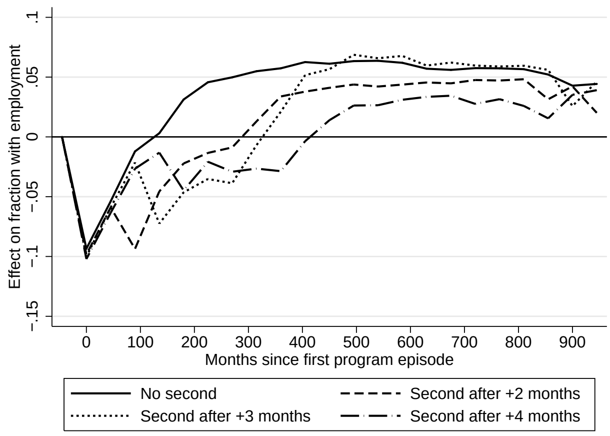
Note: The table reports cumulated re-employment rates for the first 36 months of unemployment. The comparison regime is never taking work practice. Swedish data for the period 1999-2006. The covariates used in in the weighting are gender, age number of unemployment days in the last 5 years, level of education (3 categories), indicator for UI entitlement, region of residence (6 regions), indicator for at least one child in the household, marital status, foreign born, labor income, social assistance and unemployment benefits one and two years before the start of the unemployment, and calendar year (for inflow).The threshold for the trimming rate is 4% for the single program estimates and 10% for combinations of two programs.

enrollment is confirmed by the effects on the cumulated re-employment rate in *Table 5*. For enrollment times up to 270 days, a single work practice episode leads to decreased unemployment durations, while there are no significant effects for later enrollment.

We now examine the effects of sequences of work practice. In columns 2-4 of *Table 5* we present results for the cumulated re-employment rates for specific times to the first episode and the time between the first and the second program. Note that for presentation reasons we have averaged over pre-treatment intervals. As before, the comparison is with never enrolling in work practice, so that one important comparison is between these estimates and the estimates for one single work practice episode in column 1. Moreover, note that the time between the two programs refers to the time between the start of the two programs, so that the actual time between the two programs is significantly shorter. We find that in all cases a second program episode leads to a smaller decrease in unemployment compared to only enrolling once. This holds for all the timings of the first program and for all the spacings between the first and the second program episode, even though these differences are not always significant.

We find no clear pattern when comparing the estimates for short and long spacing between the two program episodes. On average the effects are surprisingly similar for different spacings between the two program episodes. However, behind this result there are two interesting and counteracting effects. In order to illustrate this, *Figure 2* displays the survival rate estimates averaged over the first program enrollment times after 91-270 days and presented separately by time between the first and the second program episode. From this figure we obtain a number of interesting insights. The locking-in effects of the second program are greater if the time between the first and the second episode is short. At the same time, there are more pronounced positive post-program effects if the time between the first and the second episode is short. Thus, the relatively similar employment effects for a short and a long time between the two episodes is explained by the fact that the greater locking-in effects for a short time between the program episodes are counteracted by greater post-program effects.

Figure 2: Estimates of the effect of the first work practice episode on the fraction re-employed



All in all, we conclude that the work practice program reduces time spent in unemployment and that this effect is greater if the program starts before 315 days of unemployment. An analysis of the sequences of program episodes during the unemployment spell reveals that enrolling an unemployed individual twice does not lead to any additional reduction of unemployment beyond the one for a single work practice episode. This holds no matter if the second work practice episode starts shortly after the first or whether there are several months between the two program episodes.

Table 6: Cumulatative re-employment effects of work practice, training and subsidized employment. By timing of the treatment

First program after	Work practice Total (1)	0-4.5 months (2)	4.5+ months (3)	Training Total (4)	0-4.5 months (5)	4.5+ months (6)	Subsidized employment Total (7)	0-4.5 months (8)	4.5+ months (9)
91-135 days	-0.86	0.37	-1.23	-1.07	0.93	-2.00	6.92	1.72	5.20
	(0.81)	(0.11)	(0.85)	(0.12)	(0.02)	(0.12)	(0.15)	(0.02)	(0.16)
136-180 days	-1.00	0.27	-1.27	-1.51	0.77	-2.29	5.17	1.62	3.56
	(1.01)	(0.06)	(1.03)	(0.16)	(0.02)	(0.17)	(0.15)	(0.03)	(0.16)
181-225 days	-1.20	0.21	-1.41	-1.64	0.70	-2.34	3.66	1.36	2.30
	(0.68)	(0.04)	(0.69)	(0.17)	(0.03)	(0.18)	(0.20)	(0.03)	(0.22)
226-270 days	-1.45	0.09	-1.54	-1.84	0.57	-2.41	3.55	1.16	2.40
	(1.17)	(0.04)	(1.16)	(0.19)	(0.03)	(0.20)	(0.19)	(0.03)	(0.21)
271-315 days	-1.00	0.10	-1.10	-1.92	0.53	-2.45	1.36	1.04	0.32
	(0.80)	(0.07)	(0.83)	(0.20)	(0.03)	(0.21)	(0.18)	(0.03)	(0.19)
316-360 days	-0.56	0.10	-0.66	-1.79	0.51	-2.30	1.53	1.10	0.42
	(0.52)	(0.06)	(0.55)	(0.22)	(0.03)	(0.24)	(0.21)	(0.04)	(0.23)

Note: The table reports cumulated re-employment rates for the first 36 months of unemployment. Each cell reports one estimate. The comparison regime is participating in neither work practice, training nor subsidized employment. Swedish data for the period 1999-2006. The covariates used in in the weighting are gender, age number of unemployment days in the last 5 years, level of education (3 categories), indicator for UI entitlement, region of residence (6 regions), indicator for at least one child in the household, marital status, foreign born, labor income, social assistance and unemployment benefits one and two years before the start of the unemployment, and calendar year (for inflow). The threshold for the trimming rate is 6%.

