A Potential Outcome Approach to Dynamic Programme Evaluation – Part I: Identification

Michael Lechner and Ruth Miquel [*]

SIAW

Mai 2001

Date this version has been printed: 01 May 2001

Comments welcome

Addresses for correspondence

Michael Lechner

Professor of Econometrics

Swiss Institute for International Economics and Applied Economic Research (SIAW)

University of St. Gallen

Dufourstr. 48, CH-9000 St. Gallen, Switzerland

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

---

[*] The first author is also affiliated with CEPR, London, ZEW, Mannheim, and IZA, Bonn. Financial support from the Swiss National Science Foundation (projects 4043-058311 and 4045-050673) and the IAB, Nürnberg, is gratefully acknowledged. We thank participants at the Econometrics seminars in Strasbourg and St. Gallen for helpful comments and suggestions. All remaining errors are our own.

Abstract

This paper approaches the problem of an econometric evaluation of dynamic programme sequences from an potential outcome perspective. The identifying power of several different assumptions about the connection between the dynamic selection process and the potential outcomes of different programme sequences is discussed. The assumptions invoke different types of randomisation compatible with different selection regimes. Parametric forms are not involved. When participation in the sequences is decided every period depending on the success in the past, the resulting endogeneity problem destroys nonparametric identification for many parameters of interest, so that several dynamic versions of the average treatment effects on the treated parameter are not identified. However, some interesting dynamic forms of the average treatment effect are still identified. We also present a bounds analysis to learn from the data as much as possible, even when parts of the identifying assumptions are violated.

Keywords

Dynamic treatment regimes, nonparametric identification, Rubin causal model, sequential randomisation

JEL classification: C40

1 Introduction

There is considerable interest in the econometric evaluation of labour market programmes. Overviews of the activities in this field can be found in Angrist and Krueger (1999), Heckman, LaLonde and Smith (1999) and Lechner and Pfeiffer (2001). Many recent contributions to this literature use explicitly a 'causal model' of potential outcomes that is typically associated with Neyman (1923), Roy (1951) and Rubin (1974). In this model causal inference relates to the question about what would happen on average in one hypothetical situation (e.g. participating in the programme) compared to another situation (e.g. not participating in the programme[1]) for a particular population. It is an advantage of this so-called counterfactual framework that it clarifies the distinction between associations and causal statements. The latter typically require some additional knowledge for identification that is not verifiable from the data. The recent literature seems to concentrate on ways to obtain knowledge about causal effects without imposing strong parametric modelling assumptions on the link between assignment and outcomes used to measure the success of specific treatments.[2] It is an important feature of this strand of the literature that the causal framework used is basically a static one. The static framework provides a considerable simplification as compared to a dynamic setting and allows the derivation of exiting results with respect to nonparametric identification and robust estimation.

Unfortunately, although a static model may provide a good approximation in some cases, it is generally insufficient for the analysis of many active labour market (or other) policies, because many such policies consist in sequential strategies of assigning people to different subprogrammes. In many cases the effects of the subprogrammes in which the individual participated so far have an influence on the choice of the next subprogramme. For example, an unemployed is first assisted with job search. If she remains unemployed, then she is send to a training programme. If she still remains unemployed, then she is allocated to an employment programme, and so on. Obviously any static causal framework has a very hard time, i.e. needs a lot of simplifying assumptions, to even be able to define the interesting questions, not to mention the ability

---

1 To stick to the terminology of that literature that has strong links to statistics and biometrics we will use the term *treatment* from now on as a substitute for *programme*.

2 Recent examples are papers by Imbens and Angrist (1994) for the identifying power of instrumental variables, Heckman and Vytlacil (2001b) and Vytlacil (1999) for nonparametric selection models, and by Rubin (1977) and Lechner (2001b) for the conditional independence assumption.

to discuss identification of parameters of interest, namely the effects of different strategies (sequences of treatments).

One way to address these issues would be to 'return' to structural parametric econometric models for panel data, like duration models (see van den Berg, 2000) or regression models of some sort (see Baltagi, 1995). These models can in principle accommodate dynamics and selectivity in a fairly direct way. It appears however that they have not (yet) been used to tackle the problem of causal effects of dynamic treatment regimes explicitly. When they are used (e.g. Lalive, van Ours, and Zweimüller, 2000), there is no explicit definition of the causal effects being estimated. In the light of the literature on nonparametric identification of static causal models, the dependence of these models on parametric functional form assumptions to achieve identification is a undesirable feature of this approach. Yet, once identification is achieved nonparametrically, these models can play an important part in reducing the dimensionality problem in estimating various conditional expectations of observables variables in dynamic treatment settings. Using parametric models only at the estimation step has the advantage that the fundamental economic relations that do lead or do not lead to identification are not obscured by some functional form assumptions that always come with using models at the identification stage.

There is a related literature in epidemiology and biostatistics that uses counterfactual outcomes explicitly in a dynamic setting (e.g. Robins, 1986, 1987, 1997, Robins, Greenland, and Hu, 1999). These papers typically assume some sort of sequential randomisation, which will also play a role in this paper as a benchmark case. However, their assumptions (and also their notation) are motivated from applications in biostatistics, do not always correspond to assumptions that would seem to be appropriate for many econometric applications. We reformulate and extend their framework to arrive at new and interesting results.

In this paper we generalise the econometric counterfactual static causal evaluation model to a dynamic context. By drawing on recent extension of the static model to multiple treatments (Imbens, 2000, Lechner, 2001a), and on the literature on bounding treatment effects (Balke and Pearl, 1997, Heckman and Vytlacil, 2001a, Lechner, 1999, Manski, 1989, 1990, Robins, 1989, Pearl, 2000), we provide a framework to discuss different definitions of the effects of interest and their nonparametric identification under various assumptions relating to the content of the information available to the researcher about the connection between the assignment decision and the outcomes of the treatments. In doing so we explicitly avoid parametric or semiparametric modelling assumptions to achieve identification and explore how far our extensions of the nonparametric literature based on the static model will lead us without having to rely on parametric models for identification.

Specifically, we explore the potential of several assumptions of the sequential randomisation variant. These assumptions basically suppose that the current assignment to a programme can be treated as independent of the (potential) outcomes conditional on some information set. They differ with respect to the important notion of which information set to condition on. In fact in any application the information set required to make these assumptions plausible defines the variables that need to be observed in the sample. We show what kind of effects are identified using assumptions that could be plausible for example in rich data sets (and institutional circumstances) sometime found in labour economics.[3] Furthermore, we also derive bounds for cases where specific assumption are partially violated.

The following section defines the notation as well as effects of interest in a dynamic treatment setting. Section 3 proposes several identification strategies and discusses their identifying power with respect to the parameter defined in Section 2. In Section 4 nonparametric bounds are developed for some of the cases when these assumptions are not sufficient to identify particular parameters of interest. Finally, Section 5 summarises our main findings and concludes. Appendix A contains the proofs of the lemma and theorems. Appendix B and C contain further expressions and derivations for bounds derived under possible violations of the identifying assumptions not considered in Section 4.

2 The notation of the dynamic causal model

2.1 The variables

We consider a world with a finite number of $T + 1$ discrete time periods indexed by t ($t = 0,1,...,T$). The particular treatment received by the population of interest in the beginning of

[3] In fact we plan to apply these identification results to the estimation of the effects dynamic regimes in active labour market policies in Switzerland and Germany. For these countries rich administrative data sets that allow to address these issues are available.

each period is described by a $T + 1$ dimensional vector of random variables $S = (S_0, ..., S_T)$.[4] Starting in period one each treatment in the particular period can take a finite number of M values ($m = 1, ..., M$). In period 0 everybody receives the same treatment $S = (0, S_1, ..., S_T)$. A particular realisation of S in period t is denoted s_t. Furthermore denote the history of variables up to period t by a bar below a variable, e.g. $\underline{s}_t = (0, s_1, ..., s_t)$.[5] Therefore, for each period t there are M^t (t is not an index, but an exponent) different possibilities of sequences of treatments so far. In the last period a member of the population is observable in exactly one of M^T treatment sequences. This notation allows also to specify incomplete or partial sequences by considering sequence $\underline{s}_t$ with $t < T$. The importance of this distinction becomes clear during the discussion of the definition of the effects of the treatments and their identification. To sum up, we consider $\sum_{t=1}^{T} M^t = \frac{M(M^T - 1)}{M - 1}$ different potential outcomes.[6] Every individual belongs to exactly T potential outcomes that overlap over time.

Variables used to measure the effects of the treatment, i.e. the potential outcomes, are also indexed by the treatments and denoted by $Y^{\underline{s}_t} = (Y_0^{\underline{s}_t}, Y_1^{\underline{s}_t}, ..., Y_T^{\underline{s}_t})$. After t and before $t + 1$, i.e. after the treatment of period t and before the next treatment, one of those potential outcomes is observable and denoted by Y_t. The resulting observation rule is defined in equation (1):[7]

$$Y_t = \sum_{j=1}^{M} \underline{1}(\underline{s}_1 = j) Y_t^{\underline{s}_1^j} = ... = \sum_{j=1}^{M^t} \underline{1}(\underline{s}_t = j) Y_t^{\underline{s}_t^j} = ... = \sum_{j=1}^{M^T} \underline{1}(\underline{s}_T = j) Y_t^{\underline{s}_T^j} . \tag{1}$$

Having defined treatment status and potential outcomes it remains to define variables that influence treatment selection and (or) potential outcomes, so called *attributes* and *potential confounders*, denoted by X. In general we allow the treatment status to influence the values of these variables (introducing some *endogeneity*, a term to be defined later on). Therefore, there are

---

4 We avoid the technical term *units* for members of the population. In resemblance to our motivating application we call them *individuals* instead. Generally the notation set-up in this section follows the spirit of Rubin (1974) and others.

5 To differentiate between different sequences sometimes it will be convenient to assign a specific value (e.g. j) to the specific sequence or to index a sequence like $\underline{s}_t^j$.

6 Note that this is a geometric sequence.

7 The *observation rule* is similar to the *consistency condition* used by Robins in several papers.

potential values of the X variables depending on the treatment sequences, so that they have also to be indexed by the treatments ($X^{\underline{s}_\tau} = (X_0^{\underline{s}_\tau}, X_1^{\underline{s}_\tau}, ..., X_T^{\underline{s}_\tau})$). The K dimensional vector X_t can be observed at the same time as Y_t. The corresponding observation rule for X_t is analogous to the one given in equation (1). It is convenient to collect the potentially endogenous variables Y and X, as well as their potential counterparts in *(K+1) x T* dimensional matrices denoted by $H = (H_0, H_1, ..., H_T)$, $H_t = (Y_t, X_t')'$, and $H^{\underline{s}_\tau} = (H_0^{\underline{s}_\tau}, H_1^{\underline{s}_\tau}, ..., H_T^{\underline{s}_\tau})$.

