Identification of the Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions

Michael Lechner and Ruth Miquel

Swiss Institute for
Empirical Economic Research

University of St.Gallen

Abstract

This paper approaches the dynamic analysis of the effects of training programs for the unemployed in West Germany, or in general the effects of sequences of interventions, from a potential outcome perspective. The identifying power of different assumptions concerning the connection between the dynamic selection process and the outcomes of different sequences is discussed. When participation in the particular sequence of programs is decided period by period depending on its success so far, many parameters of interest are no longer identified. Nevertheless, some interesting dynamic forms of the average treatment effect are identified by a sequential randomization assumption. Based on this approach, we present some new results on the effectiveness of West-German training programs.

Address for correspondence

Michael Lechner (corresponding author), Ruth Miquel

University of St. Gallen, Swiss Institute for Empirical Economic Research (SEW)

Varnbüelstr. 14, CH-9000 St. Gallen, Switzerland

Michael.Lechner@unisg.ch, www.sew.unisg.ch/lechner

1   Introduction*

The empirical and methodological literature on econometric program evaluation places great emphasis on learning causal effects of economic interventions from empirical correlations by understanding selection processes. Many recent contributions to this literature explicitly use a 'causal model' of potential outcomes that is typically associated with Neyman (1923), Wilks (1932), Roy (1951), Cochran and Chambers (1965), and Rubin (1974). This type of causal inference relates to the question of what would happen in one hypothetical situation like participating in a training program, compared to another situation like not participating in such a program. The static treatment model is the workhorse of empirical evaluation studies (we use the term *treatment* as a substitute for *program* or *intervention* to stick to the terminology of this literature, which has strong links to statistics and biometrics). The static model is very explicit about problems and possible solutions of selective treatment participation. It allows the derivation of exciting results with respect to nonparametric identification and robust estimation allowing for very general effect heterogeneity. Examples are papers by Imbens and Angrist (1994) for the identifying power of instrumental variables, Heckman and Vytlacil (2006), and Vytlacil (2002) for nonparametric selec-

tion models, and Rubin (1977) for the conditional independence assumption. Furthermore, see the comprehensive surveys by Heckman and Robb (1985), Heckman, LaLonde, and Smith (1999), and Imbens (2004), as well as the textbook by Cameron and Trivedi (2005). The static model is, however, silent about selection that occurs while treatment is already in progress (dropout). It is also not helpful to address selection problems that occur during a sequence of treatments when interest is in the effect of the full sequence.

Consider the problem of evaluating individual labor market effects of training programs for the unemployed. In many countries that use such policies, unemployed might participate in a sequence of courses, instead of only one course. For example, an unemployed person is assisted with job search. If she remains unemployed, then she is sent to a short training program. If she remains unemployed, then she participates in a longer and more extensive training program, and so on. Usually, the effects of the programs, in which the individual participated so far, influence the next program participation (or her attrition from the planned program sequence). Similar issues arise when the start dates of the programs vary individually, when there are different program durations, or in any combination of such dynamic phenomena. Obviously, any static causal framework needs many simplifying assumptions to be able to define the interesting questions, not to mention the ability to discuss identification of causal parameters of interest.

An explicit dynamic treatment framework has the advantage that questions relating to the definition of parameters and selection biases that occur while the treatment (the sequence) is in operation can be addressed explicitly. Thus, we can derive explicitly the conditions required to identify such parameters from experimental and non-experimental data by allowing for dynamic selection processes that depend on the success of the treatment received so far. This type of selection comes on top of the 'static' selection process in operation when deciding which treatment sequences to start with. Its dependence on the intermediate outcomes of the treatment complicates

2

identification of the causal effects. This type of endogeneity bias is the key issue we tackle in the methodological part of this paper.[1]

In recent econometric evaluation studies of labor market programs, several authors have addressed dynamic causal issues by using ad-hoc modifications of the static causal framework. For example, Bergemann, Fitzenberger, and Speckesser (2004) evaluate training program sequences, Lechner (1999) and Sianesi (2004) propose procedures to deal with participants entering labor market programs at different points in their unemployment spell (different 'waiting times'). In a related setting, Crépon and Kramarz (2002) use different 'start times' to analyze the effects of the introduction of a policy to reduce standard working hours in France. A similar problem is the issue of program duration as analyzed by Behrman, Sengupta, and Todd (2005) in the context of a school subsidy experiment. Because these papers use static models of potential outcomes, it is difficult to define the desired causal effect in a way such that the impact of the (implicit) assumptions about the dynamic selection process on the estimand becomes apparent.[2]

Applications of the explicit dynamic causal framework based on potential outcomes are very rare in econometrics so far. Ding and Lehrer (2003) use such a framework to evaluate a sequentially randomized class size study using difference-in-difference-type estimation methods. The paper

---

[1]   See Rosenbaum (1984), Rubin (2004, 2005), and Lechner (2008) on how the fact that the treatment influences control variables (and thus renders them endogenous) may bias the usual estimators of static evaluation models.

[2]   There are further connections to other strands of econometrics: For example, the literature on dynamic panel data models that are identified by sequential moment conditions (e.g. Chamberlain, 1987, 1992) is related (see discussion paper version of the paper for details). Another connection is with the literature on social learning. In particular, Manski (2004) is concerned with dynamic selection problems from one cohort to the next. However, he assumes that the outcome distribution is stationary over time, which is in sharp contrast to our modeling of the outcomes.

by Lechner (2004) suggests different estimators, subjects them to a Monte Carlo study, and applies some of them to Swiss labor market data. Based on artificial data, Lechner (2008) discusses practical issues that may arise in applications using those matching-type estimators.

The suggested approach builds on work in epidemiology and biostatistics that uses dynamic counterfactual outcomes explicitly to define the effect of treatments in discrete time (e.g. Robins, 1986, 1989, 1997, 1999, Robins, Greenland, and Hu, 1999, for discrete treatments; Robins, 1998, Gill and Robins, 2001, for continuous treatments). Identification is achieved by sequential randomization assumptions (see the summaries by Abbring, 2003, and Abbring and Heckman, 2008, as well as Heckman and Navarro, 2005, and Taber, 2000, for more structural approaches). The effects are typically estimated using parametric models to contain the dimensionality problem that is common in such a set-up. There is also a similarity to Murphy (2003). She proposes estimators for optimal treatment rules that specify how the treatment changes over time depending on how covariates change.

In this paper, we use an explicit dynamic causal model of potential outcomes to discuss conditions for identifying different causal parameters that are of interest in program evaluation and other causal studies. Similar to Robins and co-authors we find that observing the information set that influences the allocation to the next treatment in a treatment sequence as well as the outcome of interest is sufficient to identify average population treatment effects nonparametrically even if this information is influenced by the past of the treatment sequence. In addition to the biometrics literature, we show that treatment effects are not only identified for the (average) population, but also for some subpopulations experiencing parts, or all, of the treatment sequence. These parameters are called treatment-on-the-treated treatment effects and are popular for example in labor market evaluations as they can be used to analyze effect heterogeneity among different populations characterized by participation status. However, for many such subpopulations, it turns out

that this assumption is not strong enough. The reason is that the subpopulation of interest (the participants who completed the sequence) has evolved (been selected) based on the realized intermediate outcomes of the sequence. In many cases, the sequential randomization assumption must be strengthened by imposing (exogeneity) conditions on the joint distribution of potential outcomes and conditioning variables to obtain identification of treatment on the treated parameters. Intuitively, this additional assumption rules out any influence of intermediate outcomes on the future selection processes.

The paper is organized as follows: The next section outlines the general setup of our empirical example that is concerned with the evaluation of German training programs for unemployed and based on a large administrative database. We use this example in all later sections to clarify ideas and to show the usefulness and feasibility of our approach in applied work. In Section 3, we define the notation as well as the effects of interest in a dynamic treatment setting. Only for ease of exposition, we concentrate on the simplest version of the dynamic model with two periods and two treatments per period. The general model with a finite number of periods and a finite number of treatments per period is available in a previous discussion paper version that is available from the website of the authors. Section 4 proposes identification strategies based on sequentially applying conditional independence assumptions and discusses their identifying power for the effects defined in Section 3. Section 5 briefly sketches possible estimation procedures. Section 6 contains the results for the empirical example and Section 7 summarizes our main findings and concludes. The Appendix contains the proofs of the theorems stated in the main part of the paper. Finally, additional technical material, as well as material concerning the details of the empirical application is available on the internet at www.siaw.unisg.ch/lechner/lm_2005.

2   Empirical example: Government sponsored training in West Germany

Germany runs a considerable active labor market policy to combat its high unemployment. In this example, we concentrate on the training part of the policy in West Germany. Between 1991 and 1997, West Germany spent about 3.6 bn Euro per year on such training programs. Lechner, Miquel, and Wunsch (2005, LMW in the following) evaluate the effects of the different training programs. More precisely, they evaluate the effects of beginning the first program participation spell in 1993 or 1994 based on an informative new administrative database. LMW find that the programs have different short-term and long-term effects. However, the causal effects they are estimating neither take account of dropout nor do they include the additional effects of subsequent program participation.

We use the same administrative data as LMW and refer the reader to LMW and the internet appendix for more details concerning the data. We focus on a subsample of individuals who enter unemployment between January 1992 and December 1993 and receive unemployment benefits or unemployment assistance. It is the first month of unemployment within this window that we define as period zero - our reference period. We are interested in comparing three different types of treatments: (i) remaining unemployed and receiving benefits and services from the employment offices (denoted by *U*); (ii) participating in a vocational training program paid for by the employment office (*T*); and (iii) participating in a retraining program paid for by the employment office (*R,* the aim of such programs is to obtain a vocational degree in a different occupation). Since in the data there is not enough variation over time to analyze monthly movements in and out of *R* and *T*, we aggregate the monthly information into quarterly information. We are interested in the effects of participating for four quarters in different types of programs (*TTTT* vs. *RRRR*). Furthermore, we consider the effects of participating in either of those programs com-

pared to remaining in open unemployed (*TTTT* vs. *UUUU*, *RRRR* vs. *UUUU*).[3] Our outcome of interest is whether the individual is employed two, respectively four, years later.