Sequences of different programs

Table 6 presents results for a single program episode of each of the three programs. As before, we focus on cumulative re-employment rates for the first 36 months of unemployment. Columns 1, 4 and 7 give the total effect of the programs. The other columns report the size of the locking-in effect during the first 4.5 months after the start of the program and the remaining effect defined as the difference between the total effect and the locking-in effect.[9] The results in columns 1-3 for the work practice program confirm what we have already seen. After an initial locking-in period, the work practice program increases the re-employment rate, leading to a total reduction of the average unemployment durations.

From columns 4-6 we see that the training program also leads to shorter unemployment durations and that the total effect is of the same magnitude as for the work practice program. This holds even if the training program is associated with a greater locking-in effect compared to the work practice program; an expected result as the training program

[9]Note that the remaining effect is a combination of the post-program effect and a selection effect due to dynamic selection.

is usually substantially longer than the work practice program. We also find interesting heterogeneous effects with respect to the timing of the training program. In general is later enrollment in training associated with larger employment effects. Next, the results in columns 7-9 show that the effects of subsidized employment are highly dependent on the timing of the program. For enrollment before 271 days, the program has large negative employment effects, mainly due to very large locking-in effects. Enrollment after this point in time has much smaller negative employment effects. However, note that one explanation for these negative employment effects is that we treat the participants as unemployed while iun the program.

In *Table 7* we study the effects of different sequences of the three programs. Columns 1-4 present the results for when the first program is work practice and columns 5-7 for when training is the first program in the sequences analyzed.[10] For presentation reasons columns 2 and 5 restate the results when the sequence of interest is one single program episode. Panel A presents the results when work practice is the second program. Thus, columns 1-4 of panel A give the results for combinations of two work practice episodes, which essentially mimic the results of *Table 5*. The only difference is that here we aggregate the results to a somewhat higher level. In panels B and C we show the results for when training and subsidized employment, are the second program in the sequence. From a comparison of the results in column 1 for only one single episode of work practice and the results in the other columns, we see that in almost all cases a second program episode leads a less favorable employment effect than one single work practice episode. This holds both when the work practice program is started early and late in the unemployment spell, for short and long spacing between programs as well as no matter if the second program is work practice, training or subsidized employment.

We also obtain a number of interesting results from the estimates in columns 4-7.[11] First, from the results in column 5 of panel A, we find some evidence that work practice after completed training could reduce unemployment more than just one single training

[10] Because the subsidized employment scheme has fewer participants we are unable to consider sequences where the subsidized employment comes first followed by some other program.

[11] Since the training program normally lasts for several months we are unable to consider sequences where the second program starts less than 136 days after the start of the training program.

Table 7: Treatment regimes and cumulative employment rates. By type of programs and timing and spacing of the programs

	Work practice first program				Training first program		
Time to first	No second	Time between programs 45-135 days	136-225 days	226-315 days	No second	Time between programs 136-225 days	226-315 days
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Work practice second program							
91-180 days	-0.93	0.26	0.09	0.17	-1.44	-1.81	-0.93
	(0.14)	(0.74)	(0.44)	(0.39)	(0.13)	(1.04)	(0.75)
181-270 days	-1.32	-0.48	-0.75	-0.39	-1.86	-3.34	-2.38
	(0.21)	(0.85)	(0.54)	(0.70)	(0.17)	(0.93)	(2.08)
271-360 days	-0.78	0.44	0.27		-1.90	-2.48	
	(0.20)	(0.89)	(0.91)		(0.21)	(2.37)	
Training second program							
91-180 days	-0.93	0.16	3.17	-0.53	-1.44	0.11	0.37
	(0.14)	(0.70)	(0.68)	(0.62)	(0.13)	(0.81)	(0.74)
181-270 days	-1.32	0.70	2.12	0.45	-1.86	0.56	-0.40
	(0.21)	(1.12)	(0.72)	(1.19)	(0.17)	(1.12)	(1.25)
271-360 days	-0.78	2.36	1.08		-1.90	-1.17	
	(0.20)	(1.24)	(1.14)		(0.21)	(1.95)	
Subsidized job second program							
91-180 days	-0.93	4.06	3.16	0.92	-1.44	1.24	-1.11
	(0.14)	(0.74)	(0.57)	(0.44)	(0.13)	(1.26)	(0.71)
181-270 days	-1.32	0.85	-1.28	-0.51	-1.86	0.35	-1.93
	(0.21)	(1.55)	(0.55)	(0.56)	(0.17)	(1.38)	(1.46)
271-360 days	-0.78	-0.73	-0.36		-1.90	1.72	
	(0.20)	(0.69)	(0.73)		(0.21)	(3.00)	

Note: The table reports cumulated re-employment rates for the first 36 months of unemployment. The comparison regime is participating in no program. Swedish data for the period 1999-2006. The covariates used in in the weighting are gender, age number of unemployment days in the last 5 years, level of education (3 categories), indicator for UI entitlement, region of residence (6 regions), indicator for at least one child in the household, marital status, foreign born, labor income, social assistance and unemployment benefits one and two years before the start of the unemployment, and calendar year (for inflow). The threshold for the trimming rate is 4% for the single program estimates and 10% for combinations of two programs.

episode. For instance, training after 181-217 months shortens unemployment by on average 1.86 months. If such a training episode is quickly followed by a work practice episode, this effect increases to 3.34 months. One reason for this might be that work practice in the occupation that you have been trained for could serve as a quick way into a new type of occupation or into a new type of labor market.

Second, from panel B we learn that repeated episodes of training seem to be a waste of resources. This holds no matter whether the time between two episodes is short or long. One possible explanation is that another training episode only gives rise to an additional locking-in period and this is not counteracted by greater positive post-program effects of the second training episode. Thirdly, a similar pattern is obtained for subsidized employment. Evidently, subsidized employment in some cases could be beneficial to unemployed individuals who have never previously participated in a program, but not for unemployed individuals who have already participated in either work practice or training.

We conclude that in most cases starting a second program episode will not shorten the unemployment spell more than the first single program episode. One exception to this could be work practice straight after training.

