



Universität St.Gallen

## Matching Estimating of Dynamic Treatment Models: Some Practical Issues

Michael Lechner

January 2006, Discussion Paper no. 2006-03

Editor: Prof. Jörg Baumberger  
University of St. Gallen  
Department of Economics  
Bodanstr. 1  
CH-9000 St. Gallen  
Phone +41 71 224 22 41  
Fax +41 71 224 28 85  
Email [joerg.baumberger@unisg.ch](mailto:joerg.baumberger@unisg.ch)

Publisher: Department of Economics  
University of St. Gallen  
Bodanstrasse 8  
CH-9000 St. Gallen  
Phone +41 71 224 23 25  
Fax +41 71 224 22 98

Electronic Publication: <http://www.vwa.unisg.ch>

# Matching Estimating of Dynamic Treatment

## Models: Some Practical Issues

Michael Lechner

Author's address:

Professor Dr. Michael Lechner  
University of St. Gallen  
SIAW-HSG  
Bodanstrasse 8  
CH-9000 St. Gallen  
Tel. +41 71 224 23 50  
Fax +41 71 224 22 98  
Email [michael.lechner@unisg.ch](mailto:michael.lechner@unisg.ch)  
Website [www-siaw.unisg.ch](http://www-siaw.unisg.ch)

**Abstract**

Lechner and Miquel (2001) approached the causal analysis of sequences of interventions from a potential outcome perspective based on selection on observable type of assumptions (sequential conditional independence assumptions). Lechner (2004) proposed matching estimators for this framework. However, many practical issues that might have substantial consequences for interpretation of the results have not been thoroughly investigated so far. This paper discusses some of these practical issues. The discussion is related to estimates based on an artificial data set for which the true values of the parameters are known and that shares many features of data that could be used for an empirical dynamic matching analysis.

**Keywords**

Dynamic treatment regimes, nonparametric identification, causal effects, sequential randomisation, programme evaluation, treatment effects, dynamic matching, panel data.

**JEL Classification**

C31, C41

## 1 Introduction\*

The goal of the paper is to address practical issues that come with the non- or semiparametric estimation of dynamic treatment models that are identified by sequential selection-on-observables (or conditional independence) assumptions.

While the effects of dynamic selection bias and the impact of sequential interventions received little attention in the applied econometrics literature so far, there is a substantial literature about the estimation of average ‘causal effects’ of interventions using large micro data based on a static causal model. Angrist and Krueger (1999) and Heckman, LaLonde, and Smith (1999) provide a comprehensive overview over this vast literature. Robins (1986) is the first one who suggests an explicitly dynamic causal framework based on potential outcomes that allows to define causal effects of dynamic interventions and to systematically address this type of selection problem. His approach was applied subsequently in epidemiology and biostatistics. The estimators used in these applications typically involve parametric assumptions not required for identification, like parametric duration modelling. Gill and Robins (2001) extend his approach to the case of continuous covariates and treatments. Recently, Lechner and Miquel (2001, LM01 further on) extend Robins’ (1986) framework to comparisons of more general sequences, different parameters and selection processes, and to identifying assumptions that are more relevant in typical microeconomic studies. Focussing on the case when all elements that influence selection and outcome at each stage of the sequence are observable, LM01 discuss different identification conditions required for particular dynamic causal effects. Since the assumptions used in LM01 bear some similarity to the selection on observables or conditional independence assumption

---

\* I have further affiliations with CEPR, London, ZEW, Mannheim, IZA, Bonn, and PSI, London. This paper benefited considerably from previous work with Ruth Miquel about sequential treatment models, as well as from the consistency checks she performed with the artificial data. I thank Conny Wunsch for careful proof reading of a previous version of this paper.

(CIA) that is prominent in the static evaluation literature, Lechner (2004, L04 further on) proposed matching and inverse probability weighting estimators that are dynamic extensions of similar estimators used in the static model. These estimators retain most of the flexible and convenient properties of the static methods that made them the workhorse in empirical evaluation studies (see the excellent survey by Imbens, 2004).

Since there are no experiences so far with this type of estimation for these models (other than the two illustrative examples in LM01 and L04), this paper discusses several issues that come up when the dynamic approach is applied in practice. To begin with, three examples are chosen to show that the dynamic model can be fruitfully used to address questions that surface in applied evaluation studies and that are hard to address within a static framework, because the latter is not able to handle selection problems that occur while a particular treatment is in operation.

The first example concerns the effects of sequences of programmes. The breakdown of the static model occurs when selection into the second and any subsequent programme is influenced by the outcome of the previous programmes, thus particular control variables become endogenous in particular ways with respect to the complete sequence. This is called an intermediate outcome for which the static model does not provide any way to handle it.

The second example concerns the effects of earlier or later starts of the programme. Although there have been attempts to estimate such effects in the evaluation literature by Sianesi (2004),<sup>1</sup> her adjusted static framework does not allow to clearly spell out the causal contrasts estimated as well as to explicitly define the exogeneity conditions required for the control variables to identify the underlying causal effect. The latter deficiency is also shared by papers that try to mitigate the problem of different starting dates in evaluation studies by randomly drawing start dates (see Lechner, 1999, Gerfin and Lechner, 2002, and the critique of this procedure by Fredriksson and Johansson, 2003).

---

<sup>1</sup> See also the related approach by Li, Rosenbaum, and Propert (2001) that appeared in the statistics literature.

The third example is the effect of the actual duration of the programme. The problem with the actual duration of a programme is that it could be endogenous in the following sense: For example, if the effect of the programme comes from the signal that participation sends, then it is very likely that people leave the programme while it is under way. This attrition is, however, an effect of beginning and staying up to that point in the programme. So far, empirical evaluation studies circumvented this problem by considering the effects of planned duration only (e.g. Lechner, Miquel, and Wunsch, 2004), which estimates a different parameter that may or may not be of interest in the particular situation.

The conditional independence assumption that justifies matching estimation in the static context is sometimes called a data hungry identification and estimation strategy. To remain in this picture, if static matching is data hungry, then dynamic matching is starving for data. This relates to the number of observations necessary in the particular sequences to obtain precise inference, to the time-varying variables required to obtain identification, and to the heterogeneity of the characteristics observed in the particular treatments. Taken together, this could lead to the undesirable situation that the price to pay for using the much more informative dynamic approach is that the resulting parameters are very noisy estimates on a common support that has no policy relevance. Therefore, these issues are considered in more depth from different perspectives: (i) a comparison to static matching;<sup>2</sup> (ii) the relation of the common support and the length of specified sequences; (iii) the number of regressors included in the propensity score estimation.

This paper does not derive new analytical results. The discussion is based on known properties of the estimators, as well as on the performance of the estimation procedures in the data. This data comes from a rather elaborate attempt to generate artificial data that might have found in true evaluation studies of European type active labour market programmes (many covariates, 40 periods with autocorrelation, 4 programmes with different start dates and lock-in effects, etc.). Since in terms of

---

<sup>2</sup> There is a bias issue in the comparison to static matching as well. However, Lechner (2004) discusses this topic extensively. Therefore, this paper does not take it up again.

computing time, its generation is far too expensive for a Monte Carlo study, only one replication is used. For the limited illustrative goals of this paper, this suffices.

The paper proceeds as follows: Section 2 outlines the dynamic causal framework suggested by LM01 and L04. The notation is introduced and the basic identification conditions are restated. The estimation problem is explained in Section 3 and sequential matching as proposed by L04 is reviewed. Section 4 details the data. Section 5 presents the three empirical examples. Section 6 covers the brief comparison to static matching. The issues of additional variables and the relation between length of sequence and common support are discussed in Section 7. Section 8 concludes and Appendix A contains some descriptive statistics concerning the distribution of the true values of the potential outcomes.

## 2 The dynamic causal model - notation, effects, and identification

This section briefly repeats the definition of the dynamic causal model as well as the identification results derived by Lechner and Miquel (2001) for the case of sequential selection on observables. To ease the notational burden, I use a three-period-two-treatments model to discuss the most relevant issues that distinguish the dynamic from the static model, although in the application more periods and more treatments are considered. As usual in the econometric evaluation literature, I use the standard statistics terminology based on treatments and potential outcomes to define causal effects.

### 2.1 Basic structure of the model

Suppose that there is an initial period in which everybody is in the same treatment, plus two subsequent periods in which different treatment states are realised. The periods are indexed by  $t$  or  $\tau$  ( $t, \tau \in \{0, 1, 2\}$ ). The treatment defined over all periods is described by a vector of Bernoulli random variables (RV),  $S = (S_1, S_2)$ . For notational convenience, the RV of the initial period ( $S_0$ ) is not mentioned explicitly. A particular realisation of  $S_t$  is denoted by  $s_t \in \{0, 1\}$ . Denote the history of



variables up to period  $t$  by a bar below a variable, i.e.  $\underline{S}_2 = (S_1, S_2)$ .<sup>3</sup> Since we are not restricting effect heterogeneity over time, it makes sense to define potential outcomes in terms of sequences of potential states of the world. Thus, in period one, an individual (or a firm, country, or any other unit of interest) is observed in exactly one of two treatments. In period two, the same treatment in period 2 will be captured by two potential outcomes depending on what happened in period 1. Therefore, she participates in one of four treatments, defined by the sequences  $(0,0), (1,0), (0,1), (1,1)$ . Thus, every individual participates in exactly one sequence defined by  $S_1$  and another sequence defined by the same value  $S_1$  and a value of  $S_2$ . To sum up, in the two (plus one)-period-two-treatments example we consider six different overlapping potential outcomes corresponding to two mutually exclusive states defined by treatment status in period 1 only, plus four mutually exclusive states defined by treatment status in period 1 and 2 together.

Variables used to measure the effects of the treatment in period  $t$ , i.e. the potential outcomes, are indexed by treatments and denoted by  $Y_t^{\underline{S}_1}$  ( $t \geq 1$ ) or  $Y_t^{\underline{S}_2}$  ( $t \geq 2$ ). They are measured at the end of each period, whereas treatment status is measured in the beginning of each period. For each length of a sequence (1 or 2 periods), one of the potential outcomes is observable and denoted by  $Y_t$ . The resulting two observation rules are defined in equation (1):

$$Y_t = S_1 Y_t^1 + (1 - S_1) Y_t^0 = S_1 S_2 Y_t^{11} + S_1 (1 - S_2) Y_t^{10} + (1 - S_1) S_2 Y_t^{01} + (1 - S_1) (1 - S_2) Y_t^{00}; \quad t = 0, 1, 2. \quad (1)$$

Finally, variables that may influence treatment selection and (or) potential outcomes are denoted by  $X$ .

The  $K$ -dimensional vector  $X_t$  may contain functions of  $Y_t$  and is observable at the same time as  $Y_t$ .

---

<sup>3</sup> To differentiate between different sequences, sometimes a letter (e.g.  $j$ ) is used to index a sequence like  $\underline{S}_t^j$ . As a further convention, capital letters usually denote random variables, whereas small letters denote specific values of the random variable. When we deviate from this convention, it will be obvious.