We call a variable *endogenous*, if its potential outcomes differ across treatment states (e.g. V is called *endogenous* if $V_t \neq V_t^{\underline{s}_\tau}$ for some t and $\underline{s}_\tau$). In other words, at least one of the treatments has a causal effect on an endogenous variable. Similarly, we denote a variable as *strictly exogenous* if no treatment has any effect on it (e.g. V is called strictly exogenous if $V_t = V_t^{\underline{s}_\tau}$ for all t, τ and all $\underline{s}_\tau$). An equivalent statement is that every treatment is non-causal for a strictly exogenous variable.

Example

The following restricted version of the general framework with $M = T = 2$ should help clarifying the ideas of the paper. The treatments variables are now defined as follows:

$$S = (0, S_1, S_2),\ \underline{s}_1 = (0, s_1),\ \underline{s}_2 = (0, s_1, s_2),\ \underline{s}_1 \in \{(0,0),(0,1)\},\ \underline{s}_2 \in \{(0,0,0),(0,1,0),(0,0,1),(0,1,1)\}.$$

In total there are two (2^1) different treatment sequences in the first period, and four (2^2) different treatment sequences in the second period. Thus there are $\frac{2(2^2-1)}{2-1} = 6$ potential outcomes. The observation rule simplifies to:

$$Y_t = s_1 Y_t^{1\cdot} + (1-s_1) Y_t^{0\cdot} = s_1 s_2 Y_t^{11} + s_1(1-s_2) Y_t^{10} + (1-s_1)(1-s_2) Y_t^{00} + (1-s_1) s_2 Y_t^{01}.$$

The changes in the remaining variables $Y^{\underline{s}_\tau}$, $X^{\underline{s}_\tau}$, $H^{\underline{s}_\tau}$, Y, X, H are obvious.

Before defining causal effects in detail we assume that all random variables introduced so far have finite support.

Assumption 1-I (finite support)

The joint support of all random variables is defined on a closed set.

As a short hand notational convention the upper and lower bounds of the support of a vector of random variables A is denoted by ${}_{L}A$ and ${}^{U}A$. In practise this assumption is only in rare cases restrictive, because one could usually assume that the lower bound is a very large, but finite negative number, and the upper bound is a very large, but finite positive number. ASSUMPTION 1-I has the virtue that we need not be concerned about the existence of (unconditional) moments of the various random variables.

2.2 The effects

The purpose of the desired empirical analysis is to estimate the mean causal effect of a sequence of treatments defined up to period τ ($\underline{s}_{\tau}^{k}$) compared to another sequence of the same length ($\underline{s}_{\tau}^{l}$) for a particular population and for a period *t*, denoted by $\theta_{t}^{\underline{s}_{\tau}^{k},\underline{s}_{\tau}^{l}}$ $(t \geq \tau)$. Note that we consider only pair-wise effects of sequences that have the same length. Comparing sequences of different lengths does not appear to be attractive, because it amounts to a comparisons of some well-defined sequence with another sequence that is only partially specified and left undetermined for some periods before period τ (each sequence contains all possible treatments - including the state of no treatment - up to that period). The other restriction implied in the notation is to only consider effects of treatments after they have happened. This is not restrictive at all however, because the vector of potential treatments could be redefined such as to include future participation as a current treatment status (for example to allow anticipatory effects of future treatments on current outcomes, like the famous Ashenfelter's, 1978, dip story).

The question remains for what subpopulations to consider mean causal effects. Trivially, mean causal effects can be constructed for all subpopulations defined by variables that do not depend on the treatment (i.e. they are *strictly exogenous*). All what follows is also valid in strata defined by these variables. The other extreme, namely a causal statement that is conditional (partially) on the effects of the treatments, does not make much sense. Thus, conditioning on *H* in general is inappropriate without exogeneity assumptions about *H*. However, conditioning on treatment status (instead of treatment outcome) is appropriate, because it allows the typical comparisons of (many different) effects of treatment on the treated compared to treatment on the nontreated. For consistency reasons we consider only subpopulations defined by treatment status going not beyond the last period of the specified treatment sequence. In other words, to avoid further complications we do not consider the effects of a treatment in the past conditional on future treatment status. If a comparison for such a group is desired one may always prolong the specified sequence of treatments to match the period used for defining the group. The definition of the average causal effects is given in equation (2):

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

$$\tilde{\tau} \le \tau \le t \le T,\ k \ne l,\ k,l \in \{1,...,M^\tau),\ j \in \{1,...,M^{\tilde{\tau}}).\qquad (2)$$

Note that for $\tilde{\tau} = 0$ we obtain the average effect for the population. To ensure that equation (2) can be interpreted as a causal effect the standard assumptions of the Rubin (1974) model are also necessary, like SUTVA. They mainly imply that the effects of treatment on person *i* does not depend on the treatment choice of other people. Note the resemblance of the effects to effects that are typically of interest in the static evaluation literature, namely the average treatment effect (ATE) and the average treatment effect on the treated (ATET). Here, we may call $\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}$ the dynamic average treatment effect (DATE). Accordingly, $\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_\tau^k)$ as well as $\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_\tau^l)$ may be termed dynamic average treatment effects on the treated (DATET) and the nontreated (i.e. persons with $\underline{S}_\tau = \underline{s}_\tau^l$).

Table 1 summarises the notation as well as definitions introduced so far. Obviously there are many cases in between when the condition set consists of a shorter sequence than the ones evaluated, but we refrain from inventing other names for these effects and will call them all DATET instead.

Table 1: Summary of notation and definitions

Symbol	Meaning	Timing
$t = 0,1,...,T$	time periods	--
$S = (0, S_1,..., S_T)$	RV: treatment	begin of period
$\underline{s}_t = (0, s_1,..., s_t)$	specific sequence of treatments until period *t*	begin of period
$s_t \in \{1,..., M\}$	*M* exclusive treatments in each period	begin of period
$Y^{\underline{s}_t} = (Y_0^{\underline{s}_t}, Y_1^{\underline{s}_t},..., Y_T^{\underline{s}_t})$	RV: potential outcomes	end of period
$Y = (Y_0, Y_1,..., Y_T)$	RV: observable outcomes	end of period
$X^{\underline{s}_t} = (X_0^{\underline{s}_t}, X_1^{\underline{s}_t},..., X_T^{\underline{s}_t})$	RV: potential confounders	end of period
$X = (X_0, X_1,..., X_T)$	RV: observable confounders	end of period
$H^{\underline{s}_t} = (H_0^{\underline{s}_t}, H_1^{\underline{s}_t},..., H_T^{\underline{s}_t})$, $H_t^{\underline{s}_t} = (Y_t^{\underline{s}_t}, X_t^{\underline{s}_t}{}')'$	RV: collection of potential outcomes and confouders	end of period
$H = (H_0, H_1,..., H_T)$, $H_t = (Y_t, X_t{}')'$	RV: collection of observable outcomes and confouders	end of period
$\theta_t^{\underline{s}_\tau^k, \underline{s}_\tau^l}(\underline{s}_{\tilde{\tau}}^j)$	mean causal effect of sequence $\underline{s}_\tau^k$ compared to sequence $\underline{s}_\tau^l$ for those participating in $\underline{s}_{\tilde{\tau}}^j$	end of period

RV: Random variable.

2.3 Connection of effects and outcomes in different periods and sequences

Before using this notation to discuss identification of the effects, it is useful to derive some connection between the different treatment effects $\theta_t^{\underline{s}_\tau^k, \underline{s}_\tau^l}(\underline{s}_{\tilde{\tau}}^j)$. It is of particular interest to see the relation between different lengths of the conditioning sets $\underline{s}_{\tilde{\tau}}^j$ and $\underline{s}_{\tilde{\tau}+\delta}^j$ ($\delta \in \mathbb{N}^+, (\tilde{\tau}+\delta) \le t$), when both sets coincide until period $\tilde{\tau}$.

Lemma 1 (connection of treatment effects defined for different lengths of conditioning set)

$$\theta_t^{\underline{s}_\tau^k, \underline{s}_\tau^l}(\underline{s}_{\tilde{\tau}}^j) = \sum_{m_1=1}^{M} \dots \sum_{m_\delta=1}^{M} \theta_t^{\underline{s}_\tau^k, \underline{s}_\tau^l}(\underline{s}_{\tilde{\tau}}^j, \underbrace{m_1}_{s_{\tilde{\tau}+1}}, ..., \underbrace{m_\delta}_{s_{\tilde{\tau}+\delta}}) P[\underline{S}_{\tilde{\tau}+\delta} = (\underline{s}_{\tilde{\tau}}^j, \underbrace{m_1}_{s_{\tilde{\tau}+1}}, ..., \underbrace{m_\delta}_{s_{\tilde{\tau}+\delta}}) \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j].$$

The proof is direct by applying the definitions of the treatment effect.

Because treatments are observable, $P(\underline{S}_{\tilde{\tau}+\delta} = \underline{s}_{\tilde{\tau}+\delta}^j \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j)$ is identified. Thus, LEMMA 1 implies that if the treatment effects are identified on 'finer subpopulations' given by $\underline{s}_{\tilde{\tau}+\delta}^j$, they are also identified on the 'coarser subpopulation' defined by $\underline{s}_{\tilde{\tau}}^{j}$ and can be computed as a weighted average.

Before similar relationships are derived for the same conditioning set but different treatments (again coinciding until period τ), the following relations among potential outcomes from different periods are useful:[8]

$$Y_t^{\underline{s}_\tau^k} = \sum_{m=1}^{M} \underline{1}[(S_{\tau+1} = m)] Y_t^{(\underline{s}_\tau^k, m)} = \sum_{m_2=1}^{M} \sum_{m_1=1}^{M} \underline{1}[(S_{\tau+2} = m_2)] \underline{1}[(S_{\tau+1} = m_1)] Y_t^{(\underline{s}_\tau^k, m_1, m_2)} = \ldots \tag{3}$$

Equation (3) states that the individual potential outcome of a shorter sequence specified up to period τ is the same as the potential outcome of a sum of possible potential outcomes for longer sequences consisting of the same subsequence of treatments up to period *t*.

Equation (3) can be used to establish a connection between the treatment effects over time ($\theta_t^{\underline{s}_\tau^k, \underline{s}_\tau^l}(\underline{s}_{\tilde{\tau}}^j)$, $\theta_t^{\underline{s}_{\tau+\delta}^k, \underline{s}_{\tau+\delta}^l}(\underline{s}_{\tilde{\tau}}^j, \cdot, s_{\tau+1}, \ldots, s_{\tau+\delta})$) in LEMMA 2-I.