6 Conclusions

In this paper, we have introduced a new framework for analyzing the effects of sequences of ALMP programs for the unemployed on time in unemployment. The new framework is applicable in all settings in which the treatments are given while in an initial state and when the outcome of interest is the time in unemployment. Besides sequences of ALMP programs, other applications could concern, for example, sequences of medical treatments. One conclusion is that under sequential unconfoundedness and without parametric restrictions it is possible to compare treatment regimes in terms of protocols stipulating when each treatment should start. That is, the effect considered in this paper is the effect of being enrolled in and following a specific sequence until employment is found, and not the effect of staying unemployed until the entire treatment sequence is finished. Such protocols are also policy relevant since they could be manipulated by policymakers. Another advantage is that it take into account re-employment effects both during and after the sequence of treatments. Moreover, it handles issues with dynamic selection as it avoids having to condition on survival in unemployment until the entire sequence of treatments has been completed.

This paper provides detailed identification results and introduces an inverse probability weighting estimator. The main assumption is that conditional on possible time-varying characteristics, treatment assignments among those still in the initial state are random. Another important assumption is no-anticipation. The new estimator re-weights the outcomes using a series of propensity scores, and given the estimates of these scores, the estimator imposes no parametric restrictions. Inference could be based on bootstrapping or using a joint estimation procedure that estimates the scores and the effect jointly. The properties of the new estimator are supported by a simulation study.

The new estimator is implemented using data on unemployed individuals in Sweden. The application provides many interesting results on the effects of different sequences of a work practice program, a training program and a subsidized employment scheme. One result is that the effect of the work practice program and the subsidized employment scheme to a large extent dependent on the timing of the programs. Work practice leads to more favorable results if the program starts relatively early in the unemployment spell, while subsidized employment is more efficient if it starts later in the unemployment spell. The training program shortens time in unemployment by as much as the work practice program, but for training the timing seems to be less important. All three programs are associated with locking-in effects that are closely related to the duration of the different programs.

Another key result is that enrolling an unemployed individual twice in the same program or in two different programs in most cases leads to longer unemployment spells compared to only participating in a single program once. This holds for most timings of the first program and most spacings between the two program episodes, as well as for most combinations of the three programs. This is explained by the fact that a second program episode gives rise to an additional locking-in period and in most cases this is not counteracted by larger post-program effects. One exception is that for some pre-treatment durations work practice straight after training has favorable employment effects. One reason for this could be that work practice allows the former training participants to quickly gain experience in a new occupation and to establish new employer contacts.

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 Myly. and P. Sevestre (Eds.), *The Econometrics of Panel Data*, Chap. 24, Berlin Heidelberg: Springer Verlag, 796–863.

Van den Berg G. (2001), "Duration models: specification, identification and multiple durations", Handbook of Econometrics, in: J.J. Heckman and E.E. Leamer (ed.), 3381–3460 Elsevier.

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.

Busso M., J. DiNardo and J.McCrary (2009), "Finite Sample Properties of Semiparametric Estimators of Average Treatment Effects", mimeo

Gill R.D. and J.M. Robins (2001), "Causal Inference for Complex Longitudinal Data: The Continuous Case", *Annals of Statistics 29*, 1785–1811.

Carling K. and K. Richardson K. (2001), "The Relative Efficiency of Labor Market Programs: Swedish Experience From the 1990", *Labour Economics 11(3)*, 335–354.

Crump R., J. Hotz, G. Imbens and O. Mitnik (2009), "Dealing With Limited Overlap in Estimation of Average Treatment Effects", *Biometrika 96*, 197–199.

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

Frölich M. (2004), "Finite Sample Properties of Propensity-Score Matching and Weighting Estimators", *Review of Economics and Statistics*, 77–90.

Gill R.D. and J.M. Robins (2001), “Causal Inference for Complex Longitudinal Data: The Continuous Case”, *Annals of Statistics 29*, 1785–1811.

Heckman J.J, H. Ichimura and P. Todd (1998), “Matching As An Econometric Evaluation Estimator”, *Review of Economic Studies 65*, 261–294.

Heckman J.J and J. Smith (1999), “The Pre-programme Earnings Dip and the Determinants of Participation in a Social Programme. Implications for Simple Programme Evaluation Strategies”, *Economic Journal 109*, 313–348.

Hernan M.A, B. Brumback and J.M. Robins (2001), “Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments”, *Journal of the American Statistical Association 96*, 440–448.

Huber M., M. Lechner and C. Wunsch (2013), “The Performance of Estimators Based on the Propensity Score”, *Journal of Econometrics 175*, 1–21.

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.

Lechner M. and C. Wunsch (2013), “Sensitivity of matching-based program evaluations to the availability of control variables,” *Labour Economics 21*, 111–121.

Robins J.M. (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*, 1393–1512.

Robins J.M. (1998), "Marginal Structural Models", *In 1997 Proceedings of the American Statistical Association, Section on Bayesian Statistical Science*, 1–10.

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

Appendix

In the remainder of the appendix we use the following notation

$$S^{\overline{d}_t} = \left\{ Y_t^{\overline{d}_t} = ... = Y_1^{d_1} = 0 \right\}$$

Identification of $\mathrm{SATE}(\overline{d}_t, \overline{d}_t^*)$

Under assumptions $A.1$ and $A.2$ we have for treatment regime $\overline{d}_t$ (the same reasoning applies for $\overline{d}_t^*$)

$$\begin{aligned}
\Pr(\overline{Y}_t^{\overline{d}_t} = 0) &= E_X \left[\Pr(\overline{Y}_t^{\overline{d}_t} = 0|X) \right] \\
&= E_X \left[\prod_{k=1}^{t} \Pr(Y_k^{\overline{d}_k} = 0|X, S^{\overline{d}_{k-1}}) \right] \\
&\overset{A.1}{=} E_X \left[\prod_{k=1}^{t} \Pr(Y_k^{\overline{d}_k} = 0|X, S^{\overline{d}_{k-1}}, D_1 = d_1) \right] \\
&= E_X \left[\prod_{k=2}^{t} \{ \Pr(Y_k^{\overline{d}_k} = 0|X, S^{\overline{d}_{k-1}}, D_1 = d_1) \} \Pr(Y_1^{d_1} = 0|X, D_1 = d_1) \right] \\
&\overset{A.1}{=} E_X \left[\prod_{k=2}^{t} \{ \Pr(Y_k^{\overline{d}_k} = 0|X, S^{\overline{d}_{k-1}}, \overline{D}_2 = \overline{d}_2) \} \Pr(Y_1^{d_1} = 0|X, D_1 = d_1) \right] \\
&\overset{A.1}{=} E_X \left[\prod_{k=1}^{t} \Pr(Y_k^{\overline{d}_k} = 0|X, S^{\overline{d}_{k-1}}, \overline{D}_k = \overline{d}_k) \right] \\
&\overset{obs.rule.}{=} E_X \left[\prod_{k=1}^{t} \Pr(Y_k = 0|X, \overline{Y}_{k-1} = 0, \overline{D}_k = \overline{d}_k) \right],
\end{aligned}$$

and it is assumption $A.2$ that makes this averaging over X feasible.

