Universität St.Gallen

# Identification of Effects of Dynamic Treatments with a Difference-in-Differences Approach

Ruth Miquel

February 2003 Discussion paper 2003-06

Department of Economics University of St. Gallen

| | |
|---|---|
| Editor: | Prof. Jörg Baumberger<br>University of St. Gallen<br>Department of Economics<br>Bodanstr. 1<br>CH-9000 St. Gallen<br>Phone ++41 71 224 22 41<br>Fax ++41 71 224 28 85<br>Email joerg.baumberger@unisg.ch |
| Publisher: | Forschungsgemeinschaft für Nationalökonomie<br>an der Universität St. Gallen<br>Dufourstrasse 48<br>CH-9000 St. Gallen<br>Phone ++41 71 224 23 00<br>Fax ++41 71 224 26 46 |
| Electronic Publication: | www.fgn.unisg.ch/public/public.htm |

Identification of Effects of Dynamic Treatments

with a Difference-in-Differences Approach[1]

Ruth Miquel

Author's address:

Ruth Miquel
SIAW
Dufourstrasse 48
9000 St. Gallen
Tel. ++41 71 224 27 67
Fax ++41 71 224 22 98
Email ruth.miquel@unisg.ch
Website www.siaw.unisg.ch\lechner\miquel

[1] I am thankful to Michael Lechner, Markus Frölich and Conny Wunsch for helpful discussions. All remaining errors are mine.

**Abstract**

This paper examines the power of a conditional difference-in-differences approach to nonparametrically identify the causal effects of sequences of interventions. In the classical difference-in-differences case, a period previous to the implementation of the intervention is used as a comparison period to get rid of common trends for treated and nontreated. When sequences of programs are studied, two different utilizations of the available information are possible to construct the comparison period. The first one only uses the information available in the period previous to the first possible intervention. The second one uses the information sequentially. Here, we use the information available between the interventions in addition to the period previous to the first intervention. Furthermore, we investigate if we still can achieve identification when participation to the sequences is decided sequentially on the intermediate results. Identification can be obtained in each case for all parameters of interest.

**Keywords**

dynamictreatment effects, sequential difference-in-differences, sequences of programs

**JEL Classification**

C40

## 1) Introduction

Often, the question of interest when investigating the effect of some policy intervention is to evaluate the consequences of the policy on the units or populations subject to this policy. We want to find out if those populations are better off after they have been subject to the policy. To answer this question, we need to compare two outcomes: the outcome these populations would have if the policy would not have been implemented and the outcome they would have if the policy would have been implemented. It is well known that such analysis suffers from a missing data problem because we cannot simultaneously observe both outcomes for the same population. If the population is subject to the policy, we observe only the second outcome. Conversely, if the population is not subject to the policy, we observe only the first outcome. With non-experimental data, this missing data problem is compensated by making certain hypothesis. These assumptions allow us to combine the observable information in a way that enables the estimation of the effects of interest (Heckman and Robb, 1985 and Heckman, Lalonde and Smith, 1999). For example, if we want to evaluate the effect of a vocational training program on the employability of the participants in the program – the effect of the treatment on the treated – we need to compute the potential outcome of the participants if they would not have been treated. Under the *Conditional Independence Assumption* (CIA) (Rubin, 1977), the selection of the individuals is supposed to be made on observable characteristics and thus, conditioning on those variables, the potential outcomes and the participation status are independent. Therefore, the conditional average of the outcome for the nontreated is used instead of the conditional average of the non-observable potential outcome for the treated. However, the validity of this assumption is questionable if the data is not rich enough. The *difference-in-differences* (DiD) approach, another widely used approach, (Meyer, Viscusi and Durbin, 1995, Blundell, Duncan and Meghir, 1998, Angrist and Krueger, 2000, for a survey, Hujer, Caliendo and Radic, 2002) allows selection on unobservable characteristics and is less data consuming (it does not require to observe all variables simultaneously influencing the participation decision and the potential outcomes). Using the terminology of treatment, treated and nontreated, the "classical" difference-in-differences approach assumes that over time, the no-treatment average outcome for the treated follows the same path as the no-treatment average outcome for the nontreated. Consequently, common trends are eliminated by differencing the before-after change in the outcome of the treated and the before-after change in the outcome of the nontreated. Two criticisms have been formulated against this approach. Ashenfelter (1978) regards it as inappropriate for investigating the effect of a treatment or a program on earnings, because the earnings of participants use to decline in the last period prior to the beginning of the program. Depending on the permanency of this decline, the estimated effect could be upward biased. Moreover, in this conventional difference-in-differences approach the effects are mostly estimated parametrically. Donald and Lang (2001), and Bertrand, Duflo and Mullainathan (2001) show that even if the estimators are unbiased, the standard errors are often not correctly estimated leading to wrong conclusions on the impact of the interventions. Heckman and Smith (1999) investigate the problem mentioned by Ashenfelter and find that to resolve it, it is important to control for the determinants of the participation decision, principally to control for the labor force dynamics.

Heckman, Ichimura and Todd (1997) and Heckman, Ichimura, Smith and Todd (1998) propose a conditional difference-in-differences estimator based on a previous proposal made by Heckman and Robb (1985, 1986). This nonparametric approach allows the introduction of covariates in a more flexible way and does not suffer if a misspecified parametric form is employed. Moreover, the assumption of the difference-in-differences estimator may be more credible when it is expressed for different populations defined by covariates. This nonparametric approach has not yet been extensively used in empirical studies compared to the conventional difference-in-differences approach. Nonetheless, several recent studies apply this method and modify it by using different estimation approaches (Abadie, 2001, Blundell, Dias, Meghir and Van Reenen, 2001, Eichler and Lechner, 2002). Athey and Imbens (2002) extend this approach further. They are able to nonparametrically identify the entire counterfactual distribution of the no-treatment outcome for the treated in the period after the program takes place. They also derive the necessary assumptions for the identification of the effect of the treatment on the nontreated.

Nevertheless, all these papers strongly simplify the problem of evaluation because they do not take into account the dynamic aspect of this problem. For example, when evaluating active labor market policies, unemployed are usually assigned to a sequence of programs. The number of programs participated in is typically influenced by the success achieved through the first, the second, the ... program of the sequence: if the persons are still unemployed after the first program, they will have to participate in a second one, and then, if they are still unemployed they have to participate in a third one, and so on. The dependence on the intermediate outcomes of the treatment complicates the identification of causal effects. This complexity has been combined to some transformed CIA assumptions by Robins (1986) and Lechner and Miquel (2001). It has been combined with instrumental variables type assumptions by Miquel (2002) to identify some effects of sequences of programs.

What happens if the CIAs presented in Robins and Lechner and Miquel are not credible with the data available? How can we estimate the effect of sequences of programs? Is it possible to identify such effects? In this paper, we provide some answers to these questions by combining the dynamic framework with a difference-in-differences approach to inspect the identification of effects of sequences on different populations characterized by their participation in different periods. We examine two possible ways to use the available information. Firstly, we study which effect is identified if only the period prior to the first possible participation can be used as base period to compute the time trend for the outcomes. The results, as well as the assumptions, mimic closely those of the static case (where only one participation is possible).

Secondly, we introduce some sequentiality and use two base periods. In addition to the information of the period previous to the first participation, we utilize the information (the observed outcomes) available after the first participation (between the interventions) to gain information on the outcomes we are interested in but that we cannot observe. In both cases, we discuss what happens to the identification if the decision to participate in a program in the future is directly influenced by the outcomes obtained after previous participations, which are endogenous (because they can be affected by the sequences under study). In both cases, using the same kind of assumptions as in the static framework, all effects we are interested in are identified.

The next section presents the framework and the notation used in the paper, as well as the effects we are interested in. Section 3 restates the results of the static case using our multi-period framework. Section 4 discusses the major point of the paper, the identification of the effect of sequences of programs. Finally, section 5 presents some concluding remarks.

## 2) Notation

As a basis we use the causal model of potential outcomes proposed by Roy (1951) and Rubin (1974) and extended to a dynamic framework by Robins (1986, 1989, 1997, 1999) and Lechner and Miquel (2001). It is a model of causal relationships. We consider two frameworks, the static framework, where one possible participation in a program (or one intervention) is considered and the dynamic one, where several possible participations (or interventions) are considered. They investigate the identification of different effects under the conditional independence assumption. Here, we consider $T$ periods, indexed by $t$ or $t'$. We are interested in the possible participation in a program or possible application of a treatment in only two periods, period 1 and period 2. To reduce the complexity of the notation needed to differentiate between the time periods, we assume that period 2 directly follows period 1. The generalization to the case where more than two participations are possible is straightforward.[1] In the periods prior to period 1, everybody has the same history (of participations in a program or of nonparticipations). All the effects and the conditions we will meet are conditioned on this history. The treatment received by a member of the population is described by a vector of random variables $\underline{S}_2 = (\underline{S}_0, S_1, S_2)$. $\underline{S}_0$ denotes the history of participation up to time 0 and will always take the value $\underline{s}_0$.[2] Since everything that we do is conditioned on this value, it will not appear further in the notation, but we have to remember that it is part of the conditioning set. To simplify the presentation we consider that in each period only one program is available. There is no additional difficulty when several programs are available (all the proofs of the paper given in the appendix are for this general case). Thus, $S_t$ can take two values. A particular value is denoted by $s_t \in \{0,1\}$. To differentiate different sequences of programs they may be indexed, e.g. $\underline{s}_2^k$. In period 1 an individual[3] can be observed in one of two programs. In period 2 this same individual can be observed in one of four sequences of programs, $(0,0),(1,0),(0,1),(1,1)$. Thus, we consider 2 states defined by treatment status in period 1, and 4 states defined by treatment status in period 1 and period 2 together. Hence, we can evaluate the effects of sequences of treatments of different lengths.

For each individual several outcomes, that is the variable we are interested in are defined: one outcome in each state of the world defined by the sequences of programs. Of course, all but one of these outcomes are unobservable as for a person it is only possible to follow one sequence. Therefore, they are called potential outcomes and indexed by the sequences, $Y_1^{\underline{s}_t}, Y_2^{\underline{s}_t}, ..., Y_{t'}^{\underline{s}_t}$. The observable outcome is denoted by $Y_t$ and is related to the potential outcomes by the following rules:

---

[1] We choose to study only two periods where participation is possible, because it allows us to understand the essence of the problem while keeping the notation relatively simple. Later on, we briefly sketch what happens when additional participation is possible (more than two periods).

[2] Small letters denote specific values of the random variables.

[3] We use the term of "individual" instead of "unit" although the treatment can affect some entities like firms.

$$Y_t = S_1 Y_t^1 + (1-S_1)Y_t^0 = S_1 S_2 Y_t^{1,1} + S_1(1-S_2)Y_t^{1,0} + (1-S_1)S_2 Y_t^{0,1} + (1-S_1)(1-S_2)Y_t^{0,0}, \quad t = 1,2,...,T .$$

These outcomes are observed at the end of each period, whereas treatment status is measured at the beginning. Finally, there are attributes or confounders, $X$. These variables may influence the decision process of participation and / or the potential outcomes. It is possible that participation status may affect the values of these variables, for example the outcome of previous periods. To simplify the presentation of the problem, we will assume that the variables denoted by $X$ are not affected by the treatments and we will explicitly add the observed outcomes in the conditioning set when the case of endogeneity (the confounders are influenced by the treatments) is investigated. If some other variables (as the outcomes) are also influenced by the participation status, all we have to do is to handle these endogenous variables in the same way as the outcomes in the conditioning set. The attributes, $X_t$, are observable at the end of the period, i.e. at the same time as the outcome.