## 2.2 Average causal effects

As in the static model, the potential outcomes are used to define several average causal effects (for period  $t$ ) of a sequence of treatments defined up to period 1 or 2 ( $\tau, \tau'$ ) compared to an alternative sequence of the same or a different length for a population defined by one of those sequences or a third sequence is defined in equation (2):

$$\theta_t^{s_\tau^k, s_{\tau'}^l}(s_\tau^j) = E(Y_t^{s_\tau^k} | \underline{s}_\tau = s_\tau^j) - E(Y_t^{s_{\tau'}^l} | \underline{s}_{\tau'} = s_{\tau'}^j),$$

$$0 \leq \tilde{\tau}; \quad 1 \leq \tau, \tau' \leq 2, \quad \tilde{\tau} \leq \tau', \tau; \quad k \neq l, \quad k, l \in (1, \dots, 2^\tau), \quad j \in (1, \dots, 2^{\tilde{\tau}}). \quad (2)$$

LM01 call  $\theta_t^{s_\tau^k, s_{\tau'}^l}$  the dynamic average treatment effect (DATE). Accordingly,  $\theta_t^{s_\tau^k, s_{\tau'}^l}(s_\tau^k)$ , as well as  $\theta_t^{s_\tau^k, s_{\tau'}^l}(s_{\tau'}^l)$  are termed DATE on the treated (DATET) and DATE on the nontreated. There are cases in-between, like  $\theta_t^{s_2^k, s_2^l}(s_1^l)$  for which the conditioning set is defined by a sequence shorter than the ones defining the causal contrast. Note that the effects are symmetric for the same population ( $\theta_t^{s_\tau^k, s_{\tau'}^l}(s_\tau^k) = -\theta_t^{s_{\tau'}^l, s_\tau^k}(s_{\tau'}^l)$ ), which, however, does not restrict effect heterogeneity across individuals ( $\theta_t^{s_\tau^k, s_{\tau'}^l}(s_\tau^k) \neq \theta_t^{s_{\tau'}^k, s_\tau^l}(s_{\tau'}^l)$ ).

## 2.3 Identification

Assume that a large sample  $\{s_{1i}, s_{2i}, x_{0i}, x_{1i}, x_{2i}, y_{1i}, y_{2i}\}_{i=1:N}$  of size  $N$  is available, randomly drawn from a very large population of interest. The latter is characterised by the corresponding random variables  $(S_1, S_2, X_0, X_1, X_2, Y_1, Y_2)$ .<sup>4</sup> Furthermore, assume that all conditional expectations that are of interest in

<sup>4</sup> To simplify the notation further, we consider period 2 as the only period relevant for the outcome of interest. However, for all what follows  $X_2$  and  $Y_2$  should be considered as measured at some point in time after treatment

the remainder of this paper exist. LM01 show that if we can observe the variables that jointly influence selection at each stage as well as the outcomes some average treatment effects are identified (weak conditional independence assumptions):<sup>5</sup>

Weak dynamic conditional independence assumption (W-DCIA) <sup>6</sup>

- a)  $Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \perp\!\!\!\perp S_1 \mid X_0 = x_0;$
- b)  $Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \perp\!\!\!\perp S_2 \mid \underline{X}_1 = \underline{x}_1, S_1 = s_1;$
- c)  $1 > P(S_1 = 1 \mid X_0 = x_0) > 0, 1 > P(S_2 = 1 \mid \underline{X}_1 = \underline{x}_1, S_1 = s_1) > 0, \forall \underline{x}_1 \in \underline{\mathcal{X}}_1, \forall s_1 : s_1 \in \{0, 1\}.$

Part a) of W-DCIA states that potential outcomes are independent of treatment choice in period 1 ( $S_1$ ) conditional on  $X_0$ . This is the standard version of the static CIA (e.g. Rubin, 1974). Part b) states that conditional on the treatment in period 1, on observable outcomes of period 1 (which may be part of  $X_1$ ) and on confounding variables of period 0 and 1 ( $\underline{X}_1$ ), potential outcomes are independent of participation in period 2 ( $S_2$ ).

To see whether such an assumption is plausible in a particular application, we have to ask which variables influence potential changes in treatment status as well as outcomes and whether they are observable. If the answer to the latter question is yes, and if there is common support (defined in part c) of W-DCIA), then we have identification, even if some or all conditioning variables in period 2 are influenced by the outcome of the treatment in period 1. LM01 show that, for example, quantities like  $E(Y_2^{11})$ ,

---

2 occurred. The exact timing is determined by the substantive interest of the researcher conducting the empirical study.

<sup>5</sup> The following assumptions relate to identification of all treatment effects that could possibly be defined by the notation in Section 2. If the desired comparison involves fewer periods, the required changes are obvious.

<sup>6</sup>  $A \perp\!\!\!\perp B \mid C = c$  means that *each element* of the vector of random variables  $B$  is independent of the random variable  $A$  conditional on the random variable  $C$  taking a value of  $c$  in the sense of Dawid (1979).

$E(Y_2^{11} | S_1 = 0)$ ,  $E(Y_2^{11} | S_1 = 1)$ , or  $E[Y_2^{11} | \underline{S}_2 = (1, 0)]$  are identified, but that  $E[Y_2^{11} | \underline{S}_2 = (0, 0)]$  or  $E[Y_2^{11} | \underline{S}_2 = (0, 1)]$  are not identified. Thus,  $\theta_2^{\underline{s}_2^k, \underline{s}_2^l}$ ,  $\theta_2^{\underline{s}_2^k, \underline{s}_2^l}(s_1^j)$  are identified,  $\forall s_1^k, s_2^k, s_1^l, s_2^l, s_1^j, s_2^j \in \{0, 1\}$ , but  $\theta_2^{\underline{s}_2^k, \underline{s}_2^l}(s_2^j)$  are not identified if  $s_1^l \neq s_1^k$ . This result states that pairwise comparisons of all sequences are identified, but only for groups of individuals defined by their treatment status in period 0 or 1. The relevant distinction between the populations defined by treatment states in the first and subsequent periods is that in the first period, treatment choice is random conditional on exogenous variables, which is the result of the initial condition stating that  $S_0 = 0$  holds for everybody. However, in the second period, randomisation into these treatments is conditional on variables already influenced by the first part of the treatment. W-DCIA has an appeal for applied work as a natural extension of the static framework.

LM01 show that to identify all treatment parameters, W-DCIA must be strengthened by essentially imposing that the confounding variables used to control selection into the treatment of the second period are not influenced by the selection into the first-period treatment. This can be summarised by an independence condition like  $Y_2^{\underline{s}_2} \perp\!\!\!\perp \underline{S}_2 | \underline{X}_1$  (LM01 call this the strong conditional dynamic independence assumption, S-DCIA). Note that the conditioning set includes the outcome variables from the first period. This is the usual conditional independence assumption used in the multiple treatment framework (with four treatments). In other words, when the control variables are not influenced by the previous treatments, the dynamic problem collapses to a static problem of four treatments with selection on observables.

Finally, in the applied static matching literature balancing scores are a popular device to reduce the dimension of the estimation problem (see Rosenbaum and Rubin, 1983). LM01 show that similar properties hold for the dynamic model as well.

Balancing score property for W-DCIA

If the conditions of W-DCIA hold, then:

a)  $Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \perp\!\!\!\perp S_1 \mid b_1(X_0) = b_1(x_0)$  holds for all  $b_1(x_0)$  such that

$$E[p^{s_1}(x_0) \mid b_1(X_0) = b_1(x_0)] = p^{s_1}(x_0), \quad p^{s_1}(x_0) := P(S_1 = 1 \mid X_0 = x_0).$$

b)  $Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \perp\!\!\!\perp S_2 \mid b_2(\underline{X}_1, S_1) = b_2(\underline{x}_1, s_1)$  holds for all  $b_2(\underline{x}_1, s_1)$  such that

$$E[p^{s_2|s_1}(\underline{x}_1) \mid b_2(\underline{X}_1, S_1) = b_2(\underline{x}_1, s_1)] = p^{s_2|s_1}(\underline{x}_1), \quad p^{s_2|s_1}(\underline{x}_1) := P(S_2 = 1 \mid \underline{X}_1 = \underline{x}_1, S_1 = s_1).$$

A low-dimensional choice for balancing scores suggested by LM01 are the conditional transition probabilities in combination with the variable indicating the selection in the previous period (which of course can be ignored in the first period):  $b_1(x_0) = p^{s_1}(x_0)$ ,  $b_2(\underline{x}_1, s_1) = [p^{s_2|s_1}(\underline{x}_1), s_1]$ .

### 3 Estimation

#### 3.1 Structure of sequential estimators

Lechner (2004) shows that these scores are convenient for constructing sequential propensity score matching estimators to correct for selection bias under W-DCIA. The focus on this particular estimator is because of its simplicity and because it is the workhorse of empirical evaluation studies. Other static matching-type estimators can be adapted to the dynamic context in a similar way (see Imbens, 2004, for an overview of available estimators). I refrain from discussing estimation based on the S-DCIA explicitly, because estimation under S-DCIA is essentially a static problem, with an increased number of treatments. Such estimation problems for multiple treatments are discussed by Imbens (2000) and Lechner (2001, 2002) and need not be explained here. Nevertheless, the suggested estimators are consistent under S-DCIA as well. Thus, a comparison of estimators that are consistent under both DCIA (sequential matching) and those that are consistent under S-DCIA (static matching) only could be a basis

for checking the plausibility of S-DCIA (or the endogeneity of some covariates). However, this issues is not developed further in this paper.

Using the balancing scores suggested above the following estimand (quantity to be estimated by sample analogues of observables) results for quantities identified under W-DCIA:

$$E(Y_2^{s_2^k} | S_1 = s_1^j) = E_{p^{s_1^k}(X_0)} \left\{ E_{p^{s_2^k|s_1^k}(X_1)} [E(Y_2 | \underline{S}_2 = \underline{s}_2^k, \underline{p}^{s_2^k|s_1^k}(X_1)) | S_1 = s_1^k, p^{s_1^k}(X_0)] | S_1 = s_1^j \right\},$$

$$\underline{p}^{s_2^k|s_1^k}(X_1) := [p^{s_2^k|s_1^k}(X_1), p^{s_1^k}(X_0)], \quad s_1^k, s_2^k, s_1^j, s_1 \in \{0, 1\}.$$

(3)

To learn the counterfactual outcome for the population participating in  $s_1^j$  (the target population) had they participated in sequence  $\underline{s}_2^k$ , we need to reweight the participants in  $\underline{s}_2^k$  to make them comparable to the target population ( $s_1^j$ ). The dynamic, sequential structure of the causal model restricts the possible ways to do so. Intuitively, for the participants in the target population, we should reweight participants in the first element of the sequence of interest ( $s_1^k$ ) such that they have the same distribution of  $p^{s_1^k}(X_0)$  as the target distribution. Call this artificially created group comparison group 1. Yet, to estimate the effect of the full sequence, the outcomes of participants in  $\underline{s}_2^k$  instead of  $s_1^k$  are required. Thus, an artificial subpopulation of participants in  $\underline{s}_2^k$  that has the same distribution of characteristics of  $p^{s_1^k}(X_0)$  and  $p^{s_2^k|s_1^k}(X_1)$  as the artificially created comparison group 1 is required. The same principle applies for average dynamic treatment effects in the population (DATE).