Inverse probability weigthing estimator for $\mathrm{SATE}(\overline{d}_t, \overline{d}_t^*)$

We have

$$p \lim_{N \to \infty} \widehat{\mathrm{SATE}}(\overline{d}_t, \overline{d}_t^*) = p \lim_{N \to \infty} \prod_{k=1}^{t} \left[1 - \frac{\sum_{i=1}^{N} w_{\overline{d}_{k,i}} Y_{k,i} 1(\overline{Y}_{k-1,i} = 0) 1(\overline{D}_{k,i} = \overline{d}_k)}{\sum_{i=1}^{N} w_{\overline{d}_{k,i}} 1(\overline{Y}_{k-1,i} = 0) 1(\overline{D}_{k,i} = \overline{d}_k)} \right] - \quad \text{(A.1)}$$

$$p \lim_{N \to \infty} \prod_{k=1}^{t} \left[1 - \frac{\sum_{i=1}^{N} w_{\overline{d}_{k,i}^*} Y_{k,i} 1(\overline{Y}_{k-1,i} = 0) 1(\overline{D}_{k,i} = \overline{d}_k^*)}{\sum_{i=1}^{N} w_{\overline{d}_{k,i}^*} 1(\overline{Y}_{k-1,i} = 0) 1(\overline{D}_{k,i} = \overline{d}_k^*)} \right],$$

with

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

$$\prod_{k=1}^{t}\left[1-\frac{p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}w_{\overline{d}_{k,i}}Y_{k,i}1(\overline{Y}_{k-1,i}=0)1(\overline{D}_{k,i}=\overline{d}_{k})}{p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}w_{\overline{d}_{k,i}}1(\overline{Y}_{k-1,i}=0)1(\overline{D}_{k,i}=\overline{d}_{k})}\right]=$$

$$\prod_{k=1}^{t}\left[1-\frac{E[w_{\overline{d}_{k}}Y_{k}1(\overline{Y}_{k-1}=0)1(\overline{D}_{k}=\overline{d}_{k})]}{E[w_{\overline{d}_{k}}1(\overline{Y}_{k-1}=0)1(\overline{D}_{k}=\overline{d}_{k})]}\right]=$$

$$\prod_{k=1}^{t}\left[1-\frac{E_X\{E[w_{\overline{d}_{k}}Y_{k}1(\overline{Y}_{k-1}=0)1(\overline{D}_{k}=\overline{d}_{k})|X]\}}{E_X\{E[w_{\overline{d}_{k}}1(\overline{Y}_{k-1}=0)1(\overline{D}_{k}=\overline{d}_{k})|X]\}}\right],$$

and using similar reasoning

$$p\lim_{N\to\infty}\prod_{k=1}^{t}\left[1-\frac{\sum_{i=1}^{N}w_{\overline{d}^*_{k,i}}Y_{k,i}1(\overline{Y}_{k-1,i}=0)1(\overline{D}_{k,i}=\overline{d}^*_{k})}{\sum_{i=1}^{N}w_{\overline{d}^*_{k,i}}1(\overline{Y}_{k-1,i}=0)1(\overline{D}_{k,i}=\overline{d}^*_{k})}\right]= \tag{A.3}$$

$$\prod_{k=1}^{t}\left[1-\frac{E_X\{E[w_{\overline{d}_{k}}Y_{k}1(\overline{Y}_{k-1}=0)1(\overline{D}_{k}=\overline{d}^*_{k})|X]\}}{E_X\{E[w_{\overline{d}_{k}}1(\overline{Y}_{k-1}=0)1(\overline{D}_{k}=\overline{d}^*_{k})|X]\}}\right].$$

Next, consider $k=1$. If assumption $A.1$ holds and using (1) we have

$$\begin{aligned}
E\left[w_{d_1}Y_1\mathbf{1}(D_1=d_1)|X\right] &= \frac{1}{p_{d_1}(X)}E\left[Y_1^{d_1}1(D_1=d_1)|X\right] \\
&= \frac{p_{d_1}(X)}{p_{d_1}(X)}E\left[Y_1^{d_1}(d_1)|X,D_1=d_1\right] \\
&\overset{A.1}{=} E\left[Y_1^{d_1}|X,D_1=d_1\right]=\Pr\left(Y_1^{d_1}=1|X\right),
\end{aligned} \tag{A.4}$$

and using similar reasoning

$$\begin{aligned}
E\left[w_{d_1}1(D_1=d_1)|X\right] &= 1 \\
E\left[w_{d_1^*}Y_1 1(D_1=d_1^*)|X\right] &= \Pr\left(Y_1^{d_1^*}=1|X\right) \\
E\left[w_{\overline{d}_1^*}1(D_1=d_1^*)|X\right] &= 1.
\end{aligned} \tag{A.5}$$