Lemma 2-I (connection of treatment effects defined for different lengths of treatments)

$$\theta_t^{\underline{s}_\tau^k, \underline{s}_\tau^l}(\underline{s}_{\tilde{\tau}}^j) = \sum_{m=1}^{M} P(S_{\tau+1} = m \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j)\, \theta_t^{(\underline{s}_\tau^k, m), (\underline{s}_\tau^l, m)}(\underline{s}_{\tilde{\tau}}^j, \cdot, s_{\tau+1} = m) =$$

$$= \sum_{m_1=1}^{M} \ldots \sum_{m_\delta=1}^{M} P(S_{\tau+1} = m_1, \ldots, S_{\tau+\delta} = m_\delta \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j)\, \theta_t^{(\underline{s}_\tau^k, m_1, \ldots, m_\delta), (\underline{s}_\tau^l, m_1, \ldots, m_\delta)}(\underline{s}_{\tilde{\tau}}^j, \cdot, s_{\tau+1} = m_1, \ldots, s_{\tau+\delta} = m_\delta).$$

The proof is contained in Appendix A.1.

Note there is a minor complication for cases when τ and $\tilde{\tau}$ do not coincide (i.e. $\tilde{\tau} < \tau$), because then the effect of the sequence is conditioned on an unspecified part. In slight abuse of our previously introduced notation the unspecified part is denoted by bullets in LEMMA 2.

[8] Note that these relations hold for all elements of *H*, not just the potential outcomes, but also the potential confounders.

LEMMA 2 implies that we can infer the effects of shorter sequences from comparisons of all longer sequences that are exactly the same after period τ weighted by their probability of occurrence. This is intuitive, because if a pair of sequences is the same after period τ, the causal effect of their difference must be entirely due to differences of the sequences until period τ.

The properties derived so far are useful to obtain the identification results as well as the bounds in the next sections.

Example (continued - 1)

The following 48 pair-wise effects can be defined for the example:

i) $\theta_1^{(0,0,\cdot),(0,1,\cdot)}(0)$, $\theta_2^{(0,0,\cdot),(0,1,\cdot)}(0)$

$\theta_2^{(0,0,0),(0,1,0)}(0)$, $\theta_2^{(0,0,0),(0,0,1)}(0)$, $\theta_2^{(0,0,0),(0,1,1)}(0)$, $\theta_2^{(0,0,1),(0,1,0)}(0)$, $\theta_2^{(0,0,1),(0,1,1)}(0)$, $\theta_2^{(0,1,0),(0,1,1)}(0)$;

(ii) $\theta_1^{(0,0,\cdot),(0,1,\cdot)}(\underline{s}_1)$, $\theta_2^{(0,0,\cdot),(0,1,\cdot)}(\underline{s}_1)$

$\theta_2^{(0,0,0),(0,1,0)}(\underline{s}_1)$, $\theta_2^{(0,0,0),(0,1,1)}(\underline{s}_1)$, $\theta_2^{(0,0,0),(0,1,1)}(\underline{s}_1)$, $\theta_2^{(0,0,1),(0,1,0)}(\underline{s}_1)$, $\theta_2^{(0,0,1),(0,1,1)}(\underline{s}_1)$,

$\theta_2^{(0,1,0),(0,1,1)}(\underline{s}_1)$, $\quad \underline{s}_1 \in \{(0,0,\cdot),(0,1,\cdot)\}$;

(iii) $\theta_2^{(0,0,0),(0,1,0)}(\underline{s}_2)$, $\theta_2^{(0,0,0),(0,1,1)}(\underline{s}_2)$, $\theta_2^{(0,0,0),(0,1,1)}(\underline{s}_2)$, $\theta_2^{(0,0,1),(0,1,0)}(\underline{s}_2)$, $\theta_2^{(0,0,1),(0,1,1)}(\underline{s}_2)$,

$\theta_2^{(0,1,0),(0,1,1)}(\underline{s}_2)$, $\quad \underline{s}_2 \in \{(0,0,0),(0,1,0),(0,0,1),(0,1,1)\}$.

With respect to different conditioning sets used for the definition of the various effects, the connection between the effects simplify as follows:

$$\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(0) = \theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(0,0)P[S_1 = 0] + \theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(0,1)P[S_1 = 1];$$

$$\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_1) = \theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_1,0)P[S_2 = 0 \mid \underline{S}_1 = \underline{s}_1] + \theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_1,1)P[S_2 = 1 \mid \underline{S}_1 = \underline{s}_1].$$

Similar simplifications can be obtained for the connection of different potential outcomes using the observation rule:

$$Y_t^{\underline{s}_1} = S_2 Y_t^{(\underline{s}_1,1)} + (1 - S_2) Y_t^{(\underline{s}_1,0)}.$$

Hence, we get the following relations for the expected potential outcomes:

$$E(Y_t^{\underline{s}_1^k} \mid \underline{S}_1 = \underline{s}_1^j) = P(S_2 = 1 \mid \underline{S}_1 = \underline{s}_1^j) E(Y_t^{(\underline{s}_1^k,1)} \mid \underline{S}_1 = \underline{s}_1^j, S_2 = 1) + [1 - P(S_2 = 1 \mid \underline{S}_1 = \underline{s}_1^j)] E(Y_t^{(\underline{s}_1^k,0)} \mid \underline{S}_1 = \underline{s}_1^j, S_2 = 0).$$

This result connects the different effects:

$$\theta_t^{\underline{s}_1^k,\underline{s}_1^l}(\underline{s}_1^j) = P(S_2 = 1 \mid \underline{S}_1 = \underline{s}_1^j)\theta_t^{(\underline{s}_1^k,1),(\underline{s}_1^l,1)}(\underline{s}_1^j,1) + [1 - P(S_2 = 1 \mid \underline{S}_1 = \underline{s}_1^j)]\theta_t^{(\underline{s}_1^k,0),(\underline{s}_1^l,0)}(\underline{s}_1^j,0) .$$

---

3 Identification

Suppose we have an indefinitely large random sample from the population and observe $\{s_i, y_i, x_i\}_{i=1,...,N}$.[9] In this section we explore different assumptions that identify various treatment effects with such data and assumption that might be appropriate in some econometric evaluation studies based on very informative data.

3.1 Nondynamic treatment regimes

Let us assume that in addition to ASSUMPTION 1 the researcher has enough information in the beginning of the initial period so that assignment to the treatment in every period can be treated as random conditional on that information. In other words, this case is equivalent to a scheme where the assignment of all treatments is made in the initial period and is subsequently *not* changed when new information arrives.

Assumption 2-I (full conditional independence assumption, F-CIA)

a) $Y^{\underline{s}_t},...,Y^{\underline{s}_\tau} \coprod S_t \mid X_0 = x_0$; [10]

b) $1 > P(\underline{S}_t = \underline{s}_t \mid X_0 = x_0) > 0$; $\quad \forall x_0 \in \chi_0$, $\forall \underline{s}_t \in \mathbb{S}_t$, $\forall t : 1 \le t \le \tau$, $\tau \le T$.

Part a) of ASSUMPTION 2-I states that conditional on some exogenous variables X_0 all potential outcomes are independent of all assignments up to the last period for which the treatment is specified. This assumption is valid within a set defined by exogenous characteristics χ_0. Part b)

---

9 As a convention, capital letters usually denote random variables, whereas small letters denote specific values of the random variable. When we deviate from this convention it will be obvious (like for *T* and *t*).

10 $A \coprod B \mid C = c$ means that each element of the vector of random variables *A* is independent of the random variable *B* conditional on the random variable C taking a value of c.

is the usual *common support requirement* (CSR), basically stating that all sequences to be evaluated must have a positive probability of occurring in all strata defined by the values of x_0 that are in the set of interest χ_0.[11] Obviously, only sequences that are feasible can be evaluated (otherwise they would have zero probability anyway and violate the support condition). Such sequences defined up to period *t* are elements of the set $\underline{\mathbb{S}}_t$. Assumption 2-I implies that subsequent treatment choices do not depend on the outcomes of the treatments in the previous periods, because the complete treatment sequence is chosen initially based on the information contained in X_0. New information revealed in later periods does not play any role for the selection into the treatment. In the language of biometrics as suggested for example by Murphy, van der Laan, Robins, and CPPRG (2000), such a mechanism is appropriately called a NONDYNAMIC TREATMENT REGIME. Note also that this assumption corresponds to the CIA assumption of the static multiple treatment framework as discussed by Imbens (2000) and Lechner (2001a), where the treatment assignment is independent of the potential outcomes conditional on some exogenous variables. In these models assumptions like 2-I identify all effects.[12]

The full conditional independence assumption (F-CIA) can also be written in a sequential way, so that the potential outcomes are independent of treatment choice conditional on previous treatments. An interpretation is that the assignment decision for the next period is made in the beginning of the period taking into account treatment history and the initial information (and also of course other information would be allowed if it is conditionally independent of the potential outcomes).

Assumption 2-II (F-CIA – sequential version)

a) $Y^{\underline{S}_1},...,Y^{\underline{S}_T} \coprod S_t \mid \underline{S}_{t-1} = \underline{s}_{t-1}, X_0 = x_0$;

b) $1 > P(\underline{S}_t = \underline{s}_t \mid X_0 = x_0) > 0$; $\quad \forall x_0 \in \chi_0$, $\forall \underline{s}_t \in \underline{\mathbb{S}}_t$, $\forall t: 1 \leq t \leq \tau$, $\tau \leq T$.

---

[11] Strictly speaking all effects discussed below are defined on the set χ_0 (support). However, for notational convenience this dependence will not be made explicit in the following. For some considerations about the relevance of this assumption see Lechner (2001c).

[12] The multiple treatments in their framework correspond to the particular sequences here.

Again, this assumption implies a nondynamic treatment regime because any new information related to the outcomes that might be revealed after period 0 does not play any role in the selection process. Thus, it is not surprising that LEMMA 3 states that both assumptions are statistically identical.

Lemma 3

ASSUMPTIONS 2-I and 2-II are statistically identical.

The proof of LEMMA 3 is given in Appendix A.2.