All proposed estimators in L04 have the same structure, in the sense that they are computed as weighted means of the outcome variables observed in subsample  $\underline{S}_2 = \underline{S}_2^k$ . The weights depend on the specific effects of interest and are functions of the balancing scores.

$$\widehat{E(Y_2^{\underline{S}_2^k} | S_1 = s_1^j)} = \sum_{i \in \underline{S}_2^k} w_i^{s_2^k, s_1^j} [p^{s_2^k | s_1^k, s_1^j}(\underline{x}_{1,i}), s_1^k] y_i; \quad w_i^{s_2^k, s_1^j} \geq 0; \quad \sum_{i \in \underline{S}_2^k} w_i^{s_2^k, s_1^j} = 1; \quad (4)$$

$$\widehat{E(Y_2^{\underline{S}_2^k})} = \sum_{i \in \underline{S}_2^k} w_i^{s_2^k} [p^{s_2^k | s_1^k}(\underline{x}_{1,i}), s_1^k] y_i; \quad w_i^{s_2^k} \geq 0; \quad \sum_{i \in \underline{S}_2^k} w_i^{s_2^k} = 1 \quad (5)$$

Note that in the case of more than two treatments, the balancing scores for (4) and (5) will differ with respect to the probability of participating in the first period. For equation (4), the required quantity is

$P(S_1 = s_1^k | X_0 = x_0, S_1 \in \{s_1^k, s_1^l\})$ , whereas in equation (5), in which all of the population is the target,

$P(S_1 = s_1^k | X_0 = x_0)$  is appropriate.

### 3.2 Sequential matching estimators (SM)

LM01 proposes to extend the simple pair-matching estimators that are highly popular in applied evaluation studies to the dynamic context. The idea is to perform the required adjustments by sequentially choosing close pairs of observations in the various steps, so as to mimic the sequential conditional expectations appearing in expressions (4) and (5). The first step is the same for both effects and consists in finding for every member of  $S_1 = s_1^k$  a member of  $\underline{S}_2 = \underline{S}_2^k$  with very similar (the same) values of  $p^{s_2^k | s_1^k}(\underline{x}_{1,i})$  and  $p^{s_1^k}(x_{0,i})$ . Note that matching must be with replacement, because the target population may be larger than the treatment population. In the second step, every member of  $S_1 = s_1^j$  (equation (4)) or  $S_0 = 0$  (equation (5)) is to be paired with a member of  $S_1 = s_1^k$  with very similar (same) values of  $p^{s_1^k}(x_{0,i})$ . The positive weights that are attached to some or all members of  $\underline{S}_2 = \underline{S}_2^k$

coming from step 1 are then updated depending on how often an observation in  $\underline{S}_2 = \underline{S}_2^k$  is matched to an observation of the target population via the intermediate matching step. This procedure leads to the following weights:

$$w_i^{\underline{S}_2^k, \underline{S}_1^j} = \frac{1}{N^{\underline{S}_1^j}} \sum_{n \in \underline{S}_1^j} \sum_{m \in \underline{S}_1^k} v_1[p^{\underline{S}_1^k}(x_{0,n}), p^{\underline{S}_1^k}(x_{0,m}); \cdot] v_2[\underline{p}^{\underline{S}_2^k | \underline{S}_1^k}(\underline{x}_{1,m}), \underline{p}^{\underline{S}_2^k | \underline{S}_1^k}(\underline{x}_{1,i}), s_1^k; \cdot] ; \forall i \in \underline{S}_2 = \underline{S}_2^k ; \quad (6)$$

$$w_i^{\underline{S}_2^k} = \frac{1}{N} \sum_{n=1}^N \sum_{m \in \underline{S}_1^k} v_1[p^{\underline{S}_1^k}(x_{0,n}), p^{\underline{S}_1^k}(x_{0,m}); \cdot] v_2[\underline{p}^{\underline{S}_2^k | \underline{S}_1^k}(\underline{x}_{1,m}), \underline{p}^{\underline{S}_2^k | \underline{S}_1^k}(\underline{x}_{1,i}), s_1^k; \cdot] ; \quad \forall i \in \underline{S}_2 = \underline{S}_2^k . \quad (7)$$

$N^{\underline{S}_1^j}$  denotes the number of observations for which  $\underline{S}_1 = \underline{S}_1^j$ . The function  $v_1[p^{\underline{S}_1^k}(x_{0,n}), p^{\underline{S}_1^k}(x_{0,m}); \cdot]$  is defined to be one if  $p^{\underline{S}_1^k}(x_{0,m})$  is closest to  $p^{\underline{S}_1^k}(x_{0,n})$  among all observations belonging to the subsample defined by  $\underline{S}_1 = \underline{S}_1^k$ , and zero otherwise. Similarly,  $v_2[\underline{p}^{\underline{S}_2^k | \underline{S}_1^k}(\underline{x}_{1,m}), \underline{p}^{\underline{S}_2^k | \underline{S}_1^k}(\underline{x}_{1,i}), s_1^k; \cdot]$  is one if observation  $i$  is closest to observation  $m$  (with  $\underline{S}_{1,m} = \underline{S}_1^k$ ) in terms of  $\underline{p}^{\underline{S}_2^k | \underline{S}_1^k}(\underline{x}_{1,i})$  and  $\underline{p}^{\underline{S}_2^k | \underline{S}_1^k}(x_{0,i})$ , and zero otherwise. The Mahalanobis metric is a frequently used measure for similarity. Note that the weight of observation  $i$  is 0, if it is not matched to any member of the target population. On the other extreme, if observation  $i$  would be matched to every member of the target population its weight would be 1. A specific variant of this estimator is shown in Table 1 for the example of estimating  $\theta_t^{\underline{S}_2^1, \underline{S}_2^0}(s_1^1)$ .

Some remarks about this protocol that are already contained in L04 are worth repeating: First, matching is with replacement. Every step of the matching sequence is essentially the same as for matching in a static framework. However, sequential propensity score matching involves several probabilities in the second period matching step. Second, some issues arise from the sequential nature of matching. By choosing observations as matches with similar values of the probabilities instead of the same values (because such observations may not be available), it may happen that the probabilities attached to



observations in early matching steps (relating to transitions in early periods) change over different sequential matching steps due to imprecise matching. To prevent this from happening, every matched comparison observation in period 2 is recorded with the values  $\hat{p}_i^{s_1^l}$  of the observation it was matched to in period 1, instead of its own. Hence, the ‘history’ of the match, or, in other words, the characteristics of the reference distribution, does not change when the next match occurs in the subsequent period.

Third, to compute  $E(Y^{s_2^k} | S_1 = s_1^l)$  the only information that is needed for the  $N^{s_1^l}$  participants in  $s_1^l$  is  $\hat{p}_i^{s_1^k}$ . Similarly, for participants in  $s_2^k$ , all probabilities of the type  $\hat{p}_i^{s_2^k | s_1^k, s_1^k}$  are required. For participants in  $s_1^k$  but not in  $s_2^k$  only  $\hat{p}_i^{s_1^k}$  is needed, and so on. To estimate  $E(Y^{s_2^l} | S_1 = s_1^l)$  instead of  $E(Y^{s_2^k} | S_1 = s_1^l)$ , the only change in the matching protocol is that the initial matching step on  $\hat{p}_i^{s_1^l}$  is redundant. When interest is in the average effect in the population ( $E(Y^{s_2^k})$ ), then the whole population plays the role of the first reference group (instead of  $s_1^l$ ). In this case, in the matching step based on  $\hat{p}_i^{s_1^k}$ , all participants in  $s_1^k$  are matched to themselves, as well as selected participants in  $s_1^k$  are matched to participants in the remaining treatments in the first period.

Table 1: A sequential matching estimator for  $\theta_{t^{s_2^1, s_2^0}}(s_1^1)$  based on propensity scores

<b>Step 0: Sample reduction</b>		Delete all observations not belonging to $s_1^1$ , $s_2^1$ , or $s_2^0$
<b>Step A: Match <math>s_2^0 = (s_1^0, s_2^0)</math> to <math>s_1^1</math></b> $(E(Y_{t^{s_2^0}}   S_1 = s_1^1))$	A.1.0	Define a weight $w_i^{s_2^0} = 0$ for every observation in $s_2^0$ .
	A.1.P	Estimate a probit for $P(S_1 = s_1^0   X_0 = \underline{x}_0) \rightarrow p^{s_1^0}(x_{0,i}) = p_i^{s_1^0}$
	A.1.CS	Delete all obs. of $s_1^1$ with lower or higher values of $p_i^{s_1^0}$ than obs. in $s_2^0$
	A.1.M	For every obs. in $s_1^1$ not deleted in A.1.CS find the obs. in $s_1^0$ that is closest in terms of $p_i^{s_1^0}$ (a match).
	A.1.C	For the matched obs. keep the value of $p_i^{s_1^0}$ of the obs. in $s_1^1$ they have been matched to. Some obs. in $s_1^0$ may appear many times in this matched sample.
	A.2.R	Define a sample of obs. in $s_1^0$ .
	A.2.P	Estimate a probit for $P(S_2 = s_2^0   S_1 = s_1^0, X_1 = \underline{x}_1) \rightarrow p^{s_2^0 s_1^0}(\underline{x}_{1,i}) = p_i^{s_2^0 s_1^0}$
	A.2.CS	Delete all obs. of the matched comparison sample of $s_1^1$ (defined in A.1.C) (as well as the corresponding elements of the target population $s_1^1$ ) with lower or higher values of $p_i^{s_1^0}$ and $p_i^{s_2^0 s_1^0}$ than obs. in $s_2^0$ .
	A.2.M	For every obs. in the matched comparison sample of $s_1^1$ not deleted in A.2.CS find an obs. in $s_2^0$ that is closest in terms of $p_i^{s_2^0 s_1^0}$ and $p_i^{s_1^0}$ using the Mahalanobis metric (covariance computed in $s_1^1$ ).
	Every time an obs. in $s_2^0$ is matched, its weight $w_i^{s_2^0}$ is increased by 1.	
	<b>Step B: Match <math>s_2^1 = (s_1^1, s_2^1)</math> to <math>s_1^1</math></b> $(E(Y_{t^{s_2^1}}   S_1 = s_1^1))$	B.1.0 Define a weight $w_i^{s_2^1} = 0$ for every obs. in $s_2^1$ .
	B.2.R	Reduce sample to participants in $s_1^1$ still in the common support.
	B.2.P	Estimate a probit for $P(S_2 = s_2^1   S_1 = s_1^1, X_1 = \underline{x}_1) \rightarrow p^{s_2^1 s_1^1}(\underline{x}_{1,i}) = p_i^{s_2^1 s_1^1}$
	B.2.CS	Delete all obs. of $s_1^1$ with lower or higher values of $p_i^{s_2^1 s_1^1}$ than obs. in $s_2^1$ .
	B.2.M	For every obs. in $s_1^1$ not deleted in B.2.CS find the member of $s_2^1$ that is closest in terms of $p_i^{s_2^1 s_1^1}$ .
	Every time an obs. in $s_2^1$ is matched, its weight $w_i^{s_2^1}$ is increased by 1.	
<b>Step C: Joint common support</b>		C.1 Reduce $w_i^{s_2^1}$ by 1 for every obs. i matched to an obs. in $s_1^1$ deleted in A.1.CS or A.2.CS.
		C.2 Reduce $w_i^{s_2^0}$ by 1 for every obs. i matched to an obs. in $s_1^1$ deleted in B.2.CS.
<b>Step D: Estimation of <math>\theta_{t^{s_2^1, s_2^0}}(s_1^1)</math></b>		D.1 $\hat{\theta}_{t^{s_2^1, s_2^0}}(s_1^1) = \frac{1}{\sum_{i \in s_2^1} w_i^{s_2^1}} \sum_{i \in s_2^1} w_i^{s_2^1} y_i - \frac{1}{\sum_{i \in s_2^0} w_i^{s_2^0}} \sum_{i \in s_2^0} w_i^{s_2^0} y_i$

Table 1 to be continued.