Second, if assumption $A.1$ holds and using (1) we have for $k > 1$

$$E\left[w_{\overline{d}_k} Y_k 1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k)|X\right] = \frac{E\left[Y_k 1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k)|X\right]}{p_{d_1}(X)\prod_{m=2}^{k} p_{d_m}(X,\overline{d}_{m-1})} = \tag{A.6}$$
$$E\left[Y_k^{\overline{d}_k}|X,S^{\overline{d}_{k-1}},D_k=d_k\right]\times$$
$$\frac{\prod_{m=1}^{k-1}\left[\Pr(Y_m^{\overline{d}_m}=0|X,S^{\overline{d}_{m-1}},D_m=d_m)p_{d_m}(X,\overline{d}_{m-1})\right]\Pr(Y_1^{d_1}=0|X,D_1=d_1)p_{d_1}(X)}{p_{d_1}(X)\prod_{m=2}^{k} p_{d_m}(X,\overline{d}_{m-1})} =$$
$$E\left[Y_k^{\overline{d}_k}|X,S^{\overline{d}_{k-1}},d_k\right]\prod_{m=1}^{k-1}\left[\Pr(Y_m^{\overline{d}_m}=0|X,S^{\overline{d}_{m-1}},d_m)\right]\Pr(Y_1^{d_1}=0|X,d_1) =$$
$$E\left[Y_k^{\overline{d}_k}|X,S^{\overline{d}_{k-1}},d_{k-1}\right]\prod_{m=1}^{k-1}\left[\Pr(Y_m^{\overline{d}_m}=0|X,S^{\overline{d}_{m-1}},d_m)\right]\Pr(Y_1^{d_1}=0|X,d_1) =$$
$$E\left[Y_k^{\overline{d}_k}|X,S^{\overline{d}_{k-1}},d_1\right]\prod_{m=1}^{k-1}\left[\Pr(Y_m^{\overline{d}_m}=0|X,S^{\overline{d}_{m-1}},d_1)\right]\Pr(Y_1^{d_1}=0|X,d_1) =$$
$$E\left[Y_k^{\overline{d}_k}|X,S^{\overline{d}_{k-1}}\right]\prod_{m=1}^{k-1}\left[\Pr(Y_m^{\overline{d}_m}=0|X,S^{\overline{d}_{m-1}}\right]\Pr(Y_1^{d_1}=0|X) =$$
$$\Pr\left(Y_k^{\overline{d}_k}=1,\overline{Y}_{k-1}^{\overline{d}_{k-1}}=0|X\right),$$

and using similar reasoning

$$\begin{aligned} E\left[w_{\overline{d}_k} 1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k)|X\right] &= \Pr\left(\overline{Y}_{k-1}^{\overline{d}_{k-1}}=0|X\right) \\ E\left[w_{\overline{d}_k^*} Y_k 1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k^*)|X\right] &= \Pr\left(Y_k^{\overline{d}_k^*}=1,\overline{Y}_{k-1}^{\overline{d}_{k-1}^*}=0|X\right) \\ E\left[w_{\overline{d}_k^*} 1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k^*)|X\right] &= \Pr\left(\overline{Y}_{k-1}^{\overline{d}_{k-1}^*}=0|X\right). \end{aligned} \tag{A.7}$$

Then, using (A.1)-(A.7) we have

$$p\lim_{N\to\infty}\widehat{\text{SATE}}(\overline{d}_t,\overline{d}_t^*) =$$

$$\prod_{k=1}^{t}\left[1-\frac{\Pr\left(Y_k^{\overline{d}_k}=1,\overline{Y}_{k-1}^{\overline{d}_{k-1}}=0\right)}{\Pr\left(\overline{Y}_{k-1}^{\overline{d}_{k-1}}=0\right)}\right]-\prod_{k=1}^{t}\left[1-\frac{\Pr\left(Y_k^{\overline{d}_k^*}=1,\overline{Y}_{k-1}^{\overline{d}_{k-1}^*}=0\right)}{\Pr\left(\overline{Y}_{k-1}^{\overline{d}_{k-1}^*}=0\right)}\right] =$$

$$\prod_{k=1}^{t}\Pr\left(Y_k^{\overline{d}_k}=0|\overline{Y}_{k-1}^{\overline{d}_{k-1}}=0\right)-\prod_{k=1}^{t}\Pr\left(Y_k^{\overline{d}_k^*}=0|\overline{Y}_{k-1}^{\overline{d}_{k-1}^*}=0\right)=\text{SATE}(\overline{d}_t,\overline{d}_t^*).$$

Identification and estimation for $\mathrm{SATE}(\overline{d}_t, \overline{d}_t^*)$ **under censoring**

First, concerning identification, under assumptions $A.1 - A.4$ we have

$$
\begin{aligned}
&\Pr(\overline{Y}_t^{\overline{d}_t} = 0) = E_X\left[\Pr(\overline{Y}_t^{\overline{d}_t} = 0|X)\right] \\
&= E_X\left[\prod_{k=1}^{t} \Pr(Y_k^{\overline{d}_k} = 0|X, S^{\overline{d}_{k-1}})\right] \\
&\overset{A.1,A.3}{=} E_X\left[\prod_{k=1}^{t} \Pr(Y_k^{\overline{d}_k} = 0|X, S^{\overline{d}_{k-1}}, D_1 = d_1, C_1 = 0)\right] \\
&= E_X\left[\prod_{k=2}^{t} \{\Pr(Y_k^{\overline{d}_k} = 0|X, S^{\overline{d}_{k-1}}, D_1 = d_1, C_1 = 0)\} \Pr(Y_1^{d_1} = 0|X, D_1 = d_1, C_1 = 0)\right] \\
&\overset{A.1,A.3}{=} E_X\left[\prod_{k=2}^{t} \{\Pr(Y_k^{\overline{d}_k} = 0|X, S^{\overline{d}_{k-1}}, \overline{D}_2 = \overline{d}_2, \overline{C}_2 = 0)\} \Pr(Y_1^{d_1} = 0|X, D_1 = d_1, C_1 = 0)\right] \\
&\overset{A.1,A.3}{=} E_X\left[\prod_{k=1}^{t} \Pr(Y_k^{\overline{d}_k} = 0|X, S^{\overline{d}_{k-1}}, \overline{D}_k = \overline{d}_k, \overline{C}_k = 0)\right] \\
&\overset{obs.rule}{=} E_X\left[\prod_{k=1}^{t} \Pr(Y_k = 0|X, \overline{Y}_{k-1} = 0, \overline{D}_k = \overline{d}_k, \overline{C}_k = 0)\right].
\end{aligned}
$$

and similiar reasoning for $\Pr(\overline{Y}_t^{\overline{d}_t^*} = 0)$ gives

$$
\mathrm{SATE}(\overline{d}_t, \overline{d}_t^*) = \mathbb{E}_X\left[\prod_{k=1}^{t} \Pr(Y_k = 0|X, \overline{D}_k = \overline{d}_k, \overline{Y}_{k-1} = 0, \overline{C}_k = 0)\right] -
$$

$$
\mathbb{E}_X\left[\prod_{k=1}^{t} \Pr(Y_k = 0|X, \overline{D}_k = \overline{d}_k^*, \overline{Y}_{k-1} = 0, \overline{C}_k = 0)\right].
$$