Table 1 shows descriptive statistics for selected variables. The statistics give an indication about differences over time as well as across subsamples defined by treatment status (note that e.g. *UUUU* is a subsample of *UUU,* which is a subsample of *UU,* which is a subsample of *U*).

-- Table 1 about here ---

It is important to distinguish two types of variables, those that are time constant and, thus, cannot be influenced by the treatment (but may influence selection decisions), and those that are time varying and may be influenced by the treatment. The first panel gives some examples of such variables (like age, sex, nationality). The remaining parts of this table contain examples of time varying variables, like the receipt of unemployment benefits, earnings, and the remaining claim to unemployment benefits. Note that time variation in some earnings and particular unemployment benefit claim variables is generated by short interruptions of the unemployment spell as well as by specific events within the unemployment spell, like training or benefit sanctions. There is additional information about education, position in last job, last occupation, industrial sector, region, and information about the last employer (sector, size), employment and unemployment histories and benefit entitlement.

---

[3]   There are many other effects that could be defined and estimated using this framework, like the effect of entering programs at different times (e.g. *UTTT* compared to *TTTT,* or *UTTT-TTT*), like the effect of different lengths of programs (e.g. *T-TTTT*, *TU-TTTT*, etc.), but for the sake of brevity they will be discussed elsewhere. Note that in our application there is an additional state beginning in period 2, which is defined by neither participating in the programs nor being registered as unemployed (i.e. anything else that is not covered by the sequences of interest).

The about 27,000 individuals who are unemployed in the first quarter dominate the sample. 19,500 of them are coded as receiving unemployment benefits every quarter (and did not participate in R or T). A similar decline is observed for vocational training (T) that has a mean duration below one year (500 $\rightarrow$ 120), whereas the number of participants in retraining (R) with a mean duration of almost two years remains relatively stable (175 $\rightarrow$ 150).

The first three columns reveal some information about the initial selection into the treatments. We see that the main differences occur with respect to age. The nonparticipants are 4 - 8 years older on average, had higher earnings and higher remaining benefit claims (both positively correlated with age) before getting unemployed. Over time the older unemployed are more likely to remain unemployed, thus those remaining unemployed over one year are on average five years older than those who are unemployed in the first period. Similar changes over time do not occur with the other groups. The raw estimates, i.e. unadjusted for any differences in observable covariates, of the employment impacts four years later suggest a large positive effect of training and retraining compared to unemployment ($TTTT$ - $UUUU$ = 39%-points; $RRRR$ - $UUUU$ = 52%-points) and a positive effect of at least one year of retraining compared to at least one year of training of about 13%-points. In Section 6, we present the results that correct for selection effects and will find that the effects are generally smaller for the comparison with unemployment and more in favor of $RRRR$ when compared to $TTTT$.

A more sophisticated way to analyze differences in covariates in the different subsamples is to use sequential binary probit analysis, using as covariates time-constant and predetermined time-varying variables (with their full history across the quarters) that may be considered to influence selection in each step of each sequence as well as the outcome variables. Table 2 shows the results of these probits for selected variables (see the Internet Appendix for the results with all covariates and for standard errors). Note that there is one particular specification for every transition

we are interested in. Therefore, Table 2 displays results for 12 probits. The specification is intended to be similar across different programs. However, in some cases (typically related to *R*, a program with not much dropout and not much variation in the planned program duration), there is not much variation in the dependent variable. For example, 164 observations are observed in *RRR*, of which only 15 do not appear in *RRRR*. In these cases, drastically fewer variables are used as independent variables in the respective probit estimation.

-- Table 2 about here --------------------------------------

The table shows many variables influencing the different transitions in a statistically significant way. The coefficients seem to confirm the impressions obtained from the descriptive statistics discussed above. Since history of the time-varying variables is included in all specifications (for reasons that will become apparent after the next section), some multicollinearity problems appear. They either lead to alternating signs of coefficients for different lags of these variables (in the case of *U*) or lead to serious instability or break down of the estimation for some variables in *T*. In case of breakdown, some of the variables are omitted. Finally, note that due to small movement over time for retraining, it appears to be hard to find significant variables for these transitions (other than the constant term), which suggests that selection bias (conditional on having chosen *R* in the first period) is a minor issue for *RR*, *RRR* and *RRRR*.

3   The notation of the dynamic model of potential outcomes

3.1   The variables

In the previous section, we considered a setup with three types of treatments that could occur in four different periods (plus the initial period in which everybody became unemployed). However, to focus ideas we present the formal model for the 'minimal' case of two treatments and two pe-

riods only (to compare *TT* vs. *RR* in our example). Although, the necessary extension is not entirely trivial, the key ideas become apparent with this stylized version.[4]

Consider a world with an initial period in which everybody is in the same treatment state (*U* in our example), plus two subsequent periods in which different treatment states are realized. The periods are indexed by t or τ ($t, \tau \in \{0,1,2\}$). A vector of random variables $S = (S_0, S_1, S_2)$ describes the treatment received by members of the population.[5] Later on, for notational convenience, the initial period is not mentioned explicitly. Starting in period 1, S_t can take two values (e.g. *T* or *R*). A particular realization of S_t is denoted by $s_t \in \{0,1\}$. Furthermore, denote the history of variables up to period t by a bar below a variable, i.e. $\underline{s}_2 = (0, s_1, s_2)$ (e.g. *URR*).[6] In period 1, a member of the population can be observed in exactly one of two treatments. In period 2, she participates in one of four treatment sequences ($(0,0),(1,0),(0,1),(1,1)$), depending on what happened in period 1. This notation allows us to specify shorter (partial) sequences by considering effects of sequences s_1 instead of $\underline{s}_2$. Therefore, every individual 'belongs' to exactly one se-

---

[4] The notation and most of the proofs for the general model are contained in a previous discussion paper version of this paper that is available from the website of the authors (Lechner and Miquel, 2001). See also Robins (1986), etc., for the identification of the average population effects.

[5] We avoid the technical term *units* for members of the population. Given our motivating application, we call them *individuals* instead. Generally, the notational setup in this section follows the spirit of Rubin (1974) and others.

[6] To differentiate between different sequences, sometimes a letter (e.g. *j*) is used to index a sequence like $\underline{s}_t^j$. Furthermore, since all sequences are identical for the base period, we ignore that period in the following when denoting different sequences. As a further convention, capital letters usually denote random variables, whereas small letters denote specific values of the random variable. Deviations from this convention will be obvious.

quence defined by s_1 and another sequence defined by s_2.[7] To sum up, in the three-periods-two-treatments example we consider six different overlapping potential outcomes corresponding to two mutually exclusive states defined by treatment status in period 1 (e.g. *T, R*), plus four mutually exclusive states defined by treatment status in period 1 and 2 together (e.g. *TT, TR, RT, TT*), thus allowing us to evaluate treatments of different lengths.

The variables used to measure the effects of the treatment, i.e. the potential outcomes, are indexed by treatments, and denoted by $Y_t^{s_2}$ (e.g. employment two or four years after the beginning of unemployment). Potential outcomes are measured at the end of (or just after) each period, whereas treatment status is measured in the beginning of each period. For each length of a sequence (1 or 2 periods), there are corresponding potential outcomes. For each length of the sequence, one of the potential outcomes is observable and denoted by Y_t. The resulting observation rules are defined in equation (1):

$$Y_t = S_1 Y_t^1 + (1-S_1)Y_t^0; \quad t \geq 1$$

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

In words, for those who participated in training in the first period, we observe Y_t^T. For those remaining in training for two periods, we observe Y_t^{TT}. It remains to define the variables *X*, that

---

[7] There are different ways to understand such shorter and longer sequences. As suggested by a referee, one may rightly call the effect for the longer sequences the 'structural parameter' as it measures the total effect of the intervention. In that case, the effect for the shorter sequence may be called a reduced form parameter. However, there are contexts in which the later is of primary interest as well and thus becomes structural (e.g. when we are interested in the effect of a case worker sending an unemployed into a program without forcing the unemployed to stay on the program, as the later would be impossible in many countries).

may influence treatment selection and (or) potential outcomes. Because we do not rule out that treatment status influences the values of these variables, there are potential values of these variables ($X^{\underline{s}_{\tilde{\tau}}} = (X_0^{\underline{s}_{\tilde{\tau}}}, X_1^{\underline{s}_{\tilde{\tau}}})$) that we call dynamic confounders. $X_t^{\underline{s}_{\tilde{\tau}}}$ may contain $Y_t^{\underline{s}_{\tilde{\tau}}}$ or functions of it. The K-dimensional vector X_t is observable at the same time as Y_t. The corresponding observation rule for X_t is analogous to the one for the potential outcomes given in equation (1).

3.2   The effects

The purpose of the intended empirical analysis is to estimate the mean causal effect denoted by $\theta_t^{\underline{s}_{\tilde{\tau}}^k ; \underline{s}_{\tilde{\tau}}^l}$ ($t \geq \tau$) in period t of a sequence of treatments defined up to period 1 or 2 (or further) (s_1^k or $\underline{s}_2^k$), e.g. *T* or *TT*, compared to another sequence of the same length (s_1^l or $\underline{s}_2^l$), e.g. *R* or *RR*,. For notational convenience, we consider only pair-wise effects of sequences having the same length. Effects may be heterogeneous across participants in different sequences. For obvious reasons, we are only interested in subpopulations defined by treatment status not specified beyond the last period of the specified treatment sequence. In other words, we do not consider the effects of the treatment in period 1 for the population of participants in a treatment sequence defined for periods 1 and 2.