Table 1: Continued

<b>Step E:</b>	D.2
Estimation of $Var[\hat{\theta}_{t_i^1, s_2^0}(s_1^1)]$	$\widehat{Var}[\hat{\theta}_{t_i^1, s_2^0}(s_1^1)] = \frac{\sum_{i \in \mathcal{S}_2^1} [(w_i^{s_2^1})^2 \widehat{Var}(Y_i   S = s_2^1, w = w_i^{s_2^1})]}{\sum_{i \in \mathcal{S}_2^1} (w_i^{s_2^1})^2} + \frac{\sum_{i \in \mathcal{S}_2^0} [(w_i^{s_2^0})^2 \widehat{Var}(Y_i   S = s_2^0, w = w_i^{s_2^0})]}{\sum_{i \in \mathcal{S}_2^0} (w_i^{s_2^0})^2}$ $\widehat{Var}(Y_i   S = s_2^j, w = w_i^{s_2^j}) = \frac{1}{N^{I(s_2^j, w_i^{s_2^j})}} \left[ \sum_{i \in I(s_2^j, w_i^{s_2^j})} y_i - \bar{y}^{I(s_2^j, w_i^{s_2^j})} \right]^2,$ $N^{I(s_2^j, w_i^{s_2^j})} = \sum_{i \in I(s_2^j, w_i^{s_2^j})} 1, \quad \bar{y}^{I(s_2^j, w_i^{s_2^j})} = \frac{1}{N^{I(s_2^j, w_i^{s_2^j})}} \sum_{i \in I(s_2^j, w_i^{s_2^j})} y_i.$ <p>The set <math>I(s_2^j, w_i^{s_2^j})</math> is determined as the <math>2\sqrt{N}</math> closest neighbours w.r.t. the value of <math>w_i^{s_2^j}</math> of observations with same treatment as observation 'i'</p>
Note:	$t > 1$ . Changes required for the case $t=1$ are obvious. The number of neighbours in the k-NN estimation of the conditional variance is the one used in L04. In the empirical application below, the results appear not to be sensitive to small deviations from this value.

When matching is on the propensity score instead of directly on the confounding variables, there is the issue of selecting a probability model. It seems that so far even in the static model the literature has not addressed this thoroughly. However, the consensus seems to be that a flexibly specified (and extensively tested) parametric model is sufficiently rich and that the choice of the model does not really matter (e.g. the Monte Carlo results by Zhao, 2004).

Next, there is the issue of consistent estimation of the standard errors that is not yet resolved for the static matching literature. Based on the simulation results presented in L04, the standard errors are computed conditional on the weights allowing for heteroskedasticity of the distribution of the outcomes conditional on the weights.

The final remark about the matching protocol concerns the common support. The region of common support - defined on the reference distribution for which the effect is desired - has to be adjusted period by period with respect to the conditioning variables of this period. The matching estimator allows easily tracing back the impact of this procedure on the reference distribution.<sup>7</sup>

<sup>7</sup> In the application the support is defined slightly more conservatively than given in Table 1. It is the region between the *second* largest and second smallest value of the respective propensity scores in the treated sample.

### 3.3 Multiple treatments and many periods

The main issue concerns the specification of the propensity scores: For example, when specifying the probability of participating in  $S_2^k$  conditional on participating in  $S_1^k$ , is it necessary to take account of the fact that not participating in  $S_2^k$  implies a range of possible other states in period 2? The answer is no, because in each step the independence assumption relates only to a binary comparison, e.g.  $Y_2^k \perp\!\!\!\perp \mathbb{1}(S_2 = s_2^k) | S_1 = s_1^k, X_1 = x_1$ , and  $Y_2^k \perp\!\!\!\perp \mathbb{1}(S_1 = s_1^k) | S_1 \in \{s_1^j, s_1^k\}, X_0 = x_0$  ( $s_1^j$  being the target population as before). Therefore, the conditional probabilities of not participating in the event of interest conditional on the history are sufficient.<sup>8</sup> Hence, as already noted  $P(S_2 = s_2^k | S_1 = s_1^k, X_1 = x_1)$ ,  $P[S_1 = s_1^k | X_0 = x_0, S_1 \in \{s_1^l, s_1^k\}]$  and  $P[S_1 = s_1^k | X_0 = x_0, S_1 \in \{s_1^l, s_1^k\}]$  may be used in the matching step in period 1. The multiple treatment feature of the problem does not add to the dimension of the propensity scores.

## 4 Data

The artificial data are generated to look similar to individual panel data found in true evaluation studies of European type active labour market programmes (like in Gerfin and Lechner, 2002, or Lechner, Miquel, Wunsch, 2004), although the exact properties of those data sets used for static evaluation studies are not exactly reflected in the artificial data. In this data, a period is rationalised as a quarter. In that sense, there is detailed information about many employment related variables on a quarterly basis for 9 years (quarter 1(1) to quarter 9(4)). In addition, there are summary measures supposed to capture the events before the data is recorded in quarterly intervals (the quarterly records start in the first quarter of year 1).

---

<sup>8</sup> Imbens (2000) and Lechner (2001) develop the same argument to show that in the static multiple treatment models conditioning on appropriate one-dimensional scores is sufficient.

The sample is selected to contain unemployed in the last quarter of year 2. Starting in the first quarter in year 3, individuals may participate in active labour market programmes. If not having done so before, they may start a programme every quarter up to and including 4(2). In addition, in 4(2) they may start a new programme, even if they already participated in a programme completed before 4(2).

The programmes consist of two employment programmes (treatments '3' and '4') and two training programmes (treatments '1' and '2'). The main difference between those types is that employment programmes have a smaller lock-in effect and that the positive medium-run effects all programmes have, deteriorate at a faster rate than for training programmes.

We consider (potential) outcome processes for all sequences of different programmes that relate to employment status (employed, unemployed, out-of-the-labour force) and earnings. All processes show considerable state dependence, time trends and are influenced by several covariates. Shorter programmes with about a mean of 5 months for training and 6 months (standard deviation 2 and 4 months) for the employment programmes have much shorter lock-in effects than the longer programmes of the same category (mean duration 20, 18; std.: 2, 4), but the (positive) effects also depreciate faster. However, for the sake of brevity the latter two are not contained in the descriptive statistics given in Table 2.

Table 2 contains both descriptive statistics as well as a characterisation of the type of variable for the most important time-dependent and time-independent variables. They are the usual types of variables with typical codings, means, and standard errors. A full set of statistics for all variables are available on request from the author.

Table 2: Descriptive statistics for selected variables and selected subsamples

Variable	Type	Subsample											
		0		000000		1		1111		2		222222	
		mean	std.	mean	std.	mean	std.	mean	std.	mean	std.	mean	std.
Monthly earnings in EUR													
Last 10 years	C	1349	1789	1030	1089	4673	4071	4851	4074	5449	3883	5432	3895
1(1)	C	1095	1732	786	1028	4323	4137	4386	4156	5009	4040	4953	4067
2(4)	C	0	0	0	0	0	0	0	0	0	0	0	0
4(1)	C	358	1034	525	832	1337	3107	722	2246	1278	3093	444	1943
9(4)	C	1241	1933	768	1038	3964	4270	4040	4345	5229	4533	5091	4512
Unemployment in %													
Months last 10 y.	D	12	7	11	7	11	7	12	7	12	7	13	7
1(1)	I	10	-	9	-	6	-	7	-	5	-	6	-
2(4)	I	100	-	100	-	100	-	100	-	100	-	100	-
4(1)	I	54	-	23	-	60	-	65	-	60	-	78	-
9(4)	I	10	-	21	-	5	-	5	-	7	-	9	-
Employment in %													
Months last 10 y.	D	85	20	86	22	89	19	87	20	89	18	88	18
1(1)	I	78	-	76	-	90	-	87	-	90	-	89	-
2(4)	I	0	-	0	-	0	-	0	-	0	-	0	-
4(1)	I	31	-	49	-	32	-	27	-	35	-	17	-
9(4)	I	68	-	58	-	86	-	87	-	83	-	80	-
Labour market prospects as assessed by case worker (1, 2, 3, 4)													
2(4)	D	2.8	1.1	3.1	1.0	2.7	1.1	2.6	1.1	2.4	1.2	2.5	1.2
3(1)	D	2.3	1.1	2.7	1.1	2.0	1.1	1.9	1.1	1.9	1.0	1.9	1.0
4(1)	D	2.6	1.1	3.4	.8	2.5	1.1	2.3	1.1	2.4	1.2	2.1	1.1
Regional unemployment rate in %-points (85 regions)													
2(4)	D	12	5	12	5	12	5	12	5	12	5	12	5
3(1)	D	13	5	13	5	13	5	13	5	13	5	13	5
4(1)	D	15	27	14	5	14	5	15	5	14	5	14	5
Other variables													
Age in 1(1), years	D	40	6	39	6	41	6	41	6	40	6	40	6
Women	I	41	-	45	-	36	-	39	-	27	-	28	-
Schooling (8-12)	D	10	1.3	10	1.2	11	1.3	11	1.3	12	.9	12	.9
Vocational degree (0,1, 2)	D	.8	.5	.8	.5	1.3	.7	1.4	.7	1.6	.6	1.6	.6
Nationality (1 -5)	D	1.6	1.1	1.6	1.1	1.6	1.1	1.6	1.1	1.6	1.1	1.6	1.1
Regional share of service sector	C	58	13	58	14	60	13	59	13	63	12	62	12
Regional share of production sector	C	30	11	29	11	29	11	30	11	28	10	28	10
Sectoral UE rate	C	12	4	12	18	12	19	12	4	13	4	13	4
Profess. UE rate	C	12	21	11	5	11	21	12	5	13	4	13	4
Observations		69951	-	16871	-	8997	-	888	-	8665	-	3404	-

Note: I: Binary indicator variable (0, 1); D: Discrete variable; C: Continuous variable. For indicator variables the share of ones in % is given. The number of observations in the complete sample is 100.000. For treatment 1 we show a subsample based on a sequence of 4 periods only instead of a sequences of 6 periods as for treatments 0 and 2, because for this short programme there would be only 34 observations in the latter group. Descriptive statistics for the variables and subsamples not included are available on request from the author.

Selection is based on an index model (multinomial probit) with five alternatives. Choices depend on observables that also appear in the outcome processes as well as normally distributed unobservables that are mutually dependent but independent of the observables and unobservables appearing in the outcome

equations. With the exception of nationality, all variables appearing in Table 2 influence selection in each period, but only earnings, the employment states, and the assessment by the case worker are influenced by the treatment and are thus considered outcome variables, or intermediate outcomes for those variables that relate to periods in which the sequences are not yet completed. Comparing the values of the covariates and intermediate outcomes across sequences reveals considerable selection biases as well as considerable differences in the moments of the outcome variables.

The actual probit models underlying the results presented in the next sections are subject to some misspecification, but one that remains largely undetected by conventional specification tests. The misspecification relates to the functional form (only single models instead of the underlying multiple index models are used) as well as to omitting some covariates that are highly correlated with covariates included in the sample. In this respect as well, the artificial data seem to exhibit the same problems and questions as real data sets usually do.

The effects of the programmes are heterogeneous depending on type and duration of programme, as well as on several covariates. They show lock-in effects that depend on programme duration. The effects of the programmes all depreciate, but with different speeds. The autocorrelation in the outcome process may increase the effect of the lock-in effect and dampen the effect of the depreciation. Table A.1 in Appendix A, as well as Tables 6, 7 and 8, show mean and standard deviations for some outcomes of selected sequences and subsamples.