Second, concerning estimation, for $k > 1$ we have

$$
\begin{aligned}
&E\left[w_{\overline{d}_k} Y_k 1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k)1(\overline{C}_k=0)|X\right] = \\
&\frac{E\left[Y_k 1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k)|X\right]}{p_{d_1}(X)\prod_{m=2}^{k} p_{d_m}(X,\overline{d}_{m-1})\prod_{m=2}^{k} c_m(X,\overline{d}_{m-1})} \\
&\frac{E\left[Y_k^{\overline{d}_k}|X,S^{\overline{d}_{k-1}},D_k=d_k,C_k=0\right]\times \Pr(Y_1^{d_1}=0|X,D_1=d_1,C_1=0)p_{d_1}(X)\times \prod_{m=1}^{k-1}\left[\Pr(Y_m^{\overline{d}_m}=0|X,S^{\overline{d}_{m-1}},D_m=d_m,C_m=0)p_{d_m}(X,\overline{d}_{m-1})c_m(X,\overline{d}_{m-1})\right]}{p_{d_1}(X)\prod_{m=2}^{k} p_{d_m}(X,\overline{d}_{m-1})\prod_{m=2}^{k} c_m(X,\overline{d}_{m-1})} = \\
&E\left[Y_k^{\overline{d}_k}|X,S^{\overline{d}_{k-1}},D_k=d_k,C_k=0\right]\times \\
&\prod_{m=1}^{k-1}\left[\Pr(Y_m^{\overline{d}_m}=0|X,S^{\overline{d}_{m-1}},D_k=d_k,C_k=0)\right]\Pr(Y_1^{d_1}=0|X,D_k=d_k,C_k=0) = \\
&E\left[Y_k^{\overline{d}_k}|X,S^{\overline{d}_{k-1}}\right]\prod_{m=1}^{k-1}\left[\Pr(Y_m^{\overline{d}_m}=0|X,S^{\overline{d}_{m-1}}\right]\Pr(Y_1^{d_1}=0|X) = \\
&\Pr\left(Y_k^{\overline{d}_k}=1,\overline{Y}_{k-1}^{\overline{d}_{k-1}}=0|X\right),
\end{aligned}
$$

and using similar reasoning

$$
\begin{aligned}
E\left[w_{\overline{d}_k} 1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k)1(\overline{C}_k=0)|X\right] &= \Pr\left(\overline{Y}_{k-1}^{\overline{d}_{k-1}}=0|X\right) \\
E\left[w_{\overline{d}_k^*} Y_k 1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k^*)1(\overline{C}_k=0)|X\right] &= \Pr\left(Y_k^{\overline{d}_k^*}=1,\overline{Y}_{k-1}^{\overline{d}_{k-1}^*}=0|X\right) \\
E\left[w_{\overline{d}_k^*} 1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k^*)1(\overline{C}_k=0)|X\right] &= \Pr\left(\overline{Y}_{k-1}^{\overline{d}_{k-1}^*}=0|X\right).
\end{aligned}
$$

Then, using similar reasoning as above for $k = 1$ and collecting all results as above yields $p\lim_{N\to\infty}\widehat{\text{SATE}}(\overline{d}_t,\overline{d}_t^*)$ also under censoring.

Identification and estimation for $\text{SATE}(\overline{d}_t,\overline{d}_t^*)$ with time-varying covariates

First, concerning identification, under assumption and $A.5$ and a corresponding overlap condition we have

$$
\begin{aligned}
&\Pr(\overline{Y}_2^{\overline{d}_2}=0)= \\
&=E_{X_1}\left[\Pr(\overline{Y}_2^{\overline{d}_2}=0|X_1)\right] \\
&=E_{X_1}\left[\Pr(Y_2^{\overline{d}_2}=0|X_1,S^{d_1})\Pr(Y_1^{d_1}=0|X_1)\right] \\
&\overset{A.5}{=}E_{X_1}\left[\Pr(Y_2^{\overline{d}_2}=0|X_1,S^{d_1},D_1=d_1)\Pr(Y_1^{d_1}=0|X_1,D_1=d_1)\right] \\
&=E_{X_1}\left[E_{X_2|X_1,S^{d_1}}\left[\Pr(Y_2^{\overline{d}_2}=0|X_2,S^{d_1},D_1=d_1)\right]\Pr(Y_1^{d_1}=0|X_1,D_1=d_1)\right] \\
&\overset{A.5}{=}E_{X_1}\left[E_{X_2|X_1,S^{d_1}}\left[\Pr(Y_2^{\overline{d}_2}=0|X_2,S^{d_1},\overline{D}_2=\overline{d}_2)\right]\Pr(Y_1^{d_1}=0|X_1,D_1=d_1)\right] \\
&\overset{obs.rule}{=}E_{X_1}\left[E_{X_2|X_1,D_1=d_1,Y_1=0}\left[\Pr(Y_2=0|X_2,Y_1=0,\overline{D}_2=\overline{d}_2)\right]\Pr(Y_1=0|X_1,D_1=d_1)\right],
\end{aligned}
$$

and using similar reasoning for $\Pr(\overline{Y}_t^{\overline{d}_t^*}=0)$ we have

$$
\text{SATE}(\overline{d}_t,\overline{d}_t^*)=
$$

$$
\begin{aligned}
&E_{X_1}\left[E_{X_2|X_1,D_1=d_1,Y_1=0}\left[\Pr(Y_2=0|X_2,Y_1=0,\overline{D}_2=\overline{d}_2)\right]\Pr(Y_1=0|X_1,D_1=d_1)\right]- \\
&E_{X_1}\left[E_{X_2|X_1,D_1=d_1^*,Y_1=0}\left[\Pr(Y_2=0|X_2,Y_1=0,\overline{D}_2=\overline{d}_2^*)\right]\Pr(Y_1=0|X_1,D_1=d_1^*)\right].
\end{aligned}
$$

