DISCUSSION PAPER SERIES

IZA

IZA DP No. 1062

Dynamic Treatment Assignment - The Consequences for Evaluations Using Observational Data

Peter Fredriksson
Per Johansson

March 2004

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Dynamic Treatment Assignment – The Consequences for Evaluations Using Observational Data

Peter Fredriksson
Uppsala University, IFAU and IZA Bonn

Per Johansson
Uppsala University and IFAU

Discussion Paper No. 1062
March 2004

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

IZA Discussion Paper No. 1062
March 2004

ABSTRACT

Dynamic Treatment Assignment – The Consequences for Evaluations Using Observational Data*

This paper discusses the evaluation problem using observational data when the timing of treatment is an outcome of a stochastic process. We show that the duration framework in discrete time provides a fertile ground for effect evaluations. We suggest easy-to-use nonparametric survival function matching estimators. These estimators can be used to estimate the time profile of the treatment. We apply one of the estimators to evaluate the effects of an employment subsidies program. The longer run program effects are positive. The estimated time profile suggests locking-in effects while participating in the program and a big upward jump in the employment hazard upon program completion.

JEL Classification: C14, C41

Keywords: treatment effects, dynamic treatment assignment, program evaluation, method of matching

Corresponding author:

Peter Fredriksson
Department of Economics
Uppsala University
Box 513
751 20 Uppsala
Sweden
Tel.: +46 18 471 1113
Email: peter.fredriksson@nek.uu.se

* Earlier versions of this paper have been circulated as "Program Evalution and Random Program Starts". Thanks to Kenneth Carling, Markus Frölich, Paul Frijters, Xavier de Luna, Jeffrey Smith, and Gerard van den Berg for very useful comments. Comments from seminar participants at the conference on "The Evaluation of Labour Market Policies" (Amsterdam, October 2002), Department of Statistics, Umeå University, and IFAU are also gratefully acknowledged. Fredriksson acknowledges the financial support from the Swedish Council for Working Life and Social Research (FAS).

1 Introduction

The prototypical evaluation problem is cast in a framework where treatment is offered only once. Thus treatment assignment is a static problem and the information contained in the timing of treatment is typically ignored; see Heckman et al. (1999) for an overview of the literature. This prototype concurs rather poorly with how most real-world programs work. Often it makes more sense to think of the assignment to treatment as a dynamic process, where the start of treatment is the outcome of a stochastic process.

There are (at least) two important implications of taking the timing of events into account. First of all, the timing of events contains additional information which is useful for identification purposes. Indeed, Abbring and van den Berg (2003) have shown that one can identify a causal effect[1] non-parametrically in the Mixed Proportional Hazard model from single-spell duration data without conditional independence assumptions (i.e. without assuming that all factors that jointly determine the treatment assignment and the outcome are observed). Second of all, the dynamic assignment process has serious implications for the validity of the conditional independence assumptions usually invoked to estimate effects such as treatment on the treated.

The main objective of this paper is to substantiate the second of the above claims. In particular we discuss program evaluations when (i) there are restrictions on treatment eligibility, (ii) no restrictions on the timing of the individual treatment, and (iii) the timing of treatment is linked to the outcome of interest. For instance, this evaluation problem arises when unemployment is a precondition for participation in a labor market program, programs may start at any time during the unemployment spell, and we are interested in employment outcomes. Employment outcomes have increasingly become the focus of the labor market evaluation literature so our analysis should have wide applicability.[2] We choose to focus on employment outcomes for illustrative purposes but our analysis has implications for all situations when points (ii) and (iii) apply. For instance, it follows immediately that the points we raise should be taken into consideration in analyses of earnings

[1] At this stage, we are deliberately vague on what causal effect this really is.

[2] The prime candidate for the shift in emphasis is that the ultimate goal of many labor market programs is to raise the reemployment probability rather than increasing the productivity of the participants. Also, the targets that government agencies responsible for, e.g., training, should fulfill are usually formulated in terms of employment rather than wages. For instance, one of the key targets for evaluating the performance of the Swedish labor market board is that at least 70 percent of participants in labor market training should be regularly employed one year after the end of treatment.

outcomes.

A second objective of the paper is to bridge some of the gap that exists between the literature on matching and the literature using hazard regressions. In the matching literature one typically considers, e.g., the probability of employment some fixed time period after treatment; Gerfin and Lechner (2002) is a recent example. By assumption, unobserved heterogeneity is not an issue. In the hazard regressions literature, the focus in on the timing of the outflow to a state of interest (e.g. employment). Usually, there is more structure imposed on the form of the hazard but there is also greater concern about unobserved heterogeneity; van den Berg et al. (2004) is an example. Clearly, the outcomes considered are intimately related and to us the division of the literature seems rather superficial. For instance, with rich data, one might well think of applying a matching approach to estimate the hazard to employment.

Here we assume that we can construct the counterfactual outcome using the method of matching. We take this approach for illustrative purposes – not because we are strong believers in the matching approach. To convey our basic messages as clearly as possible we want to avoid the complications arising from unobserved heterogeneity. Moreover, we want to refrain from making assumptions about the appropriate bivariate distribution for the timing of events. If one is prepared to make assumptions about the functional form of the bivariate distribution, this is an alternative way of attacking the particular evaluation problem that we are considering.

The rest of this paper is structured in the following way. In section 2, we present the evaluation framework. We discuss the potential outcomes of interest, possible estimands, and the specific problem associated with dynamic treatment assignment. Section 3 considers some alternative estimators available in the literature. We also propose an estimator of the effect of treatment on the treated. In section 4 we apply this estimator to the problem of evaluating the employment effects of a Swedish employment subsidy program. Section 5, finally, concludes.

2 The evaluation framework

Consider a pool of openly unemployed individuals who are eligible for the program as long as they are unemployed.[3] We assume that these individuals are identical conditional on the date of unemployment entry and observed covariates. For expositional convenience, we suppress the observed covariates

[3]The openly unemployed refers to the unemployed who do not participate in a labor market program.

for most of our analysis. At time τ, these individuals are exposed to two kind of risks: either they get a job offer with instantaneous probability $\lambda(\tau, 0)$, or they get an offer to participate in a treatment (a program) with probability $\gamma(\tau)$ per unit time. The instantaneous probability of being offered a job if treated at s is $\chi_s(\tau, 1)$. The hazard rates, $\lambda(\tau, 0)$ and $\chi_s(\tau, 1)$, are potential outcomes. They are potential in the sense that they are not observed for everyone. Let us also denote the potential unemployment duration if not treated by $T(0)$.

It is important to realize that this set-up implies that the duration until treatment *start* (S) is stochastically dependent on $T(0)$. Whether we observe S or not depends on such things as the behavior of case workers, the characteristics of the unemployed, but also on whether the individual had the luck of receiving a job offer before receiving an offer to participate in treatment. In particular, we observe individuals taking treatment ($D = 1$) if their unemployment duration is longer than the duration until program start:

$$D = I(T(0) > S) \tag{1}$$

This may appear to be tautologically true, but this specification also embodies an important assumption: individuals must not anticipate their actual date of treatment start nor anticipate future jobs (i.e. the realization of $T(0)$).[4] If they knew exactly when treatment will start their behavior will be altered prior to entering treatment, i.e., there will be a pre-treatment effect.

Let us be clear on what we can observe in the data. For non-treated individuals we observe the potential duration in unemployment if not treated and thus $T = T(0)$. For individuals who are treated at $S = s$ we observe

$$T = s + T^s(D = 1)$$

where $T^s(D = 1)$ is the post-treatment duration for those treated at s. The evaluation problem consists in not observing the post-treatment duration without treatment for the treated, i.e., we do not observe $T^s(0)$ for the $D = 1$ sample. Since

$$T^s(0) = [(T(0) - s)|(T(0) > s)]$$

the dynamic assignment process implies that there is a stochastic dependence between the post-treatment and pre-treatment duration even if there is no treatment effect. In other words $T(0)$ need to be independent of D in order

[4]This assumption is also invoked by Abbring and van den Berg (2003). They refer to it as a "no anticipation" assumption.

for the potential unemployment duration post s if not treated, $T^s(0)$, to be independent of D.

Now that we have introduced some notation let us define the notational convention that we will adopt throughout the paper. Stochastic variables are denoted by upper-case roman letters (e.g. T, S, and T^s), realizations of the stochastic processes are lower-case roman letters (e.g. t, s and t^s), and potential outcomes are indicated by 0 and 1 (e.g. $T(0)$, $T^s(0)$, $\lambda(\tau, 1)$ and $\chi_s(\tau, 1)$).

2.1 Objects of evaluation with random assignment

To fix ideas, it is useful to digress to the case with random assignment into treatment. What is the virtue of random assignment and what are sensible objects of evaluation?

Suppose that treatment is randomly assigned among the individuals in the unemployment pool. Random assignment takes place at time ψ, implying that the pre-treatment duration is a function of ψ, $s = s(\psi)$. Since treatment status (D^R) is assigned randomly, the pre-treatment duration distributions, F_S^R, are equal for the two samples consisting of the treated and the non-treated respectively, i.e., $F_S^R(s|D^R) = F_S^R(s)$. Using the notation of Dawid (1979), we thus have

$$S \perp\!\!\!\perp D^R \tag{2}$$

Since random assignment balances the pre-treatment duration distributions there is no need for information on unemployment duration prior to treatment entry.[5] In other words, random assignment creates samples of treated and non-treated that are *duration matched*.