The definition of the average causal effects is given in equation (2):

$$\theta_t^{\underline{s}_{\tilde{\tau}}^k ; \underline{s}_{\tilde{\tau}}^l}(\underline{s}_{\tilde{\tau}}^j) := E(Y_t^{\underline{s}_{\tilde{\tau}}^k} \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j) - E(Y_t^{\underline{s}_{\tilde{\tau}}^l} \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j),$$

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

Note that all of what follows is also valid in strata defined by the dynamic attributes that are not influenced by the treatment conditional on $\underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}$.

In our examples, if we were interested in the effect of two periods of retraining compared to training in the first period and retraining in the second period for those participating in retraining in both periods on the outcome in period 3, we denote this effect as $\theta_3^{TT;RT}(RR)$. If the relevant subpopulation consists only of those receiving retraining in period 1, then $\theta_3^{TT,RT}(R)$ is the relevant parameter. For $\tilde{\tau}=0$, we obtain the average effect for the population (which is defined by their status in period 0), e.g. $\theta_3^{TT,RT}$. To interpret $\theta_t^{\underline{s}_\tau^k;\underline{s}_\tau^l}(\underline{s}_{\tilde{\tau}}^j)$ as a causal effect, the standard assumptions of the potential outcome framework, like the Rubin (1974) Stable Unit Treatment Value Assumption (SUTVA) have to be invoked as well. They imply that the effects of treatment on person i do not depend on the treatment choices of other people, as is already implied by the specific potential outcome notation. SUTVA is already implicit in the observation rules.

There is a close resemblance of the effects defined in equation (2) to effects that are typically of interest in the static evaluation literature, namely the average treatment effect (ATE, e.g. $\theta_3^{T,R}$) and the ATE on the treated (ATET, e.g. $\theta_3^{T,R}(T)$). Here, we call $\theta_t^{\underline{s}_\tau^k;\underline{s}_\tau^l}$ the dynamic average treatment effect (DATE). Accordingly, $\theta_t^{\underline{s}_\tau^k;\underline{s}_\tau^l}(\underline{s}_\tau^k)$, e.g. $\theta_3^{TT,RT}(TT)$, as well as $\theta_t^{\underline{s}_\tau^k;\underline{s}_\tau^l}(\underline{s}_\tau^l)$, e.g. $\theta_3^{TT,RT}(RT)$ are termed DATE on the treated (DATET) and DATE on the nontreated.[8] There are cases in-between, like $\theta_t^{\underline{s}_\tau^k;\underline{s}_\tau^l}(s_1^l)$, e.g. $\theta_3^{TT,RT}(R)$, for which the conditioning set is defined by a sequence shorter than the ones that are evaluated. Furthermore, note that the effects are symmetric in the sense of $\theta_t^{\underline{s}_\tau^k;\underline{s}_\tau^l}(\underline{s}_{\tilde{\tau}}^j) = -\theta_t^{\underline{s}_\tau^l;\underline{s}_\tau^k}(\underline{s}_{\tilde{\tau}}^j)$, but also that $\theta_t^{\underline{s}_\tau^k;\underline{s}_\tau^l}(\underline{s}_\tau^k) \neq -\theta_t^{\underline{s}_\tau^l;\underline{s}_\tau^k}(\underline{s}_\tau^l)$. The appendix

---

[8] Note that this notation can be used to capture an additional year for some program compared to some alternative, like $\theta_3^{TT,RT}(RT)$. Behrman, Cheng, and Todd (2004) call such effects 'marginal effects'.

provides results concerning the connection of effects defined for different lengths of treatments and conditioning sets.

--- Figure 1 about here --------------------------------

--- Table 3 about here ---------------------------------

Table 3 and Figure 1 summarize the notation as well as the definitions introduced so far, whereas Figure 1 clarifies the timing of the different potential and observable variables, as well as how they relate to the potential treatment paths. To understand the effect of the latter is the ultimate goal of an evaluation exercise based on a dynamic causal framework.

To simplify our notational burden and increase readability of the paper, we will not consider comparisons for which the dynamic approach does not provide new insights, because they are essentially static. These include all comparisons for a treatment specified over one period only, as well as those comparisons of sequences that are defined for two periods but for which period one coincides for the two sequences as well as for the population under investigation, like $\theta_3^{IT,TR}(T)$.

3.3 Sampling and regularity conditions

To complete the framework, assume that a large random sample $\{s_{1i}, s_{2i}, x_{0i}, x_{1i}, x_{2i}, y_{1i}, y_{2i}\}_{i=1:N}$ is at disposal, drawn from a large population of participants in $S_0 = 0$. This population is characterized by the corresponding random variables $(S_1, S_2, X_0, X_1, X_2, Y_1, Y_2)$. Furthermore, all conditional expectations that are of interest in the remainder of this paper shall exist. Taken together these assumptions allow phrasing the questions of identification of the causal parameters as to whether it is possible to express the various expectations of the potential outcomes in terms of expectations of observable outcomes, which, in principle, can be estimated consistently.

4   Identification of the effects of dynamic treatment regimes

4.1   Introduction

From the considerations in the previous section, it is obvious that the data alone could not identify the effects. Like with any causal model based on potential outcomes, there are three general routes to identification. The first option is to choose a particular parametric specification (up to a finite number of unknown parameters) of the joint distribution of the potential outcome and selection variables conditional on attributes. A major criticism of this approach is that, usually, particular specifications are hard to justify with behavioral, institutional, and data related arguments, in particular when they are not a result of a structural behavioral model.

Most of the modern program evaluation literature focuses on nonparametric identification (e.g. Heckman, LaLonde, and Smith, 1999). Within this group, there are two different approaches. The first one relies on having access to variables that influence treatment choice but do not influence potential outcomes, thus fulfilling an exclusion restriction. The causal implications of these so-called instrumental variable (IV) approaches in a nonparametric setting with effect heterogeneity have been explored first by Imbens and Angrist (1994), Angrist, Imbens and Rubin (1996), and Heckman and Vytlacil (1999). Based on various types of exclusion restrictions Miquel (2002, 2003) developed identification strategies for the dynamic causal model presented in the previous section. Ding and Lehrer (2003) apply these ideas in their empirical work.

The second group of nonparametric identification strategies also assumes the existence of an instrument, but does not require observing it. Instead, it supposes that all variables that jointly influence selection and outcome are observed (and the potentially unobserved instrument causes some additional variation of treatment status to be exploited). Thus, conditional on the values of these variables, called confounding variables in the statistics literature, we are in an experimental

15

situation and we can learn the (unobservable) nontreatment outcomes of the treated from the (observable) nontreatment outcome of the nontreated and vice versa (Rubin, 1974, 1977). This assumption is called 'selection-on-observables' or conditional independence assumption (CIA) and gives rise to matching type estimators (see the excellent survey by Imbens, 2004). The current surge in the use of matching estimation is probably due to *better* data becoming available, in particular from government sources (e.g. Gerfin and Lechner, 2002, or Angrist, 1998). *Better* means more informative about outcome and selection variables as well as more observations. The additional information is crucial for making the identifying assumptions plausible, whereas a large number of observations permit application of nonparametric estimation techniques.

The database used in our empirical example falls in this category. In this case, substantial efforts have been made to compile these data from government sources with the particular intention of gathering selection information. Therefore, in the remainder of this section we explore the identifying power of two different versions of the selection on observables assumption in the dynamic context. Compared to the static approach, the major complication is that the outcomes of the treatments experienced so far may influence the variables used to correct for the selection effects.

4.2 Dynamic conditional independence assumptions

In LMW, we argue extensively that the data used in the empirical example is very rich in covariates (like socio-demographic variables, regional variables, employment histories, etc.). Therefore, we proceed under the assumption that the data contains the variables that *jointly* influence the selection process as well as the outcome variables (see LMW for an extensive discussion of the available variables, the institutional details underlying the selection process, the selection process itself, and the plausibility of the static CIA). Of course, this implies that we suppose that unobserved factors, like discount rates for example, have only a negligible influence conditional on observables. ASSUMPTION 1 stating the WEAK DYNAMIC CONDITIONAL INDEPENDENCE AS-

SUMPTIONS (W-DCIA) formalizes this idea using sequential statements about the conditional in-

dependence of outcomes and selection variables given the values of the confounders (also called

sequential randomization assumption in various papers by Robins).[9]

Assumption 1: *Weak dynamic conditional independence assumption (W-DCIA)* [10]

a) $\quad Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \amalg S_1 \mid X_0 = x_0 ;$

b) $\quad Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \amalg S_2 \mid \underline{X}_1 = \underline{x}_1, S_1 = s_1 ;$

c) $\quad 1 > P(S_1 = 1 \mid X_0 = x_0) > 0 , \quad 1 > P(S_2 = 1 \mid \underline{X}_1 = \underline{x}_1, S_1 = s_1) > 0 ;$

$$\forall \underline{x}_1 \in \underline{\chi}_1, \quad \forall s_1 : s_1 \in \{0,1\} .$$

Part a) of ASSUMPTION 1 states that conditional on X_0 potential outcomes are independent of

assignment in period 1 (S_1). This is the standard version of the static CIA. Part b) states that

conditional on the treatment, on observable outcomes (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 an application, we have to

think about which variables influence *changes* in treatment status as well as outcomes. It is likely

that time-varying confounders and outcomes from previous periods play some role. For example,

in our empirical example, there are several variables relating to or derived from events that occur

after the beginning of period 1 and before its end (like changes in claims to unemployment bene-

fits, employment status, etc.). Thus, again, in this example, the assumption that we can control for

---

[9]   The following assumptions relate to identification of all treatment effects defined in Section 2. If the desired

   comparison involves fewer potential outcomes, then the required changes will be obvious.