In terms of estimation we have

$$
\begin{aligned}
p\lim_{N\to\infty}\widehat{\text{SATE}}(\overline{d}_t,\overline{d}_t^*)=p\lim_{N\to\infty}\prod_{k=1}^{t}\left[1-\frac{\sum_{i=1}^{N}w_{\overline{d}_{k,i}}Y_{k,i}1(\overline{Y}_{k-1,i}=0)1(\overline{D}_{k,i}=\overline{d}_k)}{\sum_{i=1}^{N}w_{\overline{d}_{k,i}}1(\overline{Y}_{k-1,i}=0)1(\overline{D}_{k,i}=\overline{d}_k)}\right]- \\
p\lim_{N\to\infty}\prod_{k=1}^{t}\left[1-\frac{\sum_{i=1}^{N}w_{\overline{d}_{k,i}^*}Y_{k,i}1(\overline{Y}_{k-1,i}=0)1(\overline{D}_{k,i}=\overline{d}_k^*)}{\sum_{i=1}^{N}w_{\overline{d}_{k,i}^*}1(\overline{Y}_{k-1,i}=0)1(\overline{D}_{k,i}=\overline{d}_k^*)}\right],
\end{aligned}
\tag{A.8}
$$

with

$$
\begin{aligned}
&p\lim_{N\to\infty}\prod_{k=1}^{t}\left[1-\frac{\sum_{i=1}^{N}w_{\overline{d}_{k,i}}Y_{k,i}1(\overline{Y}_{k-1,i}=0)1(\overline{D}_{k,i}=\overline{d}_k)}{\sum_{i=1}^{N}w_{\overline{d}_{k,i}}1(\overline{Y}_{k-1,i}=0)1(\overline{D}_{k,i}=\overline{d}_k)}\right]= \\
&\prod_{k=1}^{t}\left[1-\frac{p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}w_{\overline{d}_{k,i}}Y_{k,i}1(\overline{Y}_{k-1,i}=0)1(\overline{D}_{k,i}=\overline{d}_k)}{p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}w_{\overline{d}_{k,i}}1(\overline{Y}_{k-1,i}=0)1(\overline{D}_{k,i}=\overline{d}_k)}\right]= \\
&\prod_{k=1}^{t}\left[1-\frac{E[w_{\overline{d}_k}Y_k1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k)]}{E[w_{\overline{d}_k}1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k)]}\right].
\end{aligned}
\tag{A.9}
$$

First, for $k = 1$, if assumption $A.5$ holds we have using similar reasoning as with only time-invariant characteristics (in the first period assumptions $A.1$ and $A.5$ are equivalent)

$$\begin{aligned} E\left[w_{d_1} Y_1 \mathbf{1}(D_1 = d_1)\right] &= E_{X_1}\{E\left[w_{d_1} Y_1 \mathbf{1}(D_1 = d_1)|X_1\right]\} = \\ E_{X_1}\{\Pr\left(Y_1^{d_1} = 1|X_1\right)\} &= \Pr\left(Y_1^{d_1} = 1\right) \end{aligned}$$

and

$$\begin{aligned} E\left[w_{d_1} 1(D_1 = d_1)\right] &= 1 \\ E\left[w_{d_1^*} Y_1 1(D_1 = d_1^*)\right] &= \Pr\left(Y_1^{d_1^*} = 1\right) \\ E\left[w_{\overline{d}_1^*} 1(D_1 = d_1^*)\right] &= 1. \end{aligned}$$

Second, under assumption $A.5$ we have for $k = 2$

$$\begin{aligned} & E[w_{\overline{d}_2} Y_2 1(Y_1 = 0) 1(\overline{D}_2 = \overline{d}_2)] = \\ & E_{X_1}\{E[w_{\overline{d}_2} Y_2 1(Y_1 = 0) 1(\overline{D}_2 = \overline{d}_2)|X_1]\} = \\ & E_{X_2|X_1}\{E_{X_1}\{E[w_{\overline{d}_2} Y_2 1(Y_1 = 0) 1(\overline{D}_2 = \overline{d}_2)|X_1, X_2]\}\}. \end{aligned}$$

Next,

$$\begin{aligned} & E[w_{\overline{d}_2} Y_2 1(Y_1 = 0) 1(\overline{D}_2 = \overline{d}_2)|X_2, X_1] = \frac{E\left[Y_2 1(\overline{Y}_1 = 0) 1(\overline{D}_2 = \overline{d}_2)|X_2, X_1\right]}{p_{d_1}(X_1) p_{d_2}(X_2, d_1)} \\ & E\left[Y_2^{\overline{d}_2}|X, S^{\overline{d}_1}, \overline{D}_2 = \overline{d}_2\right] \frac{p_{d_2}(X_2, d_1) \Pr(Y_1^{d_1} = 0|X, D_1 = d_1) p_{d_1}(X)}{p_{d_1}(X_1) p_{d_2}(X_2, d_1)} \\ & E\left[Y_2^{\overline{d}_2}|X_2, S^{\overline{d}_1}, \overline{D}_2 = \overline{d}_2\right] \Pr(Y_1^{d_1} = 0|X_1, D_1 = d_1) = \\ & \Pr(Y_2^{\overline{d}_2} = 1, Y_1^{d_1} = 0|X_2, X_1), \end{aligned}$$

then,

$$\begin{aligned} & E[w_{\overline{d}_2} Y_2 1(Y_1 = 0) 1(\overline{D}_2 = \overline{d}_2)] = E_{X_2|X_1}\{E_{X_1}\{\Pr(Y_2^{\overline{d}_2} = 1, Y_1^{d_1} = 0|X_2, X_1)\}\} \\ & \Pr(Y_2^{\overline{d}_2} = 1, Y_1^{d_1} = 0). \end{aligned}$$