Example (continued - 2)

Part a) of ASSUMPTION 2-I (F-CIA) translates to the example as follows:[13]

$$Y^{(0,0)}, Y^{(0,1)}, Y^{(0,0,0)}, Y^{(0,1,0)}, Y^{(0,0,1)}, Y^{(0,1,1)} \amalg S_1 \mid X_0 \text{ and } Y^{(0,0,0)}, Y^{(0,1,0)}, Y^{(0,0,1)}, Y^{(0,1,1)} \amalg S_2 \mid X_0 .$$

Part a) of ASSUMPTION 2-II translates in the same way:

$$Y^{(0,0)}, Y^{(0,1)}, Y^{(0,0,0)}, Y^{(0,1,0)}, Y^{(0,0,1)}, Y^{(0,1,1)} \amalg S_1 \mid X_0 \text{ and } Y^{(0,0,0)}, Y^{(0,1,0)}, Y^{(0,0,1)}, Y^{(0,1,1)} \amalg S_2 \mid \underline{S}_1, X_0 .$$

The proof of identification is shown only for one example, that could be considered a worst case. It is one of the 'most counterfactual' cases, i.e. it considers the counterfactual outcome in both periods:

$$F[Y^{(0,1,0)} \mid \underline{S}_2 = (0,0,1)] = \underset{X_0 \mid \underline{S}_2}{E}\{F[Y^{(0,1,0)} \mid \underline{S}_2 = (0,0,1), X_0] \mid \underline{S}_2 = (0,0,1)\} =$$

$$= \underset{X_0 \mid \underline{S}_2}{E}\{F[Y^{(0,1,0)} \mid \underline{S}_2 = (0,1,0), X_0] \mid \underline{S}_2 = (0,0,1)\} =$$

$$= \underset{X_0 \mid \underline{S}_2}{E}\{F[Y \mid \underline{S}_2 = (0,1,0), X_0] \mid \underline{S}_2 = (0,0,1)\} .$$

When we use the sequential version of the full randomisation ASSUMPTION (2-II) the proof of identification is of course the same as before. It might nevertheless be illustrative to see the proof of the equivalence of the sequential version

[13] For the sake of brevity any reference to the sets $\mathcal{X}_0$, $\mathcal{X}$ and $\underline{\mathbb{S}}_t$ is omitted in the example.

and the nonsequential version of CIA for this example. Here, we verify that $Y^{(0,1,0)}$ is indeed independent of (S_1, S_2) conditional on X_0, so that $F(Y^{(0,1,0)}, S_1, S_2 \mid X_0) = F(Y^{(0,1,0)} \mid X_0)\, F(S_1, S_2 \mid X_0)$ holds:

$$F(Y^{(0,1,0)}, S_1, S_2 \mid X_0) = F(Y^{(0,1,0)}, S_2 \mid S_1, X_0)\, F(S_1 \mid X_0) =$$

$$= F(Y^{(0,1,0)} \mid S_1, X_0)\, F(S_2 \mid S_1, X_0)\, F(S_1 \mid X_0) =$$

$$= F(Y^{(0,1,0)} \mid X_0)\, F(S_2 \mid S_1, X_0)\, F(S_1 \mid X_0) = F(Y^{(0,1,0)} \mid X_0)\, F(S_2, S_1 \mid X_0).$$

Next, assume that assignment is decided each period based on initial information, as well as on treatment history and new information that is revealed up to that period. However, the information revealed has not been caused by past treatment, thus the observed X_t is independent of the treatment status in the same period S_t given past information about X and treatment status. The participation decision may be based on the values of time varying confounders observable at the beginning of the period, but they are not influenced by the treatments of this period (note that S_t is observed before X_t). Thus, we call such information STRICTLY EXOGENOUS (We use only a week version of strict exogeneity for which $t = \tau$). Note that this is still a nondynamic treatment regime, because functions of the outcomes of the treatments do not appear in the conditioning set.

Assumption 3 (sequential conditional independence assumption, S-CIA)

a) $Y^{\underline{s}_1}, ..., Y^{\underline{s}_T} \amalg S_t \mid \underline{S}_{t-1} = \underline{s}_{t-1}, \underline{X}_{t-1} = \underline{x}_{t-1}$;

b) $X_1 \amalg S_1 \mid X_0 = x_0$ and $X_{t-1} \amalg S_{t-1} \mid \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2} = \underline{x}_{t-2}$, $(t > 2)$;

c) $1 > P(\underline{S}_t = \underline{s}_t \mid \underline{X}_{t-1} = \underline{x}_{t-1}) > 0$; $\forall \underline{x}_t \in \underline{\chi}_t$, $\forall \underline{s}_t \in \underline{\mathbb{S}}_t$, $\forall t : 1 \leq t \leq \tau$, $\tau \leq T$.

Note that ASSUMPTION 3b coincides with a different formulation where X_{t-1} and the treatment history $\underline{S}_{t-1}$ are independent conditional on $\underline{X}_{t-2}$. It does however not imply $X_{t-1} \amalg S_t \mid \underline{S}_{t-1} = \underline{s}_{t-1}, \underline{X}_{t-2} = \underline{x}_{t-2}$, so that selection into treatment in period t could still be based on information acquired in period *t-1*. ASSUMPTION 3 again holds for all feasible sequences, as well as for all values of $\underline{x}_t$ that are in the set $\underline{\chi}_t$.

Part b) of ASSUMPTION 3 can be reformulated in terms of potential confounders: The treatment in period t is not causal for the value of X observed in this period given the history (note that S_t is observed before X_t is observed). Thus, the distributions of potential confounders are identical conditional on observed treatment and observed confounder history.

Lemma 4 (noncausality condition)

ASSUMPTION 3b) is equivalent to the following noncausality condition:

$F(X_1^{(0,s_1^k)} \mid S_1 = s_1^k, X_0) = F(X_1^{(0,s_1^l)} \mid S_1 = s_1^l, X_0)$ and

$F(X_{t-1}^{(\underline{s}_{t-2}, s_{t-1}^k)} \mid S_{t-1} = s_{t-1}^k, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2}) = F(X_{t-1}^{(\underline{s}_{t-2}, s_{t-1}^l)} \mid S_{t-1} = s_{t-1}^l, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2})$ $(t > 2)$.

The proof is contained in Appendix A.3.

Theorem 1

ASSUMPTION 1 is satisfied. If ASSUMPTION 2 and / or ASSUMPTION 3 holds, all treatment effects up to period τ are identified.

The proof is contained in Appendix A.4.

Example (continued - 4)

Now we consider for simplicity only the potential outcome $Y^{(0,1,0)}$. Parts a) and b) of the S-CIA assumption simplify in the example as follows:

a) $Y^{(0,1,0)} \amalg S_1 \mid X_0$; $Y^{(0,1,0)} \amalg S_2 \mid \underline{S}_1, \underline{X}_1$;

b) $X_1^{0,0} = X_1^{0,1} = X_1^0 = X_1$; $X_1 \amalg S_1 \mid X_0$.

Note that $X_1 \not\amalg S_2 \mid X_0$ and $X_1 \not\amalg S_2 \mid S_1, X_0$ is explicitly allowed for. Identification is shown as follows:

$$F(Y^{(0,1,0)} \mid S_1 = 0, S_2 = 1) = \underset{X_1, X_0}{E}[F(Y^{(0,1,0)} \mid S_1 = 0, S_2 = 1, X_0, X_1) \mid S_1 = 0, S_2 = 1]$$

$$\overset{a}{=} \underset{X_1, X_0}{E}[F(Y^{(0,1,0)} \mid S_1 = 0, X_0, X_1) \mid S_1 = 0, S_2 = 1].$$

So $\underset{X_1, X_0}{E}[F(Y^{(0,1,0)} \mid S_1 = 0, S_2 = 1, X_0, X_1)$ does not depend on S_2. It remains to show that it does not depend on S_1 either.

$$F(Y^{(0,1,0)}, S_1 = 0 \mid X_0, X_1) = \frac{F(Y^{(0,1,0)}, S_1 = 0, X_1 \mid X_0)}{F(X_1 \mid X_0)} \overset{a,b}{=} \frac{F(Y^{(0,1,0)}, X_1 \mid X_0) F(S_1 = 0 \mid X_0)}{F(X_1 \mid X_0)}.$$

Because $\dfrac{F(Y^{(0,1,0)}, X_1 \mid X_0)}{F(X_1 \mid X_0)} = F(Y^{(0,1,0)} \mid X_1, X_0)$ and $F(S_1 = 0 \mid X_0) \overset{b}{=} F(S_1 = 0 \mid X_0, X_1)$, we get:

$$F(Y^{(0,1,0)}, S_1 = 0 \mid X_0, X_1) = F(Y^{(0,1,0)} \mid X_0, X_1) F(S_1 = 0 \mid X_0, X_1).$$

Thus $Y^{(0,1,0)}$ and S_1 are independent conditional on (X_0, X_1), hence all effects are identified:

$$F(Y^{(0,1,0)} \mid S_1 = 0, S_2 = 1) = \underset{X_1, X_0}{E}[F(Y^{(0,1,0)} \mid X_0, X_1) \mid S_1 = 0, S_1 = 1]$$

$$= \underset{X_1, X_0}{E}[F(Y \mid S_1 = 1, S_2 = 0, X_0, X_1) \mid S_1 = 0, S_2 = 1].$$

3.2 Dynamic treatment regimes

Next we rebalance the assumptions by weakening the conditions on the conditioning set and strengthening the assumptions on the randomisation process. Assume that in addition to ASSUMPTION 1 the researcher knows enough in the beginning of each period so that the assignment to the treatment in every period can be treated as independent from potential outcomes conditional on that information. In contrast to the previous assumptions the relevant information set may contain variables related to realisation from observed outcome variables of previous periods (with the exception of period 0). However, the conditional independence assumptions applying to outcomes are also assumed to hold for the potential values of these ’endogenous’ confounding variables. This set-up constitutes the first and in fact most restrictive DYNAMIC TREATMENT REGIME we consider in this paper.

Assumption 4 (sequential endogenous conditional independence assumption – SE-CIA)

a) $H^{\underline{s}_t},...,H^{\underline{s}_\tau} \coprod S_t \mid \underline{H}_{t-1} = \underline{h}_{t-1}, \underline{S}_{t-1} = \underline{s}_{t-1}$;

b) $1 > P(S_t = s_t \mid \underline{H}_{t-1} = \underline{h}_{t-1}, \underline{S}_{t-1} = \underline{s}_{t-1}) > 0$, $\forall \underline{h}_{t-1} \in \underline{\mathbf{H}}_{t-1}$, $\forall \underline{s}_t \in \underline{\mathbb{S}}_t$, $\forall t: 1 \leq t \leq \tau$, $\tau \leq T$.

The set $\forall \underline{h}_{t-1} \in \underline{\mathbf{H}}_{t-1}$ is defined analogously to the set $\underline{\chi}_{t-1}$.

Theorem 2

If ASSUMPTIONS 1 and 4 are satisfied, then all treatment effects up to period τ are identified.

The proof is contained in Appendix A.5.