## **5 Interesting causal effects and estimation results**

The static model of potential outcomes may also be used to define potential states of the world for dynamic phenomena and estimate the effects by the usual econometric methods. Using a static model, each possible sequence of treatments corresponds to one 'static' treatment. The limitation of the static models relates to the way how restrictions on the joint distribution of selection variables and potential outcome variables have to be formulated that can then identify some average causal effects. By the

nature of the static model, those restrictions cannot take into account selection effects based on intermediate outcomes. For example, if a conditional independence (selection on observables) assumption is deemed plausible, then endogenous variables should not appear in the conditioning set, thus ruling out intermediate outcomes. Thus, it would be straightforward to accommodate dynamic phenomena based on the strong dynamic conditional independence assumption (S-DCIA). This is not the case for the weak dynamic conditional independence assumption (W-DCIA) that allows some specific endogeneity of the covariates.

The following sub-sections give some examples (sequences of programmes, waiting for the start of a programme, and duration of a programme) when such considerations appear to be particularly relevant, discuss potential set-ups of the estimation problems and present some results. A nice by-product of the dynamic approach compared to the static approach is that one has to be explicit about what the alternative treatment state is, i.e. do we compare two periods of treatments to one period of no treatments, or two periods of no treatment, or what else. In many empirical evaluation papers, the no-treatment state is not clearly defined. Therefore, below there are some examples for different comparisons and their interpretation.

### **5.1 *Programme sequences***

Consider an example coming from the literature evaluating training programmes in which interest is not in the effect of one particular programme, but in a sequence of programmes. However, if the first course of such a sequence is very effective, many unemployed individuals may find that their employment chances have drastically increased afterwards and may not want to attend the next course as originally intended. If interest is not in the first course, but in the sequence of courses, such behaviour creates a selection problem that cannot be addressed in the static model. For example, controlling for pre-training variables does not work for obvious reasons in the static model, because an important selection variable, i.e. the outcome of the first participation, is missing. However, controlling for variables realised after the



first training course that influence selection into the second course bears the potential problem that they may be influenced by the first part of the training sequence. Thus, they are ‘endogenous’ in the static model and thus ruled out as control variables. If it is true that selection in each period is based on what is known about the unemployed so far, and that this information is observed, then W-DCIA holds.

Table 3 shows some comparisons of the (monthly) earnings effects of different sequences of programmes. To differentiate between short-run and medium- to long-run effects, the columns of Table 3 (as well as Tables 4 and 5) provide estimates for the periods 4(4), 7(4), and 9(4) for the different sequences under investigation. Whereas in period 4(4) some of the programmes of interest may still be running, in period 9(4) the different lengths of the sequences and programmes are not important anymore, because all potential outcomes are close to their long-run trends. The rows of these tables contain further information about the sample sizes of the target populations after common support, as well as for the levels of the two counterfactual outcomes and the resulting effects. The cells that contain the latter are shaded. Estimates of the standard errors are provided below the corresponding estimates of outcomes and effects (in brackets).

Table 3: Earnings effects of programme sequences (sequential matching)

Earnings in year (quarter)												
(1)	(2)	(3)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(11)	(12)	(13)
	4(4)	7(4)	9(4)	4(4)	7(4)	9(4)	4(4)	7(4)	9(4)	4(4)	7(4)	9(4)
Sequences $\underline{s}^1 - \underline{s}^0$	111002-000000			222001 - 000000			111002-222001			444003-000000		
Population $S(N_s)$	1 (5638)			2 (5790)			1 (7098)			4 (6434)		
$E(Y_t^{\underline{s}^1}   S = s)$	1043 (210)	2597 (250)	2845 (315)	837 (83)	3746 (281)	3746 (281)	2168 (288)	4420 (343)	4950 (354)	217 (31)	897 (70)	697 (66)
$E(Y_t^{\underline{s}^0}   S = s)$	1505 (162)	2122 (182)	2167 (182)	1921 (258)	2341 (271)	2537 (292)	1327 (106)	4535 (317)	4757 (316)	393 (38)	621 (61)	626 (59)
$\theta_t^{\underline{s}^0, \underline{s}^1}(s)$	-462 (265)	678 (364)	678 (364)	-1084 (271)	<b>1404</b> (390)	<b>1311</b> (393)	<b>840</b> (307)	-115 (468)	192 (475)	<b>-175</b> (49)	<b>276</b> (92)	71 (88)
Population $S(N_s)$	0 (48050)			0 (49121)			2 (7732)			0 (16861)		
$E(Y_t^{\underline{s}^1}   S = s)$	509 (78)	1549 (87)	1482 (99)	246 (101)	1937 (185)	1971 (186)	2668 (377)	5074 (418)	5678 (415)	284 (56)	1032 (105)	950' (116)
$E(Y_t^{\underline{s}^0}   S = s)$	729 (29)	1008 (34)	1054 (34)	742 (28)	1016 (32)	1065 (34)	1608 (128)	5152 (275)	5526 (302)	601 (28)	843 (33)	875 (35)
$\theta_t^{\underline{s}^0, \underline{s}^1}(s)$	<b>-219</b> (83)	<b>540</b> (94)	<b>427</b> (105)	<b>-496</b> (104)	<b>920</b> (188)	<b>906</b> (189)	<b>1059</b> (399)	-78 (501)	152 (513)	<b>-317</b> (63)	188 (110)	75 (121)
Sequences $\underline{s}^1 - \underline{s}^0$	111000-000000			222000-000000			111000-222000			444000-000000		
Population $S(N_s)$	1 (6773)			2 (5893)			1 (7844)			4 (7056)		
$E(Y_t^{\underline{s}^1}   S = s)$	1080 (88)	2691 (152)	2324 (151)	771 (86)	3654 (320)	3033 (289)	2331 (191)	4125 (202)	3715 (209)	271 (30)	606 (42)	680 (54)
$E(Y_t^{\underline{s}^0}   S = s)$	1382 (134)	1943 (146)	1992 (152)	1869 (237)	2277 (251)	2469 (272)	1291 (198)	4064 (430)	3506 (397)	420 (39)	643 (58)	655 (57)
$\theta_t^{\underline{s}^0, \underline{s}^1}(s)$	-302 (160)	<b>748</b> (211)	332 (214)	-1097 (252)	<b>1377</b> (407)	564 (398)	<b>1039</b> (275)	60 (475)	209 (449)	<b>-149</b> (49)	-37 (72)	23 (79)
Population $S(N_s)$	0 (67827)			0 (56100)			2 (7219)			0 (31642)		
$E(Y_t^{\underline{s}^1}   S = s)$	572 (34)	1432 (51)	1125 (54)	350 (142)	1620 (285)	1497 (217)	3085 (321)	5133 (302)	4481 (320)	383 (139)	721 (156)	859 (190)
$E(Y_t^{\underline{s}^0}   S = s)$	617 (23)	857 (27)	900 (27)	693 (25)	955 (28)	1001 (29)	1525 (136)	4589 (327)	3837 (303)	564 (20)	795 (25)	840 (26)
$\theta_t^{\underline{s}^0, \underline{s}^1}(s)$	-45 (41)	<b>574</b> (58)	<b>225</b> (60)	-343 (144)	665 (287)	496 (219)	<b>1559</b> (349)	543 (445)	644 (441)	-180 (140)	-73 (158)	19 (192)

Note: The first element of each sequence refers to period 3(1). Standard errors in parentheses. Bold : Effect is significant at 1% level. Italics: Effect significant at 10% level.  $N_s$  is the sample size after imposing common support.

For the interpretation of the estimated effects and counterfactual outcomes, it is important to note that whenever the potential state '0' appears, it does not necessarily mean that the individual is unemployed, but merely that she is not participating in some programme in that period. The first practical problem that appears is that there are many possible comparisons. Here, we focus on different programmes of a length of 3 periods that may or may not be followed by a start of a further programme in period 4(2). The duration of the latter programme is not part of the explicit definition of the effect. To capture the effects of having or not having a second programme in period 4(2), all sequences have to be fully

specified for at least 6 periods. Thus, a sequence like 111002 should be interpreted as participating in programme 1 for three quarters and then starting programme 2 in 4(2). The dynamic causal model allows to fix the duration of the first programme (of course the duration of the second programme could be fixed as well, but that would lead to a too small number of remaining participants), and thus isolate the effects of different programme durations (see next section) from sequences of programmes.

Table 3 shows a variety of potentially interesting comparisons for different target populations, namely those who participated in a programme in the first period and the non-participants in that period. Since W-DCIA has only limited identifying power, the reference populations are based on the treatment status in the first period. Compared to nonparticipation, the results show lock-in effects of different sizes, and - with the exception of programme 4 - considerable positive effects thereafter (although there is sometimes a lack of precision), a finding that is line with the true values that can be found in Appendix A. There is considerable effect heterogeneity across target populations (note that their effective sample sizes and composition after common support depend on the sequences under investigation; thus computing the effects of 222001 compared to 222000 for participants in 2 by using the triangular results of 222000 and 222001 compared to 000000 is not strictly valid). Furthermore, the estimates for the comparison of the different programmes are too noisy to precisely pin down any effect.

The precision of the estimates depends on the number of 'useful' observations (i.e. observations that are comparable to those in the target populations) in the sequences, which is of course related to the length of the specified sequence - the longer the more precise the meaning of the causal contrast but the fewer observations. Furthermore, an increased number of observations in the target population increases precision as well (comparing the results for different target populations used for the comparison of the same sequences).

## **5.2 *Delayed programme start - the effects of waiting***

In the previous section, the specification of the target quantity for which the causal contrast is desired appeared to be obvious. When interest is in the effect of waiting for the start (or delaying the start) of a programme, there are different ways to state the causal parameter. The first possibility is to just concentrate on the beginning of the programme and to take no account of the fact that programmes that may start at different points in time may differ for example with respect to their duration. Such a comparison is displayed in the upper part of Table 4. Note that the required sequences have different lengths. It appears that for programme 1 the long-run effects of delaying are small in this set-up, which is again in line with the true values.

The alternative (or the complement) to this approach is to require some minimum programme duration, as is shown in the second panel for programme 3. For that programme the estimation results do not change much, when different minimum lengths are considered. The price to pay for specifying longer sequences is of course as before a reduced precision of the estimator, although since programme 3 is a longer programme, there is not so much attrition in the first three periods considered in Table 4 (which is also the reason why the results do not change much when increasing the length of the sequences). Note that the comparisons of the different effects for the same target population is sometimes hampered by the fact that the common support may shrink drastically when the length of the sequence increases, because the participants tend to become more homogenous the longer the sequences. This issue that is taken up again below.