[10] $A \amalg B \mid C = c$ 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).

the confounding variables that are related to treatment and selection in period 2 (given the treatment history) appears likely to hold. Note that ASSUMPTION 1 does not impose any further restrictions on X. In particular, X_1 may be influenced by the treatment in period 1.

These assumptions are supposed to hold for all values of x_0 and x_1 in a given set $\underline{\chi}_1$ for which we want to learn the effects. To make the necessary comparisons for all elements in this set, there must be a positive probability everywhere in this set to observe individuals in all relevant sequences. Part c) formalizes the condition that is called the *common support requirement* (CSR). It is not strictly necessary when estimation (and identification) is based on parametric models that allow extrapolating into the no-support regions. THEOREM 1 shows that several interesting causal effects are identified.

Theorem 1:   Identification based on W-DCIA

IF ASSUMPTION 1 holds, then $\theta_2^{s_2^k;s_2^l}$ and $\theta_2^{s_2^k;s_2^l}(s_1^j)$ are identified, $\forall s_1^k, s_2^k, s_1^l, s_2^l, s_1^j, s_2^j \in \{0,1\}$.

The proof of THEOREM 1 is given in Appendix A.2.[11]

THEOREM 1 states that pair-wise comparisons of all sequences are identified, but only for groups of individuals defined by their treatment status in period 0 or 1 (like $\theta_2^{RR,TT}(\cdot)$ or $\theta_2^{RR,TT}(T)$; $\theta_2^{RR,TT}(TT)$ is not identified). The relevant distinction between the populations defined by treatment state 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 that $S_0 = 0$

---

[11] Note that the assumption supposes conditional independence as opposed to conditional mean independence. Although the latter is sufficient for identification, the former has the virtue of being valid for all transformations of the dependent variable.

holds for everybody. However, in the second period, randomization into these treatments is conditional on variables already influenced by the first part of the treatment.

Although the appendix contains the formal proof of THEOREM 1, we use the empirical example to understand how to obtain identification. Suppose we are interested in $\theta_2^{TT,RR}(R)$. In this case, we identify $E(Y_2^{TT} \mid S_1 = R)$ and $E(Y_2^{RR} \mid S_1 = R)$ by applying ASSUMPTION 1, the law of iterated expectations (IE), and the observations rule (OR) given in equation (1):

$$
E(Y_2^{RR} \mid S_1 = R) \stackrel{IE}{=} \underset{X_1 \mid S_1 = R}{E} E(Y_2^{RR} \mid \underline{X}_1 = \underline{x}_1, S_1 = R) \stackrel{A.1b}{=} \underset{X_1 \mid S_1 = R}{E} E(Y_2^{RR} \mid \underline{X}_1 = \underline{x}_1, \underline{S}_2 = RR)
$$

$$
\stackrel{OR}{=} \underset{X_1 \mid S_1 = R}{E} E(Y_2 \mid \underline{X}_1 = \underline{x}_1, \underline{S}_2 = RR). \qquad\qquad [\underline{X}_1 = (X_1, X_0)]
$$

$$
E(Y_2^{TT} \mid S_1 = R) \stackrel{IE}{=} \underset{X_0 \mid S_1 = R}{E} E(Y_2^{TT} \mid X_0 = x_0, S_1 = R) \stackrel{A.1a}{=} \underset{X_0 \mid S_1 = R}{E} E(Y_2^{TT} \mid X_0 = x_0, S_1 = T)
$$

$$
\stackrel{IE}{=} \underset{(X_0 \mid S_1 = R)}{E} \underset{(X_1 \mid X_0, S_1 = T)}{E} E(Y_2^{TT} \mid X_0 = x_0, X_1 = x_1, S_1 = T)
$$

$$
\stackrel{A.1b}{=} \underset{(X_0 \mid S_1 = R)}{E} \underset{(X_1 \mid X_0, S_1 = T)}{E} E(Y_2^{TT} \mid \underline{X}_1 = \underline{x}_1, \underline{S}_2 = TT)
$$

$$
\stackrel{OR}{=} \underset{(X_0 \mid S_1 = R)}{E} \underset{(X_1 \mid X_0, S_1 = T)}{E} E(Y_2 \mid \underline{X}_1 = \underline{x}_1, \underline{S}_2 = TT).
$$

The symbol above the equality sign denotes the assumption or the statistical property (IE: iterated expectations; OR: observation rule) used to derive the results on the right hand side of the equality sign. This example shows how to reweigh the observations in *TT* and *RR* successively to learn the counterfactual outcome distribution of *TT* and *RR* for those in the target population of interest (R). Thus, the causal effects of interest can be expressed in terms of random variables for which realizations are observable:

$$
\theta_2^{TT,RR}(R) = \underset{(X_0 \mid S_1 = R)}{E} \underset{(X_1 \mid X_0, S_1 = T)}{E} E(Y_2 \mid \underline{X}_1 = \underline{x}_1, \underline{S}_2 = TT) - \underset{X_1 \mid S_1 = R}{E} E(Y_2 \mid \underline{X}_1 = \underline{x}_1, \underline{S}_2 = RR).
$$

W-DCIA has an appeal for applied work as a natural extension of the static framework used so far. Based on W-DCIA additional parameters are identified using assumptions that are essentially not much more demanding than in the static case. Therefore, the result for the empirical example presented below will be based on this assumption.

Using our example again, it becomes apparent that ASSUMPTION 1 is not powerful (restrictive) enough to obtain a similar identification result for $\theta_2^{TT,RR}(TT)$, $\theta_2^{TT,RR}(RR)$, $\theta_2^{TT,RR}(TR)$, $\theta_2^{TT,RR}(RT)$, although counterfactuals like $E(Y_2^{RR} \mid \underline{S}_2 = RT)$, $E(Y_2^{TR} \mid \underline{S}_2 = TT)$, $E(Y_2^{RT} \mid \underline{S}_2 = RR)$, and $E(Y_2^{TT} \mid \underline{S}_2 = TR)$ are identified:

$$E(Y_2^{TT} \mid \underline{S}_2 = RR) \overset{IE}{=} \underset{X_1 \mid \underline{S}_2 = RR}{E} E(Y_2^{TT} \mid \underline{X}_1 = \underline{x}_1, \underline{S}_2 = RR) \overset{A.1b}{=} \underset{X_1 \mid \underline{S}_2 = RR}{E} E(Y_2^{TT} \mid \underline{X}_1 = \underline{x}_1, S_1 = R) = \;?$$
.

The problem is that $E(Y_2^{TT} \mid \underline{X}_1 = \underline{x}_1, S_1 = R)$ cannot be rearranged to obtain an expression that is a function of the observable outcome (Y_2) only, because Y_2^{TT} is independent of S_1 conditional on X_0, but not conditional on (X_0, X_1). Nor is it independent of X_1 conditional on (S_1, X_0), because X_1 contains part of the effect of S_1 on Y_2^{TT}. For $\theta_2^{TT,RR}(R)$ that did not matter, but for $\theta_2^{TT,RR}(RR)$ it does matter because X_1 determines the population of interest in the second period. If we are prepared to restrict this dependence, for example by assuming the full sequences are conditionally randomized in the first period, or that there are no time varying confounders, then all effects are identified (strong dynamic conditional independence assumption, S-DCIA; see internet appendix for details). However, the problem is now essential static and static evaluation methods for multiple treatments (each possible sequence constitutes one treatment) can be directly applied.

4.3   The relation of the assumptions to the empirical example

What do these assumptions imply for our empirical example? S-DCIA is only plausible if the time-varying confounding variables are not influenced by the evolvement of the treatment over time. Clearly, considering the types of confounders used and required in our application, this assumption is not plausible. However, in cases where the new information X_1 does influence outcomes as well as the choice of treatment in the next period, and this new information comes as a surprise (or at least is not influenced by the evolvement of the treatment history so far), then S-DCIA may be plausible (and powerful and convenient with respect to estimation).

In our application W-DCIA appears to be plausible. It seems likely that all important variables that influence selection are observed. Furthermore, although intermediate outcomes play a role as potentially confounding variables, it appears likely that the unemployed does not react prior to the participation decision, as long as the date of the referral to the programs and the starting date of the program (which is the variable that must be used at the beginning of the program, since referral is unobserved) are sufficiently close. Thus, it appears to be plausible that the exogeneity condition required for W-DCIA is fulfilled as well.

5   A note on estimation

An in-depth discussion of estimation and inference in complex dynamic treatment studies is beyond the scope of this paper for reasons of space. We confine ourselves to brief considerations about how to use the sample information to obtain consistent estimators for the causal effects. We discuss the estimation of the causal effects for the different populations in turn.[12]

---

[12]   For the sake of brevity, this paper concentrates on linking the observable random variables to the causal effects. An in-depth discussion of estimators would considerably extend an already long paper. Therefore, readers who are interested in the properties of actual estimation methods that may be implemented in this framework are

When interest is in DATE, the parameter that always features prominently in the epidemiological literature using this approach (see the already mentioned papers by Robins and coauthors), i.e. the average effect for the population $S_0 = 0$, W-DCIA is sufficient for identification. Using the previous arguments, we obtain the following relation between the expected observable and expected potential outcomes:

$$E(Y_2^{s_2^k}) = \underset{X_0}{E} \underset{(X_1 | X_0 = x_0, S_1 = s_1^k)}{E} E(Y_2 \mid \underline{X}_1 = \underline{x}_1, \underline{S}_2 = \underline{s}_2^k); \quad s_1^k, s_2^k \in \{0,1\}. \tag{3}$$

This is the so-called ***G-computation algorithm*** proposed by Robins (1986) and may be implemented as an inverse-probability-of-selection estimator (see Robins, Hernan, and Brumback, 2000).