Using similar reasoning for $k > 1$ we have

$$\begin{aligned}
E\left[w_{\overline{d}_k} Y_k 1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k)\right] &= \Pr\left(\overline{Y}_k^{\overline{d}_k}=1, \overline{Y}_{k-1}^{\overline{d}_{k-1}}=0\right) \\
E\left[w_{\overline{d}_k} 1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k)\right] &= \Pr\left(\overline{Y}_{k-1}^{\overline{d}_{k-1}}=0\right) \\
E\left[w_{\overline{d}_k^*} Y_k 1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k^*)\right] &= \Pr\left(Y_k^{\overline{d}_k^*}=1, \overline{Y}_{k-1}^{\overline{d}_{k-1}^*}=0\right) \\
E\left[w_{\overline{d}_k^*} 1(\overline{Y}_{k-1}=0)1(\overline{D}_k=\overline{d}_k^*)\right] &= \Pr\left(\overline{Y}_{k-1}^{\overline{d}_{k-1}^*}=0\right).
\end{aligned} \tag{A.10}$$

Then, using (A.8),(A.9) and (A.10) we have

$$p \lim_{N\to\infty} \widehat{\text{SATE}}(\overline{d}_t, \overline{d}_t^*) =$$

$$\prod_{k=1}^{t}\left[1-\frac{\Pr\left(Y_k^{\overline{d}_k}=1, \overline{Y}_{k-1}^{\overline{d}_{k-1}}=0\right)}{\Pr\left(\overline{Y}_{k-1}^{\overline{d}_{k-1}}=0\right)}\right]-\prod_{k=1}^{t}\left[1-\frac{\Pr\left(Y_k^{\overline{d}_k^*}=1, \overline{Y}_{k-1}^{\overline{d}_{k-1}^*}=0\right)}{\Pr\left(\overline{Y}_{k-1}^{\overline{d}_{k-1}^*}=0\right)}\right]=$$

$$\prod_{k=1}^{t}\Pr\left(Y_k^{\overline{d}_k}=0|\overline{Y}_{k-1}^{\overline{d}_{k-1}}=0\right)-\prod_{k=1}^{t}\Pr\left(Y_k^{\overline{d}_k^*}=0|\overline{Y}_{k-1}^{\overline{d}_{k-1}^*}=0\right)=\text{SATE}(\overline{d}_t, \overline{d}_t^*).$$

Publication series published by IFAU – latest issues

Rapporter/Reports

2014:20 Sibbmark Kristina ”Arbetsmarknadspolitisk översikt 2013”

2014:21 Nordlund Madelene and Mattias Strandh ”Selektivitet och jobbchanser bland arbetslösa”

2014:22 Angelov Nikolay and Marcus Eliason ”Vilka arbetssökande kodas som funktionshindrade av Arbetsförmedlingen?”

2014:23 Angelov Nikolay and Marcus Eliason ”Friställd och funktionsnedsatt”

2014:24 Angelov Nikolay and Marcus Eliason ”Lönebidrag och skyddat arbete: en utvärdering av särskilda insatser för sökande med funktionshinder”

2014:25 Holmlund Helena, Josefin Häggblom, Erica Lindahl, Sara Martinson, Anna Sjögren, Ulrika Vikman and Björn Öckert ”Decentralisering, skolval och friskolor: resultat och likvärdighet i svensk skola”

2014:26 Lindgren Karl-Oskar, Sven Oskarsson and Christopher Dawes ”Kan politisk ojämlikhet utbildas bort?”

2015:1 Albrecht James, Peter Skogman Thoursie and Susan Vroman ”Glastaket och föräldraförsäk-ringen i Sverige”

Working papers

2014:20 Johansson Per, Lisa Laun and Mårten Palme “Pathways to retirement and the role of financial incentives in Sweden”

2014:21 Andersson Elvira, Petter Lundborg and Johan Vikström “Income receipt and mortality – evidence from Swedish public sector employees”

2014:22 Felfe Christina and Rafael Lalive “Does early child care help or hurt children’s development?”

2014:23 Nordlund Madelene and Mattias Strandh “The relation between economic and non-economic incentives to work and employment chances among the unemployed”

2014:24 Mellander Erik “Transparency of human resource policy”

2014:25 Angelov Nikolay and Marcus Eliason “Factors associated with occupational disability classi-fication”

2014:26 Angelov Nikolay and Marcus Eliason “The differential earnings and income effects of involuntary job loss on workers with disabilities”

2014:27 Angelov Nikolay and Marcus Eliason “The effects of targeted labour market programs for job seekers with occupational disabilities”

2014:28 Carlsson Mikael, Julián Messina and Oskar Nordström Skans “Firm-level shocks and labor adjustments”

2014:29 Lindgren Karl-Oskar, Sven Oskarsson and Christopher T. Dawes “Can political inequalities be educated away? Evidence from a Swedish school reform”

2015:1 Avdic Daniel “A matter of life and death? Hospital distance and quality of care: evidence from emergency hospital closures and myocardial infarctions”

2015:2 Eliason Marcus “Alcohol-related morbidity and mortality following involuntary job loss”

2015:3 Pingel Ronnie and Ingeborg Waernbaum “Correlation and efficiency of propensity score-based estimators for average causal effects”

2015:4 Albrecht James, Peter Skogman Thoursie and Susan Vroman “Parental leave and the glass ceiling in Sweden”

2015:5 Vikström Johan “Evaluation of sequences of treatments with application to active labor market policies”

Dissertation series

2014:1 Avdic Daniel “Microeconometric analyses of individual behaviour in public welfare systems”

2014:2 Karimi Arizo “Impacts of policies, peers and parenthood on labor market outcomes”

2014:3 Eliasson Tove “Empirical essays on wage setting and immigrant labor market opportunities”

2014:4 Nilsson Martin “Essays on health shocks and social insurance”

2014:5 Pingel Ronnie “Some aspects of propensity score-based estimators for causal inference”

2014:6 Karbownik Krzysztof “Essays in education and family economics”