Table 4: The earnings effects of waiting (sequential matching)

Earnings in year (quarter)												
(1)	(2)	(3)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(11)	(12)	(13)
	4(4)	7(4)	9(4)	4(4)	7(4)	9(4)	4(4)	7(4)	9(4)	4(4)	7(4)	9(4)
Sequences $\frac{s^1}{s^0}$	1-01			1-0001			1-00001			1-000001		
Population $S(N_s)$	1 (8833)			1 (7816)			1 (5172)			1 (3405)		
$E(Y_t^{\frac{s^1}{s^0}}   S = s)$	1805 (36)	3903 (43)	3862 (44)	1381 (32)	3255 (42)	3190 (44)	815 (23)	2066 (34)	1956 (36)	631 (20)	1587 (28)	1452 (30)
$E(Y_t^{\frac{s^0}{s^1}}   S = s)$	1698 (100)	3935 (103)	3818 (109)	1963 (220)	3304 (203)	3246 (228)	1214 (155)	2019 (140)	1802 (143)	555 (69)	1198 (92)	1179 (101)
$\theta_t^{\frac{s^0}{s^1}}(s)$	106 (106)	-31 (112)	44 (118)	<b>-581</b> (222)	-49 (207)	-56 (232)	-398 (157)	47 (144)	153 (148)	76 (72)	<b>388</b> (96)	272 (106)
Population $S(N_s)$	0 (69316)			0 (69219)			0 (57471)			0 (44463)		
$E(Y_t^{\frac{s^1}{s^0}}   S = s)$	506 (13)	1362 (17)	1209 (19)	490 (12)	1328 (18)	1174 (19)	460 (12)	1256 (17)	1101 (19)	422 (12)	1170 (17)	1004 (20)
$E(Y_t^{\frac{s^0}{s^1}}   S = s)$	463 (15)	1374 (19)	1212 (19)	564 (33)	1476 (38)	1313 (44)	606 (43)	1392 (61)	1257 (58)	447 (55)	1021 (87)	972 (91)
$\theta_t^{\frac{s^0}{s^1}}(s)$	43 (19)	-12 (25)	-2 (27)	-74 (35)	<b>-151</b> (41)	<b>-139</b> (48)	<b>-145</b> (45)	-136 (63)	<b>-156</b> (61)	-25 (56)	149 (88)	31 (93)
Sequences $\frac{s^1}{s^0}$	3 - 0003			33-00033			333-000333			4-000004		
Population $S(N_s)$	3 (4902)			3 (4895)			3 (4861)			4 (4340)		
$E(Y_t^{\frac{s^1}{s^0}}   S = s)$	611 (10)	749 (13)	707 (13)	622 (10)	760 (13)	713 (13)	649 (13)	768 (15)	727 (17)	262 (9)	782 (13)	705 (14)
$E(Y_t^{\frac{s^0}{s^1}}   S = s)$	516 (24)	845 (27)	728 (37)	526 (25)	867 (30)	744 (40)	541 (30)	908 (35)	749 (38)	261 (289)	624 (295)	809 (446)
$\theta_t^{\frac{s^0}{s^1}}(s)$	<b>95</b> (26)	<b>-95</b> (29)	-21 (39)	96 (27)	<b>-106</b> (33)	-31 (42)	<b>108</b> (32)	<b>-139</b> (38)	-22 (42)	0 (289)	158 (296)	-104 (447)
Population $S(N_s)$	0 (65673)			0 (65572)			0 (64177)			0 (20887)		
$E(Y_t^{\frac{s^1}{s^0}}   S = s)$	663 (22)	808 (32)	782 (32)	639 (23)	812 (32)	783 (31)	687 (22)	820 (31)	815 (33)	311 (55)	884 (100)	781 (100)
$E(Y_t^{\frac{s^0}{s^1}}   S = s)$	544 (21)	902 (26)	792 (28)	542 (23)	912 (27)	795 (32)	557 (25)	932 (28)	788 (32)	345 (116)	599 (116)	620 (174)
$\theta_t^{\frac{s^0}{s^1}}(s)$	<b>89</b> (31)	-93 (41)	-10 (42)	<b>95</b> (32)	-101 (42)	-12 (45)	<b>128</b> (33)	-112 (42)	27 (47)	-34 (128)	285 (153)	161 (201)

Note: See note below Table 3.

Of course, many more contrasts that are interesting could be considered, like requiring an exact length of both programmes, for example. However, they are beyond the illustrative nature and the space constraints of this paper.

### 5.3 The effects of the duration of a programme

In this section, I take up the issue of how to measure the effects of different durations of the programmes. The comparison of interest is the effect of an extension of a programme. Table 5 shows the results for extension of 1, 2, or 3 periods. The results differ according to whether the extensions

concerns only the minimum duration (upper panel) or the actual duration (lower panel). In the long run there are positive effects of programme duration that are reflected in the estimates, although, as observed before, precision becomes an issues when the specified sequence gets longer.

Table 5: The earnings effects of the duration of the programme (sequential matching)

Earnings in year (quarter)												
(1)	(2)	(3)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(11)	(12)	(13)
	4(4)	7(4)	9(4)	4(4)	7(4)	9(4)	4(4)	7(4)	9(4)	4(4)	7(4)	9(4)
Sequences $\underline{s}^1 - \underline{s}^0$	1-11			1-111			1-1111			11-1111		
Population $S(N_s)$	1 (8989)			1 (8957)			1 (8813)			1 (8810)		
$E(Y_t^{\underline{s}^1}   S = s)$	1885 (36)	4002 (43)	3965 (45)	1885 (36)	4003 (44)	3965 (45)	1882 (36)	4006 (44)	3971 (45)	1955 (45)	4133 (54)	4036 (56)
$E(Y_t^{\underline{s}^0}   S = s)$	1958 (44)	4101 (53)	4002 (55)	2073 (71)	4248 (82)	4110 (87)	1948 (174)	4341 (205)	4290 (212)	2008 (166)	4349 (194)	4151 (199)
$\theta_t^{\underline{s}^0, \underline{s}^1}(s)$	-72 (57)	-99 (69)	-37 (71)	-187 (79)	<b>-245</b> (93)	-144 (97)	-65 (178)	-334 (209)	-319 (217)	-53 (172)	-215 (201)	-114 (207)
Sequences $\underline{s}^1 - \underline{s}^0$	10-110			110-1110			10-11110			110-11110		
Population $S(N_s)$	1 (8973)			1 (8935)			1 (8757)			1 (8763)		
$E(Y_t^{\underline{s}^1}   S = s)$	1570 (83)	3583 (111)	3767 (114)	1852 (68)	3915 (83)	3858 (85)	1628 (91)	3514 (112)	3726 (117)	1855 (67)	3950 (82)	3876 (83)
$E(Y_t^{\underline{s}^0}   S = s)$	1942 (71)	3936 (82)	3900 (84)	2056 (77)	4244 (94)	4135 (98)	1995 (174)	4318 (204)	4251 (212)	1973 (166)	4288 (198)	4179 (203)
$\theta_t^{\underline{s}^0, \underline{s}^1}(s)$	<b>-371</b> (109)	<b>-352</b> (138)	-133 (142)	-203 (103)	<b>-328</b> (125)	-276 (130)	-336 (196)	<b>-803</b> (233)	-524 (243)	-117 (179)	-338 (215)	-302 (219)

Note: See note below Table 3.

Note that the static evaluation literature typically refrains from estimating the effects of actual duration, because the programmes themselves can cause attrition and thus actual duration cannot be used to differentiate the programmes. As long as relevant factors are observable, so that the W-DCIA holds, this issue is not a problem here and effects can be estimated directly.

## 6 A comparison to static matching

Having discussed some of the questions that can be fruitfully addressed using the dynamic treatment approach, this section is concerned with a comparison of the sequential estimators and static approximations of the dynamic problem. For the latter, there are (at least) two possible implementations: The first one ignores the role of the intermediate outcomes as selection variables for the

reason that they are endogenous for the beginning of the sequence. The second approximation includes those variables in the list of covariates.

There are two a priori reasons why static and dynamic matching estimators may deviate. First, the static version is biased if S-DCIA fails. Second, the estimates may be based on different populations because of different common supports. Different common supports become particularly an issue when the static approach does not use the intermediate outcomes. Since the bias of the static estimator is extensively documented in L04, I refrain from searching for an example for which the bias is sufficiently substantial on the same common support. Instead, I chose the same example that will appear below when discussing the common support problem.

In Table 6 column (1) defines the relevant comparisons and the target population. Column (2) gives the sample sizes for the treatment groups as well as for the target populations after having imposed common support. The latter depends on the particular sequences under consideration. Column (3) shows the (unadjusted) sample means of the outcome variables in the various subpopulations. Comparing these sample means to the true values of the potential outcomes given in column (4) gives some indication of the selection bias that should be corrected for by matching. Note however that the means for the sequences are based on all observations, whereas the means for the target population as well as for the true values are based on the common support for the dynamic matching. Columns (6) to (9) contain the results for the static matching, whereas columns (10) and (11) contain the corresponding results for sequential matching.

Table 6: The comparison with static matching – earnings and employment in period 9 (4)

Population	Obs.	Sample means		True Values		Static matching (own support)				Dynamic matching	
$\underline{s}^1$	$N_{\underline{s}^1}$	$E(Y_t   \underline{S} = \underline{s}^1)$		$E(Y_t^{\underline{s}^1}   S = s)$		$E(Y_t^{\underline{s}^1}   S = s)$				$E(Y_t^{\underline{s}^1}   S = s)$	
$\underline{s}^0$	$N_{\underline{s}^0}$	$E(Y_t   \underline{S} = \underline{s}^0)$		$E(Y_t^{\underline{s}^0}   S = s)$		$E(Y_t^{\underline{s}^0}   S = s)$				$E(Y_t^{\underline{s}^0}   S = s)$	
$\underline{s}$	$N_{\underline{s}}$	$E(Y_t   S = s)$		$\theta_t^{\underline{s}^0, \underline{s}^1}(s)$		$\theta_t^{\underline{s}^0, \underline{s}^1}(s)$				$\theta_t^{\underline{s}^0, \underline{s}^1}(s)$	
						exogenous X		endogenous X			
		mean	(std)	mean	std	mean	(std)	mean	(std)	mean	(std)
(1)	(2)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Earnings											
1	8997	3964	(45)	3922	4045	3943	(45)	3953	(45)	3862	(44)
01	7994	1913	(31)	3875	4128	3923	(98)	3778	(101)	3818	(109)
1	8833	3862	(45)	53	-	20	(108)	175	(111)	44	(118)
1	8997	3964	(45)	1590	1751	2598	(40)	1244	(27)	1452	(30)
000001	321	1269	(88)	1513	1873	2194	(415)	1186	(188)	1179	(101)
1	3405	1452	(31)	77	-	404	(417)	57	(190)	272	(106)
1	8997	3964	(45)	1344	1690	1916	(41)	1729	(42)	1209	(19)
01	7994	1914	(31)	1282	1707	1915	(31)	1804	(29)	1212	(19)
0	69316	1250	(7)	62	-	1	(51)	-78	(51)	-2	(27)
1	8997	3964	(45)	1130	1180	1339	(97)	961	(164)	1004	(20)
000001	321	1269	(87)	1033	1222	1274	(89)	1031	(95)	972	(91)
0	44463	1020	(6)	97	-	65	(131)	-70	(190)	31	(93)
Employment rate in %											
1	8997	86	(.4)	87	-	86	(.4)	86	(.4)	86	(.4)
01	7994	82	(.4)	87	-	86	(.8)	84	(.9)	86	(.8)
1	8833	86	(.4)	0	-	0	(.8)	2	(1)	0	(.9)
1	8997	86	(.4)	84	-	84	(.4)	81	(.6)	83	(.7)
000001	321	85	(.2)	87	-	89	(.4)	80	(.8)	82	(.3)
1	3405	83	(.6)	-3	-	-4	(.4)	1	(.8)	1	(.3)
1	8997	86	(.4)	81	-	83	(.7)	80	(.9)	80	(.8)
01	7994	82	(.4)	78	-	82	(.4)	82	(.4)	78	(.8)
0	69316	69	(.2)	3	-	1	(.9)	-2	(1)	1	(1)
1	8997	86	(.5)	79	-	83	(.2)	77	(.6)	79	(.9)
000001	321	85	(.3)	83	-	85	(.2)	80	(.3)	80	(.4)
0	44463	68	(.2)	-4	-	-2	(.3)	-3	(.7)	-1	(.4)

Note: See footnote on Table 3. Standard deviation of estimator in brackets. (std) denotes the standard deviation of the estimator, whereas std (without brackets) denotes the standard deviation in the sample (not given for the binary outcome). True values and number of observations of the target population are given for the common support of the dynamic matching estimator.