The condition (2) also implies that potential unemployment duration post assignment – $T^s(j) = T(j) - s$, $j = 0, 1$ – is independent of being assigned to treatment or not. Hence

$$(T^s(1), T^s(0)) \perp\!\!\!\perp D^R.$$

An estimand of interest is the difference in the post-treatment duration if treated and non-treated

$$\Delta^R = \mathrm{E}_s \mathrm{E} T^s(1) - \mathrm{E}_s \mathrm{E} T^s(0) \tag{3}$$

where $\mathrm{E}_S(.)$ is the expectation with respect to the pre-treatment duration. If durations are not censored, this can easily be estimated using the differences

[5]This is not to say that information on the pre-treatment duration is uninteresting in the experimental setting. With it one could ask the highly interesting question such as: Does the treatment effect vary by prior unemployment duration?

in the sample means. However, in almost every situation censoring is present. So it is more appropriate to define the estimand of interest in terms of the survival function (or the hazard rate). Let the survival function, averaged over pre-treatment durations, be denoted $\overline{F}^R(t^s, j)$. Then the difference in average survival rate by t^s is defined by

$$\Delta^R(t^s) = \overline{F}^R(t^s, 1) - \overline{F}^R(t^s, 0),\ t^s \in (0, \infty). \tag{4}$$

The two average survival functions are given by

$$\begin{aligned} \overline{F}^R(t^s, j) &= \int_0^\infty \overline{F}_s^R(t^s, j) dF_S^R(s) \\ &= \int_0^\infty \left\{ \exp(-\int_0^{t^s} \chi_s(\tau, j) d\tau) \right\} dF_S^R(s),\ j = 0, 1 \end{aligned}$$

where $\chi_s(\tau, 0) = \lambda(\tau + s, 0)(s/e_s(0)), \tau > 0$ is the stock sampling hazard and $e_s(0)$ is the expected total duration if not treated for an eligible individual given survival up to s (see e.g. Lancaster, 1990, ch. 5).[6] Also, $\chi_s(\tau, 1) = \lambda_s(\tau, 1)(s/e_s(0))$ where $\lambda_s(\tau, 1)$ is the hazard rate if treated at s.

The estimand defined in (4) is more "fundamental" than (3) since the difference in survival functions integrates to the difference in mean duration, i.e.,

$$\int_0^\infty \Delta^R(t^s) dt^s = \mathrm{E}_s \mathrm{E} T^s(1) - \mathrm{E}_s \mathrm{E} T^s(0) = \Delta^R \tag{5}$$

Calculating $\overline{F}^R(t^s, j)$, of course, requires information on unemployment duration prior to treatment. Such information is not always available and, therefore, it is useful to define an object of evaluation that is less demanding on the data. Our purpose now is to introduce an estimand that conveniently averages the hazards over s (even if the information is unavailable) and τ (to get at the "total" effect). In the special case where there is no duration dependence this estimand has the additional virtue of being equal to $\Delta^R(t^s)$ defined in (4).

Consider, therefore, the estimand

$$\Delta^{R,1}(t^s) = \overline{F}^{R,1}(t^s, 1) - \overline{F}^{R,1}(t^s, 0),\ t^s \in (0, \infty) \tag{6}$$

[6] If we randomly select an individual at s from the stock of unemployed individuals, then the stock sampling hazard is equal to $\chi_s(\tau, 0)$ and the density of participation is

$$f(s) = \frac{\exp(-\int_0^s \lambda(\tau, 0) d\tau)}{\mathrm{E} T(0)}.$$

where $\overline{F}^{R,1}(t^s, j)$ denotes the survival function for the average treated and untreated individual. Hence, $\overline{F}^{R,1}(t^s, j)$ is defined by

$$\overline{F}^{R,1}(t^s, j) = \exp(-\int_0^{t^s} \chi(\tau, j) d\tau)$$

where $\chi(\tau, j) = \int_0^\infty \chi_s(\tau, j) dF_S^R(s)$ is the average hazard rate for the treated and non-treated population. Thus, $\Delta^{R,1}(t^s)$ measures the differences in survival rate for the average individual while $\Delta^R(t^s)$ measures the differences in the average survival rate. Notice that, in general, $\overline{F}^{R,1}(t^s, j) \neq \overline{F}^R(t^s, j)$. The only situation when they are equal is when there is no duration dependence. Then we have $\Delta^{R,1}(t^s) = \Delta^R(t^s)$.

An alternative to basing the evaluation on the difference in the survival functions is to estimate hazard regressions. Assume for instance that the "structural" treatment effect is of the proportional variety and that it is independent of the date of treatment, hence $\lambda(\tau, 1) = \lambda(\tau, 0)\exp(\delta)$. This implies that the average stock sampling hazard is proportional as well: $\chi(\tau, 1) = \chi(\tau, 0)\exp(\delta)$. Thus, estimating a proportional hazard regression model using the stock sample population provides us with an estimate of δ.[7] Notice that this estimate enables us to estimate $\Delta^{R,1}(t^s)$ rather than $\Delta^R(t^s)$.

In sum, we think that there are two main lessons to be gleaned from this section. First, even with experimental data there are many possible treatment effects. Although the fundamental input for all treatment effects is the hazard rates if treated and non treated, different ways of aggregating the information contained in the hazards, results in different effect sizes. Second, even if the magnitude of the estimates vary with the pre-treatment duration, random assignment works for the obvious reason that it balances the pre-treatment duration among the treated and the non-treated. We now turn to the more complicated situation with observational data.

2.2 Objects of evaluation with observational data

Consider the case with observational data. What makes the observational study special is that we lack treatment start dates for those not treated. Hence, it is impossible to use the post treatment duration for the untreated to estimate the counterfactual mean $ET^s(0)$ for the treated population. This is different than in the experimental situation, where treatment for the stock

[7] Notice that the aggregation of the effects is achieved by the functional form assumption, i.e., that the treatment effect is of the proportional variety.

of eligibles is offered at some fixed point in time, and the fairly uncommon situation where a program starts after a fixed duration.[8]

The estimand defined in (4) is generically the most informative object of evaluation. However, in the observational setting we must focus on the treatment of the treated variant of (4) since only the pre-treatment duration for the treated, $F_S(s|D=1)$ is observed. Consider, therefore, the estimand

$$\Delta(t^s|D=1) = \overline{F}(t^s,1) - \overline{F}(t^s,0),\ \ t^s \in (0,\infty) \tag{7}$$

The survival functions, $\overline{F}(t^s,j)$, are again averaged over the pre-treatment duration distribution:

$$\overline{F}(t^s,j) = \int_0^\infty \overline{F}_s(t^s,j)F_S(s|D=1),\ j=0,1 \tag{8}$$

The difference in the survival functions integrates to the difference in mean duration for the treated and in the limiting case without censoring we obtain

$$\int_0^\infty \Delta(t^s|D=1)dt^s = \mathrm{E}_S\mathrm{E}T^s(1|D=1) - \mathrm{E}_S\mathrm{E}T^s(0|D=1) = \Delta_1 \tag{9}$$

Another option is to form the treatment of the treated variant of (6), hence

$$\Delta^1(t^s|\,D=1) = \overline{F}^1(t^s,1) - \overline{F}^1(t^s,0),\ \ t^s \in (0,\infty) \tag{10}$$

where $\overline{F}^1(t^s,1)$ and $\overline{F}^1(t^s,0)$ are the survival functions for the average treated individual if treated and non-treated respectively. Hence, $\overline{F}^1(t^s,j)$ is defined by

$$\overline{F}^1(t^s,j) = \exp(-\int_0^{t^s} \chi(\tau,j)d\tau),\ j=0,1$$

where $\chi(\tau,j) = \int_0^\infty \chi_s(\tau,j)dF_S(s|D=1)$ is the average hazard rate if treated and not treated for the treated.

In an observational study it is absolutely crucial to have information on the pre-treatment duration. Without such information it is impossible to balance the pre-treatment duration for the treated and non-treated. In the experimental setting, on the other hand, random assignment creates samples of treated and non-treated that are duration matched so information on the pre-treatment duration is not required.

[8]Of course there are some treatments that start after a fixed point in time. The expiration of UI benefits is a prototypical example.

3 Potential estimators

In this section we consider alternative strategies to estimate the parameters of interest. We begin by examining the properties of matching estimators that impose a binary treatment indicator in a setting where the assignment to treatment is really a dynamic stochastic process. We then move on to a proportional hazard model. Finally, we consider matching with a time-varying treatment indicator.

3.1 Matching with a binary treatment indicator

Here we follow the typical approach to evaluating an on-going program. As indicated above, researchers usually impose a "binary framework" even though the timing of events varies. To implement the idea that the assignment to treatment occurs only at a "single point in time" there is typically a classification window of some length (w). Individuals that take up treatment within, say, the first six months of the unemployment spell are defined as the treated ($D(w) = 1$) while those that do not are defined as the non-treated ($D(w) = 0$). Then the typical outcome would be something like the employment status one year after treatment entry (s). Thus the starting point for measuring the effect of treatment occurs before the end of the classification window ($s < w$).

A practical problem is that those who had the luck of finding a job quickly are more likely to be found in the non-treated group. Thus some trimming of the left-tail of the duration distribution seems to be called for. Here we follow an approach that is akin to the one suggested by Lechner (1999). Before matching on the covariates he proposes a procedure to trim the duration distribution of the non-treated such that he obtains a duration matched comparison sample.

To illustrate the approach as clearly as possible, let us consider the extreme case where $w \to \infty$. Now, duration matching is an attempt to estimate average effect of treatment on the treated, e.g., in terms of the post-treatment duration

$$\Delta_1 = \mathrm{E}_S \mathrm{E} T^s(1|D=1) - \mathrm{E}_S \mathrm{E} T^s(0|D=1)$$

The expectation $\mathrm{E}_S\mathrm{E}(T^s(1|D=1)$ can be estimated as

$$\hat{t}^s = \frac{1}{n}\sum_{i=1}^{n}(t_i - s_i)$$

where n is the number of treated individuals. An estimator of the counterfactual outcome, $\mathrm{E}_S\mathrm{E}(T^s(0|\, D=1)$, is based on random sampling from the inflow distribution, $F_S(s|D=1)$. For a random draw, s_i, an individual from the comparison sample is matched if the unemployment duration for this randomly assigned individual satisfies $t_c > s_i$. Applying this procedure we get a duration matched comparison sample (consisting of n matches) and may calculate

$$\hat{t}^s_c = \frac{1}{n}\sum_{i=1}^{n}(t_{c_i} - s_i), \tag{11}$$

where $t_{c_i} - s_i$ is the observed unemployment duration after s_i for a (randomly assigned) matched individual. The treatment effect is then estimated as

$$\widehat{\Delta}_1 = \hat{t}^s - \hat{t}^s_c \tag{12}$$

Proposition 1 *a) Conditional on s, the assignment is not ignorable for the remaining duration if not treated*: $T^s(0) \not\perp D|(T(0) > s)$. *b) When there is no treatment effect, the estimator $(\widehat{\Delta}_1)$ is positively biased*

Proof. Consider individuals treated at $S = s$. For these individuals we know that $T(0) > s$. For potential comparison individuals we have $s < T(0) < S$ since they were never treated. Thus

$$\mathrm{E}[T^s(0)|(D=1)] = \mathrm{E}[(T(0)-s)|T(0) > s] \tag{13}$$

$$\mathrm{E}[T^s(0)|(D=0, T(0) > s)] = \mathrm{E}[(T(0)-s)|(S > T(0) > s] \tag{14}$$

Since $\mathrm{E}[(T(0)-s)|T(0) > s] > \mathrm{E}[(T(0)-s)|(S > T(0) > s)]$, $T^s(0) \not\perp D|(T(0) > s)$. ■

Notice that the result hold for any specification generating the two processes $T(0)$ and S. In particular, the duration matched estimator is biased even though the hazards to employment and treatment are constant.