This assumption appears to be attractive, because it allows to account for the outcome of previous treatments (*predetermined endogenous* variables) in the conditioning set, but still is strong enough to give identification of all effects. However, the additional assumption that not only potential outcomes, but also potential confounders are randomised could be a problem in practise. It means that once a variable is discovered that influences both the participation decision as well as the potential outcomes, one can only use that variable in the conditioning set, if the remaining conditioning variables explain all systematic correlations between the potential value of that variable and the participation process. Obviously once many variables are needed to make sure that potential outcomes are conditionally independent of the assignment process, these conditions are very hard to verify, and thus undesirable.

Example (continued - 5)

Part a) of the SE-CIA assumption (ASSUMPTION 4) simplifies in the example as follows:

$$Y^{(0,1,0)}, X^{(0,1,0)}, Y^{(0,1)}, X^{(0,1)} \coprod S_1 \mid X_0, \quad Y^{(0,1,0)}, X^{(0,1,0)} \coprod S_2 \mid \underline{S}_1, \underbrace{Y_1, X_1, X_0}_{\underline{H}_1}$$

The above conditions imply restrictions for the following expectations:

$$E(Y^{(0,1,0)} \mid S_1 = 1, X_0) = E(Y^{(0,1,0)} \mid S_1 = 0, X_0) \quad [= E(Y^{(0,1,0)} \mid X_0)];$$

$$E(X^{(0,1,0)} \mid S_1 = 1, X_0) = E(X^{(0,1,0)} \mid S_1 = 0, X_0) \quad [= E(X^{(0,1,0)} \mid X_0)];$$

$$E(Y^{(0,1,0)} \mid S_1 = s_1, S_2 = 0, \underline{H}_1) = E(Y^{(0,1,0)} \mid S_1 = s_1, S_2 = 1, \underline{H}_1), \qquad \forall s_1 \in \{0,1\};$$

$$E(X^{(0,1,0)} \mid S_1 = s_1, S_2 = 0, \underline{H}_1) = E(X^{(0,1,0)} \mid S_1 = s_1, S_2 = 1, \underline{H}_1), \qquad \forall s_1 \in \{0,1\};$$

$$E(Y^{(0,1)} \mid S_1 = 1, X_0) = E(Y^{(0,1)} \mid S_1 = 0, X_0);\ E(X^{(0,1)} \mid S_1 = 1, X_0) = E(X^{(0,1)} \mid S_1 = 0, X_0);$$

These equations could be used to establish identification of the counterfactual event $Y^{(0,1,0)} \mid \underline{S}_2 = (0,0,1)$:

$$F(Y^{(0,1,0)} \mid \underline{S}_2 = (0,0,1)) = \underset{\underline{H}_1}{E}\{F[Y^{(0,1,0)} \mid \underline{S}_2 = (0,0,1), \underline{H}_1] \mid \underline{S}_2 = (0,0,1)\}$$

$$= \underset{\underline{H}_1}{E}\{F[Y^{(0,1,0)} \mid \underline{S}_1 = (0,0), S_2 = 1, \underline{H}_1] \mid \underline{S}_2 = (0,0,1)\}$$

$$\overset{4}{=} \underset{\underline{H}_1}{E}\{F[Y^{(0,1,0)} \mid \underline{S}_1 = (0,0), S_2 = 0, \underline{H}_1] \mid \underline{S}_2 = (0,0,1)\}$$

$$\overset{4}{=} \underset{\underline{H}_1}{E}[F(Y^{(0,1,0)} \mid \underline{S}_1 = (0,0), \underline{H}_1] \mid \underline{S}_1 = (0,0,1)\}.$$

The observation rules can be used at that stage:

$$Y_1 = \underline{1}[\underline{S}_2 = (0,0,0)]Y_1^{(0,0,0)} + \underline{1}[\underline{S}_2 = (0,0,1)]Y_1^{(0,0,1)} + \underline{1}[\underline{S}_2 = (0,1,0)]Y_1^{(0,1,0)} + \underline{1}[\underline{S}_2 = (0,1,1)]Y_1^{(0,1,1)};$$

$$X_1 = \underline{1}[\underline{S}_2 = (0,0,0)]X_1^{(0,0,0)} + \underline{1}[\underline{S}_2 = (0,0,1)]X_1^{(0,0,1)} + \underline{1}[\underline{S}_2 = (0,1,0)]X_1^{(0,1,0)} + \underline{1}[\underline{S}_2 = (0,1,1)]X_1^{(0,1,1)}.$$

Given $\underline{S}_2 = (0,0,1)$, we obtain $Y_1 = Y_1^{(0,0,1)}$, $X_1 = X_1^{(0,0,1)}$, implying $\underline{H}_1 = (X_0, H_1^{(0,0,1)})$. These equalities can be used to complete the proof:

$$\underset{\underline{H}_1}{E}[F(Y^{(0,1,0)} \mid \underline{S}_1 = (0,0), \underline{H}_1) \mid \underline{S}_2 = (0,0,1)] = \underset{X_0, H_1^{(0,0,1)}}{E}\{F[Y^{(0,1,0)} \mid \underline{S}_1 = (0,0), X_0, H_1^{(0,0,1)}] \mid \underline{S}_2 = (0,0,1)\}.$$

Now we use the assumption that the potential confounders are randomised as well:

$$F(Y^{(0,1,0)} \mid \underline{S}_1 = (0,0), X_0, H_1^{(0,0,1)}) = \frac{F(Y^{(0,1,0)}, H_1^{(0,0,1)} \mid \underline{S}_1 = (0,0), X_0)}{F(H_1^{(0,0,1)} \mid \underline{S}_1 = (0,0), X_0)}$$

$$\stackrel{4}{=}\frac{F(Y^{(0,1,0)},H_1^{(0,0,1)}\mid \underline{S}_1=(0,1),X_0)}{F(H_1^{(0,0,1)}\mid \underline{S}_1=(0,1),X_0)}$$

$$\stackrel{4}{=}\frac{F(Y^{(0,1,0)},H_1^{(0,0,1)}\mid X_0)}{F(H_1^{(0,0,1)}\mid X_0)}$$

$$=F(Y^{(0,1,0)}\mid X_0,H_1^{(0,0,1)}) = F(Y^{(0,1,0)}\mid \underline{S}_1=(0,1),X_0,H_1^{(0,0,1)}).$$

Collecting the different partial results shows identification:

$$F(Y^{(0,1,0)}\mid \underline{S}_2=(0,0,1)) = \underset{\underline{H}_1}{E}[F(Y^{(0,1,0)}\mid \underline{S}_1=(0,0),\underline{H}_1]\mid \underline{S}_2=(0,0,1)\} =$$

$$= \underset{X_0,H_1^{(0,0,1)}}{E}\{[F(Y^{(0,1,0)}\mid \underline{S}_1=(0,1),X_0,H_1^{(0,0,1)})]\mid \underline{S}_2=(0,0,1)\}$$

$$= \underset{X_0,H_1}{E}\{[F(Y^{(0,1,0)}\mid \underline{S}_1=(0,1),X_0,H_1)]\mid \underline{S}_2=(0,0,1)\}$$

$$= \underset{X_0,H_1}{E}\{[F(Y^{(0,1,0)}\mid \underline{S}_1=(0,1),S_2=0,X_0,H_1)]\mid \underline{S}_2=(0,0,1)\}$$

$$= \underset{X_0,H_1}{E}\{[F(Y\mid \underline{S}_2=(0,1,0),X_0,H_1)]\mid \underline{S}_2=(0,0,1)\}. \quad \text{q.e.d.}$$

This give us the identification of $F(Y^{(0,1,0)}\mid \underline{S}_2=(0,0,1))$. Denote $F(Y^{(0,1,0)}\mid \underline{S}_1=(0,1),X_0,H_1^{(0,0,1)})$ as function of $H_1^{(0,0,1)}$ ($g(X_0,H_1^{(0,0,1)})$). The key insight is that taking expectations of this function with respect to the same conditioning set as the one that denotes the potential confounders is the same as taking expectations with respect to the distribution of the observable confounders in the respective subset:

$$\underset{X_0,H_1^{(0,0,1)}}{E}[g(X_0,H_1^{(0,0,1)})\mid S_2=(0,0,1)] = \underset{X_0,H_1}{E}[g(X_0,H_1)\mid S_2=(0,0,1)].$$

Compared to ASSUMPTION 3 (S-CIA) the additional restrictiveness of SE-CIA as defined in ASSUMPTION 4 is that the same randomisation assumed to hold for the potential outcomes also holds for the potential confounders. This part is relaxed in ASSUMPTION 5.

Assumption 5 (partial sequential endogenous conditional independence assumption – PSE-CIA)

a) $Y^{\underline{s}_1},...,Y^{\underline{s}_T} \coprod S_t \mid \underline{H}_{t-1}=\underline{h}_{t-1}, \underline{S}_{t-1}=\underline{s}_{t-1}$;

b) $1 > P(S_t = s_t \mid \underline{H}_{t-1} = \underline{h}_{t-1}, \underline{S}_{t-1} = \underline{s}_{t-1}) > 0$; $\forall \underline{h}_{t-1} \in \underline{\mathbf{H}}_{t-1}$, $\forall \underline{s}_t \in \mathbb{S}_t$, $\forall t: 1 \le t \le \tau$, $\tau \le T$.

ASSUMPTION 5 is weaker than ASSUMPTION 4, because it does not require randomisation of the confounding variables, and still allows to condition on endogenous variables, i.e. observable variables that are functions of the outcomes of previous treatments. Note that ASSUMPTION 4 as well as ASSUMPTION 5 contain implicitly a typical (exogenous) initial condition for period 0 by assuming that the conditioning set in period one does not contain any endogenous variable (i.e. $H_0^{\underline{s}_t} = H_0$, $\forall \underline{s}_t$). Without that initial condition (or without other restrictive assumptions that could be imposed alternatively) it does not appear to be possible to obtain identification of interesting causal parameters.

Theorem 3

Assumptions 1 and 5 are satisfied.

a) All treatment effects of the form $\theta_t^{\underline{s}_\tau^k, \underline{s}_\tau^l}(\underline{s}_1^j)$ and $\theta_t^{\underline{s}_\tau^k, \underline{s}_\tau^l}$, $\tau \le t \le T$ are identified.

b) All treatment effects of the form $\theta_t^{(\underline{s}_{\tau-1}^k, s_\tau^k),(\underline{s}_{\tau-1}^k, s_\tau^l)}(\underline{s}_{\tau-1}^k, s_\tau^j)$ and $\theta_t^{(\underline{s}_{\tau-1}^k, s_\tau^k),(\underline{s}_{\tau-1}^k, s_\tau^l)}(\underline{s}_{\tau-1}^k)$, $\tau \le t \le T$, are identified.

c) All treatment effects of the form $\theta_t^{(\underline{s}_{\tau-v}^j, s_{\tau-v+1}^k, \ldots, s_\tau^k),(\underline{s}_{\tau-v}^j, s_{\tau-v+1}^l, \ldots, s_\tau^l)}(\underline{s}_{\tau-w}^j)$, $\forall v: 0 < v < \tau$, $\forall w: \tau \ge w \ge v$, $\tau \le t \le T$, are identified.