The first observation is that the results for static matching may differ substantially depending on whether the intermediate outcomes are included, or not. In the latter case, the results are typically close to those of the dynamic matching that in turn appear not too far away from the true values. The second observation is that the common support issue does play a major role as can be seen by the changing number of observations, by the sample means of the outcome variable for the target population as well as for the different estimates for the shortest sequence '1' that is not subject to the bias due to ignored



dynamics. Since the common support issue resurfaces in almost every discussion of the results so far, it is discussed explicitly in the next section.

## **7 Some potential problems**

### ***7.1 Common support***

It has already been pointed out that the comparisons of the different effects for the same target population may be hampered by the fact that the common support may shrink drastically when the length of the sequences increases, because the participants tend to become more homogenous the longer the sequences.

Table 7, which has a similar structure as Table 6, addresses this issue explicitly for two comparisons. First, the effect of waiting by increasing waiting time (and thus the length of the comparison sequence) from one period to five periods is considered. It appears that the sample size of the target population (participants in one-period treatment) drops from 8800 to only 3400 for the sequence defined over 6 periods. Comparing the changing values of the true potential outcomes in the target populations before and after imposing common support, shows that the populations are indeed very different. Whereas in the initial target population mean potential earnings for at least one period of programme 1 was 4022 EUR, within the common support for the comparison of 1 to 000001 this value for the mean potential outcome is less than half that value (1452 EUR).

A similar picture arises in the bottom panel of Table 7, where different minimum durations of participation in programme 2 are compared to spells of non-participation of the same minimum duration, although the drop is much less dramatic.

Table 7: The issue of common support– earnings in 9 (4)

Popul. $\underline{s}^1$ $\underline{s}^0$ $S$	Obs. $N_{\underline{s}^1}$ $N_{\underline{s}^0}$ $N_s$	Sample means $E(Y_t   \underline{S} = \underline{s}^1)$ $E(Y_t   \underline{S} = \underline{s}^0)$ $E(Y_t   S = s)$		True Values $E(Y_t^{\underline{s}^1}   S = s)$ $E(Y_t^{\underline{s}^0}   S = s)$ $\theta_t^{\underline{s}^0, \underline{s}^1}(s)$				Dynamic matching $E(Y_t^{\underline{s}^1}   S = s)$ $E(Y_t^{\underline{s}^0}   S = s)$ $\theta_t^{\underline{s}^0, \underline{s}^1}(s)$	
				before common sup.		after common support			
		mean	(std)	mean	std	mean	std	mean	(std)
(1)	(2)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	8997	3964	(45)	4022	4087	3922	4045	3862	(44)
01	7994	1913	(31)	3975	4169	3875	4128	3818	(109)
1	8833	3862	(45)	47	-	53	-	44	(118)
1	8997	3964	(45)	4022	4087	3705	3938	3943	(45)
001	5921	1597	(30)	4065	4382	3746	4239	3830	(198)
1	8509	3643	(45)	-43	-	-41	-	-186	(203)
1	8997	3964	(45)	4022	4087	3256	3679	3190	(44)
0001	4385	1392	(28)	3982	4397	3199	3992	3246	(228)
1	7816	3190	(44)	40	-	57	-	-56	(232)
1	8997	3964	(45)	4022	4087	1590	1751	1452	(30)
000001	321	1269	(88)	3923	4263	1513	1873	1179	(101)
1	3405	1452	(31)	99	-	77	-	<b>272</b>	(106)
2	8665	5229	(49)	5151	4348	5055	4327	5134	(49)
0	69951	1241	(7)	4672	4223	4574	4200	4777	(92)
2	8502	5134	(49)	479	-	481	-	<b>357</b>	(104)
22	8359	5249	(50)	5162	4363	4855	4268	4968	(51)
00	47265	1091	(7)	4630	4247	4376	4140	4678	(234)
2	8177	4934	(49)	532	-	479	-	290	(240)
222	7615	5242	(52)	5182	4392	4661	4229	4778	(57)
000	31021	968	(8)	4555	4227	4040	4071	4198	(291)
2	7789	4714	(50)	627	-	621	-	580	(296)
2222	5474	4986	(61)	5232	4469	3784	3778	3972	(77)
0000	20405	841	(8)	4507	4179	3046	3542	2818	(263)
2	6172	3839	(50)	725	-	738	-	<b>1154</b>	(274)

Note: See footnote on Table 3. (std) denotes the standard deviation of the estimator, whereas std (without brackets) denotes the standard deviation in the sample (not given for the binary outcome).

Again, these examples point out that there is a price to pay for getting a more precise definition of treatment and the selection effects within a nonparametric framework, because only relatively small parts of the sample may contain useful information for those particular dynamic sequences. The problem of small sample size can be remedied in the two usual ways: The first option is to change the parameter of interest by specifying shorter sequences. This corresponds to aggregating over some longer sequences. The second option is to use parametric assumptions to add additional smoothness to the estimation problem. This is, however, not the subject of this paper.

Note that these findings are fairly robust when the common support is chosen more conservatively, like for example to be the region between the 10th largest / smallest value of the respective scores as opposed to the second largest / smallest value as in the baseline estimation.

## ***7.2 Too many covariates may be needed to control selection***

A potential drawback of the sequential matching approach is that the number of theoretically required covariates increases with the length of the sequence specified, because all the past intermediate outcomes observable at each node of a sequence should be included. When matching is on the propensity score, it implies that the dimension of the matching problem is increasing even if a parametric propensity score is used. The reason is that all past scores should be included as well. Furthermore, estimating a propensity score becomes more demanding when moving along the sequences, as the number of variables increases. Both potential problems could become more severe due to fact that the number of observations decreases when moving along the sequence. Finally, increasing the number of control variables could also potentially decrease the number of observations that remain in the target population and thus change the estimand considerably.

Table 8: The importance and danger of including a rich set of covariates – earnings and employment in 9 (4)

Population	Obs.	Sample means		True Values				Dynamic matching			
$\underline{s}^1$	$N_{\underline{s}^1}$	$E(Y_t   \underline{S} = \underline{s}^1)$		$E(Y_t^{\underline{s}^1}   S = s)$				$E(Y_t^{\underline{s}^1}   S = s)$			
$\underline{s}^0$	$N_{\underline{s}^0}$	$E(Y_t   \underline{S} = \underline{s}^0)$		$E(Y_t^{\underline{s}^0}   S = s)$				$E(Y_t^{\underline{s}^0}   S = s)$			
$\underline{s}$	$N_{\underline{s}}$	$E(Y_t   S = s)$		$\theta_t^{\underline{s}^0, \underline{s}^1}(s)$				$\theta_t^{\underline{s}^0, \underline{s}^1}(s)$			
				few X		all X		few X		all X	
		mean	(std)	mean	std	mean	std	mean	(std)	mean	(std)
(1)	(2)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Earnings											
1	8997	3964	(45)	3983	4071	3922	4045	3925	(45)	3862	(44)
01	7994	1913	(31)	3936	4151	3875	4128	3663	(100)	3818	(109)
1 (all X)	8833	3862	(45)	47	-	53	-	<b>261</b>	(110)	44	(118)
1	8997	3964	(45)	1852	2372	1590	1751	1745	(36)	1452	(30)
000001	321	1269	(88)	1770	2604	1513	1873	1320	(140)	1179	(101)
1 (all X)	3405	1452	(31)	82	-	77	-	<b>425</b>	(144)	272	(106)
1	8997	3964	(45)	1349	1696	1344	1690	1324	(23)	1209	(19)
01	7994	1913	(31)	1287	1713	1282	1707	1272	(21)	1212	(19)
0 (all X)	69316	3862	(45)	62	-	62	-	52	(31)	-2	(27)
1	8997	3964	(45)	1140	1254	1130	1180	1065	(21)	1004	(20)
000001	321	1269	(88)	1251	1038	1033	1222	1080	(120)	972	(91)
0 (all X)	44463	1452	(31)	-111	-	97	-	-15	(122)	31	(93)
Employment rate in %											
1	8997	86	-	87	-	87	-	86	(.3)	86	(.4)
01	7994	82	-	85	-	87	-	85	(.8)	86	(.8)
1 (all X)	8833	86	-	2	-	0	-	1	(.9)	0	(.9)
1	8997	86	-	84	-	84	-	83	(.5)	83	(.7)
000001	321	85	-	86	-	87	-	78	(4)	82	(3)
1 (all X)	3405	83	-	-2	-	-3	-	4	(5)	1	(3)
1	8997	86	-	81	-	81	-	80	(.8)	80	(.8)
01	7994	82	-	78	-	78	-	79	(.7)	78	(.8)
0 (all X)	69316	69	-	3	-	3	-	1	(1)	1	(1)
1	8997	86	-	79	-	79	-	79	(.9)	79	(.9)
000001	321	85	-	83	-	83	-	77	(4)	80	(4)
0 (all X)	44463	68	-	-4	-	-4	-	2	(4)	-1	(4)

Note: See footnote on Table 3. (std) denotes the standard deviation of the estimator, whereas std (without brackets) denotes the standard deviation in the sample (not given for the binary outcome). Sample means and sample sizes after common support are those relating to the common support obtained for the estimation with the full set of covariates.

Table 8 documents some of those problems by comparing two sequential matching estimators based on different specifications of the covariates and compares the results for a long and a short sequence: "All X" defines a specification in which all variables that are theoretically required are included (up to 28 for sequence 01 and up to 39 for 000001).<sup>9</sup> "Few X" defines a specification where only the last information

<sup>9</sup> For the longer sequence the intermediate outcomes of only every second period is included in the probits, because otherwise it would not converge for the last node in the sequence.

about the employment history and intermediate outcomes is included, leading to only 22 variables in both cases.<sup>10</sup>

When considering the results for employment rates and earnings, there appears to be no significant differences for the employment rates. For the longer sequences and their effect on earnings, the target populations change to some extent in this example (compare the true values of the target population adjusted for the common support given by the two different estimators). However, the results for both specifications are still close to the true values. The fact that both specifications are close is probably because the observed intermediate outcomes are highly correlated and including only the last one is (almost) sufficient. On the other hand, the specification with all the variables included does not point to any problem with the estimation (indeed it seems to be preferable to the one that leaves out some variables), so that, in this example, it seems safe to include all variables in the specification that are required by theory.

## 8 Conclusions

This paper attempted to show that dynamic matching estimation can be a useful additional tool for empirical researchers because in many important cases it allows to define the causal parameter of interest more precisely and allows to address selection problems that occur while the treatment under consideration is in progress.

However, there is a price for getting a more precise definition of treatment and the selection effects within this nonparametric framework, because only relatively small parts of the sample may contain useful information for the particular dynamic sequences of interest (dynamic matching is an identification and estimation strategy starving for data). This problem of too small sample sizes can be addressed by

---

<sup>10</sup> For the comparison of 1 to 01, the specification with fewer covariates captures the employment history before 2(4) only with variables from period 2(2) as well as with the usual time-constant variables, whereas the full specification uses a much richer set of covariates.

aggregating over some longer sequences. This however changes the parameter of interest and brings the dynamic model closer to the static one. Alternatively, one may consider parametric assumptions to add additional smoothness to the estimation problem. In the future, this option may be explored in more depth.