ASSUMPTION 1: (EXOGENEITY OF THE CONFOUNDERS) $X_t \amalg \underline{S}_\tau, \ \forall t, \tau$.

As already mentioned, several different effects are of interest. In the static case where only one participation is possible, three different effects are typically defined, the average treatment effect (ATE), the average treatment effect on the treated (ATET) and the average treatment effect on the nontreated. In this multi-participations framework, even more effects can be considered depending on the sequences compared, on their length and on the population for which we want to compute the effect. We only compare sequences that have the same length and for populations defined by their treatment status.[4] The average treatment effects are defined in the following way:

[4] For a detailed discussion on the choice of the population see Lechner and Miquel (2001).

$\theta_t^{\underline{s}_\tau^k, \underline{s}_\tau^l}(\underline{s}_{\tilde{\tau}}^j) := E(Y_t^{\underline{s}_\tau^k} \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j) - E(Y_t^{\underline{s}_\tau^l} \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j), \qquad 0 \le \tilde{\tau} \le \tau \le 2,\ k \ne l.$[5]

The first sequence, $\underline{s}_\tau^k$, defines the treated population, the second one, $\underline{s}_\tau^l$, the nontreated, and the last one, $\underline{s}_\tau^j$, the population for which we want to compute the effect. When $k = j$, the computed effect corresponds to the average treatment effect on the treated, and is called the *Dynamic Average Treatment Effect on the Treated* (DATET) by Lechner and Miquel (2001).[6] They also name the other effects the *Dynamic Average Treatment Effect* (DATE) and the *Dynamic Average Treatment Effect on the nontreated*.

Finally, we need to mention the fact that the standard assumptions of the Roy-Rubin model are also necessary. Thus, the Stable Unit Treatment Value Assumption is made. One of its implication is that the choices of other people concerning participation do not affect the effect of the treatments on one particular individual. Moreover, we need to assume that all conditional expectations of interest exist. As both assumptions (SUTVA and the existence of the moments) are required in each theorem in the following, we do not explicitly include them in the statement of the theorems (we implicitly assume that they are fulfilled).

## 3) The one-treatment difference-in-differences case using the multi-period notation

Now, we recall the "static case", when no more than one participation is possible (sequence of treatments of length 1) in which nonparametric identification using a difference-in-differences approach has been introduced by Heckman, Ichimura, Smith and Todd (1998), and Heckman, Ichimura and Todd (1997, 1998). These studies deal with the effect of the treatment on the treated. Two assumptions, a "common trend" one and an "anticipatory" one, are required to identify this effect. The assumptions affect conditional expectations. The variables used in the

[5] For $\tilde{\tau} = 0$, the effect is defined for the population (conditional on $\underline{S}_0$ with $\underline{s}_0^j = \underline{s}_0^k = \underline{s}_0^l = \underline{s}_0$).

[6] For a detail discussion on the different effects see Lechner and Miquel (2001).

conditioning set are allowed to vary over time (see, for example, Eichler and Lechner, 2002, who used such conditions to evaluate the effects of off-the-job training programs in the East German state of Sachsen-Anhalt). None of these characteristics is or can be influenced by the treatment. They are taking into account or they are needed in the conditioning set because it is often more plausible that the assumption holds after we have controlled for the characteristics influencing both the participation decision and the outcome.

ASSUMPTION 2: (COMMON TREND ASSUMPTION)[7]

$$E\left(Y_t^0 \mid X = x_t, S_1 = 1\right) - E\left(Y_{t'}^0 \mid X = x_{t'}, S_1 = 1\right) = E\left(Y_t^0 \mid X = x_t, S_1 = 0\right) - E\left(Y_{t'}^0 \mid X = x_{t'}, S_1 = 0\right)$$

$t$ denotes a time period after the period when participation is possible and $t'$ denotes a time period before it. This assumption requires a common trend in both subpopulations. Remember that we only observe the outcome of the individuals participating in the program and the outcome of the individuals not participating in the program. Thus, it is possible to compute the average of the potential outcome of interest as the sum of the averages of the outcome before the program occurs and this trend.

ASSUMPTION 3: (NO PRE-TREATMENT EFFECT ASSUMPTION)

$$E(Y_{t'}^1 - Y_{t'}^0 \mid X = x_{t'}, S_1 = 1) = 0 .$$

The "pre-treatment" assumption restricts the anticipatory effects of the program on the outcome previous to possible participation. There should be no effect of a future program on the outcome previous to possible participation. This assumption is fulfilled, for example when we want to evaluate training programs if the individuals could not anticipate their participation in the program or do not act on an anticipation in a way related to pre-treatment

[7] We should read $E\left(Y_t^0 \mid X = x_t, \underline{S}_1 = (\underline{s}_0, 1)\right) - E\left(Y_t^0 \mid X = x_t, \underline{S}_1 = (\underline{s}_0, 0)\right) = \dots$ .

outcomes. As we are interested in the effect of the treatment on the treated, the no "pre-treatment" effect assumption is defined for participants only.

ASSUMPTION 4: (COMMON SUPPORT ASSUMPTION)

$$0 < P(S_1 = 1 | X = x_{t'}) < 1, \;\; 0 < P(S_1 = 1 | X = x_t) < 1, \quad \forall\, x_{t'} \in \chi_{t'}, x_t \in \chi_t .$$[8]

It remains to describe the support condition in detail and why it is needed. Heckman, Ichimura and Todd (1997) and Heckman, Ichimura, Smith and Todd (1997, 1998) point out that due to the presence of characteristics in the conditioning set, we possibly suffer from the problem of common support.[9] To be able to compute or estimate the required expectations we need to find nontreated individuals who have the same characteristics, the same $X$, as the participants. If this is not possible, we cannot find individuals who do not participate in the program to give us information on the counterfactual outcome we are interested in for the participants. Nevertheless, this problem is less severe as in the approach using the CIA because here less confounders are required for the assumptions to hold. Under these assumptions, the treatment effect on the treated is identifiable as stated in theorem 1.

*Theorem 1*: UNDER ASSUMPTIONS 2, 3 AND 4, THE EFFECT OF THE TREATMENT ON THE TREATED AND ON THE NONTREATED, $\theta_t^{1,0}(\underline{S}_1 = 1)$ AND $\theta_t^{1,0}(S_1 = 0)$, ARE IDENTIFIED.

The proof of theorem 1 is recalled in appendix A.1. The ATET equals the difference of the observable deviation of the averaged outcome of the period after participation from the averaged outcome of the period before participation for the treated and the same observable deviation for the untreated: $t' < 1 \leq t$, $E\left(Y_t | S_1 = 1\right) - \underset{X_t|S_1=1}{E}\left\{E\left(Y_{t'} | X = x_{t'}, S_1 = 1\right)\right\}$ $- \underset{X_t|S_1=1}{E}\left\{E\left(Y_t | X = x_t, S_1 = 0\right) - E\left(Y_{t'} | X = x_{t'}, S_1 = 0\right)\right\}$, where 1 denotes the participation

[8] $\chi_{t'}, \chi_t$ denote the sets of the allowed values for the characteristics at time $t'$ and $t$ respectively.

period. For the second and the last term, the outer expectation is taken with respect to the distribution of the characteristics at time $t$ although the distribution used to select the individuals in accordance with the conditioning set of the inner expectation is the distribution of those characteristics at time $t'$. Therefore, all what is needed to estimate the effect is to estimate several conditional expectations. As shown by Heckman and Robb (1985, 1986) repeated cross-sectional data with the identity of the participants known in each period are sufficient to estimate those expectations.

The effect of the treatment on the nontreated equals the same expression of observable random variables as the effect of the treatment on the treated except that the outer expectation is with respect to the distribution of the covariates of the nontreated: $E\left(Y_t \mid S_1=0\right)-\underset{X_t \mid S_1=0}{E}\left\{E\left(Y_{t'} \mid X=x_{t'}, S_1=0\right)-E\left(Y_t \mid X=x_t, S_1=1\right)+E\left(Y_{t'} \mid X=x_{t'}, S_1=1\right)\right\}$.

Moreover, when these assumptions (2,3,4) are valid for both the treated and the nontreated, the average treatment effect for the population is identified:

$$\theta_t^{1,0}=P(S_1=1)\theta_t^{1,0}(S_1=1)+(1-P(S_1=1)\theta_t^{1,0}(S_1=0)$$

$$=P(S_1=1)\left[E\left(Y_t \mid S_1=1\right)-\underset{X_t \mid S_1=1}{E}\left\{E\left(Y_{t'} \mid X=x_{t'}, S_1=1\right)+E\left(Y_t \mid X=x_t, S_1=0\right)\right\}-E\left(Y_{t'} \mid X=x_{t'}, S_1=0\right)\right]$$

$$+(1-P(S_1=1)\left[E\left(Y_t \mid S_1=0\right)-\underset{X_t \mid S_1=0}{E}\left\{E\left(Y_{t'} \mid X=x_{t'}, S_1=0\right)+E\left(Y_t \mid X=x_t, S_1=1\right)-E\left(Y_{t'} \mid X=x_{t'}, S_1=1\right)\right\}\right].$$

As the probability to participate in period 1 is estimable from the data, all effects, ATE, ATET and ATET on the nontreated, are identified.

## 4) Sequences of programs

Now we consider two sequential treatments. Thus, participation is possible in an additional period, leading to four different sequences of programs (always participate, never participate,

[9] For a detail discussion of the common support problem, see Lechner (2001).

participate in the first period but not in the second one, and participate only in the second period) and four potential outcomes defined by these sequences of participations. Note that $t$ still indicates a period after participation, but after the last possible participation in the sequence of interest, $t \geq 2$ and that $t'$ indicates a period prior to the first possible participation in the sequence of interest, $t' \leq 0$. In this framework, different approaches are presented, depending on the degree of sequentiality (the number of base periods) used to investigate the problem. Nevertheless, the assumptions necessary for the identification of the different average treatment effects are still some common trend, some no "pre-treatment" effect and a common support condition. Whatever the approach, the same common support condition applies.

ASSUMPTION 5: (COMMON SUPPORT CONDITION) $\forall x_{t'} \in \chi_{t'}, x_t \in \chi_t, s_1 \in \{0,1\}$,

$0 < P(S_1 = 1 | X_{t'} = x_{t'}) < 1,\ \ 0 < P(S_1 = 1 | X_t = x_t) < 1,$

$0 < P(S_2 = 1 | S_1 = s_1, X_{t'} = x_{t'}) < 1,\ \ 0 < P(S_2 = 1 | S_1 = s_1, X_t = x_t) < 1$

### 4.1) No sequentiality

In this part we mimic the process used in the static case and compare the period after the last possible participation with the period prior to the first possible participation. The assumptions will involve only these periods $t$ and $t'$.