The estimation problem is such that suitably modified matching or other nonparametric regression methods, popular estimators in the static causal model, can be used. In a first stage, a regression of Y_2 on $\underline{X}_1$ in the subsample of $\underline{S}_2 = \underline{s}_2^k$ is performed, obtaining $E(Y_2 \mid \underline{X}_1 = \underline{x}_1, \underline{S}_2 = \underline{s}_2^k)$. Within each stratum of X_0 in the subpopulation $S_1 = s_1^k$, this regression function is averaged according to the distribution of X_1 in each such stratum. These averages are functions of X_0 only. Finally, this function is averaged over the distribution of X_0 in the population ($S_0 = 0$) leading to a sequential matching estimator. If we are willing to parameterize the respective conditional distributions, the various parametric or semiparametric estimation methods proposed by Robins and co-authors are relevant and frequently used alternatives.

Next, consider the DATET for the population defined by treatment in period 1 and identified by W-DCIA. The estimand is given by the following equation (4):

---

referred to Lechner (2004). This paper contains an exact description of the estimator used in the application below.

$$E(Y_2^{s_2^k} \mid S_1 = s_1^j) = \underset{(X_0|S_1=s_1^j)}{E} \underset{(X_1|X_0=x_0,S_1=s_1^k)}{E} E(Y_2 \mid \underline{X}_1 = \underline{x}_1, \underline{S}_2 = \underline{s}_2^k); \quad s_1^k, s_2^k, s_1^j \in \{0,1\}. \tag{4}$$

The previous estimation principles of the sequential matching type apply here as well. However, in the first step, the averaging of $\underset{(X_1|X_0=x_0,S_1=s_1^k)}{E} E(Y_2 \mid \underline{X}_1 = \underline{x}_1, \underline{S}_2 = \underline{s}_2^k)$ is with respect to the distribution of X_0 in $S_1 = s_1^j$.

Finally, consider DATET for populations defined by the full treatment sequences, which are only identified by the S-DCIA's, as well as alternative expression for the other effects that are valid under S-DCIA only. The estimand has the following structure given by equation (5):

$$E(Y_2^{s_2^k} \mid \underline{S}_2 = \underline{s}_2^j) = \underset{X_1,X_0|\underline{S}_2=\underline{s}_2^j}{E} E(Y_2 \mid \underline{X}_1 = \underline{x}_1, \underline{S}_2 = \underline{s}_2^k);$$

$$E(Y_2^{s_2^k} \mid S_1 = s_1^j) = \underset{X_1,X_0|S_1=s_1^j}{E} E(Y_2 \mid \underline{X}_1 = \underline{x}_1, \underline{S}_2 = \underline{s}_2^k); \tag{5}$$

$$E(Y_2^{s_2^k}) = \underset{X_1,X_0}{E} E(Y_2 \mid \underline{X}_1 = \underline{x}_1, \underline{S}_2 = \underline{s}_2^k); \qquad s_1^k, s_2^k, s_1^j, s_2^j \in \{0,1\}.$$

Apparently, this estimation problem is the same as the typical static estimation problem based on the CIA. The only difference is that it is of the multiple treatment type, because four different sequences are involved (00, 01, 10, 11). For this framework several authors including Brodaty, Crépon, and Fougère (2001), Imbens (2000), and Lechner (2001, 2002) discuss issues of non- and semiparametric estimation. The estimators consistent under W-DCIA are consistent under S-DCIA as well.

Nonparametric estimation of the estimands defined in equations (4) to (5) is subject to the curse of dimensionality problem. Following Rosenbaum and Rubin (1983) and the extension to multiple treatments in Imbens (2000) and Lechner (2001) it is common practice in the static evaluation

literature to 'solve' this problem by first estimating the participation probability conditional on the confounders, and then use the estimated conditional participation probability (the propensity score) instead of the confounders as conditioning variables. Propensity score properties are available here as well. For equation (5), they are identical to the static case and can be derived in a straightforward way. Since estimation based on S-DCIA is essentially a static multiple treatment problem, the balancing scores provided by Imbens (2000) and Lechner (2001) are directly applicable. The case of equation (4) is more complicated, because more than one propensity score is required. It is analyzed in depth in the discussion paper version of this paper. It is shown that matching can be performed in the respective subsample given by the treatment status using all conditional transition probabilities for the state in the next period:

$$E(Y_2^{s_2^k} \mid S_1 = s_1^j) = \underset{(p_1(X_0)\mid S_1=s_1^j)}{E} \underset{(p_2(\underline{X}_1,s_1^k)\mid S_1=s_1^k,p_1(X_0))}{E} [F(Y_2 \mid \underline{S}_2 = \underline{s}_2^k, \underline{p}_2(\underline{X}_1,s_1^k))], \quad s_1^k,s_2^k,s_1^j \in \{0,1\} ;$$

$$\underline{p}_2(\underline{X}_1,s_1^k) = [p_1(X_0), \underline{p}_2(\underline{X}_1,s_1^k)], \quad p_1(x_0) = P(S_1 = s_1^k \mid X_0 = x_0); p_2(\underline{X}_1,s_1^k) = P(S_2 = s_2^k \mid S_1 = s_1^k, \underline{X}_1 = \underline{x}_1) .$$

Thus, instead of sequentially conditioning on all past and current values of *X* and the treatment status, conditioning is on the conditional transition probabilities.

6   Empirical results

In this section, we assume that W-DCIA is valid and use the propensity score matching estimator discussed in Lechner (2004), to estimate the causal effects of the treatment sequences. The estimator uses a sequential one-to-one matching algorithm based on propensity scores. The target populations are defined by the states in period one (*R*, *T*, or *U*). Individuals in those target popu-

lations who find no suitably close match in the subsequent comparisons in terms of the respective propensity scores are deleted (common support[13]).

Table 4 contains the results of the estimation. Column (1) shows the treatment sequences for which we estimate an effect. Since the effects may vary across the population in an arbitrary way, column (2) gives the population to which the respective effect relates to (*target population*). The next column contains the number of observations in the groups of the two treatment sequences as well as in the target population. The number of deleted observations when imposing the common support condition adjusts the latter number. Columns (4) and (6) give the means of the outcome variables (employed after two and four years) for these three different subsamples (treated, comparison, target). For each comparison, the first two lines in columns (8) and (10) contain the estimate for the respective counterfactual mean, whereas the third row contains the estimate of the causal effect, i.e. the parameter of interest. Therefore, comparing the first two lines in columns (4) and (6) to the first two lines in columns (8) and (10) gives us an indication of the amount of selection bias that the matching estimator is adjusting. Columns (5), (7), (9), and (11) present the standard errors for the estimators used.

-- Table 4 about here --

Note that the samples of participants are small since this application with four periods and three treatments per period is rather demanding, because it generates up to 81 different possible sequences and related subsamples. Nevertheless, only the first comparison of one year of training compared to one year of unemployment leads to insignificant effects. For all other comparisons, at least the effects after four years are large enough to be determined. Comparing retraining to

---

[13] Comparing sample sizes given in Tables 1 and 4 for the respective subpopulations shows how many observations

   are removed for the particular comparisons because of common support problems.

staying unemployed, we find that after two years there is no significant difference in the labor market outcomes. However, after four years we find about a 35%-point gain in employment chances for participating in retraining rather than remaining unemployed that may well be explained by the human capital effects of retraining and by stigma effects of remaining unemployed for a year. Note that this effect is about three times as large as the findings by LMW. However, they estimated the effect of starting retraining compared to remaining unemployed. For such a long program as retraining, the most important difference between the two approaches is that, here, the counterfactual state of unemployment requires remaining unemployed for one year. In the potential states compared by LMW, the unemployed as well as retrainees are allowed to accept job offers immediately after the start (which is close to a comparison of R against T in our notation). Therefore, the effect must be smaller because the unemployed will accept job offers while the retrainees are locked in their program. This argument applies here only after one year. This example shows that the dynamic treatment approach can be used to define a wealth of parameters that are of interest in policy analysis. Here, for the sake of brevity, we present only one specific type.

Finally, there is the comparison between training and retraining. It appears that training leads to much faster integration into the labor market than retraining, whereas after some years the extensive and expensive addition of human capital, which is the core concept of retraining, leads to considerably higher employment rates. Taking into account sampling uncertainty, these findings seem to hold for beginners of training as well as of retraining.

7   Conclusion

In this paper, we take up the topic of how to identify the effects of sequential interventions that first appeared in epidemiology 20 years ago with the seminal work by Robins (1986). Following

this work, we suggest approaching the problem of an econometric evaluation of dynamic program sequences from a potential outcome perspective. We discuss the identifying power of different stylized assumptions about the connection between the dynamic selection process and the potential outcomes of the different sequences of programs. These assumptions invoke different sorts of randomization which are compatible with different types of selection and outcome regimes. They are framed such that they need to be, and potentially can be, justified by sufficient knowledge about the selection and outcome process in conjunction with sufficiently rich individual panel data. Parametric forms are not involved. Participation in the sequences is cumulative in the sense that the decision concerning what program to participate in the next period depends on the outcomes of the part of the sequence that has already been completed.

These types of so-called dynamic treatment regimes are for example prototypical for the selection mechanism in many European and North American labor market programs. They are also an inherent problem in many economic policy analyses. However, due to selection on outcomes of past treatments, not all the parameters that are usually of interest in econometric evaluation studies are identified. We show that although several types of dynamic versions of the average treatment effects on the treated are not identified in this case, dynamic versions similar to the average treatment effect for some broader population are however identified.