Proposition 1 a) follows from the observation that for all classification periods such that $s < w$ there is some conditioning on the future involved when defining the potential comparison group for an individual treated at s. Thus, the estimator is always biased. Given that there is no treatment effect we can determine the sign of the bias involved in applying this procedure. The intuition is simply that for the comparison group we know that (since the individual is not treated) the spell ends with employment, while for the treated group we do not know if the spell ends in employment. Therefore, there is an upward bias in the effect of treatment on post-treatment durations (i.e. there is a bias towards finding that the program does not work). Let us also make the (perhaps obvious) remark that Proposition 1 hold if the observations on unemployment durations are censored at, say, $\widetilde{l}$, although one would expect the bias to be reduced in magnitude.

To sum up, it is not possible to create a sample of matching individual who do not receive treatment at any point in time. In defining the treated and the comparisons, the sampling is on $T(0)$, which in turn determines the (potential) outcome $T^s(0)$. We wish to emphasize that the crux of the problem with this estimator lies in the use of a classification window; it is not due to the trimming procedure. It is the strive to transform a world where treatment assignment is the outcome of two dependent stochastic processes to an idealized world where treatment assignment and outcomes occurs at single points in time that causes the problems.

3.2 The proportional hazard model

A popular approach to estimate the treatment effect is to use the proportional hazard model; see, e.g., Crowley and Hu (1977), Lalive et al. (2002), and Richardson and van den Berg (2003). Here we examine what happens when we impose a discrete time proportional hazard model in our context.

The discrete time hazard model is given by

$$\lambda(t,1) = h(t,0)\exp(\delta D(t))$$

where $D(t) = I(T(0) > S = t)$. Thus $D(t) = 1$ for individuals who have been treated by t and $D(t) = 0$ for individuals who remain untreated at t (but may be treated in the future). If δ estimates an average treatment effect then $\lambda(t,0) = h(t,0)$ and hence $\lambda(t,1) = \lambda(t,0)\exp(\delta D(t))$.

Can we estimate an average treatment effect using this framework? The following proposition provides part of the answer.

Proposition 2 *The data generating process $D(t) = I(T(0) > S = t)$ implies that the baseline hazard for the treated is not equal to the baseline hazard in the population, i.e., $h(t,0) \neq \lambda(t,0)$.*

Proof. The appropriate baseline hazard is given by

$$h(t,0) = \lambda(t,0|D=1)\Pr(D(t)=1) + \lambda(t,0|D=0)\Pr(D(t)=0) \qquad (15)$$

Clearly, $\lambda(t,0) = h(t,0)$ requires $\lambda(t,0|\, D=1) = \lambda(t,0|\, D=0)$. Now, note that Proposition 1 implies $\mathrm{E}(T(0)|\, D=1) > \mathrm{E}(T(0)|\, D=0)$. Since this is true for any censoring point $t > 0$ the probability of survival for the treated (if not treated) is greater than the probability of survival for the non-treated, i.e. $\overline{F}(t,0|D=1) > \overline{F}(t,0|D=0)$. Now,

$$\begin{aligned}
\overline{F}(t,0|D &= 1) > \overline{F}(t,0|D=0) \Leftrightarrow \\
\ln \overline{F}(t,0|\,D &= 1) > \ln \overline{F}(t,0|D=0) \Leftrightarrow \\
\int_0^t \frac{d\ln \overline{F}(\tau,0|D=1)}{d\tau} d\tau &> \int_0^t \frac{d\ln \overline{F}(\tau,0|D=0)}{d\tau} d\tau \Leftrightarrow \\
-\int_0^t \lambda(\tau,0|D &= 1)d\tau > -\int_0^t \lambda(\tau,0|D=0)d\tau \Leftrightarrow \\
\int_0^t [\lambda(\tau,0|D &= 1) - \lambda(\tau,0|D=0)]d\tau < 0
\end{aligned}$$

This shows that $\lambda(t,0|\,D=1) \neq \lambda(t,0|\,D=0)$. ■

Thus, the mirror image of the fact that those we observe taking treatment have longer expected unemployment duration is that the integrated hazard if not treated is lower for treated individuals than non-treated individuals. Further, if $\delta > 0$ it is not possible to identify all components of the baseline hazard (15) using observational data. So estimates of the treatment effect using the proportional hazards specification will, in general, neither estimate an average treatment effect nor an effect of treatment on the treated. Can we say anything about the sign of the bias relative to the true parameter, δ? Proposition 3 outlines the results

Proposition 3 *a) If there is no treatment effect* ($\delta = 0$)*, the proportional hazards estimator* ($\hat{\delta}_{PH}$) *has the property that* $\operatorname{plim} \hat{\delta}_{PH} = 0$*. b) If* $\delta \neq 0$*, then* $\operatorname{plim} \left|\hat{\delta}_{PH}\right| < |\delta|$*.*

Proof. See appendix 1. ■

The intuition for Proposition 3b) is the following. With observational data, the risk set used for estimation includes individuals who are not treated at time t but will be treated at some future time point. The inclusion of these individuals (in addition to those who have been treated prior to t and those who are never treated) will lead to attenuation bias.

However, the inclusion of those treated in the future in the risk set is a virtue when $\delta = 0$. The inclusion of these individuals balances the bias that would arise if only the never treated were used as comparisons.

The thrust of Proposition 3 is that the proportional hazards specification is a fertile ground for testing.[9] However, the estimate will be smaller in absolute value than the treatment effect when a treatment effect exists. Notice

[9] In the absence of (observed and unobserved) heterogeneity the proportional hazard specification is a non-parametric specification of a causal parameter. When heterogeneity is present however, the parametric specification (e.g. including the covariates as a single index) is always a concern.

also that standard (Wald) tests will not give correct inference since the true model is non-proportional; see DiRienzo and Lagakos (2001).

Abbring and van den Berg (2003) show that the variation in the timing of treatment identifies a causal treatment parameter in the proportional hazard model. This is also true in our case since the model in this sub-section is really a stylized version of their more general model. Consider estimating $\delta(t)$ in

$$\lambda(t,1) = h(t,0)\exp(\delta(t)D(t))$$

It is clearly possible to estimate the causal treatment effect, $\delta(t)$, since $h(t,0)$ is also the baseline hazard for those who have not been treated at t. Thus, taking the timing of treatment seriously allows the identification of causal parameters. The interpretation of $\delta(t)$ is "the effect of the program in time-period t".

3.3 Matching with a time-varying treatment indicator

The lesson from the above sub-section is that one should take the timing of treatment seriously. However, if we believe in the assumptions that justify matching we have no reason to postulate a proportional hazard model. Instead we will introduce a non-parametric matching estimator that takes the timing of events into account but does not rely on proportionality assumption. The estimand is also easier to interpret than the $\delta(t)$ parameter discussed above.

With the treatment indicator $D(s) = I(T(0) > S = s)$, the potential unemployment durations for the two groups are:

$$T^s(0)|(D(s) = 1) = T(0) - s \tag{16}$$

$$\begin{aligned} T^s(0)|(D(s) = 0) &= (T(0) - s)I(T(0) < S) + T^s(1)I(T(0) > S) \\ &= (T(0) - s) + (T^s(1) - T^s(0))I(T(0) > S) \end{aligned} \tag{17}$$

It is straightforward to show that

Proposition 4 *a) Under the null hypothesis of no treatment effect (H_0), potential unemployment duration is independent of the treatment indicator $D(s)$. b) Under the alternative (H_1), conditional independence does not hold.*

Proof. By inspection of (16) and (17) it follows that $T^s(0) \perp\!\!\!\perp D(s)$ if $T^s(1) = T^s(0)$. If there is a treatment effect, $T^s(1) \neq T^s(0)$, then the CIA does not hold. ■

Thus, the gain of introducing the time-varying treatment indicator, $D(s)$, is immediate: under H_0, potential unemployment duration is conditionally independent of $D(s)$. However, one cost associated with this procedure is that the estimand

$$\Delta_1(s) = \mathrm{E}(T^s(1)|\, D(s) = 1) - \mathrm{E}(T^s(0)|\, D(s) = 1) \qquad (18)$$

is only useful when it comes to testing for the existence of a treatment effect. For the purpose of testing one would potentially like to average over the distribution of program starts, i.e., calculate $\mathrm{E}_S(\Delta_1(s))$.[10] However, as shown in Proposition 4, this estimand is not useful for estimating the magnitude of the causal treatment effect. Under H_1, the fact that individuals may enter at future time points affects the duration distribution and, thus, $T^s(0) \not\perp D(s)$.