When participation is possible only in one period but the choice is between more than one program or when participation is possible in several periods, one additional category of effects is of interest, namely the effect of a sequence (or of a program) compared to another sequence (or program) for individuals who have done none of the sequences compared but have followed a third sequence of programs. We want to investigate what would have happened for these individuals, if they had changed from doing one particular sequence to do a different second sequence. The realizations of both sequences ($l$ and $k$) are hypothetical because in reality these individuals do a third one ($j$). This is the most challenging task with respect to identification.

ASSUMPTION 6: (TRIANGULAR COMMON TREND ASSUMPTION)

A) $E\left(Y_t^{s_2^k} \mid X = x_t, \underline{S}_2 = \underline{s}_2^j\right) - E\left(Y_{t'}^{s_2^k} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^j\right) = E\left(Y_t^{s_2^k} \mid X = x_t, \underline{S}_2 = \underline{s}_2^k\right) - E\left(Y_{t'}^{s_2^k} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^k\right)$

B) $E\left(Y_t^{s_2^l} \mid X = x_t, \underline{S}_2 = \underline{s}_2^j\right) - E\left(Y_{t'}^{s_2^l} \mid X = X_{t'}, \underline{S}_2 = \underline{s}_2^j\right) = E\left(Y_t^{s_2^l} \mid X = x_t, \underline{S}_2 = \underline{s}_2^l\right) - E\left(Y_{t'}^{s_2^l} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^l\right)$

As none of the sequences compared can be observed for the population considered, both outcomes needed to compute the effect of interest are not observed. Therefore, we use two assumptions similar to the common trend assumption of the one-participation case to compensate for this problem. Assumption 6 states that the trend of the potential outcome for the population participating in sequence $j$ is the same as the trend of the population participating in the alternative sequence ($k$ or $l$). On the other hand, only one assumption concerning the "pre-treatment" effect is required. This is intuitive because we do not need to compare each sequence with the sequence observed, we just need to compare these sequences together. For the population of interest, the no "pre-treatment" effect assumption states that the sequences studied should have the same influence on the outcome previous to the first period (when the first participation takes place). There should be no effect, if we compare sequences with at least one participation to the sequence of no participation.

ASSUMPTION 7: (NO PRE-TREATMENT EFFECT) $E\left(Y_{t'}^{s_2^l} - Y_{t'}^{s_2^k} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^j\right) = 0$.

Of course there are as many restrictions as there are possible effects (3 effects for each subpopulations and 4 subpopulations). If we only want to identify one effect it is sufficient to impose one restriction, the one corresponding to the pair of sequences compared for the population of interest. Theorem 2 presents the identification results. It is proved in Appendix A.2.

*Theorem 2:* IF ASSUMPTIONS 5, 6 AND 7 HOLD, ALL AVERAGE TREATMENT EFFECTS ARE IDENTIFIED.

As shown in the proof of this theorem in Appendix A.2, the effects, $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(\underline{S}_2=\underline{s}_2^j)$, are equal to a combination of expectations, which are estimable from the data, similar to the one obtained in the one-period case (a difference of differences):

$$\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(\underline{S}_2=\underline{s}_2^j) = \underset{X_t|\underline{S}_2=\underline{s}_2^j}{E}\left\{E\left(Y_t \mid X=x_t,\underline{S}_2=\underline{s}_2^k\right)-E\left(Y_{t'} \mid X=x_{t'},\underline{S}_2=\underline{s}_2^k\right)\right\}$$

$$-\underset{X_t|\underline{S}_2=\underline{s}_2^j}{E}\left\{E\left(Y_t \mid X=x_t,\underline{S}_2=\underline{s}_2^l\right)-E\left(Y_{t'} \mid X=x_{t'},\underline{S}_2=\underline{s}_2^l\right)\right\}.$$

The difference with the one-period expression is that the distribution of the characteristics at time $t$ (the outer expectation) is taken conditional on the third sequence of programs and not on one of the sequences compared.

If we only want to identify the treatment effect on the treated or on the nontreated, the assumptions simplify to one condition in the common trend assumption (Assumption 6-1) and the conditioning set of the no pre-treatment condition (Assumption 7-1) is changed.[10]

ASSUMPTION 6-1: (COMMON TREND ASSUMPTION)[11]

$$E\left(Y_t^{\underline{s}_2^l} \mid X=x_t,\underline{S}_2=\underline{s}_2^k\right)-E\left(Y_{t'}^{\underline{s}_2^l} \mid X=x_{t'},\underline{S}_2=\underline{s}_2^k\right)=E\left(Y_t^{\underline{s}_2^l} \mid X=x_t,\underline{S}_2=\underline{s}_2^l\right)-E\left(Y_{t'}^{\underline{s}_2^l} \mid X=x_{t'},\underline{S}_2=\underline{s}_2^l\right).$$

As one of the outcomes used to compute the effect can directly be observed in the data, we only need one assumption to compensate the unobservability of the second outcome used to

[10] The assumptions are written for the identification of the DATET. To identify the DATE on the nontreated the same assumptions are required but with the role of $k$ and $l$ switched.

[11] $\underline{s}_2^k=(\underline{s}_0,s_1^k,s_2^k)$ ; $\underline{s}_2^l=(\underline{s}_0,s_1^l,s_2^l)$ .

compute the effect. The no pre-treatment condition states that the sequences compared should have the same influence on the outcomes in the period previous to the first possible participation. The individuals must change their behavior in anticipation of what they will do in the future in a similar manner whatever the sequence (*k* or *l*) they will follow (during the two next periods). In the special case where one of the sequences compared is the "never participate" one, we have to interpret this condition a bit differently. The programs should have no effect at all on the outcomes. The individuals should not change their behavior in anticipation of the future.

ASSUMPTION 7-1: (NO PRE-TREATMENT EFFECT) $E\left(Y_{t'}^{\underline{s}_2^l} - Y_{t'}^{\underline{s}_2^k} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^k\right) = 0$.

In this case, for example, the effect of participating in the program each period compared to no participation at all for the individuals who always participate, is given by the following expression:

$$\theta_t^{(1,1),(0,0)}(\underline{S}_2 = (1,1)) = E\left(Y_t \mid \underline{S}_2 = (1,1)\right) - \underset{X_t|\underline{S}_2=(1,1)}{E}\left\{E\left(Y_{t'} \mid X = x_{t'}, \underline{S}_2 = (1,1)\right)\right\}$$

$$- \underset{X_t|\underline{S}_2=(1,1)}{E}\left\{E\left(Y_t \mid X = x_t, \underline{S}_2 = (0,0)\right) - E\left(Y_{t'} \mid X = x_{t'}, \underline{S}_2 = (0,0)\right)\right\}.$$

Therefore, to estimate the dynamic effect, we have to estimate each of these expectations. This can be done by the use of matching methods.

---

Example: Introduction of Charter Schools.

This example is based on the study of Dee and Fu (2001).[12] They investigate the effect of the introduction of charter schools in some American states on the percentage of white non-Hispanic students and on the pupil-teacher ratio in public schools. The charter schools are independent public schools (non-selective in admission, no tuition and non-religious) that are established under an agreement with a state or local agency, and this agreement must be renewed after a fixed period (five years).[13] Then, in our general terminology, the signing of an agreement (the introduction of the charter

[12] They use a static framework and investigate the effect of a one-shot program.
[13] For a detailed description of charter schools, see Finn, Manno and Vanourek, 2000.

schools) is the program. Periods 1 and 2 of our framework (when a participation is possible) each last five years. The following sequences of programs exist: (charter schools, charter schools), (no charter schools, no charter schools), (charter schools, no charter schools), and (no charter schools, charter schools). For example, the first sequence corresponds to the sequence (1,1) in our general notation in which participation occurs in each period. Let $\underline{s}_2^k$ be the first sequence (CS, CS) and $\underline{s}_2^l$ be the second sequence (no SC, no SC). Thus, assumption 6-1 implies that the unobserved trend of the percentage of the white non-Hispanic students when no charter schools were introduced for the schools in the state that introduces charter schools equals the trend of the same percentage for the schools in a state that does not introduce the charter schools. This assumption seems to be valid if the growth of the minority share of the total population in each state is similar (if in one county this share decreases much more than in the other county, the growth of the percentage of white no-Hispanic students in the public schools will be influenced by this decrease). That is the reason why in addition to the "demographic" variables (elementary school, suburban location...), we control for the minority share of the total population for the county and year in which each school is observed. (We include this variable in the conditioning set, precisely it is one of the $X$ variables).

---

In the second period, it may be that some variables we want to use in the conditioning set are influenced by participation in the sequence investigated. If those endogenous variables are in the conditioning set, the common trend assumption has to be modified. For example if we are interested in the effect of some training programs on employment, the restriction may be fulfilled only if we have in the conditioning set the employment of the previous periods $(Y_{t-1}, Y_{t-2}...)$ and thus, at time t, the outcome of period 1 which has been possibly influenced by the sequence made. As for the characteristics at time $t'$,[14] the employment of the previous periods $(Y_{t'-1}, Y_{t'-2}...)$ is also included but they do not contain the employment of period 1 or any other variable influenced by the program (this statement will be false only if anticipatory effects exist).

[14] Note that $Y_{t'}$ should not be included in $X_{t'}$, otherwise the difference of the last two terms will be zero and the effect could be stated in the same terms of observable random variables as under a conditional independence assumption (see Lechner and Miquel, 2001).

ASSUMPTION 6-2: (COMMON TREND ASSUMPTION)

$$E\left(Y_t^{\underline{s}_2^l} \mid X = x_t, Y_{t-j} = y_{t-j}, \underline{S}_2 = \underline{s}_2^k\right) - E\left(Y_{t'}^{\underline{s}_2^l} \mid X = x_{t'}, Y_{t'-j} = y_{t'-j}, \underline{S}_2 = \underline{s}_2^k\right)$$
$$= E\left(Y_t^{\underline{s}_2^l} \mid X = x_t, Y_{t-j} = y_{t-j}, \underline{S}_2 = \underline{s}_2^l\right) - E\left(Y_{t'}^{\underline{s}_2^l} \mid X = x_{t'}, Y_{t'-j} = y_{t'-j}, \underline{S}_2 = \underline{s}_2^l\right),$$
$$\text{with } t - j = 1.$$

As seen in the proof of theorem 2, the identification of the effect does not require any restriction on the relation between the sequences or the programs and the characteristics used in the conditioning set. Consequently, all the effects of theorem 2 are still identified when some endogenous variables, the outcome of period 1 for example, are in the conditioning set. Assumption 7-1 does not need to be changed because even if some past outcomes are present in the conditioning set, those outcomes are not influenced by the sequence. (It is less probable that some anticipation effects exist. These outcomes are those of periods far away from the periods when the implementation of the sequence occurs, because they are previous to period $t'$, which itself is previous to the beginning of the first program.) The only difference between assumptions 6-1 and 6-2 is the exogeneity of the conditioning variables. The choice of these variables depends on the question, we want to study.