The proofs of the three parts of THEOREM 3 are given in Appendix A.6.

THEOREM 3 shows that the price for relaxing the assumption that the potential confounders follow the same randomisation process as the outcomes comes in three different ways: Part a) states that if identification of pair-wise comparisons of all sequences is desired, then this is still possible, but only for coarser subpopulations, namely individuals defined by their treatment choice in the first period. It is however not possible, to identify all effects for example for groups of individuals defined according to their difference in the second period.[14] The relevant distinction between the

[14] Hence the usual average treatment effects (ATE) as usually defined in the static framework are identified, whereas several average treatment effects on the treated (ATET) are not identified using ASSUMPTION 5 only.

population defined by treatment choice in the first period and subsequent periods is that in the first period treatment choice is random conditional on exogenous variables, whereas in the second and later periods, randomisation into these treatments is conditional on endogenous variables.

Part b) and c) state that specific comparisons are identified for finer subpopulations. Part b) refers to comparisons of treatment that differ only with respect to the treatment in the last period. Such comparisons are identified for subpopulation that have the same sequence as the common part of the two sequences compared (namely up to period *t-1*). The populations could be defined by any treatment state in the final period (*t*). It is basically the same result as for case a), but with time period $\tau - 1$ playing the role of time period 0 (i.e. the period for which the treatment sequence still coincides) in a). In this case the endogeneity problem is not really harmful, because the endogenous variable $\underline{H}_{\tau-1}$, which is the crucial one to condition on for identification, has been influenced by the same past treatments when comparing the two sequences that differ only with respect to the treatment in period τ.

Finally part c) states that all sequences that have a common part in the beginning are identified for subpopulations defined as participating in that common part of the sequence. Of course, the relevant subpopulations for which identification is obtained could be coarser, but not finer. Compared to a) and b) the conditioning set for the effects is 'one period shorter'. The reason is that the identification of sequences that differ for more than one period is more difficult: The endogenous conditioning variable $\underline{H}_{\tau-1}$ needed to make participants comparable to non-participants in the specific sequence, is fluenced by all events during the sequence. However, since the sequences differ, these events differ (they are same in case a) and b)) and thus the endogeneity problem leads to some additional loss of identification.

Example (continued - 6)

Part a) of the PSE-CIA assumption (ASSUMPTION 5) simplifies in the example as follows:

$$Y^{(0,1,0)} \coprod S_1 \mid X_0 , \qquad Y^{(0,1,0)} \coprod S_2 \mid \underline{S}_1, \underline{H}_1 .$$

Identification is shown next. The above conditions imply restrictions for the following expectations:

$$E(Y^{(0,1,0)} \mid S_1 = 1, X_0) = E(Y^{(0,1,0)} \mid S_1 = 0, X_0) \quad [= E(Y^{(0,1,0)} \mid X_0)]; \tag{4}$$

$$E(Y^{(0,1,0)} \mid S_1 = s_1, S_2 = 0, \underline{H}_1) = E(Y^{(0,1,0)} \mid S_1 = s_1, S_2 = 1, \underline{H}_1), \ \forall s_1 \in \{0,1\}. \tag{5}$$

These equations could be used to establish identification of some counterfactual events defined by parts a) and c) in THEOREM 3.[15]

Part b) is more or less immediate in this example. It is sufficient to show that $E(Y^{(0,1,0)} \mid S_1 = 1, S_2 = 1)$ is identified:

$$E(Y^{(0,1,0)} \mid S_1 = 1, S_2 = 1) = \underset{\underline{H}_1}{E}[\, E(Y^{(0,1,0)} \mid S_1 = 1, S_2 = 1, \underline{H}_1) \mid S_1 = 1, S_2 = 1] =$$

$$= \underset{\underline{H}_1}{E}[\, E(Y^{(0,1,0)} \mid S_1 = 1, S_2 = 0, \underline{H}_1) \mid S_1 = 1, S_2 = 1] =$$

$$= \underset{\underline{H}_1}{E}[\, E(Y \mid S_1 = 1, S_2 = 0, \underline{H}_1) \mid S_1 = 1, S_2 = 1].$$

As an analogy for part a) of THEOREM 3 it needs to be shown that $E(Y^{(0,1,0)} \mid S_1 = 0)$ is identified:

$$E(Y^{(0,1,0)} \mid S_1 = s_1) = \underset{\underline{H}_1, S_2}{E}\{[E(Y^{(0,1,0)} \mid S_1 = s_1, S_2, \underline{H}_1)] \mid S_1 = s_1\}$$

$$= \underset{\underline{H}_1}{E} \underset{S_2 \mid \underline{H}_1}{E}[E(Y^{(0,1,0)} \mid S_1 = s_1, S_2, \underline{H}_1) \mid S_1 = s_1, \underline{H}_1] \mid S_1 = s_1\}$$

$$= \underset{\underline{H}_1}{E}\{[E(Y^{(0,1,0)} \mid S_1 = s_1, S_2 = 1, \underline{H}_1)\, P(S_2 = 1 \mid S_1 = s_1, \underline{H}_1) \ +$$

$$+ \ E(Y^{(0,1,0)} \mid S_1 = s_1, S_2 = 0, \underline{H}_1)\, P(S_2 = 0 \mid S_1 = s_1, \underline{H}_1)] \mid S_1 = s_1\}, \ \forall s_1 \in \{0,1\}.$$

Using equation (5) allows to manipulate the value of S_2 inside the expectation, hence we get ($P(S_2 = 1 \mid S_1 = s_1, \underline{H}_1)$ + $P(S_2 = 0 \mid S_1 = s_1, \underline{H}_1) = 1$):

$$E(Y^{(0,1,0)} \mid S_1 = s_1) = \underset{\underline{H}_1}{E}[E(Y^{(0,1,0)} \mid S_1 = s_1, S_2 = s_2, \underline{H}_1) \mid S_1 = s_1], \qquad \forall s_1, s_2 \in \{0,1\}.$$

In particular we get identification of $E(Y^{(0,1,0)} \mid S_1 = 1, X_0)$, because:

$$E(Y^{(0,1,0)} \mid S_1 = 1, X_0) = \underset{H_1}{E}[E(Y^{(0,1,0)} \mid S_1 = 1, S_2 = 0, \underline{H}_1) \mid S_1 = 1, X_0] =$$

[15] Part c) is not really relevant in a two period framework.

$$= \underset{H_1}{E}[E(Y \mid S_1 = 1, S_2 = 0, \underline{H}_1) \mid S_1 = 1, X_0].$$

If $E(Y^{(0,1,0)} \mid S_1 = 1, X_0)$ is identified, then $E(Y^{(0,1,0)} \mid S_1 = 0, X_0)$ is identified as well, because of equation (4). Thus $E(Y^{(0,1,0)} \mid S_1 = 1) (= \underset{X_0}{E}[E(Y^{(0,1,0)} \mid S_1 = 1, X_0) \mid S_1 = 1])$, $E(Y^{(0,1,0)} \mid S_1 = 0)$ and $E(Y^{(0,1,0)})$ are also identified. Using the same arguments these properties can be shown to be valid for all potential outcomes, thus all treatment effects that are unconditional or conditional on the treatment in the first period are identified.

Although even ASSUMPTION 5 is restrictive it has appeal for practical evaluation studies. For example, in many European countries rich administrative data sets are made available recently that would allow the assignment to the first treatment to be considered as random conditional on exogenous information.[16] Furthermore, these records compiled from labour offices and pension systems typically allow to follow the individuals over time on a very detailed basis. Hence, intermediate outcomes are observed that play a critical role to make the validity of ASSUMPTION 5 plausible.

At this point we need to clarify the relation of our results presented so far to previous work in the fields of epidemiology and statistics, in particular by James Robins and co-authors.[17] Several results similar to our identification results can already be found in those papers in a similar form. Although, their assumptions are differently formulated, their spirit is similar. The interest is however not focussed on the parameters for particular subpopulations and differently defined sequences, as it is in this paper. Furthermore, in several cases they are specialised to focus on particular designs of actually implemented dynamic randomised studies in those fields, that could differ very much from designs that would be compatible with designs resulting from our ASSUMPTIONS 1 to 5 which are much more akin to the econometric evaluation literature (and could well be used to approximate situations in observable studies provided they are based on rich enough data). In particular our differentiated identification results especially resulting from THEOREM 3 do not appear in that literature (see also the discussion of the G-computation algorithm after the proof of THEOREM 3 in Appendix A.6).

[16] One example is Switzerland (see Gerfin and Lechner, 2000). Advanced efforts to collect such data in other countries we are aware of in detail include Germany, Great Britain, and Sweden.

[17] See for example Robins (1986), Robins (1989), Robins (1997), Robins, Greenland, and Hu (1999). The website of James Robins (http://biosun1.harvard.edu/~robins) gives a comprehensive list of his papers relating to this subject.

3.3 A note on estimation

We leave an in depth discussion of estimation and inference in complex dynamic treatment studies using ASSUMPTION 1 to 5 to a companion paper. Here, we are confining ourselves to some brief considerations.

When identification can be based on ASSUMPTION 2, it has already been noted that the framework is exactly the same as the multiple treatment framework introduced by Imbens (2000) and Lechner (2001a). For this framework issues of non- and semiparametric estimation are already discussed by several authors including Brodaty, Crepon, and Fougère (2001), Imbens (2000), and Lechner (2001b). The issue here is essentially to find appropriately weighted means over the regression curves $g^{\underline{s}_t}(x_0) = E(Y \mid \underline{S}_t = \underline{s}_t, X_0 = x_0)$ for all relevant sequences $\underline{s}_t$. The appropriate weights depend on the distribution of X_0 in the respective subpopulations defined by treatment status and will differ depending on the precise effect that is estimated. For different estimators that all use so-called propensity score properties to reduce the dimension of the estimation problem the reader is referred to the already mentioned papers.