3.3.1 Estimators in discrete time

What assumptions are required to estimate a causal treatment effect? One alternative is to assume that time is discrete.[11] Therefore, let $Y_s(t^s, 1)$ denote the potential employment outcome in t^s if treated at s. Thus, e.g., $Y_s(t^s, 1) = 1$ if an individual treated at s is employed at time t^s. Further, define $R_s(t^s, 1) = \sum_{u=1}^{t^s-1} Y_s(u, 1)$. Then $\chi_s(t^s, 1) = E(Y_s(t^s, 1)|R_s(t^s, 1) = 0)$ and this can be estimated using the sample taking treatment at s:

$$\chi_s(t^s, D(s) = 1) = \frac{n_s^1(t^s)}{R_s^1(t^s)} = \frac{\sum_{i=1}^{R_s^1(t^s)} y_{i,s}(t^s)}{R_s^1(t^s)},\ t^s = 1, ..., \widetilde{l} - s \qquad (19)$$

where $y_{i,s}(t^s) = 1$ if individual i, treated at s, is employed at t^s and $y_{i,s}(t^s) = 0$ otherwise, $R_s^1(t^s) = R(t^s, D(s) = 1)$ is the number of individuals, treated at s, who are still unemployed at t^s, $n_s^1(t^s) = \sum_{i=1}^{R_s^1(t^s)} y_{i,s}(t^s)$ is the number who exit to employment at t^s, and $\widetilde{l}$ is a fixed censoring date.

How should the counterfactual hazard to employment – i.e. the hazard if not treated by s – be calculated? Let $Y_s(t^s, 0)$ be the potential employment outcome in t^s if not treated at s, and $R_s(t^s, 0) = \sum_{u=1}^{t-1} Y_s(u, 0)$ then $\chi_s(t^s, 0) = E(Y_s(t^s, 0)|R_s(t^s, 0) = 0)$. Notice that those who did not take treatment in period s will not have had time to experience an outcome as treated in the following period. As a consequence the sequence of potential outcomes if not treated will be the same for the treated at s and the

[10] Sianesi (2001) has an analogous definition of the estimand of interest.

[11] For the practical application that we have in mind, i.e. unemployment duration, this assumption is not restrictive; it may be more restrictive in other circumstances and then one should bear in mind the potential bias caused by time aggregation.

non-treated at s. Thus

$$\{Y_s(t^s,0)\}_{t^s=1}^{\infty} \perp\!\!\!\perp D(s). \tag{20}$$

This then yields

$$\{R_s(t^s,0)\}_{t^s=1}^{\infty} \perp\!\!\!\perp D(s) \tag{21}$$

and

$$\{\chi_s(t^s,0)\}_{t^s=1}^{\infty} \perp\!\!\!\perp D(s). \tag{22}$$

That is, the sequence of potential hazard rates if not treated is independent of treatment status at s. We can thus compare the hazard rate for those who took treatment in time period s with those that did not, since those who did not take treatment in period s will not have had the time to experience an outcome as treated.

This implies that the estimates of the counterfactual hazard to employment for those treated by s can be based on the not yet treated at s:

$$\chi_s(t^s, D(s)=0) = \frac{n_s^0(t^s)}{R_t^0(t^s)} = \frac{\sum_{j=1}^{R_s^0(t^s)} y_{j,s}(t^s)}{R_s^0(t^s)},\ t^s = 1, ..., \widetilde{l} - s$$

where $R_s^0(t^s) = R_s(t^s, D(s)=0)$ is the number of individuals who have not entered a program at time s and who are still openly unemployed (i.e., not in the program) at time t^s and $n_s^0(t^s)$ is the number leaving for employment at t^s.

Conditioning on s, the survival function for the treated and the counterfactual survival function can be estimated as

$$\overline{F}_s^j(t^s) = \prod_{u=1}^{t^s}(1 - \chi_s(t^s, D(s)=j)),\ t = 1, ..., \widetilde{l} - s,\ j = 0, 1$$

The effect of joining the program at s can then be calculated as the difference between the two survival functions, i.e.

$$\widehat{\Delta}_s(t^s) = \overline{F}_s^1(t^s) - \overline{F}_s^0(t^s),\ t^s = 1, ..., \widetilde{l} - s. \tag{23}$$

It would in principle be possible to estimate (7) and (9) by averaging over the observed distribution of S. There are some complications associated with sampling and inference, however. First, we must observe t^s of equal length for each s and hence $\widetilde{l}$ must be made dependent on $F_S(s|D=1)$. Second, it is not clear how the variance of the estimator should be calculated. Owing to these complications, we consider two simpler possibilities.

Estimator I One option, proposed in Fredriksson and Johansson (2003a), is to calculate the difference between the two flow sample survival functions

$$\overline{F}^{c}(t,j)=\prod_{u=0}^{t}(1-\lambda(u,D(u)=j),\ j=0,1 \tag{24}$$

Here $\lambda(u,D(u)=0)$ is the estimator of the hazard rate to employment if non-treated for the population who has been treated by u, while $\lambda(u,D(u)=1)$ "estimates" the hazard rate to employment if treated by u. The interpretation of the estimand $\Delta^{c}(t)=\overline{F}^{c}(t,1)-\overline{F}^{c}(t,0)$ is the difference in survival rates for the population treated by t. This is an alternative to the proportional hazard specification.

Estimator II Another option is to device an estimator of (10) in discrete time. The potential hazard at t^{s} if treated is equal to

$$\begin{aligned}\chi(t^{s},1) &= E_{R_{s}(t^{s},1)}E(Y_{s}(t^{s},1)|R_{s}(t^{s},1)=0,D=1)\\ &= \sum_{s=0}^{\widetilde{l}}\chi_{s}(t^{s},1)\Pr(R_{s}(t^{s},1)=0,D=1)\end{aligned}$$

where $\Pr(R_{s}(t^{s},1)=0,D=1)$ is the (potential) probability to still be unemployed (remain in the risk set) at t^{s} if treated at s. This hazard is easily calculated as

$$\begin{aligned}\chi_{1}^{1}(t^{s}) &= \chi(t^{s},D=1)=\frac{\sum_{s=0}^{\widetilde{l}}\sum_{i=1}^{R_{s}^{1}(t^{s})}y_{i,s}(t^{s})}{\sum_{s=0}^{\widetilde{l}}R_{s}^{1}(t^{s})}\\ &= \sum_{s=0}^{\widetilde{l}}\chi_{s}(t^{s},D(s)=1)\times p_{s}^{1}(t^{s}),\end{aligned}$$

where $p_{s}^{1}(t^{s})=R_{s}^{1}(t^{s})/\sum_{s=0}^{\widetilde{l}}R_{s}^{1}(t^{s})$ is the sample probability to be unemployed at t^{s} if treated at s. Denote the potential hazard to employment if not treated for the treated $\chi(t^{s},0|D=1)$. Now

$$\begin{aligned}\chi(t^{s},0|D &= 1)=E_{R_{s}(t^{s},0)}E(Y_{s}(t^{s},0)|R_{s}(t^{s},0)=0,D=1)\\ &= \sum_{s=1}^{\widetilde{l}}\chi_{s}(t^{s},0)\Pr(R_{s}(t^{s},0)=0,D=1)\end{aligned}$$

where $\Pr(R_{s}(t^{s},0)=0|D=1)$ is the potential probability to still be unemployed (remain in the risk set) at t^{s} if not treated at s for the treated population. Given the independence assumption (20) $\chi(t,0|D=1)$ can estimated as

$$\begin{aligned}\chi_0^1(t^s) &= \frac{\sum_{s=1}^{\widetilde{l}}\sum_{j=1}^{R_s^0(t^s)} y_{j,s}(t^s)}{\sum_{\widetilde{s}=1}^{\widetilde{l}} R_s^0(t^s)} \\ &= \sum_{s=0}^{\widetilde{l}} \chi_s(t^s, D(s)=0) \times p_s^0(t^s)\end{aligned}$$

where $p_s^0(t^s) = R_s^0(t^s)/\sum_{s=1}^{\widetilde{l}} R_s^0(t^s)$ is the estimated probability for the population of treated to be at risk if not treated at s.

The survival function in unemployment for the average treated individual if treated and non-treated can then be estimated as

$$\overline{F}_j^1(t^s) = \prod_{u=1}^{t^s}(1-\chi_j^1(u)),\ t^s = 1,...,\widetilde{l},\ j=0,1 \tag{25}$$

Finally, the effect of joining the program is given by the difference between the two survival functions, i.e.

$$\widehat{\Delta}^1(t^s) = \overline{F}_1^1(t^s) - \overline{F}_0^1(t^s), t^s = 1,...,\widetilde{l} \tag{26}$$

Notice that $\overline{F}_1^1(t^s)$ is the maximum likelihood estimator (MLE) of $\overline{F}^1(t^s,1)$; see Kalbfleich and Prentice (1980) ch. 4. Therefore, plim$\overline{F}_1^1(t^s) = \overline{F}^1(t^s,1)$. We can now make a statement about the virtue of (26)

Proposition 5 plim $\widehat{\Delta}^1(t) = \overline{F}^1(t^s,1) - \overline{F}^1(t^s,0)$.

Proof. Since $\{Y_s(t^s,0)\}_{t^s=1}^{\infty} \perp\!\!\!\perp D(s)$, $\overline{F}_0^1(t^s)$ is the MLE of $\overline{F}^1(t^s,0)$. Hence, plim $\overline{F}_0^1(t^s) = \overline{F}^1(t^s,0)$ and the proposition follows. ■

The reasoning above clearly shows that the hazard $\chi_0^1(t^s)$ and its complement, the survival function $\overline{F}_0^1(t^s)$, define a matching estimator. To apply it, we just have to create the same distribution of entrance dates s as for those who actually enter the program. The proposed estimator thus balances the pre-treatment duration. In this sense the estimator is analogous to what random assignment accomplishes in experimental data. Random assignment then balances the pre-treatment duration (and any other covariates) for the treated and control group, which is why we do not have to condition on the length of the period prior to treatment entry when estimating the treatment effect.

3.3.2 Observed heterogeneity and inference

Here we device a matching estimator for the more realistic case with observed heterogeneity. Then, we discuss the inferential aspects of this estimator.

Denote by $\mathbf{x}_s$ the observed covariates for the population treated at s. Then it is straightforward to show that

Proposition 6 *If* $\{Y_s(t^s,0)\}_{t^s=1}^{\infty} \perp\!\!\!\perp D(s)|\mathbf{x}_s$, *the counterfactual survival function for the average treated individual* $\overline{F}^1(t^s,0)$ *can be estimated using individuals that are non-treated by* s *and have covariates* $\mathbf{x}_s$.