---

Example: Active labor market policy

We want to investigate the effect of a sequence of programs, a "job-search" program in the first period and a "subsidized jobs" program in the second period, on the employment of individuals, who are unemployed before the beginning of the sequence. In the static case, where only one program is examined, the common trend assumption (similar to assumption 6) is fulfilled if the past employment history is included in the conditioning variables (see for example Blundell, Costa Dias, Meghir and Van Reenen, 2001). They argue that we can assume that the control group and the treated group in the absence of the treatment "are subject to the same aggregate labor market trends to the extend that the human capital of the two groups is perfectly substitutable...". In our context the argument is still valid if we include in the conditioning set a variable giving the number of unemployment spells during the last five years, for example. Then, this variable at time $t$ is correlated with the outcome after the first period. This outcome can be affected by the first participation. Thus, the common trend assumption we need is assumption 6-2, where the dynamic is taken into account (past outcome in the conditioning set).

---

Hence, we have seen in this section that we can identify all effects $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(\underline{S}_2=\underline{s}_2^j)$, $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(S_1=s_1^j)$, and $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}$.

Moreover, the approach presented in this part and the results obtained are equivalent to a static multiple treatment framework identified by a difference-in-differences approach, in which only one participation is considered but where the individuals have the choice to participate in several different programs.

Table 1: Identified Effects in the nonsequential approach

| | $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(\underline{s}_2^k)$ | $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(\underline{s}_2^l)$ | $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(\underline{s}_2^j)$ | $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(s_1^j)$ | $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}$ |
|---|---|---|---|---|---|
| **A.5: Common Support** $0<p_{1\|t'}<1,\ 0<p_{1\|t}<1,\ 0<p_{2\|1,t'}<1,\ 0<p_{2\|1,t}<1$ | X | X | X | X | X |
| **A.6: Triangular Common Trend** $E\left(Y_t^{\underline{s}_2^k}\mid X=x_t,\underline{S}_2=\underline{s}_2^j\right)-E\left(Y_{t'}^{\underline{s}_2^k}\mid X=x_{t'},\underline{S}_2=\underline{s}_2^j\right)=E\left(Y_t^{\underline{s}_2^k}\mid X=x_t,\underline{S}_2=\underline{s}_2^k\right)-E\left(Y_{t'}^{\underline{s}_2^k}\mid X=x_{t'},\underline{S}_2=\underline{s}_2^k\right)$ $E\left(Y_t^{\underline{s}_2^l}\mid X=x_t,\underline{S}_2=\underline{s}_2^j\right)-E\left(Y_{t'}^{\underline{s}_2^l}\mid X=x_{t'},\underline{S}_2=\underline{s}_2^j\right)=E\left(Y_t^{\underline{s}_2^l}\mid X=x_t,\underline{S}_2=\underline{s}_2^l\right)-E\left(Y_{t'}^{\underline{s}_2^l}\mid X=x_{t'},\underline{S}_2=\underline{s}_2^l\right)$ | X | X | X | X | X |
| **A.7 No Pre-Treatment** $E\left(Y_{t'}^{\underline{s}_2^l}-Y_{t'}^{\underline{s}_2^k}\mid X=x_{t'},\underline{S}_2=\underline{s}_2^j\right)=0$ | X | X | X | X | X |
| **A.6-1 Common Trend** $E\left(Y_t^{\underline{s}_2^l}\mid X=x_t,\underline{S}_2=\underline{s}_2^k\right)-E\left(Y_{t'}^{\underline{s}_2^l}\mid X=x_{t'},\underline{S}_2=\underline{s}_2^k\right)=E\left(Y_t^{\underline{s}_2^l}\mid X=x_t,\underline{S}_2=\underline{s}_2^l\right)-E\left(Y_{t'}^{\underline{s}_2^l}\mid X=x_{t'},\underline{S}_2=\underline{s}_2^l\right)$ | X | X | | | |
| **A.7-1 No Pre-Treatment** $E\left(Y_{t'}^{\underline{s}_2^l}-Y_{t'}^{\underline{s}_2^k}\mid X=x_{t'},\underline{S}_2=\underline{s}_2^k\right)=0$ | X | X | | | |

Notes: $p_{1|t'}=P(S_1=1\mid X=x_{t'})$, $p_{2|1,t'}=P(S_2=1\mid S_1=s_1,X=x_{t'})$. The last two assumptions are written to identify the DATET. To identify the DATE on the nontreated, the sequences $k$ and $l$ have to be switched in these assumptions.

The identified effects and the assumptions given those results are summarized in Table 1, where a cross means that the assumption is needed to identify the effect. For example, for the first effect, $j = k$ and Assumption 6 simplifies to Assumption 6-1 and Assumption 7 to Assumption 7-1. Thus, to identify the treatment effect on the treated, the common support assumption, Assumption 6-1 and 7-1 should be valid. Note that for the three last effects in the table, no simplification in the assumptions occurs, therefore, Assumptions 6-1 and 7-1 are not taken into account (no cross).

### *4.2) Sequential approach*

If the interval between the periods considered is large, we can doubt about the validity of the common trend assumption as presented before. Moreover, the possible participation in the first period can lead to different trends between the different subpopulations defined by their participation status. For different reasons (e.g. external shock on the economy), the trend of the outcome between periods $t$' and 1 can be different from the trend of the outcome between period 1 and $t$. These features are not well taken into account in the previous section. However, we can do better than mimic the static one period participation approach by considering a sequential approach. We have at our disposal different periods that can be used as comparison (base period), the period before any participation and the first period of possible participation. This section investigates the necessary conditions to identify the effects when one period of reference is period 1. On top of the distinctions already made between DATE, DATET and DATE on the nontreated, other decompositions can be added. We distinguish the effects according to the kind of sequences compared. Two categories will be considered.

### 4.2.1) Common part

The first category consists of effects resulting from the comparison of sequences of programs with a common first part. We look at the effect of participating in the program in the second period comparing to not participating in it when the first participation decision is the same in both sequences compared and when the population for which we want to compute the effect has made the same decision about its participation in the first period. We are in the simple case where the first period plays the role of a base period for the second period.[15] The cases where the common part is the second part of the sequence and where the population for which we want to compute the effect has followed a sequence with a different period 1 program, are not presented here. They can be treated as sub-cases of the next section (no common part ).

The assumptions needed to identify the effects are very similar to the previous ones used. But now, the base period is period 1. The common trend assumption is presented in assumption 8.

ASSUMPTION 8: (COMMON TREND ASSUMPTION)

$$E(Y_t^{(s_1,0)} \mid X = x_t, \underline{S}_2 = (s_1,0)) - E(Y_1^{(s_1,0)} \mid X = x_1, \underline{S}_2 = (s_1,0))$$
$$= E(Y_t^{(s_1,0)} \mid X = x_t, \underline{S}_2 = (s_1,1)) - E(Y_1^{(s_1,0)} \mid X = x_1, \underline{S}_2 = (s_1,1)), \quad s_1 \in \{0,1\}.$$

The second assumption is again a no "pre-treatment" effect assumption. The program in the second period should have no effect on the outcome of the first period. As in our framework only one program is available in the second period, this assumption states that the program in the second period should not have any influence on the outcome of the first period. If more than one program were available, the interpretation of this assumption would change depending on the programs compared (see the discussion in the previous section).

ASSUMPTION 9: (NO PRE-TREATMENT EFFECT) $E(Y_1^{(s_1,1)} - Y_1^{(s_1,0)} \mid X = x_1, \underline{S}_2 = (s_1,1)) = 0$.

[15] Everyone does the same in this base period.

Then the average treatment effects on the treated and on the nontreated are identified as stated in Theorem 3. Keep in mind that this results only concerns the sequences that have a common part in the first period. Moreover, two other effects are identified, those for subpopulations defined by their behavior in period one only. (As in the previous section, the assumptions are written for the identification of the DATET. To obtain the identification of the DATE for the nontreated, the same assumptions are required with the role of 1 and 0 switched. To identify the two additional effects, the assumptions have to hold simultaneously for both treated and nontreated.)

*THEOREM 3: UNDER ASSUMPTIONS 5, 8 AND 9, THE FOLLOWING AVERAGE TREATMENT EFFECTS ARE IDENTIFIED:* $\theta_t^{(s_1,1),(s_1,0)}(\underline{S}_2=(s_1,1))$, $\theta_t^{(s_1,1),(s_1,0)}(\underline{S}_2=(s_1,0))$, $\forall s_1 \in \{0,1\}$ and $\theta_t^{(1,1),(1,0)}(S_1=1)$, $\theta_t^{(0,1),(0,0)}(S_1=0)$.

Theorem 3 is proved in appendix A.3. The expressions of the first two effects in terms of observable random variables are similar to those obtained in the previous sections. For example, for the effect on the treated of the sequence "only participate in the second period" we have:

$$\theta_t^{(0,1),(0,0)}(\underline{S}_2=(0,1)) = E(Y_t^{(0,1)} \mid \underline{S}_2=(0,1)) - \underset{X_t|\underline{S}_2=(0,1)}{E}\left\{E(Y_1 \mid X=x_1, \underline{S}_2=(0,1))\right\}$$

$$- \underset{X_t|\underline{S}_2=(0,1)}{E}\left\{E(Y_t \mid X=x_t, \underline{S}_2=(0,0)) - E(Y_1 \mid X=x_1, \underline{S}_2=(0,0))\right\}.$$

To estimate this effect we only need participants in program 0 at time $t=1$ to compute the different expectations.

---

Example: Sequences of active labor market policies

We want to investigate the effect of a two-period sequence on the income of the participants at the end of period 2. We want to compare the two-participation case to the one-participation case where the participation takes place in the first period. We have to decide which comparison period we will use (we have the choice between a period prior to the first possible participation ($t$') and period 1, when the first participation occurs) and therefore which set of assumptions we believe in (the nonsequential set or the sequential one). In the studies of the effect of training programs on the income of individuals, a drop in income of the participants commonly appears in the periods prior to the beginning of the program (this drop is called the Ashenfelter's dip or the pre-program dip). If a permanent dip occurs before the first possible participation, no contradiction appears against the use of period $t$'as a comparison period for the outcome at time 2. Thus, the nonsequential set of assumptions (assumptions 6 and 7) will be considered. Contrarily, if the dip is transitory and recovers at period 1, it seems more plausible to use the period 1-outcome as a comparison. The economic environment would be more similar as if we use period $t$'. Thus, the sequential approach (assumptions 8 and 9) is assumed to be valid.

---

### 4.2.2) No common part

The second category consists of effects of sequences that have nothing in common.[16] As the sequences compared differ in both periods and as we use the first-period outcome as a base outcome, we also need to correct for the fact that, given the same "action" in the second period of the sequence, the first-period outcome needed is not directly observable. As we are investigating under which conditions the effect of a sequence compared to another sequence for a population participating in none of those sequences is identified and as both outcomes required to compute this effect are not observed, four common trend conditions are necessary, along with four no pre-treatment effect conditions. Then, if all these average treatment effects (for each possible triplet of sequences) are identified, all the average treatment effects for subpopulations characterized only by their behavior in the first period are identified as well as the average effects for the whole population.