The reevaluation of German training programs show that the dynamic treatment approach can be used to define a wealth of parameters that are of interest in policy analysis. Although our analysis confirms previous findings based on static potential outcome models in general, it clarifies that the magnitude of the effects may change dramatically depending on the precise type of dynamic counterfactual the policy maker has in mind. Thus, our empirical example shows that these methods can be a useful tool in applied work. However, it shows as well that there is a price to pay

for this new information, because very large data sets are required to be able to address the most interesting comparisons with sufficient precision.

As mentioned above, some parts of this paper are closely related to the work by Robins and co-authors in epidemiology and statistics, in particular the idea of specifying the weak dynamic conditional assumption is terms of sequential conditional randomization conditional on the observed history of the various stochastic processes. Our contribution for this part is (only) that of translating the notation and language used in his papers to a language common in the econometric treatment evaluation literature and making some of the underlying behavioral assumptions explicit. Furthermore, we extend his approach in a dimension important for evaluation studies: we discuss the identification of parameters other than the average treatment effect for the population and show that different assumptions about the nature of the dynamic selection problem are required for different effects that are usually of interest in applied studies. The differences concerning the behavioral implications of the different assumptions are substantial. In fact, identification of average treatment effects requires much less restrictive assumptions than the identification of average treatment effects on the treated, with a couple of interesting intermediate cases.

References

Abbring, J H (2003) Dynamic Econometric Program Evaluation. IZA, Discussion Paper, 804.

Abbring, J H, Heckman J J (2008) Dynamic Policy Analysis. In: L. Matyas and P. Sevestre (eds.), *The Econometrics of Panel Data*, 3rd ed., Dordrecht: Kluwer.

Angrist J D (1998) Estimating Labor Market Impact of Voluntary Military Service Using Social Security Data. *Econometrica*, 66: 249-288.

Angrist J D, Imbens G W, Rubin D B (1996) Identification of Causal Effects Using Instrumental Variables, *Journal of the American Statistical Association*, 91: 444-472.

Behrman J, Sengupta P, Todd P (2005) Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Mexico. *Economic Development and Cultural Change,* 54/1: 237-275.

Behrman J, Cheng Y, Todd P (2004) Evaluating Preschool Programs when Length of Exposure to the Program varies: A Nonparametric Approach. *Review of Economics and Statistics*, 86: 108-132.

Bergemann A, Fitzenberger B, Speckesser S (2004) Evaluating the Dynamic Employment Effects of Training Programs in East Germany using Conditional Difference-in-Differences. ZEW, Discussion Paper 04-41.

Brodaty T, Crépon B, Fougère D (2001) Using matching estimators to evaluate alternative youth employment programmes: Evidence from France, 1986-1988. In: M. Lechner and F. Pfeiffer (eds.), *Econometric Evaluation of Labour Market Policies*, Heidelberg: Physica, 85-123.

Cameron C A, Trivedi P K (2005) *Microeconometrics*. Chapter 25, Cambridge: CUP.

Chamberlain G. (1982) The General Equivalence of Granger and Sims Causality. *Econometrica*, 50: 569-581.

Chamberlain G. (1987) Asymptotic Efficiency in Estimation with Conditional Moment Restrictions. *Journal of Econometrics*, 34: 305-334.

Chamberlain G (1992) Comment: Sequential Moment Restrictions in Panel Data. *Journal of Business & Economic Statistics*, 10: 20-26.

Cochran W G, Chambers S P (1965) The Planning of Observational Studies of Human Population. *Journal of the Royal Statistical Society, Series A*, 128: 234-266.

Crépon B, Kramarz F (2002) Employed 40 hours or Not-Employed 39: Lessons from the 1982 Workweek Reduction in France. *Journal of Political Economy*, 110: 1355-1389.

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

Ding W, Lehrer S F (2003) Estimating Dynamic Treatment Effects from Project STAR. Mimeo.

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

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

Heckman J J, Vytlacil E (1999) Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects. *Proceedings of the National Academy of Science*, April 13, 1999.

Heckman J J, Vytlacil E (2006) Econometric Evaluation of Social Programs. Forthcoming in J. J. Heckman and E. Leamer (eds.), *Handbook of Econometrics,* Vol. VI, North-Holland: Amsterdam.

Heckman J J, Robb R (1985) Alternative Methods of Evaluating the Impact of Interventions. In: J.J. Heckman and B. Singer (eds.), *Longitudinal Analysis of Labour Market Data*, New York: Cambridge University Press, 156-245.

Heckman J J, S Navarro (2005) Dynamic Discrete Choice and Dynamic Treatment Effects. IZA discussion paper 1790.

Heckman J J, LaLonde R J, Smith J A (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: 4-29.

Imbens G W, Angrist J D (1994) Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62, 446-475.

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.), *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. *The Review of Economics and Statistics,* 84: 205-220.

Lechner M (2004) Sequential Matching Estimation of Dynamic Causal Models. University of St. Gallen, Discussion paper, 2004-06. Revised and shortened version forthcoming in *Journal of Business & Economic Statistics*.

Lechner M (2008) A Note on Endogenous Control Variables in Evaluation Studies. *Statistics & Probability Letters*, 78: 190-195.

Lechner M (2008) Matching Estimation of Dynamic Treatment Models: Some Practical Issues. In D. Millimet, J. Smith, and E. Vytlacil (eds.), *Advances in Econometrics*, *Volume 21, Modelling and Evaluating Treatment Effects in Econometrics*, 289-333.

Lechner M, Miquel R, Wunsch C (2005) Long - Run Effects of Public Sector - Sponsor Training in West Germany. IAB, Discussion Paper, 3/2005.

Manski C F (2004) Social learning From Private Experiences: The Dynamics of the Selection Problem. *Journal of Economics Studie*s, 71: 443-558.

Miquel R (2002) Identification of Dynamic Treatments Effects by Instrumental Variables. University of St. Gallen, Discussion paper, 2002-11.

Miquel R (2003) Identification of Effects of Dynamic Treatments with a Difference-in-Differences Approach. University of St. Gallen, Discussion paper, 2003-06.

Murphy S A (2003) Optimal Dynamic Treatment Regimes. *Journal of the Royal Statistical Society, Series B*, 65: 331-366.

Neyman J (1923) On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9, translated in *Statistical Science* (with discussion), 1990, 5: 465-480.

Robins J M (1999) Association, Causation, and Marginal Structural Models. *Synthese,* 121: 151-179.

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.

Robins J M (1989) The Analysis of Randomized and Nonrandomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies. In Sechrest, L., H. Freeman, A. Mulley (eds.), *Health Service Research Methodology: A Focus on Aids*, 113-159, Washington, D.C., Public Health Service, National Center for Health Services Research.

Robins J M (1997) Causal Inference from Complex Longitudinal Data. Latent Variable Modelling and Applications to Causality. In M. Berkane, (ed.), *Lecture Notes in Statistics* (120), New York: Springer, 69-117.

Robins J M (1998) Marginal Structural Models. *1997 Proceedings of the American Statistical Association*, 1-10.

Robins J M, Hernan M A, Brumback B (2000) Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology*, 11 (5): 550-560.

Robins J M, Greenland S, Hu F (1999a) Estimation of the Causal Effect of a Time-varying Exposure on the Marginal Mean of a Repeated Binary Outcome. *Journal of the American Statistical Association*, 94: 687-700.

Robins J M, Greenland S, Hu F (1999b) Estimation of the Causal Effect of a Time-varying Exposure on the Marginal Mean of a Repeated Binary Outcome: Rejoinder. *Journal of the American Statistical Association*, 94: 708-712.

Rosenbaum P R (1984) The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment. *Journal of the Royal Statistical Society*, *Series A*, 147: 656-666.

Rosenbaum P R, Rubin D B (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70: 41-50.

Roy A D (1951) Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, 3: 135-146.

Rubin D B (1974) Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66: 688-701.

Rubin D B (1977) Assignment to a Treatment Group on the Basis of a Covariate. *Journal of Educational Statistics*, 2: 1-26.

Rubin D B (2004) Direct and Indirect Causal Effects via Potential Outcomes. *Scandinavian Journal of Statistics*, 31: 161-170.

Rubin D B (2005) Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Society*, 100: 322-331.

Sianesi B (2004) An evaluation of the Swedish System of active labour market programmes in the 1990s. *Review of Economics and Statistics*, 86: 133-155.

Taber C R (2000) Semiparametric identification and heterogeneity in discrete choice dynamic programming models. *Journal of Econometrics*, 96: 201-229.

Wilks S S (1932) On the distribution of statistics in samples from a normal population of two variables with matched sampling of one variable. *Metron*, 9: 87-126.

Appendix: Proof of theorem

A.1   Useful lemmas

Before providing the proofs in Appendix A.2, we consider two lemmas that connect potential outcomes and treatment effects with different lengths of treatment sequence and conditioning sequences. These relations are interesting per se, but will be particularly helpful in simplifying the identification proofs.

Lemma A.1:   Connection of treatment effects defined for conditioning sets of different lengths

$$\theta_t^{s_1^k:s_1^l}(s_1^j) = \theta_t^{s_1^k:s_1^l}(s_1^j\,1)P[\underline{S}_2 = (s_1^j\,1)\,|\,S_1 = s_1^j] + \theta_t^{s_1^k:s_1^l}(s_1^j\,0)P[\underline{S}_2 = (s_1^j\,0)\,|\,S_1 = s_1^j]\;.$$

The proof is direct by applying the definitions of the treatment effect. Because treatments are observable,

$P[\underline{S}_2 = (s_1^j 1)\,|\,S_1 = s_1^j]$ and $P[\underline{S}_2 = (s_1^j 0)\,|\,S_1 = s_1^j]$ are identified.