Proof. see Appendix 2. ■

The survival functions are:

$$\overline{F}_j^1(t^s) = \prod_{u=1}^{t^s}(1-\chi_j^1(u)),\ t^s = 1,...,\widetilde{l},\ j = 0,1 \tag{27}$$

where

$$\begin{aligned}\chi_j^1(u) &= \sum_{s=0}^{\widetilde{l}} \chi_s(u, D(s) = j|\mathbf{x}_s) \times \frac{R_s^0(u|\mathbf{x}_s)}{R^0(u|\mathbf{X}_{\widetilde{l}})} \\ &= \frac{\sum_{s=0}^{\widetilde{l}} \sum_{i=1}^{R_s^j(u|\mathbf{x}_s)} y_i(u)}{\sum_{s=0}^{\widetilde{l}} R_s^j(u|\mathbf{x}_s)},\end{aligned}$$

$\mathbf{X}_{\widetilde{l}} = \{\mathbf{x}_s\}_0^{\widetilde{l}}$ denote the covariates for the treated population and $R_s^j(u|\mathbf{x}_s) = R(u, D(s) = j|\mathbf{x}_s)$ is the number of persons with covariates $\mathbf{x}_s$ still at risk and unemployed in u.

Inference For purposes of inference, a simple, but not quite correct,[12] way of calculating $\text{Var}(\widehat{\Delta}^1(t^s))$ is: $\text{Var}(\widehat{\Delta}^1(t^s)) = \text{Var}(\overline{F}_1^1(t^s))+ \text{Var}(\overline{F}_0^1(t^s))$, where the variance for the estimated survival function is equal to (see, e.g., Lancaster, 1990)

$$\text{Var}(\overline{F}_j^1(t)) = \overline{F}_j^1(t)^2 \sum_{u=1}^{t} \frac{n^j(u)}{(R^j(u)-n^j(u))R^j(u)},\ j = 0,1 \tag{28}$$

Here $n^1(u) = \sum_{s=0}^{\widetilde{l}} \sum_{i=1}^{R_s^1(u)} y_{i,s}(u)$, $n^0(u) = \sum_{s=0}^{\widetilde{l}} \sum_{j=1}^{R_s^0(u)} y_{j,s}(u)$, $R^1(u) = \sum_{s=0}^{\widetilde{l}} R_s^1(u)$ and $R^0(u) = \sum_{s=0}^{\widetilde{l}} R_s^0(u)$.

[12]The variance calculation is not quite correct since it ignores the covariance between the estimated survival functions. Ignoring the covariance does not seem to lead to biased inference (see Fredriksson and Johansson, 2003a).

Fredriksson and Johansson (2003a) study the small sample performance of the $\Delta^c(t) = \overline{F}^c(t,1) - \overline{F}^c(t,0)$ estimator (i.e. the estimator based on the survival function (24)) and the alternative estimators that we reviewed in section 3. We find that the survival function estimator is reliable in terms of testing for a treatment effect. Under the null hypothesis of no treatment, there is a substantial negative bias (i.e. a bias towards finding that the program does not work) in the matching approach applied by, e.g., Gerfin and Lechner (2002), Larsson (2003), and Lechner (1999, 2002). The bias is, as expected, increasing in length of the observation window and the sizes of their Wald tests are too large. Therefore, the null hypothesis is rejected too often and one may even find statistically significant negative treatment effects. The estimator we propose suffers from no bias (under H_0) and a our Wald test gives the correct size.[13]

4 An empirical application

In this section we apply the estimators defined by (26) and (28) to evaluate the effects of an employment subsidy (ES) program. In addition we also estimate treatment effects that vary by entrance date, i.e., we calculate the estimator given by (23).

The ES program was introduced on January 1, 1998.[14] The subsidy was targeted at the long-term unemployed, i.e., persons at least 20 years-of-age and registered as unemployed at the public employment service (PES) for at least 12 months. The subsidy amounted to 50 percent of total wage costs and was paid for a maximum period of 6 months.[15]

We use register data from the National Labour Market Board. The database, *the unemployment register*, contains information on all individuals registering at the PES in Sweden since August 1991. The database includes information on, e.g., age, educational attainment and sex, as well as the individuals' registration date, job training activities and starting dates of participation in various labour market programs.

For each individual registered at the PES we observe an event history including the number of spells and days of unemployment. Everyone that left the register before the introduction of the ES program are dropped from

[13] The small sample performance of the two estimators $\Delta^c(t)$ and $\widehat{\Delta}^1(t^s)$ and the accompanying Wald test are very similar.

[14] For a more thorough description of the employment subsidy programmes, see Forslund et al. (2004) and the references therein.

[15] The subsidy was also capped at SEK 350 per day and could be extended to 12 months in some exceptional cases.

the data. We also exclude all individuals for which the first spell of unemployment occurred before January 1, 1992 and all registered spells shorter than 365 days. The reason for the last two exclusions is that previous labour market history is the key variable for the matching estimator and the main eligibility criteria for the program is continuous unemployment for at least 365 days.

Until December 31, 2000, individuals under 25 years-of-age had the possibility to start the ES program with a registered spell of only 90 days. Due to this exception, all unemployed persons under 25 years of age on January 1, 1998, (the starting date of the program) or later are excluded from the data set. We have also excluded those who at the month of registration at the PES were at least 63 years-of-age (15,160 persons).[16] We have also excluded all individuals who had spells with negative duration before the last spell (324 spells). Finally, because we aggregate time to monthly intervals, we have discarded 63 ES spells that ended within 29 days.

A spell of unemployment is defined as an uninterrupted period of time when an unemployed person is registered at the PES. The spell is ended if the unemployed person gets a job for a period of at least 30 days, or if he or she, for any other reason, leaves the register for a period of at least 30 days.

It is possible to have more than one spell of unemployment of at least 365 days without interruption during the time the ES program has been going on. Thus, an individual can be eligible for the program more than once. The unit of observation is chosen to be every time a person becomes eligible for the ES program. In the analysis we use information on each individual's total number of spells and days in unemployment before becoming eligible for the ES. For those who are eligible more than once the total number of days and spells is aggregated each time they become eligible. Thus, the data include only persons who have been eligible for the ES program on at least one occasion.

The individuals in the data are separated into two different groups: those who start the ES program after having become eligible and eligibles who do not start the program. Each time a person becomes eligible, the total number of days until he or she either leaves the employment service office or becomes right censored is calculated. The point in time for right censoring is 1 October 2002 or when a person leaves the register for other destinations than work. For those who enter the ES program, the duration to ES is calculated as well.

[16] This is done because retirement is imminent for these individuals. The eligibility condition of being registered for one year coupled with a program duration of at least six months leaves half a year before statutory retirement occurs (normally at age 65).

Individuals are defined as having found a job if they deregister for employment or for temporary employment for a minimum period of 30 days. A non-trivial number of persons leave the register for unknown reasons (which may well include work, see Bring and Carling, 2000). To the extent that there are systematic differences between participants and non-participants in the fraction of those who leave the register for unknown reason that actually leave it for work, we would get biased estimates of the treatment effect by using our definition of the outcome of interest.[17]

A total of 631,358 individuals, aged 25–63, were eligible for ES between January 1998 and October 2002. This population of eligibles is described in *Table 1*.[18] Three percent of the eligible spells ended in ES. The most salient feature of the eligible persons is that they on average had a long lasting relationship with the employment service: in addition to the days spent in the register in order to become eligible, the average number of days in the register was almost 500 and the average number of earlier spells in the register was almost 1.5. Approximately 40 percent of the spells ended in regular employment.

The mean characteristics of the ES participants and non-participants in the eligible population are reported in *Table 2*. A significantly higher fraction (64 percent as compared to 39 percent) of the ES participants ended up in employment. This does not indicate a positive treatment effect; it more likely reflects the fact that the program participants on average registered earlier (see T_0) at the PES and, hence, on average had spent a longer time looking for a job. Males and non-Nordic immigrants are over-represented and disabled are under-represented among the participants. Participants are younger, more educated, and have spent less time at the employment service prior to the last period of unemployment. Given reasonable priors about how these characteristics should influence the exit to employment, the participants should be expected to leave unemployment more rapidly than the non-participants.

To apply our estimators in the present setting we must check whether selection on observables and no-anticipation are reasonable assumptions. To make a long story short, we think that these assumption are palatable in the present setting but let us substantiate this claim somewhat.

In a stated preference experiment, Eriksson (1997) finds that the heterogeneity of the PES caseworker is more important for determining program

[17]This problem has been pointed out by Sianesi (2001, 2002). Forslund et al. (2004) performed a sensitivity analysis with respect to this issue. They found that, if anything, the treatment effect of ES is biased downward.

[18]Note that we have categorized a number of variables since we want to perform an exact matching procedure.

Table 1: Descriptive statistics for all eligibles

Variable	Description	Mean	Std. dev.	Min	Max
ES	=1 if in ES program	0.03	0.17	0	1
Duration	Current spell duration (months)	23.7	23.2	1	118
Employed	=1 if regularly employed	0.40	0.49	0	1
Male	=1 if male	0.41	0.49	0	1
NonNordic	=1 if non-Nordic citizen	0.14	0.35	0	1
NoUI	=1 if no unemployment insurance	0.18	0.38	0	1
Disabled	=1 if disabled	0.10	0.30	0	1
Gymnasium	=1 if upper secondary degree	0.35	0.48	0	1
University	=1 if university degree	0.12	0.33	0	1
Age_1	=1 if age ≤ 30	0.22	0.42	0	1
Age_2	=1 if $30 <$ age ≤ 40	0.31	0.46	0	1
Age_3	=1 if $40 <$ age ≤ 50	0.24	0.43	0	1
TD_1	=1 if days in register during previous spell (TD) = 0	0.38	0.48	0	1
TD_2	=1 if $0 < \text{TD} \leq 100$	0.05	0.21	0	1
TD_3	=1 if $100 < \text{TD} \leq 500$	0.20	0.40	0	1
TD_4	=1 if $500 < \text{TD} \leq 1000$	0.18	0.38	0	1
TP_1	=1 if previous number of programmes (TP) = 0	0.39	0.49	0	1
TP_2	=1 if $0 < \text{TP} \leq 5$	0.39	0.49	0	1
TP_3	=1 if $5 < \text{TP} \leq 15$	0.21	0.41	0	1
T_0	Month turning eligible, January 1998=1 October 2002=118	69.8	27.6	1	118

Table 2: Mean characterstics of participants (ES), non-participants (No ES), and exactly matched sample (Matched).