ASSUMPTION 10: (TRIANGULAR COMMON TREND ASSUMPTION)

A1) $E(Y_t^{s_2^k} \mid x_t, \underline{S}_2 = \underline{s}_2^j) - E(Y_1^{s_2^k} \mid x_1, \underline{S}_2 = \underline{s}_2^j) = E(Y_t^{s_2^k} \mid x_t, \underline{S}_2 = \underline{s}_2^k) - E(Y_1^{s_2^k} \mid x_1, \underline{S}_2 = \underline{s}_2^k)$

A2) $E(Y_t^{s_2^l} \mid x_t, \underline{S}_2 = \underline{s}_2^j) - E(Y_1^{s_2^l} \mid x_1, \underline{S}_2 = \underline{s}_2^j) = E(Y_t^{s_2^l} \mid x_t, \underline{S}_2 = \underline{s}_2^l) - E(Y_1^{s_2^l} \mid x_1, \underline{S}_2 = \underline{s}_2^l)$

B1) $E(Y_1^{(s_1^k, s_2^j)} \mid x_1, \underline{s}_2^j) - E(Y_{t'}^{(s_1^k, s_2^j)} \mid x_{t'}, \underline{s}_2^j) = E(Y_1^{(s_1^k, s_2^j)} \mid x_1, (s_1^k, s_2^j)) - E(Y_{t'}^{(s_1^k, s_2^j)} \mid x_{t'}, (s_1^k, s_2^j))$

B2) $E(Y_1^{(s_1^l, s_2^j)} \mid x_1, \underline{s}_2^j) - E(Y_{t'}^{(s_1^l, s_2^j)} \mid x_{t'}, \underline{s}_2^j) = E(Y_1^{(s_1^l, s_2^j)} \mid x_1, (s_1^l, s_2^j)) - E(Y_{t'}^{(s_1^l, s_2^j)} \mid x_{t'}, (s_1^l, s_2^j))$,

to save space the notation $X = x_.$ is simplified to $x_.$ and in the last two equations the notation $\underline{S}_2 = (s_1^k, s_2^j)$ is simplified to $(s_1^k, s_2^j)$.

Each part of the assumption looks like the previous common trend assumptions. Part A restrains the average growth of the outcome between periods 1 and $t$ and part B the average growth of the outcome between period 1 and a period previous to period 1. The growth of the outcome between period 1 and $t$ for the population participating in $\underline{s}_2^j$ if they have participated in $\underline{s}_2^k$ is the same as the growth for the population participating in $\underline{s}_2^k$. An identical conclusion is assumed for the growth of the outcome if this population has participated in $\underline{s}_2^l$, it is the same as the growth for the population really participating in $\underline{s}_2^l$. Part B is the condition needed to be able to correct the error made when we use the outcome of the first period which is also influenced by the sequence and particularly by the first participation choice of the sequence.

For a fixed choice in the second period (all sequences defining the outcomes or the populations have the same choice in the second period), the condition on the trend is similar to what is assumed in the common part case (but the periods are switched: here we consider period 1 and *t'*, and the common part of the sequence is in the second period).

[16] This case is the general case. In fact, it also covers the case of sequences with a common part in either period 1 or 2.

ASSUMPTION 11: (NO "PRE-TREATMENT" EFFECT)

A1) $E(Y_1^{s_2^k} - Y_1^{(s_1^k, s_2^j)} \mid X = x_1, \underline{S}_2 = \underline{s}_2^j) = 0$, A2) $E(Y_1^{s_2^l} - Y_1^{(s_1^l, s_2^j)} \mid X = x_1, \underline{S}_2 = \underline{s}_2^j) = 0$

B1) $E(Y_{t'}^{(s_1^k, s_2^j)} - Y_{t'}^{s_2^j} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^j) = 0$, B2) $E(Y_{t'}^{s_2^j} - Y_{t'}^{(s_1^l, s_2^j)} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^j) = 0$.

Two no "pre-treatment" effect assumptions are required. The first one implies that the programs of the second period should have the same influence on the outcome of period 1 given the same participation status in the first period. The second assumption deals with the outcome before any participation and the effects of the first-period programs on this outcome. It seems that it is possible to simplify the no "pre-treatment" effect assumption. The last two points imply that $E(Y_{t'}^{(s_1^k, s_2^j)} \mid X_{t'}, \underline{S}_2 = \underline{s}_2^j) = E(Y_{t'}^{(s_1^l, s_2^j)} \mid X_{t'}, \underline{S}_2 = \underline{s}_2^j)$. Given that in the second period the participation decision follows the decision of sequence $j$, any program in the first period should have the same effect on the outcome at time $t'$. In our framework with only one program in each period, one of these assumptions (B1 and B2) is redundant because the first period of all sequences can only be 0 or 1. Thus, the sequences $k$ and $l$ have also 0 or 1 and as we are dealing with the case where the sequences compared have nothing in common, one of these sequences has, for sure, the same first-period participation as sequence $j$.

*THEOREM 4: UNDER ASSUMPTIONS 5, 10 AND 11, ALL DYNAMIC AVERAGE TREATMENT EFFECTS ARE IDENTIFIED.*

Theorem 4 states the identification of all the effects we are looking for and is proved in appendix A.4. The expression obtained in the proof for the effect (a combination of estimable averages) may look very complicated at first sight, but it is easy to understand the different pieces: $\underset{X_t \mid \underline{S}_2 = \underline{s}_2^j}{E} \left\{ E(Y_t \mid X = x_t, \underline{S}_2 = \underline{s}_2^k) - E(Y_1 \mid X = x_1, \underline{S}_2 = \underline{s}_2^k) - E(Y_t \mid X = x_t, \underline{S}_2 = \underline{s}_2^l) \right\}$

$$+ \underset{X_t|\underline{S}_2=\underline{s}_2^j}{E}\left\{E(Y_1 \mid X = x_1, \underline{S}_2 = \underline{s}_2^l) + E(Y_1 \mid X = x_1, \underline{S}_2 = (s_1^k, s_2^j)) - E(Y_{t'} \mid X = x_{t'}, \underline{S}_2 = (s_1^k, s_2^j))\right\}$$

$$- \underset{X_t|\underline{S}_2=\underline{s}_2^j}{E}\left\{E(Y_1 \mid X = x_1, \underline{S}_2 = (s_1^l, s_2^j)) - E(Y_{t'} \mid X = x_{t'}, \underline{S}_2 = (s_1^l, s_2^j))\right\}.$$

The expression is composed of four differences. The first two originate from the unobservability of both outcomes of interest due to the programs in the second period of the sequences which define those potential outcomes (those differences are similar to the ones obtained in the previous section). The last two differences come from the unobservability of the potential outcomes of interest due to the programs in the first period of the sequences. In our framework with only four sequences, this expression simplifies because each sequence of the effect has at least one common part with the sequence characterizing the population for which we compute the effect, e.g.:

$$\theta_t^{(1,1),(0,0)}(\underline{S}_2 = (1,0)) = \underset{X_t|\underline{S}_2=(1,0)}{E}\left\{E(Y_t \mid X = x_t, \underline{S}_2 = (1,1)) - E(Y_1 \mid X = x_1, \underline{S}_2 = (1,1))\right\}$$

$$- \underset{X_t|\underline{S}_2=(1,0)}{E}\left\{E(Y_t \mid X = x_t, \underline{S}_2 = (0,0)) \ - \ E(Y_{t'} \mid X = x_{t'}, \underline{S}_2 = (0,0))\right\}$$

$$+ \ E(Y_1 \mid X = x_1, \underline{S}_2 = (1,0)) - E(Y_{t'} \mid X = x_{t'}, \underline{S}_2 = (1,0)).$$

If we only want to identify the DATET or the DATE on the nontreated, assumptions 10 and 11 can be simplified to assumptions 10-1 and 11-1, where the notation used is the one for the identification of the average treatment effect on the treated (switch the role of $k$ and $l$ to obtain the hypotheses to identify the average treatment effect on the nontreated).

ASSUMPTION 10-1: (COMMON TREND ASSUMPTION)

A) $E(Y_t^{\underline{s}_2^l} \mid x_t, \underline{S}_2 = \underline{s}_2^k) - E(Y_1^{\underline{s}_2^l} \mid x_1, \underline{S}_2 = \underline{s}_2^k) \ = E(Y_t^{\underline{s}_2^l} \mid x_t, \underline{S}_2 = \underline{s}_2^l) - E(Y_1^{\underline{s}_2^l} \mid x_1, \underline{S}_2 = \underline{s}_2^l)$

B) $E(Y_1^{(s_1^l, s_2^k)} \mid x_1, \underline{s}_2^k) - E(Y_{t'}^{(s_1^l, s_2^k)} \mid x_{t'}, \underline{s}_2^k) = E(Y_1^{(s_1^l, s_2^k)} \mid x_1, (s_1^l, s_2^k)) - E(Y_{t'}^{(s_1^l, s_2^k)} \mid x_{t'}, (s_1^l, s_2^k))$, where the notation has been also simplified to reduce the size of the equations. As one of the average outcomes of the effect can be estimated directly from the data, we only need two assumptions to compensate for the unobservability of the second outcome due to both programs of the sequence. Part A of assumption 10-1 has the same interpretation as the one of the common trend assumption of the previous section (assumption 8). The unobserved potential outcome equals a trend measured by the average growth of the outcome between periods 1 and $t$ in the population participating in sequence $l$ plus an average outcome that is not observable. Therefore, part B of the assumption also restricts this average outcome to equal a trend measured by the average growth of the outcome between periods 1 and $t'$ in the population participating in a sequence with the same period 2 program as the one of the sequence the treated participate in and the period 1 program of the comparison sequence.

ASSUMPTION 11-1: (NO "PRE-TREATMENT" EFFECT)

A) $E(Y_1^{(s_1^l, s_2^k)} - Y_1^{\underline{s}_2^l} \mid X = x_1, \underline{S}_2 = \underline{s}_2^k) = 0$, B) $E(Y_{t'}^{(s_1^l, s_2^k)} - Y_{t'}^{\underline{s}_2^k} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^k) = 0$.

A second version of assumption 10 and thus, of assumption 11 can be used to identify the DATET. It differs, principally, with respect to the influence of the first-period choice of the sequence on the outcome of period 1. Nevertheless, it is not possible anymore to interpret this assumption as the imposition of a common trend for the outcomes. It is also difficult to interpret it in an economic framework, that is why it will not be further discussed.

Moreover, a last comment on the assumptions is necessary. If we only use one common trend assumption and mimic the classical difference-in-differences estimator when we estimate the effect we are interested in (for example, for the dynamic average treatment effect on the treated of the sequence "always participate" in comparison to the base sequence "never participate"), we would use the average growth of the outcome for the treated between period $t$ and period 1 minus the similar average growth for the nontreated as the expression to estimate, $E(Y_t \mid \underline{S}_2=(1,1)) \; - \underset{X_t \mid \underline{S}_2=(1,1)}{E}\left\{E(Y_1 \mid X=x_1, \underline{S}_2=(1,1)) \; + \; E(Y_t \mid X=x_t, \, \underline{S}_2=(0,0))\right\}$ $+ \underset{X_t \mid \underline{S}_2=(1,1)}{E}\left\{E(Y_1 \mid X=x_1, \underline{S}_2=(0,0))\right\}$, and therefore, we implicitly make an undesirable assumption. The no "pre-treatment" effect assumption giving this results states that the sequences have no effect on the outcome of period 1, $E(Y_1^{s_2^k} \mid X=x_1, \underline{S}_2=\underline{s}_2^k) = E(Y_1^{s_2^l} \mid X=x_1, \underline{S}_2=\underline{s}_2^k)$. Considering that there is no "pre-treatment" effect of the program in the second period on the outcome of period 1, one plausible implication of this assumption is that the program in the first period has no effect on the outcome in this period, or in other words directly after a participation in the program in period 1, no change in the outcome due to the program will exist.