Propensity score methods are a useful tool in practice. Rosenbaum and Rubin (1983) show that for the binary static case so-called balancing score properties hold.[18] These properties imply that independence conditional on some vector of variables implies independence on specific functions of these variables. These functions are called balancing scores. The balancing score with the lowest dimension is the conditional participation probability, the so-called propensity score. The advantage of knowing balancing scores is that they provide ways to reduce the dimension of the estimation problem and help to devise feasible semi- and nonparametric two-step estimators. The first step is an estimation of the balancing score. The second step estimates the effect and uses conditioning on the balancing score instead of the (original) variables that appear in the conditioning set (see for example Hahn, 1998, Heckman, Ichimura, and Todd, 1998, Hirano, Imbens, and Ridder, 2000). Since $\underline{s}_t$ may contain many different sequences and x_0 may be high dimensional to make the identifying assumption plausible, the dimensionality problem hurting any semiparametric and nonparametric analysis is certainly an issue already in that case.

Using the more general ASSUMPTION 3 that allows for time varying confounders we essentially need estimates based on appropriately weighted regression curves depending on all regressors

[18] The results are extended to the case of multiple treatments by Imbens (2000) and Lechner (2001a).

$g^{\underline{s}_t}(\underline{x}_{t-1}) = E(Y \mid \underline{S}_t = \underline{s}_t, \underline{X}_{t-1} = \underline{x}_{t-1})$. The weights depend again on the distribution of participation probabilities conditional on $\underline{x}_{t-1}$ and on the specific parameter of interest. Similarly, the input to any estimator consistent under ASSUMPTIONS 4 and 5 will be based on $g^{\underline{s}_t}(\underline{h}_{t-1})$ = $E(Y \mid \underline{S}_t = \underline{s}_t, \underline{H}_{t-1} = \underline{h}_{t-1})$. Because ASSUMPTIONS 4 and 5 also incorporate intermediate outcomes from the programmes, conditioning sets appropriate for many applications may contain a large number of variables and devising dimension reducing balancing scores will be a precondition for using any sort of nonparametric estimation technique.

4 Bounds

4.1 Preliminaries

In this section we consider situations for which identification is not or only partially warranted. If part of a chosen assumption appears to be particularly suspicious in a specific application, the sensitivity of the results can be assessed by computing the possible range of results that could occur if the respective assumption is violated. Furthermore, ranges for effect can be obtained under milder conditions then required in ASSUMPTIONS 2 to 5 that could be relevant on their own right. However, far from being complete, we consider only a few fairly specific examples of interesting violations, because there are simply too many possible violations to provide bounds for all of them in this paper. Many more cases are considered the Appendices B and C.

Before actually obtaining the bounds, some more remarks about the information available are useful. Without further information the sample analogues identify the participation probabilities:[19]

$$P(S_t = s_t),\ P(\underline{S}_t = \underline{s}_t),\ P(S_t = s_t \mid \underline{S}_{t-1} = \underline{s}_{t-1}),\ \forall \underline{s}_t \in \underline{\mathbb{S}}_t,\ t \le T. \quad (6)$$

Assume that all sequence that belong to $\underline{\mathbb{S}}_t$ have a positive probability to occur ($0 < P(\underline{S}_t = \underline{s}_t)$, $\forall \underline{s}_t \in \underline{\mathbb{S}}_t$). If this assumption is violated, there is no way to evaluate the effect of such a sequence

[19] For notational simplicity conditioning on attributes is omitted, but all those quantities are also identified conditional on attributes. For the outcome variables this holds as long as the respective events in the conditioning set have positive probability to occur.

without extrapolating from the results obtained for some other sequence, a case not considered here.

Furthermore, if appropriate support conditions hold, all conditional expectations of the observable outcome variables are also identified, which in turn identify the following conditional expectations of the potential outcomes:

$$B_t^{\underline{s}_\tau} = E(Y_t^{\underline{s}_\tau} \mid \underline{S}_\tau = \underline{s}_\tau) = E(Y_t \mid \underline{S}_\tau = \underline{s}_\tau), \quad \forall \underline{s}_\tau \in \underline{\mathbb{S}}_\tau,\ \tau \le t \le T. \tag{7}$$

To obtain bounds along the lines suggested by Robins (1989) and Manski (1989, 1990), ASSUMPTION 1-II states that the expectations of the outcome variables are bounded:[20]

Assumption 1 - II (known bounds for conditional expectations of outcomes)

a) $P[{}_LY \le E(Y_t^{\underline{s}_\tau^k} \mid \underline{S}_{\tau'} = \underline{s}_{\tau'}^j) \le {}^UY \mid \underline{S}_{\tau'} = \underline{s}_{\tau'}^j] = 1$,

b) $P[{}_LY \le E(Y_t^{\underline{s}_\tau^k} \mid \underline{S}_{\tau'} = \underline{s}_{\tau'}^j, \underline{H}_{\tau'-1} = \underline{h}_{\tau'-1}) \le {}^UY \mid \underline{S}_{\tau'} = \underline{s}_{\tau'}^j] = 1$,

$$\forall \underline{h}_{\tau'-1} \in \underline{\mathbf{H}}_{\tau'-1},\ \underline{s}_{\tau'}^j \in \underline{\mathbb{S}}_{\tau'},\ \underline{s}_\tau^k \in \underline{\mathbb{S}}_\tau,\ \tau' \le \tau \le t \le T.$$

In the following it will be useful to have a slightly different version of LEMMA 2 available. Hence, we define LEMMA 2-II like LEMMA 2-I, except that it concerns the potential outcomes and not the effects directly:

Lemma 2-II (connection of potential outcomes defined for different lengths of treatments)

$$Y_t^{\underline{s}_\tau^k}(\underline{s}_{\tilde{\tau}}^j) = \sum_{m=1}^{M} P(S_{\tau+1} = m \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j)\, Y_t^{(\underline{s}_\tau^k, m)}(\underline{s}_{\tilde{\tau}}^j, \cdot, s_{\tau+1} = m) =$$

[20] We assume that the bounds of all expected potential outcomes are the same in all subpopulations. A generalisation is immediate, but increases the notational burden significantly. Furthermore, there seems to be no application of the bounding literature so far where such differences in the bounds are assumed to exist.

$$= \sum_{m_1=1}^{M} ... \sum_{m_\delta=1}^{M} P(S_{\tau+1} = m_1, ..., S_{\tau+\delta} = m_\delta \mid \underline{S}_{\tau} = \underline{s}_{\tau}^{j}) Y_t^{(\underline{s}_\tau^k, m_1, ..., m_\delta)}(\underline{s}_\tau^{j}, \cdot, s_{\tau+1} = m_1, ..., s_{\tau+\delta} = m_\delta) .$$

The proof is contained in Appendix A.1 (first part of the proof of LEMMA 2-I).

For the selected examples that follow below, a particular version of this lemma has interesting implications:

$$E(Y^{\underline{s}_t^k} \mid \underline{s}_t^{j}) = \sum_{m=1}^{M} E(Y^{\underline{s}_t^k, m} \mid \underline{s}_t^{j}, S_{t+1} = m) P(S_{t+1} = m \mid \underline{S}_t = \underline{s}_t^{j}), \; j \neq k. \tag{8}$$

This leads to the following reformulation:

$$E(Y^{\underline{s}_t^k, s} \mid \underline{s}_t^{j}, S_{t+1} = s) = \frac{E(Y^{\underline{s}_t^k} \mid \underline{s}_t^{j}) - \sum_{\substack{m=1 \\ m \neq s}}^{M} E(Y^{\underline{s}_t^k, m} \mid \underline{s}_t^{j}, S_{t+1} = m) P(S_{t+1} = m \mid \underline{S}_t = \underline{s}_t^{j})}{P(S_{t+1} = s \mid \underline{S}_t = \underline{s}_t^{j})} \tag{9}$$

Equation (9) is particularly useful when some assumption identifies $E(Y^{\underline{s}_t^k} \mid \underline{s}_t^{j})$, but not $E(Y^{\underline{s}_t^k, m} \mid \underline{s}_t^{j}, S_{t+1} = m)$, and information about $E(Y^{\underline{s}_t^k, s} \mid \underline{s}_t^{j}, S_{t+1} = s)$ is needed to determine the effects of interest.

4.2 No information about assignment

First, we consider the case when the researcher has no information about assignment (no other information than the regularity conditions given by ASSUMPTION 1-II and equations (6) and (7)). Hence, we may call the bounds presented in the following THEOREM 4 *worst case* bounds, because no randomisation assumption whatsoever is invoked.

Theorem 4 (worst case bounds)

Suppose ASSUMPTION 1-II holds and the information given in equations (6) and (7) is available.

a) The bounds for $\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_\tau^j)$, $\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_{\tau-1}^j)$,..., $\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_1^j)$, $\forall \underline{s}_\tau \in \underline{\mathbb{S}}_\tau, \tau \le t \le T$ are not informative, when the sequence $\underline{s}_\tau^j$ has no common element with both sequences $\underline{s}_\tau^k$ and $\underline{s}_\tau^l$. The upper bound is $({}^UY - {}_LY)$ and the lower bound is $({}_LY - {}^UY)$. The width is $2({}^UY - {}_LY)$.

b) The bounds for $\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_\tau^k)$ and $\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_\tau^l)$, $\forall \underline{s}_\tau \in \underline{\mathbb{S}}_\tau, \tau \le t \le T$ are informative. They always include zero. For $\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_\tau^k)$ the upper bound is given by $B_t^{\underline{s}_\tau^k} - {}_LY$ and the lower bound is $B_t^{\underline{s}_\tau^k} - {}^UY$. The symmetric argument applies to $\theta_t^{(\underline{s}_\tau^k),(\underline{s}_\tau^l)}(\underline{s}_\tau^k)$. The width is $({}^UY - {}_LY)$.

c) The bounds for $\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_{\tau-1}^k), \theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_{\tau-2}^k),...., \theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_1^k)$ and $\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_{\tau-1}^l)$, $\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_{\tau-2}^l),...., \theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_1^l)$, $\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}$ are sometimes informative. Their widths are between $2({}^UY - {}_LY)$ and $({}^UY - {}_LY)$ depending on $P(S_\tau = s_\tau^k \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k)$, ..., $P(S_\tau = s_\tau^k \mid \underline{S}_1 = \underline{s}_1^k)$, $P(S_\tau = s_\tau^k)$. The exact bounds are given in Table 2.