## References

- Angrist, J. D., and A. B. Krueger (1999): "Empirical Strategies in Labor Economics", in O. Ashenfelter and D. Card (Hrsg.), *Handbook of Labor Economics*, Vol. III A, Amsterdam: North-Holland, 1277-1366.
- Dawid, A. P. (1979): "Conditional Independence in Statistical Theory", *Journal of the Royal Statistical Society B*, 41, 1-31.
- Gerfin, M., and M. Lechner (2002): "Microeconomic Evaluation of the Active Labour Market Policy in Switzerland," *The Economic Journal*, 112, 854-893.
- Fredriksson, P., and P. Johansson (2003): "Program Evaluation and Random Program Starts", IFAU Discussion Paper, 2003(1), Uppsala.
- Gill, R. D., and J. M. Robins (2001): "Causal Inference for Complex Longitudinal Data: the continuous case", *The Annals of Statistics*, 2001, 1-27.
- Heckman, J. J., R. J. LaLonde, and 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. III A, Amsterdam: North-Holland, 1865-2097.
- Imbens, G. W. (2000): "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*, 87, 706-710.
- Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review", *Review of Economics and Statistics*, 86(1), 4-29.
- Lechner, M. (1999): "Earnings and Employment Effects of Continuous Off-the-job Training in East Germany after Unification," *Journal of Business & Economic Statistics*, 17, 74-90.
- Lechner, M. (2001): "Identification and estimation of causal effects of multiple treatments under the conditional independence assumption", in M. Lechner and F. Pfeiffer (eds., 2001), *Econometric Evaluation of Active Labour Market Policies*, Heidelberg: Physica, 43-58.
- Lechner, M. (2002): "Programme Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labour Market Policies", *Review of Economics and Statistics*, 84, 205-220.
- Lechner, M. (2004): "Sequential Matching Estimation of Dynamic Causal Models", University of St. Gallen, Discussion paper; revised 2005.
- Lechner, M., and R. Miquel (2001): "A Potential Outcome Approach to Dynamic Programme Evaluation – Part I: Identification", Discussion paper 2001-07, Department of Economics, University of St. Gallen; revised 2004.
- Lechner, M., R. Miquel and C. Wunsch (2004): "Long-run effects of Public Sector Sponsored Training in West Germany", Discussion Paper 2004-19, Department of Economics, University of St. Gallen.
- Li, Y. P., K. J. Propert, and P. Rosenbaum (2001): "Balanced Risk Set Matching", *Journal of the American Statistical Society*, 96, 870-882.
- 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, with 1987 Errata to "A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect." *Computers and Mathematics with Applications*, 14:917-921; 1987 Addendum to "A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect." *Computers and Mathematics with Applications*, 14:923-945; and 1987 Errata to "Addendum to 'A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect'." *Computers and Mathematics with Applications*, 18:477.
- Rosenbaum, P. R., and D. B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, 70, 41-50.
- Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 66, 688-701.

- Sianesi, B. (2004): "An Evaluation of the Swedish System of Active Labour Market Programmes in the 1990s", *The Review of Economics and Statistics*, 86, 133-155.
- Zhao, Z. (2004): "Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metric and a Monte Carlo Study ", *The Review of Economics and Statistics*, 86, 91-107.



## Appendix A: Description of outcome heterogeneity

The following Tables A.1 and A.2 contain some descriptive statistics for the true potential outcome variables that can be used to deduce the true effects. To avoid flooding the reader with numbers that are not helpful in interpreting the results that appear in the main body of the paper, we give means and standard deviations for earnings (A.1) and employment (A.2) only for sequences and subsamples that have some relation to the effects considered above.

Table A.1a: Mean and standard deviation of potential earnings in 7(4) in selected subsamples

Sequence	Subsample									
	0		1		2		3		4	
	mean	std.	m.	std.	m	std.	m.	std.	m.	std.
000000	933	1587	3613	4108	4378	4169	624	796	542	782
01	1432	1588	4055	3975	4815	4023	1121	877	1026	838
001	1506	1735	4209	4321	5000	4316	1172	959	1079	941
0001	1342	1722	4089	4320	4830	4384	1013	924	923	911
000001	1283	1720	3936	4239	4724	4309	955	935	858	889
1	1470	1581	4073	3968	4822	3920	1157	858	1062	828
11	1481	1557	4105	3933	4858	3967	1164	867	1070	838
111	1500	1588	4166	4012	4924	4042	1179	885	1084	857
1111	1510	1617	4216	4097	4977	4128	1186	902	1089	872
10	1432	1488	4005	3781	4683	3778	1129	832	1035	797
110	1498	1570	4094	3905	4784	3906	1148	853	1054	822
1110	1497	1583	4150	3995	4908	4024	1177	832	1089	872
11110	1517	1615	4297	4176	4968	4120	1194	903	1097	872
111000	1337	1661	4102	4208	4885	4236	1009	880	913	869
111002	1440	1685	3781	4157	4406	4329	1173	1074	1104	1069
2	1793	1777	4338	4214	4973	4311	1505	1152	1456	1173
22	1795	1783	4351	4232	4989	4328	2263	2625	1457	1176
222	1800	1796	4377	4266	5017	4359	1510	1162	1459	1183
2222	1802	1830	4437	4361	5086	4445	1507	1179	1454	1200
222000	1657	1723	4197	4152	4841	4252	1369	1080	1326	1107
222001	2041	1867	4567	4304	5254	4424	1726	1186	1669	1199
3	1056	1655	3829	4295	4593	4356	754	870	655	847
33	1057	1661	3838	4310	4603	4371	1570	2624	655	850
333	1065	1676	3867	4354	4635	4412	1581	2648	660	857
0003	1067	1707	3802	4398	4589	4511	1574	2671	686	873
00033	1068	1713	3812	4419	4601	4534	755	878	686	877
000333	1066	1731	3830	4469	4622	4586	751	885	682	886
4	1120	1630	3853	4292	4611	4349	824	850	748	838
000004	1124	1673	3668	4275	4418	4381	835	874	763	837
444000	1033	1642	3757	4320	4523	4388	735	832	662	833
444003	1168	1658	3850	4348	4581	4430	879	865	812	844
# of obs.	69951		8997		8665		4964		7423	

Note: m.: Mean in subsample; std.: Standard deviation in subsample.

Table A.1b: Mean and standard deviation of potential earnings in 9(4) in selected subsamples

Sequence	Subsample									
	0		1		2		3		4	
	mean	std.	m.	std.	m	std.	m.	std.	m.	std.
000000	990	1638	3696	4109	4487	4173	670	863	587	850
01	1275	1707	3975	4169	4757	4219	938	981	848	961
001	1290	1832	4065	4382	4837	4486	951	1062	851	1043
0001	1234	1819	3982	4397	4779	4512	896	1021	800	1011
000001	1242	1785	3923	4263	4720	4391	923	1019	814	977
1	1337	1690	4022	4087	4786	4131	1007	967	916	955
11	1348	1707	4054	4133	4821	4175	1015	975	924	963
111	1362	1738	4109	4216	4883	4251	1026	993	933	978
1111	1398	1775	4192	4314	4968	4339	1057	1022	963	1009
10	1297	1635	3896	3940	4646	3996	980	938	886	929
110	1331	1680	3989	4063	4752	4113	1003	960	912	952
1110	1351	1730	4083	4194	4855	4233	1016	987	923	971
11110	1405	1771	4186	4300	4959	4328	1065	1021	971	1008
111000	1275	1707	3959	4321	4762	4367	816	954	729	940
111002	1866	1856	4372	4240	5037	4355	1173	1074	1500	1215
2	1991	1025	4502	4258	5151	4348	1661	1181	1456	1173
22	1990	1843	4511	4274	5162	4363	2449	2664	1636	1222
222	1889	1853	4528	4307	5182	4392	1656	1186	1633	1225
2222	1977	1878	4567	4388	5232	4469	1639	1190	1613	1232
222000	1791	1751	4274	4147	4934	4242	1462	1073	1444	1118
222001	2060	1922	4559	4323	5233	4438	1734	1250	1686	1259
3	1025	1712	3705	4341	4467	4436	697	902	617	888
33	1026	1717	3713	4355	4477	4450	1520	2653	617	890
333	1034	1734	3743	4399	4508	4492	1532	2677	622	899
0003	1002	1752	3685	4446	4479	4589	1499	2706	606	931
00033	1003	1759	3696	4466	4491	4611	755	878	607	934
000333	1004	1774	3716	4508	4514	4653	751	885	606	939
4	1116	1715	3913	4352	4695	4396	788	912	704	899
000004	1041	1743	4502	4258	4464	4461	742	958	886	926
444000	1056	1716	3830	4344	4620	4405	724	897	649	886
444003	1047	1734	3685	4394	4413	4474	744	922	669	917
# of obs.	69951		8997		8665		4964		7423	

Note: m.: Mean in subsample; std.: Standard deviation in subsample.

Table A.2a: Share of potential employment in 7(4) in selected subsamples in %

Sequence	Subsample				
	0	1	2	3	4
000000	58	71	74	56	50
01	96	97	97	95	95
001	95	96	95	95	95
0001	92	94	96	91	91
000001	91	93	94	89	89
1	97	98	98	97	97
11	97	98	98	97	97
111	97	98	98	97	97
1111	97	98	98	97	97
10	97	98	98	97	97
110	97	98	98	97	97
1110	97	98	98	97	97
11110	97	98	98	97	97
111000	96	98	98	95	95
111002	90	90	89	90	91
2	81	82	82	82	80
22	81	82	82	81	80
222	81	83	82	82	80
2222	80	82	81	81	79
222000	79	81	80	80	78
222001	96	97	96	97	96
3	71	81	82	67	65
33	70	80	82	72	65
333	71	80	82	72	65
0003	76	83	84	77	73
00033	76	83	84	74	73
000333	76	83	84	74	73
4	75	85	86	78	75
000004	81	85	86	80	78
444000	76	82	85	68	72
444003	90	92	92	65	89
# of obs.	69951	8997	8665	4964	7423

Table A.2b: Share of potential employment in 9(4) in selected subsamples in %

Sequence	Subsample				
	0	1	2	3	4
Employment in 9(4) in %-points					
000000	58	72	75	56	50
01	78	85	87	76	74
001	79	85	86	77	76
0001	78	85	86	77	75
000001	84	89	90	83	81
1	80	87	89	81	78
11	81	87	89	81	78
111	82	87	89	82	79
1111	82	87	89	82	79
10	80	86	88	80	76
110	81	87	87	81	78
1110	81	87	89	82	79
11110	82	88	90	83	80
111000	77	85	87	77	74
111002	88	88	89	89	87
2	86	84	83	87	86
22	86	84	83	85	86
222	86	84	83	87	86
2222	86	84	83	86	85
222000	85	83	82	86	84
222001	95	94	94	94	94
3	58	70	72	55	50
33	58	70	72	59	50
333	59	70	72	59	50
0003	56	69	72	58	49
00033	56	69	72	75	49
000333	56	69	72	75	49
4	71	80	82	70	65
000004	69	78	80	67	64
444000	69	79	81	68	63
444003	66	75	77	65	61
# of obs.	69951	8997	8665	4964	7423