---

Example: The linear difference-in-differences model

Usually, a linear fixed effect approach is used to estimate the effect of interest in the static case of one possible participation. Let us make use of a similar framework to illustrate the multi-period participation case. The following linear equation is used to compute the effect of interest:

$$y_{it} = \theta_i + d_t + \gamma' X_i + \alpha(S_{i2}T_{i2}) + \beta(S_{i1}T_{i1}) + \xi(S_{i2}T_{i2})(S_{i1}T_{i1}) + \varepsilon_{it} \,,$$

$$E(\varepsilon_{it} \mid D_i, D_t, X_i, S_{i2}, T_{i2},, S_{i1}, T_{i1}) = 0$$

where $\theta$ is an individual fixed effect, $d$ reflects common effects, $D_i$ is an individual dummy and $D_t$ a time dummy. $X$ is included to take into account differences due to observable characteristics. $S_{ij}$ equals 1 if a participation occurs in period $j$ and $T_{i1}=1$ if $t \geq 1$, $=0$ otherwise and $T_{i2}=1$ if $t \geq 2$, $=0$ otherwise. $\xi$ represents the part of the effect due to a participation in both periods. It follows that, computing the above expression (resulting of theorem4), the DATET, $\theta_t^{(1,1),(0,0)}(1,1)$, equals $\alpha+\beta+\xi$, the DATET $\theta_t^{(1,1),(1,0)}(1,1)$ equals $\alpha+\xi$ and the DATET $\theta_t^{(0,1),(0,0)}(0,1)$ equals $\alpha$.

---

With regard to the average treatment effects on the nontreated for individuals with the same characteristics, we can be tempted to believe that their expressions in terms of observable random variables are similar to those of the treated, as it has been the case before. Nevertheless, it is interesting to compare those effects in detail because the results differ. As an example, we compare the sequence "always participate" to the sequence "never participate". The DATET has the following form:

$$E(Y_t \mid X = x_t, \underline{S}_2 = (1,1)) - \underset{X_t|\underline{S}_2=(1,1)}{E}\left\{E(Y_t \mid X = x_t, \underline{S}_2 = (0,0)) - E(Y_1 \mid X = x_1, \underline{S}_2 = (0,0))\right\}$$

$$- \underset{X_t|\underline{S}_2=(1,1)}{E}\left\{E(Y_1 \mid X = x_1, \underline{S}_2 = (0,1)) + E(Y_{t'} \mid X = x_{t'}, \underline{S}_2 = (1,1)) - E(Y_{t'} \mid X = x_{t'}, \underline{S}_2 = (0,1))\right\}.$$

Whereas the DATE on the nontreated equals:

$$\underset{X_t|\underline{S}_2=(0,0)}{E}\left\{E(Y_t \mid X = x_t,\ \underline{S}_2 = (1,1))\ -\ E(Y_t \mid X = X_t,\ \underline{S}_2 = (0,0)) + E(Y_1 \mid X = X_1,\ \underline{S}_2 = (1,0))\right\}$$

$$- \underset{X_t|\underline{S}_2=(0,0)}{E}\left\{E(Y_1 \mid X = x_1,\ \underline{S}_2 = (1,1))\ -\ E(Y_{t'} \mid X = X_{t'},\ \underline{S}_2 = (0,0)) + E(Y_{t'} \mid X = X_{t'},\ \underline{S}_2 = (1,0))\right\}.$$

We see that those expressions are not anymore equal for individuals having the same characteristics.

Until now we did not discuss the fact that the outcome of period 1, $Y_1$, can be in the characteristics of the conditioning set. Nevertheless, as in the section 4.1) (no sequentiality), the presence of this endogenous variable in the conditioning set does not lead to any technical problems. For example consider the case of the DATET and of $t = 2$, then assumption 10-1 will become assumption 10-2. Assumption 11-1 is changed to assumption 11-2. In spite of this result, the interpretation of the assumptions are quiet different. Three cases can be defined. In the first case, the confounding variables are not influenced by the participation in the first period and we are in the usual DiD framework with more than one treatment. In the second case, some confounders are influenced by the participation in the first period and the decision of participation is taken sequentially.[17] Finally, in the last case, in the confounders influenced by the first participation, we find the outcome of the first period leading to a dynamic decision process.

ASSUMPTION 10-2: (COMMON TREND ASSUMPTION)

A) $E(Y_2^{s_2^l} \mid X=(x_2,y_1), \underline{S}_2=\underline{s}_2^k) - E(Y_1^{s_2^l} \mid X=(x_1,y_0), \underline{S}_2=\underline{s}_2^k)$

$$= E(Y_2^{s_2^l} \mid X=(x_2,y_1), \underline{S}_2=\underline{s}_2^l) - E(Y_1^{s_2^l} \mid X=(x_1,y_0), \underline{S}_2=\underline{s}_2^l)$$

B) $E(Y_1^{(s_1^l,s_2^k)} \mid X=(x_1,y_0), \underline{s}_2^k) - E(Y_{t'}^{(s_1^l,s_2^k)} \mid X=(x_{t'},y_{t'-1}), \underline{s}_2^k)$

$$= E(Y_1^{(s_1^l,s_2^k)} \mid X=(x_1,y_0),(s_1^l,s_2^k)) - E(Y_{t'}^{(s_1^l,s_2^k)} \mid X=(x_{t'},y_{t'-1}),(s_1^l,s_2^k)) .$$

ASSUMPTION 11-2: (NO "PRE-TREATMENT" EFFECT)

A) $E(Y_1^{(s_1^l,s_2^k)} - Y_1^{s_2^l} \mid X=(x_1,y_0), \underline{S}_2=\underline{s}_2^k)=0$, B) $E(Y_{t'}^{(s_1^l,s_2^k)} - Y_{t'}^{s_2^k} \mid X=(x_{t'},y_{t'-1}), \underline{S}_2=\underline{s}_2^k)=0$ .

Despite these changes, the proof of the identification remains the same. Therefore, even if an endogenous variable is present in the conditioning set, all treatment effects are identifiable. The results of this section are summarized in table 2.

When participation is possible in more than two periods, we need similar assumptions and obtain expression in terms of observable variables for the effects similar to those presented here. For example, if individuals can now participate in a program at time $t=3$, just after the second period, the common trend assumption and the "no pre-treatment" effect assumption should be complemented by a third condition. This condition in the common trend assumption deals with the growth of the outcome between a period after the third participation and period 2 and the no pre-treatment effect assumption deals with the anticipatory effect of the third

[17] This case is not directly treated in the main text because it is included in the third case.

program. The expression we will obtain in this case (under these assumptions and a common support condition valid for the three periods of possible participation case) for the effect of the treatment on the treated would be: [18]

$$\theta_t^{(\underline{s}_3^k),(\underline{s}_3^l)}(\underline{s}_3^k) = E(Y_t \mid \underline{S}_3 = \underline{s}_3^k) - \underset{X_t \mid \underline{S}_3 = \underline{s}_3^k}{E} \left\{ E(Y_t \mid X = x_t, \underline{S}_3 = \underline{s}_3^l) \right\}$$

$$+ \underset{X_t \mid \underline{S}_3 = \underline{s}_3^k}{E} \left\{ E(Y_2 \mid X = x_2, \underline{S}_3 = \underline{s}_3^l) - E(Y_2 \mid X = x_2, \underline{S}_3 = (\underline{s}_2^l, s_3^k)) + E(Y_1 \mid X = x_1, \underline{S}_3 = (\underline{s}_2^l, s_3^k) \right\}$$

$$- \underset{X_t \mid \underline{S}_3 = \underline{s}_3^k}{E} \left\{ E(Y_1 \mid X = x_1, \underline{S}_3 = (s_1^l, s_2^k, s_3^k) - E(Y_{t'} \mid X = x_{t'}, \underline{S}_3 = (s_1^l, s_2^k, s_3^k) + E(Y_{t'} \mid X = x_{t'}, \underline{S}_3 = \underline{s}_3^k) \right\}.$$

Consequently, the generalization to several participations is straightforward.

## 5) Conclusion

In this paper, we analyze the difference-in-differences conditions needed to identify effects of sequences of programs. In such a framework several participations are possible. We follow two different directions. First, we only use the information available in the period previous to any participation (the outcomes at time *t'*) as the base period for the common trend assumptions. These assumptions are needed because we cannot observe the potential average outcomes we are interested in. Then, we use more information and also utilize the outcomes after the first participation to compute those common trend. We have two base periods at our disposal. As we use the outcome after the first participation as comparison, and as this outcome can be influenced by this participation, the number of common trend assumptions required has to be increased as well as the no "pre-treatment" effect assumption. The exact amount of assumptions needed depends on the sequences compared and on the population for which we want to compute the effect. For example, if the sequences have no part in common and we search for the dynamic average treatment effect on the treated, two common trend

[18] $t \geq 3$.

assumptions and two no "pre-treatment" effect assumptions are required. Independently of the approach used, all effects are identified.

We take into account the possible endogeneity problem due to some variables used in the conditioning set (as observable characteristics) that depend on the participations. Nevertheless, no big changes occur and no additional restriction is needed to obtain identification. All effects remain identified.

We do not discuss the estimation of the effects. A nonparametric estimation can be difficult to conduct due to the number of characteristics in the conditioning set. Heckman, Ichimura and Todd (1997), Heckman, Ichimura, Smith and Todd (1998) and Eichler and Lechner (2002) condition on the propensity score,[19] the probability to participate in the program, instead of conditioning on the vector of characteristics to decrease the dimension of the conditioning set. They use different matching methods to estimate the effects. In our multi-periods framework it will be necessary to investigate whether all these methods can be applied and if it is the case, which one is easier to implement. As almost all these methods use the propensity score, it remains to be seen what would be a propensity score in our framework. Under the conditional independence assumption, Lechner (2002) proves that more than one propensity score can be defined and used in the conditioning set. It remains to be seen if his investigations also apply for the difference-in-differences case. It is worthwhile to examine in detail the complexity of the estimation methods because contrary to some other methods (e.g. the conditional independence assumption approach (Lechner and Miquel, 2001) or the instrumental variable approach (Miquel, 2002)) all dynamic effects are identified with the difference-in-differences approach.

[19] This idea comes from the results of Rosenbaum and Rubin (1983).