Variable	ES mean	No ES mean	ES-No ES t-value	Matched mean
Duration	34.38	23.37	62.90	–
Employed	0.64	0.39	71.54	–
		Covariates		
Male	0.61	0.41	56.38	0.56
NonNordic	0.21	0.14	25.41	0.14
NoUI	0.16	0.18	-5.43	0.11
Disabled	0.06	0.10	-20.41	0.02
Gymnasium	0.43	0.35	24.08	0.42
University	0.12	0.12	-3.43	0.09
Age_1	0.26	0.22	11.34	0.24
Age_2	0.32	0.31	3.39	0.30
Age_3	0.27	0.24	7.04	0.25
TD_1	0.41	0.38	9.44	0.51
TD_2	0.05	0.05	4.28	0.02
TD_3	0.22	0.20	8.37	0.19
TD_4	0.18	0.18	1.22	0.15
TP_1	0.41	0.38	7.15	0.51
TP_2	0.42	0.39	7.82	0.35
TP_3	0.17	0.22	-18.62	0.14
T_0	58.33	70.22	-65.75	60.18

participation than the heterogeneity of the individuals. Carling and Richardson (2001) report evidence in the same vein. They compared the effects of eight different programs on the probability of finding a job. They argue that their results probably do not reflect selection by showing that program placement depended more on the employment service office that the job seeker had visited than on her observed characteristics.

For the specific ES program we are considering there is also survey evidence on the selection process (see Lundin, 2000). The survey was directed to the caseworkers at the PES. The majority of the caseworkers (55 %) reported that the first initiative to enter ES was taken by them; an additional 33 percent said that the initiative came from the employer; only 6 percent maintained that the initiative came from the eligible. In the same survey, the caseworkers were asked about the criteria that were important when suggesting ES participation. The caseworkers were allowed to give multiple responses. The following list orders the criteria by their importance: (1) the formal requirements of ES (84 %); (2) the motivation of the eligible (58 %); (3) that the eligible belonged to a high priority group (31 %); (4) that the employer has suggested to employ the eligible (13 %); (5) that the eligible has education for the job (11 %); (6) in order to extend the period on unemployment insurance (UI) for the eligible (9 %); and (7) the unemployed suggested ES (5 %).

All in all, individual self-selection to ES do not seem to be a big issue. The potential threat to the selection on observables assumption is the case workers' appreciation of the motivation of the unemployed individual. However, Eriksson (1997) shows that this appreciation is very heterogenous among caseworkers, suggesting that this may not be so problematic. Moreover, we have detailed information on the previous labour market history (presumably a good indicator of motivation). Also, there is information on the local labour market[19] where the individual is registered. We can thus control for any common component in the appreciation of the motivation of the unemployed individual; c.f. Carling and Richardson (2001).

The no-anticipation assumption requires that the unemployed neither knows the exact date of treatment start nor the exact date of the start of the job spell. There must be some randomness in these events. Therefore, it is comforting to see that only 9 percent of the caseworkers think that extending the period on UI is important, since the expiration of benefits is an event that is known for certain in advance. Anticipation is also less of a concern since we are using monthly rather than daily data.

Concerns about anticipation is not the main reason for the time aggre-

[19] There are 100 local labour markets and the definition is based on commuting patterns.

gation, however. The principal reason is that there is a good deal of measurement error in the exact day of a start of a job spell. The main reason for this measurement error is the strategy used by PES officers to obtain the information on when the job spell began. If the unemployed individual has not been in contact with the PES office for some specific time period, the unemployed is asked over the phone whether s(he) is employed or not. As noted in section 3.1, time aggregation of data may lead to a small downward bias in the estimated treatment effects.

We match on the covariates listed in *Table 2* and the local labor market where the individual is registered. In this application, we use a one-to-one exact matching estimator. Alternatively, a propensity score matching estimator could be used (see Appendix 2 and Fredriksson and Johansson, 2003b, for an application).[20]

Index the treated at $s = 0, \ldots, \widetilde{l} - 1$ by i and the comparison group at s by c. The unique match (for each s) is then given by

$$c_i = \mathbf{x}_i \equiv \mathbf{x}_c,\ c \in N(s), \tag{29}$$

where $N(s)$ is the number of individuals in the comparison group. If there is more than one individual in the comparison group with the same values of the covariates, we randomize over the potential matches. If no unique match from c is found for individual i, this individual is removed from the estimation. With complete pairs of treated and non-treated individuals, (26) is estimated as

$$\widehat{\Delta}^1(t^s) = \overline{F}_1^{1m}(t^s) - \overline{F}_0^{1m}(t^s), t^s = 1, ..., \widetilde{l} \tag{30}$$

where $\overline{F}_1^{1m}(t^s))$ and $\overline{F}_0^{1m}(t^s)$ are the Kaplan-Meier survival function if treated and non-treated for the subset of treated individual with support common to the comparison sample.

Matching is based on $7,651$ treated individuals.[21] Descriptive statistics for the matched pairs are reproduced in *Table 2*.

4.1 Results

Figure 1 displays the estimated survival functions and *figure 2* shows the estimated treatment effects ($\widehat{\Delta}^1(t^s)$) along with 95 percent confidence intervals.

[20] This would make it more easy to find comparison individuals to the treated individuals and would increase the efficiency of the estimators. On the other hand, estimation of the propensity score may introduce bias.

[21] The original sample consisted of $19,951$ ES individuals. Thus $12,300$ observations were excluded since no matching individual was found in the control group.

In both figures the upper panels is the estimate when we control for both pre ES duration and covariates while the estimate in the lower panel only control for pre ES duration. This means that the differences between the upper and the lower panels reflect the effects of observed heterogeneity.

From *figure 2* we can see that after an initial period of about 6 months with a negligible (negative) treatment effect there is a downward jump; from then on the effect gradually becomes smaller, but it is negative and significant over the rest of the follow-up horizon (57 months). This scenario is consistent with an initial period of locking-in and a subsequent period with a positive treatment effect. The sum of the effects over the whole follow-up horizon is 7.78 months, which corresponds to a decrease in unemployment duration for the average individual by 14 percent over the follow-up horizon.

Further, by comparing the upper and lower panels of *figures 1* and *2*, we can see that the "treatment effects" are reduced considerably by matching. Hence, this confirms our prior (from our inspection of Table *2*) of a positive selection to ES among the eligibles.

One advantage of our estimator is that we can compute treatment effects by pre-treatment duration; see equation (23). In *Figure 3* we plot the treatment effect for those entering during months 0–3 after eligibility, and in *Figure 4* we plot the treatment effect for those entering during months 36–39.[22] The general message is that treatment effects look rather similar irrespective of the timing of program entry; once again we see that it is imperative to control for observed heterogeneity.

A likely explanation for the downward jump in the estimated treatment effect after 6 months is that the participants simply tend to stay-on at the work place where they were employed with the subsidy. On the one hand, this is an intended effect of the program. On the other hand, this result may be seen as an indication that the program tends to displace regular employment. That is, employers use the subsidy to fill vacancies that would have been filled by hiring on the regular market in the absence of the program. This interpretation is consistent with the qualitative evidence reported in Lundin (2000).

5 Concluding remarks

In this paper we have considered the evaluation problem using observational data when the program start is the outcome of a stochastic process. We have

[22] Only 489 matched individuals begin treatment during months 36–39 and the number of matched individuals is 206.

shown that (i) matching with a binary indicator variable using a classification window will yield biased estimates of treatment on the treated effects; (ii) the duration framework in discrete time (with a time-varying treatment indicator) is a fertile ground for effect evaluations.

We have suggested easy-to-use non-parametric matching estimators of the survival functions. These estimators do not rely on strong assumptions about the functional form of the two processes generating the inflow into program and employment. We have assumed that selection is purely based on observables. Whether the conditional independence assumptions required for the estimators are reasonable depend crucially on the richness of the information in the data. Even if we assume that unobserved heterogeneity is not an issue, the evaluation problem is demanding on the data. We need longitudinal data where we can observe the duration path. Knowing the entire path is crucial as we need to screen it in order to define a comparison sample that is matched in terms of the pre-treatment duration.

We think that the issues we have raised applies fairly generally to evaluations of on-going labor market programs. The problems associated with estimating well-defined treatments effects affect all outcomes that are functions of the outflow to employment. Hence, it applies directly when the outcome of interest is employment (or annual earnings) some time after program start. Moreover, if skill loss increases with unemployment duration, as suggested by the recent analysis in Edin and Gustavsson (2001), one should be careful when estimating the effect of treatment on wages. Although it may be tempting to screen the future in order to find individuals who did not take part in the program during some window there is a definite risk associated with doing this. It is more probable that individuals who, by the luck of the dice, found employment are included in the comparison group. But if there is skill loss, this lucky draw will in turn spill over onto wages yielding a negative bias in the estimates of the treatment effects. Thus the issues we have raised here may be important also for studies examining the treatment effects on wages.

References

Abbring, J.H. and G.J. van den Berg (2003), The Non-parametric Identification of Treatment Effects in Duration Models, *Econometrica*, **71**, 1491-1518.

Bring, J. and K. Carling (2000), Attrition and Misclassification of Drop-outs in the Analysis of Unemployment Duration, *Journal of Official Statistics* **16**, 321–30.

Carling, K. and K. Richardson (2001), The Relative Efficiency of Labor market programs: Swedish experience from the 1990's, Working Paper 2001:2, Institute for Labour Market Policy Evaluation.

Crowley, J. and M. Hu (1977), Covariance Analysis of Heart Transplant Survival Data, *Journal of the American Statistical Association,* **72,** 27-36.