Lemma A.2:   Connection of treatment effects defined for treatments of different lengths

$$\theta_t^{s_1^k:s_1^l}(s_1^j) = \theta_t^{(s_1^k 1);(s_1^l 1)}(s_1^j 1)P[S_2 = 1\,|\,S_1 = s_1^j] + \theta_t^{(s_1^k 0);(s_1^l 0)}(s_1^j 0)P[S_2 = 0\,|\,S_1 = s_1^j]\;.$$

For the proof of this lemma, consider the following relations:

$$E(Y_t^{s_1^k}\,|\,S_1 = s_1^j) = E[S_2 Y_t^{(s_1^k 1)}\,|\,S_1 = s_1^j] + E[(1-S_2)Y_t^{(s_1^k 0)}\,|\,S_1 = s_1^j]\ =$$

$$=\ E[S_2 Y_t^{(s_1^k 1)}\,|\,S_2 = 1, S_1 = s_1^j]P(S_2 = 1\,|\,S_1 = s_1^j) + E[(1-S_2)Y_t^{(s_1^k 0)}\,|\,S_2 = 0, S_1 = s_1^j]P(S_2 = 0\,|\,S_1 = s_1^j)$$

$$=\ E[Y_t^{(s_1^k 1)}\,|\,S_2 = 1, S_1 = s_1^j]P(S_2 = 1\,|\,S_1 = s_1^j) + E[Y_t^{(s_1^k 0)}\,|\,S_2 = 0, S_1 = s_1^j]P(S_2 = 0\,|\,S_1 = s_1^j)\;. \qquad (6)$$

Using the result of equations (6) we obtain the desired results for the connection of $\theta_t^{s_1^k:s_1^l}(s_1^j)$ and $\theta_t^{s_2^k:s_2^l}(\underline{s}_2^j)$.

q.e.d.

A.2 Proof of Theorem 1

First note that ASSUMPTION 1 (W-DCIA) implies the following restrictions:[14]

$$F(Y_2^{s_2^k} \mid S_1 = s_1^j, X_0 = x_0) = F(Y_2^{s_2^k} \mid S_1 = s_1^k, X_0 = x_0) \; ; \tag{7}$$

$$F(Y_2^{s_2^k} \mid \underline{S}_2 = \underline{s}_2^j, \underline{X}_1 = \underline{x}_1) = F(Y_2^{s_2^k} \mid S_1 = s_1^j, S_2 = s_2^k, \underline{X}_1 = \underline{x}_1) \; . \tag{8}$$

We must show that $F(Y_2^{s_2^k} \mid S_1 = s_1^j)$ is identified. Let us consider the starting point for the proof in detail. We relate $F(Y_2^{s_2^k} \mid S_1 = s_1^j)$ to some function of the observable outcomes by sequentially applying equations (7) and (8) to conditional expectation versions of $F(Y_2^{s_2^k} \mid S_1 = s_1^j)$:

$$F(Y_2^{s_2^k} \mid S_1 = s_1^j) = \underset{X_0 \mid S_1 = s_1^j}{E}[F(Y_2^{s_2^k} \mid S_1 = s_1^k, X_0 = x_0)] \; ;$$

$$F(Y_2^{s_2^k} \mid S_1 = s_1^k, X_0 = x_0) = \underset{X_1 \mid S_1 = s_1^k, X_0 = x_0}{E}[F(Y_2^{s_2^k} \mid S_1 = s_1^k, \underline{X}_1 = \underline{x}_1)]$$

$$= \underset{X_1 \mid S_1 = s_1^k, X_0 = x_0}{E}[F(Y_2^{s_2^k} \mid \underline{S}_2 = \underline{s}_2^k, \underline{X}_1 = \underline{x}_1)] \; .$$

Thus, we obtain the following term for the counterfactual distribution:

$$F(Y_2^{s_2^k} \mid S_1 = s_1^j) = \underset{X_0 \mid S_1 = s_1^j}{E} \underset{X_1 \mid S_1 = s_1^k, X_0 = x_0}{E}[F(Y_2^{s_2^k} \mid \underline{S}_2 = \underline{s}_2^k, \underline{X}_1 = \underline{x}_1)] = \underset{X_0 \mid S_1 = s_1^j}{E} \underset{X_1 \mid S_1 = s_1^k, X_0 = x_0}{E}[F(Y_2 \mid \underline{S}_2 = \underline{s}_2^k, \underline{X}_1 = \underline{x}_1)] \; .$$

<div align="right">q.e.d.</div>

---

[14] Depending on whether the elements of the vector of random variables A are continuous or discrete or both, $F(A \mid B = b)$ denotes the distribution function, the probability mass function or a mixture of both, conditional on the event that the random variable B equals a fixed value b.

Figures and Tables

Figure 1: Notation and time line of the dynamic causal model

Period t=0   t=1   t=1   t=2   t=2

$X_1^0(Y_1^0)$
$S_1 = 0$ $X_1^{01}(Y_1^{01})$
$X_1^{00}(Y_1^{00})$

$S_2 = (0,0) \longrightarrow Y_2^{00}$
$S_2 = (0,1) \longrightarrow Y_2^{01}$

$\theta_2^{01;00}(.)$
$\theta_2^{10;00}(.)$
$\theta_2^{10;01}(.)$
$\theta_2^{11;00}(.)$

$S_0 = 0$

$X_1^1(Y_1^1)$
$S_1 = 1$ $X_1^{11}(Y_1^{11})$
$X_1^{10}(Y_1^{10})$

$S_2 = (1,0) \longrightarrow Y_2^{10}$
$S_2 = (1,1) \longrightarrow Y_2^{11}$

$\theta_2^{11;01}(.)$
$\theta_2^{11;10}(.)$

$S_0 \longrightarrow X_0 \longrightarrow S_1 \longrightarrow X_1(Y_1) \longrightarrow S_2 \longrightarrow Y_2$
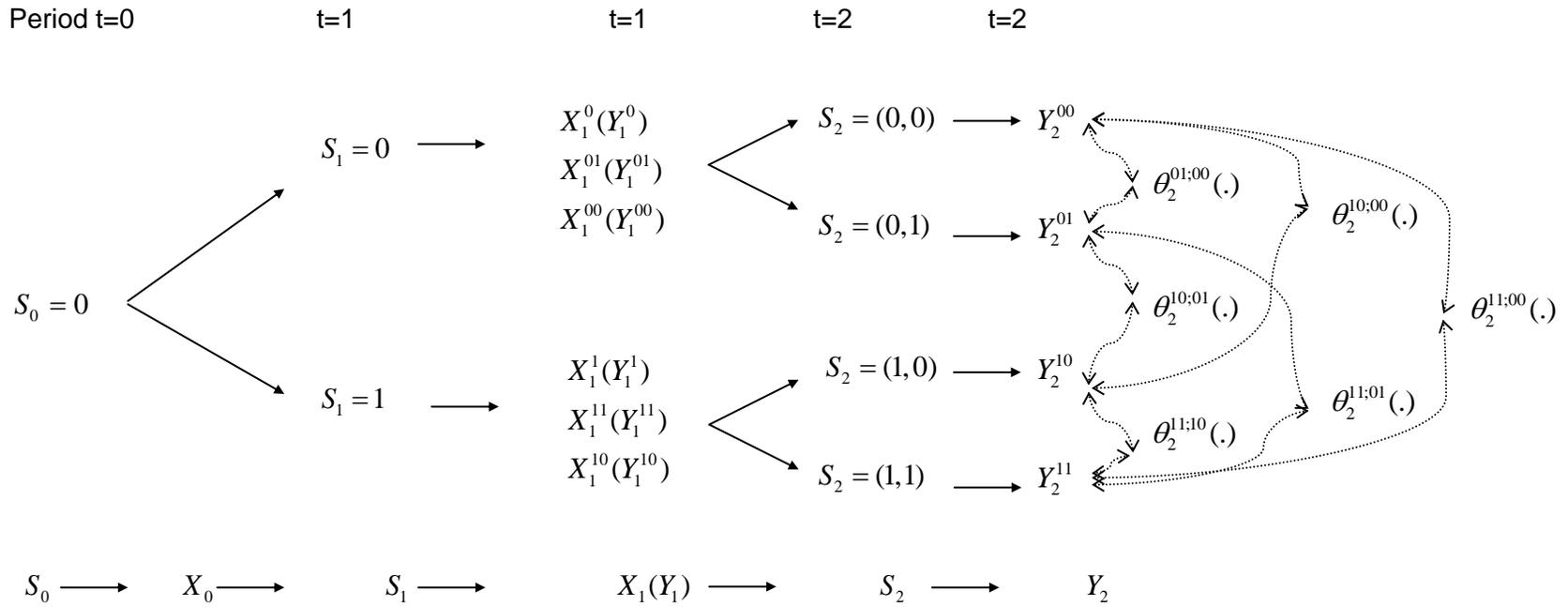
Table 1: Descriptive statistics for selected variables