Table 2: Identified Effects and required assumptions in the sequential approach

| | $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(\underline{s}_2^k)$ | $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(\underline{s}_2^l)$ | $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(\underline{s}_2^j)$ | $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(s_1^j)$ | $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}$ |
|---|---|---|---|---|---|
| **A.5: Common Support**<br>$0 < p_{1\mid t'} < 1,\ \ 0 < p_{1\mid t} < 1,\ \ 0 < p_{2\mid 1,t'} < 1,\ \ 0 < p_{2\mid 1,t} < 1$ | X | X | X | X | X |
| **A. 10: Triangular common trend**<br>$E(Y_t^{\underline{s}_2^k} \mid x_t, \underline{S}_2 = \underline{s}_2^j) - E(Y_1^{\underline{s}_2^k} \mid x_1, \underline{S}_2 = \underline{s}_2^j) = E(Y_t^{\underline{s}_2^k} \mid x_t, \underline{S}_2 = \underline{s}_2^k) - E(Y_1^{\underline{s}_2^k} \mid x_1, \underline{S}_2 = \underline{s}_2^k)$;<br>$E(Y_t^{\underline{s}_2^l} \mid x_t, \underline{S}_2 = \underline{s}_2^j) - E(Y_1^{\underline{s}_2^l} \mid x_1, \underline{S}_2 = \underline{s}_2^j) = E(Y_t^{\underline{s}_2^l} \mid x_t, \underline{S}_2 = \underline{s}_2^l) - E(Y_1^{\underline{s}_2^l} \mid x_1, \underline{S}_2 = \underline{s}_2^l)$;<br>$E(Y_1^{(s_1^k,s_2^j)} \mid x_1, \underline{s}_2^j) - E(Y_{t'}^{(s_1^k,s_2^j)} \mid x_{t'}, \underline{s}_2^j) = E(Y_1^{(s_1^k,s_2^j)} \mid x_1, (s_1^k, s_2^j)) - E(Y_{t'}^{(s_1^k,s_2^j)} \mid x_{t'}, (s_1^k, s_2^j))$;<br>$E(Y_1^{(s_1^l,s_2^j)} \mid x_1, \underline{s}_2^j) - E(Y_{t'}^{(s_1^l,s_2^j)} \mid x_{t'}, \underline{s}_2^j) = E(Y_1^{(s_1^l,s_2^j)} \mid x_1, (s_1^l, s_2^j)) - E(Y_{t'}^{(s_1^l,s_2^j)} \mid x_{t'}, (s_1^l, s_2^j))$ | X | X | X | X | X |
| **A. 11: No "pre-treatment" effect**<br>$E(Y_1^{\underline{s}_2^k} - Y_1^{(s_1^k,s_2^j)} \mid X = x_1, \underline{S}_2 = \underline{s}_2^j) = 0$; $E(Y_1^{\underline{s}_2^l} - Y_1^{(s_1^l,s_2^j)} \mid X = x_1, \underline{S}_2 = \underline{s}_2^j) = 0$;<br>$E(Y_{t'}^{(s_1^k,s_2^j)} - Y_{t'}^{\underline{s}_2^k} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^j) = 0$; $E(Y_{t'}^{\underline{s}_2^l} - Y_{t'}^{(s_1^l,s_2^j)} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^j) = 0$ | X | X | X | X | X |
| **A. 10-1: Common trend**<br>$E(Y_t^{\underline{s}_2^l} \mid x_t, \underline{S}_2 = \underline{s}_2^k) - E(Y_1^{\underline{s}_2^l} \mid x_1, \underline{S}_2 = \underline{s}_2^k) = E(Y_t^{\underline{s}_2^l} \mid x_t, \underline{S}_2 = \underline{s}_2^l) - E(Y_1^{\underline{s}_2^l} \mid x_1, \underline{S}_2 = \underline{s}_2^l)$;<br>$E(Y_1^{(s_1^l,s_2^k)} \mid x_1, \underline{s}_2^k) - E(Y_{t'}^{(s_1^l,s_2^k)} \mid x_{t'}, \underline{s}_2^k) = E(Y_1^{(s_1^l,s_2^k)} \mid x_1, (s_1^l, s_2^k)) - E(Y_{t'}^{(s_1^l,s_2^k)} \mid x_{t'}, (s_1^l, s_2^k))$ | X | X | | | |
| **A. 11-1 No "pre-treatment"**<br>$E(Y_1^{(s_1^l,s_2^k)} - Y_1^{\underline{s}_2^l} \mid X = x_1, \underline{S}_2 = \underline{s}_2^k) = 0$; $E(Y_{t'}^{(s_1^l,s_2^k)} - Y_{t'}^{\underline{s}_2^k} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^k) = 0$ | X | X | | | |

Notes: $p_{1\mid t'} = P(S_1 = 1 \mid X = x_{t'})$, $p_{2\mid 1,t'} = P(S_2 = 1 \mid S_1 = s_1, X = x_{t'})$. The last two assumptions are written to identify the DATET. To identify the DATE on the nontreated, the sequences $k$ and $l$ have to be switched in these assumptions. Due to a problem of space the notation $X = x_{.}$ has been simplified to $x_{.}$ and sometimes the notation $\underline{S}_2 = (s_1^k, s_2^j)$ has been simplified to $(s_1^k, s_2^j)$.

## References


Abadie A., 2001: " Semiparametric Difference-in-differences Estimators", unpublished work.

Angrist J. and A. Krueger, 2000: "Empirical Strategies in Labor Economics", in O. Ashenfelter and D. Card (eds.): Handbook of Labor Economics, North Holland: Elsvier, 1277-1366.

Ashenfelter O., 1978: "Estimating the Effect of Training Programs on Earnings", Review of Economic and Statistics 60, 47-57.

Athey S. and G.W. Imbens, 2002: "Identification and Inference in Nonlinear Difference-in-differences Models", unpublished work.

Bertrand M., E. Duflo and S. Mullainathan, 2001: "How Much Should We Trust Difference-in-differences Estimates?", Massachusetts Institute of Technology Working Paper 01-34 .

Blundell R., M. Costa Dias, C. Meghir and J. Van Reenen, 2001: "Evaluating the Employment Impact of a Mandatory Job Search Assistance Program", IFS Working Paper 01/20.

Blundell R., A. Duncan and C. Meghir, 1998: "Estimating Labor Supply Responses Using Tax Reforms", Econometrica 66, 827-862.

Dee, T.S. and H. Fu, 2001: "Do Charter Schools Skim Students or Drain Resources?", unpublished work.

Donald S.G. and K. Lang, 2001: "Inference with Difference-in-differences and Other Panel Data", unpublished work.

Eichler M. and Lechner M., 2002: "An Evaluation of Public Employment Programmes in the East German State of Sachsen-Anhalt", Labour Economics: An International Journal 9, 143-186.

Finn, C.E., B.V. Manno and G. Vanourek, 2000: "Renewing Public Education : Charter Schools in Action", Princeton, NJ, Princeton University Press.

Heckman J.J., H. Ichimura and P.E. Todd, 1997: "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme", Review of Economic Studies 64, 605-654.

Heckman J.J., H. Ichimura, J.A. Smith and P.E. Todd, 1998: "Characterizing Selection Bias Using Experimental Data", Econometrica 66, 1017-1098.

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, 1865-2097, North-Holland, Amsterdam.

Heckman J.J., and R. Robb, 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., and R. Robb, 1986: "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes", in H. Wainer (ed.): Drawing Inferences from Self-Selected Samples, New-York: Springer-Verlag, 63-107.

Heckman J.J., and J.A. Smith, 1999: "The Pre-Programme Earnings Dip and the Determinants of Participation in a Social Programme. Implications for Simple Programme Evaluation Strategies", Economic Journal 109, 313-348.

Hujer R., M. Caliendo and D. Radic, 2002: "Estimating the Effects of Wage Subsidies on the Labour Demand in West-Germany using the IAB Establishment Panel", unpublished work.

Lechner M., 2001: "A Note on the Common Support Problem in Applied Evaluation Studies", Working Paper 2001-01, University of St. Gallen.

Lechner M., 2002: "Sequential Matching Estimation of Dynamic Causal Models", unpublished work.

Lechner M., and R. Miquel, 2001: "A Potential Outcome Approach to dynamic Programme Evaluation – Part I: Identification", Working Paper 2001-07, University of St. Gallen, with the revised version: "Identification of Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions".

Meyer B.D., W.K. Viscusi and D.L. Durbin, 1995: "Workers' Compensation and Injury Duration: Evidence from a Natural Experiment", American Economic Review 85, 322-323.

Miquel R., 2002: "Identification of Dynamic Treatment Effects by Instrumental Variables", Working Paper 2002-11, University of St. Gallen.

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 Modeling and Applications to Causality", in M. Berkane, (ed.): Lecture Notes in Statistics (120), New York: Springer, 69-117.

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

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.

# Appendix

In all the proofs the conditioning on the history $\underline{S}_0 = \underline{s}_0$ is implicit.

### *A.1) Proof of theorem 1:*

We only prove that the treatment effect on the treated (ATET) is identified because the identification of the treatment effect on the nontreated follows directly from the following:

$$\theta_t^{1,0}(S_1=0) = E\left(Y_t^1 - Y_t^0 \mid X = x_t, S_1 = 0\right) = -E\left(Y_t^0 - Y_t^1 \mid X = x_t, S_1 = 0\right) = -\theta_t^{0,1}(S_1 = 0).$$

$$\theta_t^{1,0}(S_1=1) = E\left(Y_t^1 - Y_t^0 \mid S_1 = 1\right) = E\left(Y_t^1 \mid S_1 = 1\right) - \underset{X_t|S_1=1}{E}\left\{E\left(Y_t^0 \mid X = x_t, S_1 = 1\right)\right\}.$$

The notation $\underset{X_t|S_1=1}{E}(.)$ indicates that the expectation is taken with respect to the conditional (on $S_1=1$) distribution at time $t$ of the $X$. The first expectation can be computed directly from the data, $E\left(Y_t^1 \mid S_1=1\right)=E\left(Y_t \mid S_1=1\right)$. From Assumptions 2 and 3, we can get an expression for the second expectation only formed by quantities computable form the data:

$$E\left(Y_t^0 \mid X=x_t, S_1=1\right) \overset{ass.2}{=} E\left(Y_{t'}^0 \mid X=x_{t'}, S_1=1\right)+E\left(Y_t^0 \mid X=x_t, S_1=0\right)-E\left(Y_{t'}^0 \mid X=x_{t'}, S_1=0\right)$$

$$\overset{ass.3}{=} E\left(Y_{t'}^1 \mid X=x_{t'}, S_1=1\right)+E\left(Y_t \mid X=x_t, S_1=0\right)-E\left(Y_{t'} \mid X=x_{t'}, S_1=0\right)$$

$$=E\left(Y_{t'} \mid X=x_{t'}, S_1=1\right)+E\left(Y_t \mid X=x_t, S_1=0\right)-E\left(Y_{t'} \mid X=x_{t'}, S_1=0\right).$$

Thus, the DATET equals the following expression, computable from the data:

$$E\left(Y_t \mid S_1=1\right) \; - \underset{X_t|S_1=1}{E}\left\{E\left(Y_{t'} \mid X=x_{t'}, S_1=1\right)+E\left(Y_t \mid X=x_t, S_1=0\right)-E\left(Y_{t'} \mid X=x_{t'}, S_1=0\right)\right\}.$$