Dawid, A.P. (1979). Conditional Independence in Statistical Theory, *Journal of the Royal Statistical Society Series B*, **41**, 1-31.

DiRienzo, A.G. and S.W. Lagakos (2001), Effects of Model Misspecification on Tests of no Randomization Treatment Effect Arising from Cox's Proportional Hazard Model. *Journal of the Royal Statistical Society Series B,* **63**, 745-757.

Edin, P-A. and M. Gustavsson (2001), Time out of Work and Skill Depreciation, mimeo, Department of Economics, Uppsala University.

Eriksson, M. (1997). Placement of Unemployed into Labour Market Programs: A Quasi Experimental Study, Umeå Economic Studies 439, Umeå university.

Forslund, A., P. Johansson, and L. Lindquist (2004). Employment Subsidies – A Fast Lane From Unemployment to Work? mimeo, Institute for Labour Market Policy Evaluation.

Fredriksson, P. and P. Johansson (2003a), Program Evaluation and Random Program Starts. Working Paper 2003:1, Department of Economics, Uppsala University.

Fredriksson, P. and P. Johansson (2003b), Employment, Mobility, and Active Labor Market Programs. Working Paper 2003:5, Department of Economics, Uppsala University.

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

Heckman, J.J., R.J. Lalonde, J.A. Smith (1999), The Economics and Econometrics of Active Labor Market Programs, in O. Ashenfelter and D. Card (eds) *Handbook of Labor Economics* vol. 3, North-Holland, Amsterdam.

Kalbfleich, J.D. and R.L. Prentice (1980). *The Statistical Analysis of Failure Time Data*, New York: Wiley.

Lalive, R, J. van Ours and J. Zweimüller (2002), The Impact of Active Labor Market Programs on the Duration of Unemployment, IEW Working Paper No. **51**, University of Zurich.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data,* Cambridge: Cambridge University Press.

Larsson, L. (2003), Evaluation of Swedish Youth Labour Market programs, *Journal of Human Resources,* **38**, 891-927.

Lechner, M. (1999), Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification, *Journal of Business and Economic Statistics,* **17**, 74-90.

Lechner, M. (2002), program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labour Market Policies, *Review of Economics and Statistics,* **84**, 205-220.

Lundin, M. (2000), Anställningsstödens implementering vid arbetsförmedlingarna, Stencilserie 2000:4, Institute for Labour Market Policy Evaluation.

Richardson, K. and G.J. van den Berg (2002), The Effect of Vocational Employment Training on the Individual Transition Rate from Unemployment to Work, Working Paper 2002:8, Institute for Labour Market Policy Evaluation.

Rosenbaum, P.R. (1995). *Observational Studies* (*Springer Series in Statistics*), Springer Verlag. New York.

Rosenbaum, P.R and D.B. Rubin (1983), The Central Role of the Propensity Score in Observational Studies for Causal Effect, *Biometrika*, **70**, $41-55$.

Sianesi, B. (2001), An Evaluation of the Active Labour Market programs in Sweden, Working Paper 2001:5, Institute for Labour Market Policy Evaluation.

Sianesi (2002), Differential Effects of Swedish Active Labour market programs for Unemployed Adults During the 1990s, Working Paper 2002:5, Institute for Labour Market Policy Evaluation.

van den Berg, G.J., B. van der Klaauw, and J.C. van Ours (2004), Punitive Sanctions and the Transition from Welfare to Work, *Journal of Labor Economics*, **22**, 211–241.

Appendix 1: Proof of proposition 4

It is helpful to first consider the experimental estimate $\hat{\delta}_R$. Suppose we were to conduct an experiment where at $t = 0$ individual are randomly assigned to a treatment ($D = 1$) and a comparison (control) group ($D = 0$). To simplify the exposition, assume that we observe k unique durations after randomization. Order the k survival times such that $t_{(1)} < t_{(2)} < .... < t_{(k)}$. Associate a treatment indicator with each unique duration such that $D_{(j)} = 1$ if the individual has been treated in period $t \leq t_{(j)}$ and $D_{(j)} = 0$ otherwise. Now, consider the partial likelihood

$$L(\delta) = \prod_{j=1}^{k} \left(\frac{\exp(\delta D_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\delta D_l)} \right) = \prod_{j=1}^{k} \left(\frac{\exp(\delta D_{(j)})}{R_{(j)}(1) \exp(\delta) + R_{(j)}(0)} \right)$$

where $R_{(j)}(1)$ and $R_{(j)}(0)$ denote the number of treated and non-treated in the risk-set respectively. The maximum likelihood estimator of δ under random sampling is given as

$$\hat{\delta}_R = \ln \left(\sum_{j=1}^{k} D_{(j)} R_{(j)}(0) \right) - \ln \left(\sum_{j=1}^{k} R_{(j)}(1)(1 - D_{(j)}) \right).$$

If there is no treatment effect then

$$\begin{aligned} E(D_{(j)} R_{(j)}(0)) &= E(R_{(j)}(0) | D_{(j)} = 1) \Pr(D_{(j)} = 1) \\ &= E(R_{(j)}(0)) \Pr(D = 1) \end{aligned} \tag{31}$$

and

$$\begin{aligned} E((1 - D_{(j)}) R_{(j)}(1)) &= E(R_{(j)}(1) | D_{(j)} = 0) \Pr(D_{(j)} = 0) \\ &= E(R_{(j)}(1)) \Pr(D = 0) \end{aligned} \tag{32}$$

and hence $\hat{\delta}_R \xrightarrow{p} 0$. If $\delta > 0$ then, $R_{(j)}(1)$ and $D_{(j)}$ are no longer independent and $\Pr(D_{(j)}) \neq \Pr(D)$.

Now consider the partial likelihood in the observational setting

$$\begin{aligned} L(\delta) &= \prod_{j=1}^{k} \left(\frac{\exp(\delta D_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\delta D_l)} \right) \\ &= \prod_{j=1}^{k} \left(\frac{\exp(\delta D_{(j)})}{R_{(j)}(1) \exp(\delta) + R_{(j)}(0) + R_{(j)}(0|1)} \right) \end{aligned} \tag{33}$$

The difference compared with the partial likelihood in the experimental setting is the inclusion of $R_{(j)}(0|1)$, which is the number of individuals that have not been treated at $t \leq t_{(j)}$ but will be treated in the future. The estimator for the observational data is equal to

$$\hat{\delta}_{PH} = \ln\left(\sum_{j=1}^{k} D_{(j)}(R_{(j)}(0) + R_{(j)}(0|1))\right) - \ln\left(\sum_{j=1}^{k} R_{(j)}(1)(1 - D_{(j)})\right),$$

If there is no treatment effect (i.e. $\delta = 0$) then, as above, $\Pr(D_{(j)}) = \Pr(D)$; that is, the probability to enter treatment at duration $t_{(j)}$ is the same at the probability to enter treatment for a randomly chosen individual at $t = 0$. This means that the probability to belong to the comparison group is not dependent on the order (j) of the durations and as a result we get the same expressions as above; hence, $\text{plim}\hat{\delta}_{PH} = 0$. The inclusion of those treated in the future in the risk-set, i.e. $R_{(j)}(0|1)$, balances the bias that would result if only the never treated are used as comparisons.

If $\delta \neq 0$ then $\text{plim}\hat{\delta}_R = \delta$. This estimator is only based on the rank orders of the treated relative to the rank orders for those not treated.[23] In the observational setting the only change (from the case without a treatment effect) in rank order is for the individuals who are never treated and the estimator $\hat{\delta}_{PH}$ will be biased downwards in absolute terms; hence $\text{plim}|\hat{\delta}_{PH}| < |\delta|$.

[23] Note that the rank statistic is sufficient to yield consistent estimates of the parameters in the proportional hazards model without knowledge of $\lambda_0(\cdot)$. This is also true if the true model is of the non-proportional variety (see DiRienzo and Lagakos, 2001). Wald tests of a treatment effect are biased, however.

Appendix 2: Matching

Proof of proposition 6 The CIA

$$\{Y_s(t^s,0)\}_{t^s=1}^{\infty} \perp\!\!\!\perp D(s)|\mathbf{x}_s \tag{34}$$

implies

$$\{R_s(t^s,0)\}_{t^s=1}^{\infty} \perp\!\!\!\perp D(s)|\mathbf{x}_s,\ \forall s.$$

and

$$\{\chi_s(t^s,0)\}_{t^s=1}^{\infty} \perp\!\!\!\perp D(s)|\mathbf{x}_s, \forall s.$$

Thus

$$\chi_s(t^s,0) = \mathrm{E}_{\mathbf{x}_s}[\chi_s(t^s, D(s)=1)|\mathbf{x}_s] = \mathrm{E}_{\mathbf{x}_s}[\chi_s(t^s, D(s)=0)|\mathbf{x}_s],\ \forall t^s \text{ and } s$$

and

$$\overline{F}_s(t^s,0) = \mathrm{E}_{\mathbf{x}_s}[\overline{F}_s(t^s, D(s)=1)|\mathbf{x}_s] = \mathrm{E}_{\mathbf{x}_s}[\overline{F}_s(t^s, D(s)=0)|\mathbf{x}_s],\ \forall t^s \text{ and } s$$

Furthermore we get from using (34) that

$$\chi(t^s,0) = \mathrm{E}_{p_s(t^s|\mathbf{x}_s)}[\chi(t^s, D(s)=1|\mathbf{x}_s] = \mathrm{E}_{p_s(t^s|\mathbf{x}_s)}[\chi(t^s, D(s)=0)|\mathbf{x}_s)],\ \forall t^s$$

and

$$\overline{F}^1(t^s,0) = \mathrm{E}_{p_s(t^s|\mathbf{x}_s)}[\overline{F}_s(t^s, D(s)=1)|\mathbf{x}_s] = \mathrm{E}_{p_s(t^s|\mathbf{x}_s)}[\overline{F}_s(t^s, D(s)=0)|\mathbf{x}_s],\ \forall t^s,$$

where $\mathrm{E}_{p_s(t^s|\mathbf{x}_s)}$ is the expectation with respect to the distribution of the probability of belonging to the risk set (if not treated) given treatment at s, (i.e. $p_s(t^s|\mathbf{x}_s) = \Pr(R_s(t^s,0|\mathbf{x}_s)=0|D=1)$).