Variables		Means / shares in % in subsamples										
Subpopulations	U	T	R	UU	UUU	UUUU	TT	TTT	TTTT	RR	RRR	RRRR
Women	41	43	42	42	41	40	44	40	38	43	44	45
Age	38	33	30	40	42	43	33	33	33	30	30	30
Nationality: German	83	85	82	83	83	82	85	85	85	84	84	84
Nationality: Western European	9	5	5	10	10	11	5	5	7	4	3	4
At least one child	32	34	36	33	33	32	34	32	29	37	38	38
UE benefits in the month before the beginning of the period under consideration												
before quarter 1	90	89	86	89	89	89	90	90	91	87	86	86
before quarter 2	80	5	2	87	87	87	1	1	0	0	0	0
before quarter 3	64	19	5	79	82	82	13	1	1	2	0	0
before quarter 4	56	29	6	66	75	77	29	22	0	4	1	0
Last monthly earnings (in euros; mean of positive earnings)												
before quarter 1	1902	1605	1524	1874	1978	2069	1612	1663	1753	1532	1529	1377
before quarter 2	1846	1604	1517	1875	1978	2070	1611	1662	1756	1524	1520	1529
before quarter 3	1789	1606	1510	1875	1978	2070	1607	1662	1756	1524	1520	1529
before quarter 4	1792	1639	1512	1873	1976	2069	1646	1684	1756	1527	1519	1529
Remaining unemployment (UE) benefits claim in months												
before quarter 1	9.5	4.0	3.1	10.6	11.5	12.3	4.5	4.7	4.5	3.2	3.2	3.2
before quarter 2	6.7	6.3	5.8	8.1	9.0	9.8	7.0	6.9	7.2	5.9	5.8	5.8
before quarter 3	5.3	5.7	5.8	5.8	6.8	7.6	6.1	7.1	7.3	6.0	6.0	6.0
before quarter 4	4.4	5.2	6.1	4.4	5.0	5.8	5.6	6.2	7.6	6.4	6.4	6.5
Employment (1: employed)												
24 month after the first unemployment spell	32	50	29	26	21	17	48	46	42	27	24	24
48 month after the first unemployment spell	32	52	66	27	23	19	53	56	58	69	70	71
Sample Size	27332	494	174	19677	15417	12484	372	231	120	164	149	143

Note: See LMW and Table WWW.A.1 in the internet appendix for more details on the data.

Table 2: Estimated coefficients of sequential binary probit models for choice of state in the beginning of the different periods (selected variables only)

Variables	U vs. T	R vs. T	U vs. R	UU if U	UUU if UU	UUUU if UUU	TT if T	TTT if TT	TTTT if TTT	RR if R	RRR if RR	RRRR if RRR
	coeff	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.
Women	**0.11**	-0.05	*0.13*	0.02	-0.02	**-0.10***	0.22	-0.23	0.23	0.55	0.26	0.24
Age/10	**-0.99***	-0.06	**-1.10***	-0.14	0.02	**0.55***	1.16	0.33	2.83			
(Age/10)^2	**0.14***	-0.04	**0.19***	**0.03**	0.00	**-0.06***	-0.18	-0.04	-0.34			
Younger than 26 years	**-0.19**	**-0.58**	0.27	**-0.13***	*0.10*	0.08	-0.30	-0.26	*1.05*			
Nationality: German	-0.04	-0.01	-0.14	**-0.11***	**-0.13***	**-0.14***	0.11	0.11	-0.48			
Nationality: Western European	0.14	-0.19	0.38	-0.00	-0.06	*0.15*						
At least one child	0.03	0.11	0.02	**0.11***	0.02	0.01	-0.01	*0.34*	**-0.60**			
Position in last job (reference category: skilled worker, Master craftsman)												
Salaried Employee	**-0.22***	0.12	**-0.21**	**0.07**	-0.00	-0.01						
Part-time worker	**-0.25***	-0.27	0.03	-0.00	-0.01	-0.02						
Unskilled worker	0.01	**0.42**	*-0.18*	**0.10***	*0.06*	**0.11***						
Unemployment and employment status before the beginning of the period under consideration: (reference categories: out-of-labor, missing)												
Unemployed in the 6th. month before beg.	0.05	0.37	-0.21	*0.11*	-0.03	-0.09	-0.34	0.09	-0.75			
12th. month before beg.	-0.08	0.00	-0.07	**-0.11**	-0.10	0.10	**1.00***	-0.21	**-1.88**			
24th. month before beg.	**-0.17**	*-0.39*	0.15	**-0.11**	0.01	**0.15**	0.10	0.36	-0.90			
UE benefits in the month before the beginning of the period under consideration (reference category: UE assistance)												
before quarter 1	**-0.73***	-0.06	**-0.67***	**-0.81***	-0.07	-0.05	-0.41	0.53	-0.26			
before quarter 2				**0.92***	**-0.26***	*0.19*						
before quarter 3					**0.25***	-0.12						
before quarter 4						**-0.25***						
Remaining unemployment (UE) benefits claim in months												
before quarter 1	**0.05**	-0.03	0.04	**0.03***	**0.12***	-0.06	-0.11	**-0.31**	**0.48***	*0.10*		
before quarter 2				**0.05***	0.04	**-0.08**	**0.23***	-0.42	**-0.88***		-0.02	
before quarter 3					**-0.15***	**0.22***		**0.67***				0.06
before quarter 4						-0.00			**0.49**			

Note: Empty cell means that the respective variable is excluded from this equation. **Bold** numbers indicate significance at the 5% level, numbers in *italics* relate to the 10 % level and * to the 1 % level. The following variables are included in the estimates but omitted from this table: Constant term; education; last monthly earnings and missing information about earnings; various variables capturing future claim of unemployment; industrial sector, previous occupation before defining unemployment spell; firm size of employer; regional information; various historical un-/ out-of/employment information before the "first unemployment period". The appendix contains a table with all variables included. The detailed results including asymptotic std. errors can be found in the internet appendix. All the results for unemployment are from the comparison with training.

Table 3: Summary of notation and definitions

Symbol	Meaning	Timing within period
$t = 0, 1, 2$	Time periods	--
$S = (0, S_1, S_2)$	RV: treatment	Beginning
$\underline{s}_1 = (0, s_1)$, $\underline{s}_2 = (0, s_1, s_2)$	Specific sequence of treatments until period 1 or 2	Beginning
$s_t \in \{0,1\}$	2 exclusive treatments in each period	Beginning
$Y^{\underline{s}_t} = (Y_1^{\underline{s}_t}, Y_2^{\underline{s}_t})$	RV: potential outcomes	End
$Y = (Y_1, Y_2)$	RV: observable outcomes	End
$X^{\underline{s}_t} = (X_0^{\underline{s}_t}, X_1^{\underline{s}_t})$	RV: potential confounders	End
$X = (X_0, X_1)$	RV: observable confounders	End
$\theta_t^{\underline{s}_\tau^k ; \underline{s}_\tau^l}(\underline{s}_{\tilde{\tau}}^j)$	mean causal effect of $\underline{s}_\tau^k$ compared to $\underline{s}_\tau^l$ for those participating in $\underline{s}_{\tilde{\tau}}^j$	End

RV: Random variable.

Table 4: Results of the dynamic matching estimation

Sequences $\underline{s}_4^1$ $\underline{s}_4^0$	Target pop. S_1	Sample size $N_{\underline{s}_4^1}$ $N_{\underline{s}_4^0}$ N_{s_1}	$E(Y_t \mid \underline{S}_4 = \underline{s}_4^1)$ $E(Y_t \mid \underline{S}_4 = \underline{s}_4^0)$ $E(Y_t \mid S_1 = s_1)$				$E(Y_t^{\underline{s}_4^1} \mid S_1 = s_1)$ $E(Y_t^{\underline{s}_4^0} \mid S_1 = s_1)$ $\theta_t^{\underline{s}_4^0, \underline{s}_4^1}(s_1)$			
			ET24 (%)	(std)	ET48 (%)	(std)	ET24 (%)	(std)	ET48 (%)	(std)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
TTTT		120	41.7	(4.5)	58.3	(4.5)	28.5	(11.6)	35.7	(11)
UUUU		12484	16.5	(0.3)	19.5	(0.4)	22.9	(0.8)	27.4	(0.9
	U	16600	33.9	(0.4)	34.3	(0.4)	5.6	(11.7)	8.2	(11)
TTTT		120	41.7	(4.5)	58.3	(4.5)	36.5	(6.7)	48.8	(6.7)
UUUU		12484	16.5	(0.3)	19.5	(0.4)	28.1	(3.2)	38.3	(3.5)
	T	334	51.2	(2.7)	55.3	(2.7)	8.4	(7.4)	10.5	(7.5)
RRRR		143	24.5	(3.6)	71.3	(3.8)	25.6	(6.9)	66.8	(7.3)
UUUU		12484	16.5	(0.3)	19.5	(0.4)	25.7	(1.0)	31.8	(1.1)
	U	19088	40.1	(0.4)	40.7	(0.4)	-0.2	(7.0)	**35.0**	(7.4)
RRRR		143	24.5	(3.6)	71.3	(3.8)	25.3	(3.9)	71.3	(4.1)
UUUU		12484	16.5	(0.3)	19.5	(0.4)	25.9	(3.8)	40.2	(4.3)
	R	174	28.7	(3.4)	66.1	(3.6)	-0.6	(5.5)	**31.0**	(6.0)
RRRR		143	24.5	(3.6)	71.3	(3.8)	13.9	(7.0)	73.2	(7.3)
TTTT		120	41.7	(4.5)	58.3	(4.5)	39.7	(6.3)	54.5	(6.3)
	T	325	52.6	(2.8)	56.0	(2.8)	**-25.9**	(9.5)	*18.8*	(9.7)
RRRR		143	24.5	(3.6)	71.3	(3.8)	24.6	(4.7)	72.1	(4.9)
TTTT		120	41.7	(4.5)	58.3	(4.5)	43.4	(9.8)	54.1	(9.8)
	R	122	29.5	(4.1)	63.9	(4.4)	-18.9	(10.9)	*18.0*	(11)

Note: ET24 = employed the 24th month after the first month of unemployment, ET48 = employed the 48th month after the first month of unemployment. **Bold :** Effect is significant at 1% level. *Italics*: Effect significant at 10% level. N_{s_1} is the sample size after imposing common support, $N_{\underline{s}_4^1}$ and $N_{\underline{s}_4^0}$ are the sample sizes before imposing common support.