### *A.2) Proof of theorem 2:*

First, we prove the identification of the following kind of effect, $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(\underline{S}_2=\underline{s}_2^j)$ $= \underset{X_t|\underline{S}_2=\underline{s}_2^j}{E}\left\{E\left(Y_t^{\underline{s}_2^k}-Y_t^{\underline{s}_2^l} \mid X=x_t, \underline{S}_2=\underline{s}_2^j\right)\right\}$. We have to proof that the inner expectation is identified. Then the effect is obtained by averaging this expectation with respect to the conditional distribution of the characteristics at time $t$. From Assumption 6, we obtain the following expressions for the conditional expectations of the potential outcomes:

$$E\left(Y_t^{\underline{s}_2^k} \mid X=x_t, \underline{S}_2=\underline{s}_2^j\right)=E\left(Y_{t'}^{\underline{s}_2^k} \mid X=x_{t'}, \underline{S}_2=\underline{s}_2^j\right)+E\left(Y_t^{\underline{s}_2^k} \mid X=x_t, \underline{S}_2=\underline{s}_2^k\right)-E\left(Y_{t'}^{\underline{s}_2^k} \mid X=x_{t'}, \underline{S}_2=\underline{s}_2^k\right),$$

$$E\left(Y_t^{\underline{s}_2^l} \mid X=x_t, \underline{S}_2=\underline{s}_2^j\right)=E\left(Y_{t'}^{\underline{s}_2^l} \mid X=x_{t'}, \underline{S}_2=\underline{s}_2^j\right)+E\left(Y_t^{\underline{s}_2^l} \mid X=x_t, \underline{S}_2=\underline{s}_2^l\right)-E\left(Y_{t'}^{\underline{s}_2^l} \mid X=x_{t'}, \underline{S}_2=\underline{s}_2^l\right).$$

Building the difference between both terms, the inner expectation equals

$$E\left(Y_t^{\underline{s}_2^k} \mid X=x_t, \underline{S}_2=\underline{s}_2^k\right)-E\left(Y_{t'}^{\underline{s}_2^k} \mid X=x_{t'}, \underline{S}_2=\underline{s}_2^k\right)-E\left(Y_t^{\underline{s}_2^l} \mid X=x_t, \underline{S}_2=\underline{s}_2^l\right)+E\left(Y_{t'}^{\underline{s}_2^l} \mid X=x_{t'}, \underline{S}_2=\underline{s}_2^l\right)$$

$+E\left(Y_{t'}^{\underline{s}_2^k} - Y_{t'}^{\underline{s}_2^l} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^j\right)$. From Assumption 7, the last expectation is zero. Therefore, all these expectations can be computed from the data, and the treatment effect is identified because it is identical to $\underset{X_t|\underline{S}_2=\underline{s}_2^j}{E}\left\{E\left(Y_t \mid X = x_t, \underline{S}_2 = \underline{s}_2^k\right) - E\left(Y_{t'} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^k\right)\right\}$ $-\underset{X_t|\underline{S}_2=\underline{s}_2^j}{E}\left\{E\left(Y_t \mid X = x_t, \underline{S}_2 = \underline{s}_2^l\right) - E\left(Y_{t'} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^l\right)\right\}$.

Secondly, we prove that if assumptions 6 and 7 simultaneously hold for all possible comparisons of sequences, $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(S_1 = s_1^j)$, $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}$ are identified. We can estimate the probability to participate in the second period given the participation in the first period directly from the data and use the following relation $\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(\underline{S}_1 = \underline{s}_1^j) = P(S_2 = 1 \mid \underline{S}_1 = \underline{s}_1^j)\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(\underline{S}_2 = (\underline{s}_1^j, 1)) + (1 - P(S_2 = 1 \mid \underline{S}_1 = \underline{s}_1^j))\theta_t^{\underline{s}_2^k,\underline{s}_2^l}(\underline{S}_2 = (\underline{s}_1^j, 1))$. To identify the average effect for the population we only have to compute a weighted sum of the previous effects using the probability to participate in the program in the first period as weight.

### *A.3) Proof of theorem 3:*

Firstly, we only prove that the treatment effect on the treated is identified. The proof that the treatment effect on the nontreated is identified is based on the same relation as the one in Appendix A.1.

$$\theta_t^{(s_1,1),(s_1,0)}(\underline{S}_2 = (s_1,1)) = E(Y_t^{(s_1,1)} \mid \underline{S}_2 = (s_1,1)) - \underset{X_t|\underline{S}_2=(s_1,1)}{E}\left\{E(Y_t^{(s_1,0)} \mid X = x_t, \underline{S}_2 = (s_1,1))\right\}.$$

The first expectation is estimable directly from the data, $E(Y_t^{(s_1,1)} \mid \underline{S}_2 = (s_1,1)) = E(Y_t \mid \underline{S}_2 = (s_1,1))$. Then, under assumption 8, $E(Y_t^{(s_1,0)} \mid X = x_t, \underline{S}_2 = (s_1,1)) = E(Y_1^{(s_1,0)} \mid X = x_1, \underline{S}_2 = (s_1,1)) + E(Y_t^{(s_1,0)} \mid X = x_t, \underline{S}_2 = (s_1,0))$ $-E(Y_1^{(s_1,0)} \mid X = x_1, \underline{S}_2 = (s_1,0))$. From assumption 9, the first term of the right hand side of the equation equals $E(Y_1^{(s_1,1)} \mid X = x_1, \underline{S}_2 = (s_1,1))$. Therefore, the inner expectation is identical to $E(Y_1 \mid X = x_1, \underline{S}_2 = (s_1,1)) + E(Y_t \mid X = x_t, \underline{S}_2 = (s_1,0)) - E(Y_1 \mid X = x_1, \underline{S}_2 = (s_1,0))$ and is estimable.

Secondly, the identification of the last two effects of the theorem are proved: $\forall s_1 \in \{0,1\}$,

$$\theta_t^{(s_1,1),(s_1,0)}(S_1 = s_1) = P(S_2 = 1 \mid S_1 = s_1)\theta_t^{(s_1,1),(s_1,0)}(\underline{S}_2 = (s_1,1)) + (1 - P(S_2 = 1 \mid S_1 = s_1))\theta_t^{(s_1,1),(s_1,0)}(\underline{S}_2 = (s_1,0)).$$

As the effects on the right hand side of the equation are estimable and as the conditional probability is also estimable from the data, the effect on the participants in $s_1$ is identified.

### *A.4) Proof of theorem 4:*

For the same reason as in the previous proof, as soon as the effects $\theta_t^{s_2^k, s_2^l}(\underline{s}_2^j)$ are identified, the effects $\theta_t^{s_2^k, s_2^l}(\underline{s}_1^j)$ and $\theta_t^{s_2^k, s_2^l}$ are identified. Therefore, we only prove the identification of the first category of effects, $\theta_t^{s_2^k, s_2^l}(\underline{s}_2^j) = \underset{X_t \mid \underline{S}_2 = \underline{s}_2^j}{E} \left\{ E(Y_t^{s_2^k} - Y_t^{s_2^l} \mid X = x_t, \underline{S}_2 = \underline{s}_2^j) \right\}$.

From part A of assumption 10, we obtain:

$E(Y_t^{s_2^k} \mid X = x_t, \underline{S}_2 = \underline{s}_2^j) = E(Y_1^{s_2^k} \mid X = x_1, \underline{S}_2 = \underline{s}_2^j) + E(Y_t^{s_2^k} \mid X = x_t, \underline{S}_2 = \underline{s}_2^k) - E(Y_1^{s_2^k} \mid X = x_1, \underline{S}_2 = \underline{s}_2^k)$ and

$E(Y_t^{s_2^l} \mid X = x_t, \underline{S}_2 = \underline{s}_2^j) = E(Y_1^{s_2^l} \mid X = x_1, \underline{S}_2 = \underline{s}_2^j) + E(Y_t^{s_2^l} \mid X = x_t, \underline{S}_2 = \underline{s}_2^l) - E(Y_1^{s_2^l} \mid X = x_1, \underline{S}_2 = \underline{s}_2^l)$.

The first term of the right hand side of each equation is unobservable. Nevertheless, from part A of assumption 11, the first equation can be written as

$$E(Y_t^{s_2^k} \mid X = x_t, \underline{S}_2 = \underline{s}_2^j) = E(Y_1^{(s_1^k, s_2^j)} \mid X = x_1, \underline{S}_2 = \underline{s}_2^j) + E(Y_t \mid X = x_t, \underline{S}_2 = \underline{s}_2^k) - E(Y_1 \mid X = x_1, \underline{S}_2 = \underline{s}_2^k)$$

and the second equation as

$$E(Y_t^{s_2^l} \mid X = x_t, \underline{S}_2 = \underline{s}_2^j) = E(Y_1^{(s_1^l, s_2^j)} \mid X = x_1, \underline{S}_2 = \underline{s}_2^j) + E(Y_t \mid X = x_t, \underline{S}_2 = \underline{s}_2^l) - E(Y_1 \mid X = x_1, \underline{S}_2 = \underline{s}_2^l).$$

Then, using part B of assumption 10,

$$E(Y_t^{s_2^k} \mid X = x_t, \underline{S}_2 = \underline{s}_2^j) = E(Y_{t'}^{(s_1^k, s_2^j)} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^j) + E(Y_1^{(s_1^k, s_2^j)} \mid X = x_1, \underline{S}_2 = (s_1^k, s_2^j))$$

$$- E(Y_{t'}^{(s_1^k, s_2^j)} \mid X = x_{t'}, \underline{S}_2 = (s_1^k, s_2^j)) + E(Y_t \mid X = x_t, \underline{S}_2 = \underline{s}_2^k) - E(Y_1 \mid X = x_1, \underline{S}_2 = \underline{s}_2^k) \text{ and}$$

$$E(Y_t^{\underline{s}_2^l} \mid X = x_t, \underline{S}_2 = \underline{s}_2^j) = E(Y_{t'}^{(s_1^l, s_2^j)} \mid X = x_{t'}, \underline{S}_2 = \underline{s}_2^j) + E(Y_1^{(s_1^l, s_2^j)} \mid X = x_1, \underline{S}_2 = (s_1^l, s_2^j))$$

$$- E(Y_{t'}^{(s_1^l, s_2^j)} \mid X = x_{t'}, \underline{S}_2 = (s_1^l, s_2^j)) + E(Y_t \mid X = x_t, \underline{S}_2 = \underline{s}_2^l) - E(Y_1 \mid X = x_1, \underline{S}_2 = \underline{s}_2^l) .$$

Finally, building the difference and using the relations of part B of assumption 11,

$$E(Y_t^{\underline{s}_2^k} - Y_t^{\underline{s}_2^l} \mid X = x_t, \underline{S}_2 = \underline{s}_2^j) = E(Y_t \mid X = x_t, \underline{S}_2 = \underline{s}_2^k) - E(Y_1 \mid X = x_1, \underline{S}_2 = \underline{s}_2^k)$$

$$+ E(Y_1 \mid X = x_1, \underline{S}_2 = (s_1^k, s_2^j)) - E(Y_{t'} \mid X = x_{t'}, \underline{S}_2 = (s_1^k, s_2^j)) - E(Y_t \mid X = x_t, \underline{S}_2 = \underline{s}_2^l)$$

$$+ E(Y_1 \mid X = x_1, \underline{S}_2 = \underline{s}_2^l) - E(Y_1 \mid X = x_1, \underline{S}_2 = (s_1^l, s_2^j)) + E(Y_{t'} \mid X = x_{t'}, \underline{S}_2 = (s_1^l, s_2^j)) .$$

Thus, the effects are identified, because they are functions only of observable quantities.