Propensity score matching Let the conditional probability for the population at risk of being treated at s given $\mathbf{x}_s$ be given by $e(\mathbf{x}_s) = \Pr(D(s) = 1|\mathbf{x}_s)$ and let $0 < e(\mathbf{x}_s) < 1$ for all $\mathbf{x}_s$.[24] By the conditional independence assumption (34) it holds that (see Rosenbaum and Rubin, 1983)

$$\mathbf{x}_s \perp\!\!\!\perp D(s)|e(\mathbf{x}_s)$$

Thus, the counterfactual can be estimated as

$$\begin{aligned}\overline{F}^1(t^s,0) &= \mathrm{E}_{p_s(t^s|e(\mathbf{x}_s))}[\overline{F}_s(t^s, D(s)=1)|e(\mathbf{x}_s)] \\ &= \mathrm{E}_{p_s(t^s|e(\mathbf{x}_s))}[\overline{F}_s(t^s, D(s)=0)|e(\mathbf{x}_s)],\ \forall t^s.\end{aligned}$$

[24] This means that for each $\mathbf{x}_s$ satisfying the CIA there must be individuals in both states.

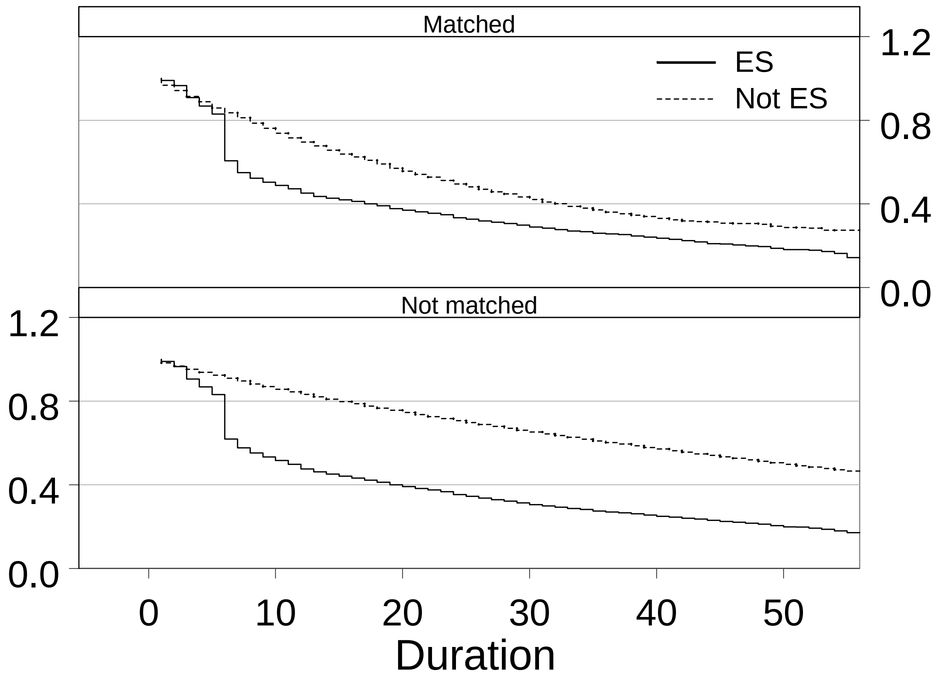

Figure 1: Survival functions to employment for participants and eligible non-participants

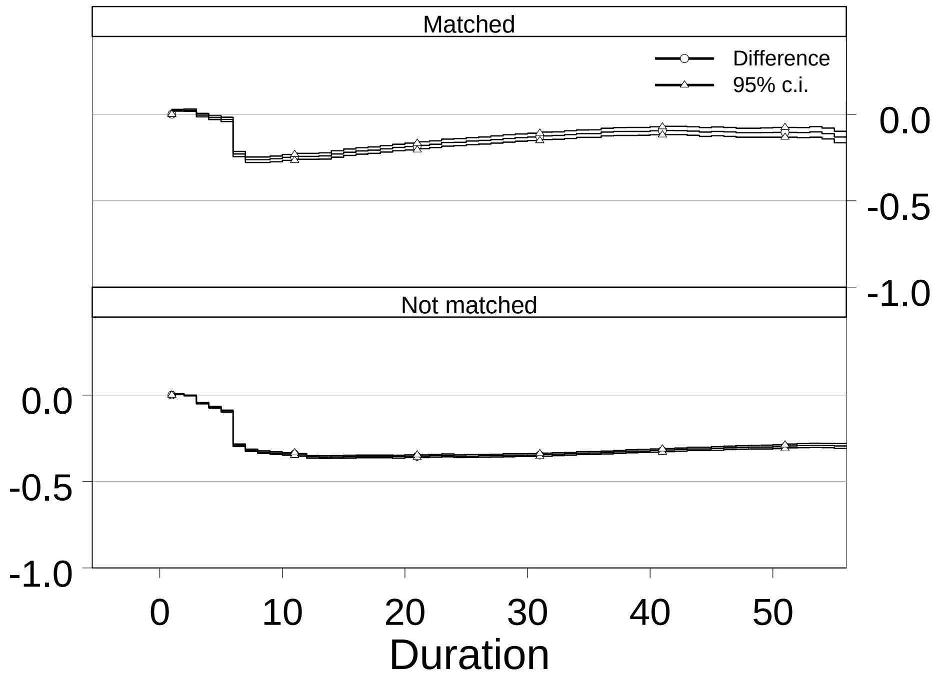

Figure 2: Estimated treatment effect

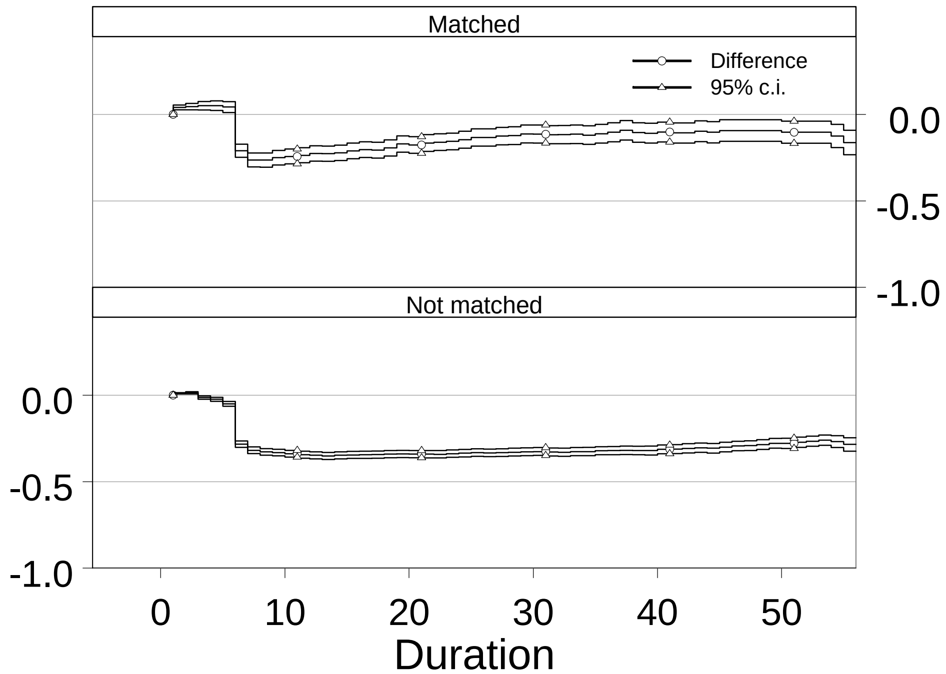

Figure 3: Estimated treatment effect for participants entering during months $t = 0, \ldots, 3$

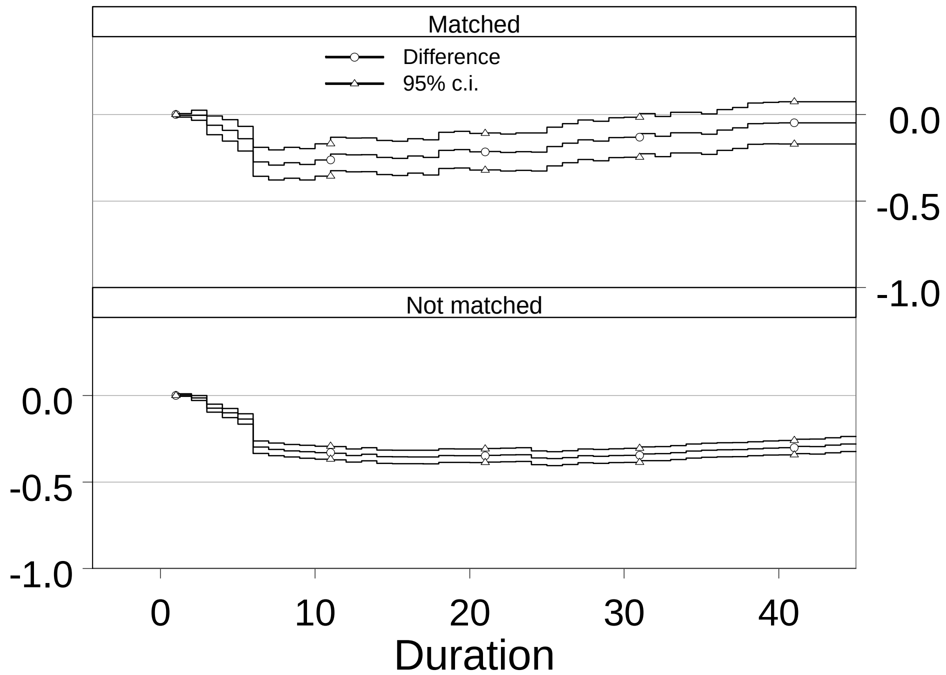

Figure 4: Estimated treatment effect for participants entering during months $t = 36, \ldots, 39$