Table 2: Exact bounds for the cases considered in THEOREM 4

Effect	Upper bound	Lower bound	Width
$\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_\tau^j)$	$({}^UY - {}_LY)$	$({}_LY - {}^UY)$	$2({}^UY - {}_LY)$
$\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_\tau^k)$	$(B_t^{\underline{s}_\tau^k} - {}_LY)$	$(B_t^{\underline{s}_\tau^k} - {}^UY)$	$({}^UY - {}_LY)$
$\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_{\tau-1}^k)$	${}^UY(1-P^{s_\tau^k \mid \underline{s}_{\tau-1}^k})+$ $+B_t^{\underline{s}_\tau^k}P^{s_\tau^k \mid \underline{s}_{\tau-1}^k} - {}_LY$	${}_LY(1-P^{s_\tau^k \mid \underline{s}_{\tau-1}^k})+$ $+B_t^{\underline{s}_\tau^k}P^{s_\tau^k \mid \underline{s}_{\tau-1}^k} - {}^UY$	$2({}^UY - {}_LY)-({}^UY - {}_LY)P^{s_\tau^k \mid \underline{s}_{\tau-1}^k}$
...	...	...	...
$\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_1^k)$	${}^UY(1-P^{s_\tau^k \mid \underline{s}_1^k})+$ $+B_t^{\underline{s}_\tau^k}P^{s_\tau^k \mid \underline{s}_1^k} - {}_LY$	${}_LY(1-P^{s_\tau^k \mid \underline{s}_1^k})+$ $+B_t^{\underline{s}_\tau^k}P^{s_\tau^k \mid \underline{s}_1^k} - {}^UY$	$2({}^UY - {}_LY)-({}^UY - {}_LY)P^{s_\tau^k \mid \underline{s}_1^k}$
$\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}$	${}^UY(1-P^{s_\tau^k})+$ $+B_t^{\underline{s}_\tau^k}P^{s_\tau^k} - {}_LY$	${}_LY(1-P^{s_\tau^k})+$ $+B_t^{\underline{s}_\tau^k}P^{s_\tau^k} - {}^UY$	$2({}^UY - {}_LY)-({}^UY - {}_LY)P^{s_\tau^k}$

$P^{A|B} := P(S_\tau = A \mid B)$.

The proof of THEOREM 4 is given in Appendix A.7.

Part a) of THEOREM 4 states the intuitively obvious that without further assumptions, the sample does not contain any information about treatment effects in subpopulations where neither of the sequences under consideration are observable $(\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_\tau^j)$, $\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_{\tau-1}^j),\ldots,\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_1^j)$, $\underline{s}_\tau^j \neq \underline{s}_\tau^k$, $\underline{s}_\tau^j \neq \underline{s}_\tau^l$). Therefore, the upper bound on these effects are given by $({}^UY - {}_LY)$ and the lower bounds are $({}_LY - {}^UY)$, so that the width (*W*) of these non-informative intervals for the true value is $2({}^UY - {}_LY)$. Note that this marks a difference compared to the comparable bounds for static treatment effects (Manski, 1990), for which the sample information reduces the bounds always by half to a width of $({}^UY - {}_LY)$.

Part b) concerns the opposite case compared to part a): one of the sequences used to define the effects is exactly equivalent to the sequence used in the conditioning set. In that case the intervals shrink.

Part c) concerns the intermediate case, where one of the sequences used to define the effects is equivalent to the sequence used in the conditioning set but do not have the same length. The reduction depends on how large the subsample for which we identify the expectation is compared to the other subsample considered. Depending on this probability ($P^{s_\tau^k|\underline{s}_{\tau-1}^k}$), the width of the bounds vary between $({}^UY - {}_LY)$ and $2({}^UY - {}_LY)$.

Note that none of these bounds exclude a zero effect, because the upper bound is always nonnegative, whereas the lower bound is also nonpositive. Thus, as observed for the static model as well, the *worst-case* bounds are of limited value in any policy analysis.

Example (continued – 7)

Let us again consider the case of binary outcome variable with two treatments each period as well as the comparison between sequences (0,0,1) and (0,1,0). The entries in Table 2 simplify considerably:

Table 2-E: Exact bounds for the cases considered in THEOREM 4 (example)

Effect	Upper bound	Lower bound	Width
$\theta_t^{(0,0,1),(0,1,0)}(0,1,1)$	1	-1	2
$\theta_t^{(0,0,1),(0,1,0)}(0,0,1)$	$P(Y=1\|0,0,1)$	$-1+P(Y=1\|0,0,1)$	1
$\theta_t^{(0,0,1),(0,1,0)}(0,0)$	$1-$ $[1-P(Y=1\|0,0,1)]P^{(0,0,1)\|(0,0)}$	$-1+P(Y=1\|0,0,1)P^{(0,0,1)\|(0,0)})$	$2-P^{(0,0,1)\|(0,0)}$
$\theta_t^{(0,0,1),(0,1,0)}(0)$ = $\theta_t^{(0,0,1),(0,1,0)}$	$1-$ $[1-P(Y=1\|0,0,1)]P^{(0,0,1)}$	$-1+P(Y=1\|0,0,1)P^{(0,0,1)}$	$2-P^{(0,0,1)}$

4.3 Lack of information for the last period

Consider now the case in which the researcher can plausibly assume any of the ASSUMPTIONS 3 to 5 to hold, but only until period $\tau-1$. Thus, she has no reliable information about the last assignment in period τ, for which the desired effect is defined. In other words, respective treatment effects of sequences defined up to period $\tau-1$ are identified, whereas those defined up to period τ are not.[21] The focus of this section is to assess the value of the information contained in this '**o**ne-**p**eriod-**s**hort' (OPS) randomisation assumption on the effects of interest.

First, let us consider a 'one-period-short' version of ASSUMPTION 3:

Assumption 6-3 (sequential conditional independence assumption OPS, S-CIA-OPS)

a) $Y^{\underline{s}_t},...,Y^{\underline{s}_T} \amalg S_t \mid \underline{S}_{t-1}=\underline{s}_{t-1}, \underline{X}_{t-1}=\underline{x}_{t-1}$;

b) $X_1 \amalg S_1 \mid X_0 = x_0$ and $X_{t-1} \amalg S_{t-1} \mid \underline{S}_{t-2}=\underline{s}_{t-2}, \underline{X}_{t-2}=\underline{x}_{t-2}$, $(t>2)$;

c) $1 > P(\underline{S}_t=\underline{s}_t \mid \underline{X}_{t-1}=\underline{x}_{t-1}) > 0$; $\forall \underline{x}_t \in \underline{\chi}_t$, $\forall \underline{s}_t \in \underline{\mathbb{S}}_t$, $\forall t: 1 \le t \le \tau-1$, $\tau \le T$.

Note that the only difference compared to ASSUMPTION 3 is that ASSUMPTION 6-3 holds only for $\forall t: 1 \le t \le \tau-1$, but not for $t=\tau$. ASSUMPTION 6-4 is a similar reformulation of ASSUMPTION 4:

[21] An example: There are several evaluations studies where researchers have access to good data explaining the assignment to the first element of a sequence, but intermediate results are missing or are less reliable (see Gerfin and Lechner, 2000, for such a case).

Assumption 6-4 (sequential endogenous conditional independence assumption OPS, SE-CIA-OPS)

a) $H^{\underline{s}_t},...,H^{\underline{s}_\tau} \amalg S_t \mid \underline{H}_{t-1} = \underline{h}_{t-1}, \underline{S}_{t-1} = \underline{s}_{t-1}$;

b) $1 > P(S_t = s_t \mid \underline{H}_{t-1} = \underline{h}_{t-1}, \underline{S}_{t-1} = \underline{s}_{t-1}) > 0$, $\forall \underline{h}_{t-1} \in \underline{\mathbf{H}}_{t-1}$, $\forall \underline{s}_t \in \underline{\mathbb{S}}_t$, $\forall t : 1 \leq t \leq \tau - 1$, $\tau \leq T$.

The bounds for these cases are given by THEOREM 5-3/4:

Theorem 5-3/4 (bounds under S-CIA-OPS) for $\theta^{(\underline{s}^k_{\tau-1},s),(\underline{s}^l_{\tau-1},s)}(\underline{s}^k_{\tau-1}, s)$

Given ASSUMPTION 1-II and either ASSUMPTION 6-3 or ASSUMPTION 6-4 hold, then Table 3a and 3b contain bounds for $\theta^{(\underline{s}^k_{\tau-1},s),(\underline{s}^l_{\tau-1},s)}(\underline{s}^k_{\tau-1}, s)$ and their corresponding widths. Their particular expression depend on (observable) constellations of the data.

Table 3a: Upper and lower bounds for $\theta^{(\underline{s}^k_{\tau-1},s),(\underline{s}^l_{\tau-1},s)}(\underline{s}^k_{\tau-1}, s)$

Cases	$P^{s_\tau \mid \underline{s}^k_{\tau-1}} > \frac{O^{\underline{s}^l_{\tau-1}}(\underline{s}^k_{\tau-1}) - {}_LY}{{}^UY - {}_LY}$	$P^{s_\tau \mid \underline{s}^k_{\tau-1}} = \frac{O^{\underline{s}^l_{\tau-1}}(\underline{s}^k_{\tau-1}) - {}_LY}{{}^UY - {}_LY}$	$P^{s_\tau \mid \underline{s}^k_{\tau-1}} < \frac{O^{\underline{s}^l_{\tau-1}}(\underline{s}^k_{\tau-1}) - {}_LY}{{}^UY - {}_LY}$
$P^{s_\tau \mid \underline{s}^k_{\tau-1}} > \frac{{}^UY - O^{\underline{s}^l_{\tau-1}}(\underline{s}^k_{\tau-1})}{{}^UY - {}_LY}$	$[B^{(\underline{s}^k_{\tau-1},s)} - \frac{O^{\underline{s}^l_{\tau-1}}(\underline{s}^k_{\tau-1}) - (1 - P^{s_\tau \mid \underline{s}^k_{\tau-1}})\,{}_LY}{P^{s_\tau \mid \underline{s}^k_{\tau-1}}},$ $B^{(\underline{s}^k_{\tau-1},s)} - \frac{O^{\underline{s}^l_{\tau-1}}(\underline{s}^k_{\tau-1}) - (1 - P^{s_\tau \mid \underline{s}^k_{\tau-1}})\,{}^UY}{P^{s_\tau \mid \underline{s}^k_{\tau-1}}}]$	$[B^{(\underline{s}^k_{\tau-1},s)} - {}^UY,$ $B^{(\underline{s}^k_{\tau-1},s)} - {}^UY + \frac{(1 - P^{s_\tau \mid \underline{s}^k_{\tau-1}})({}^UY - {}_LY)}{P^{s_\tau \mid \underline{s}^k_{\tau-1}}}]$	$[B^{(\underline{s}^k_{\tau-1},s)} - {}^UY,$ $B^{(\underline{s}^k_{\tau-1},s)} - \frac{O^{\underline{s}^l_{\tau-1}}(\underline{s}^k_{\tau-1}) - (1 - P^{s_\tau \mid \underline{s}^k_{\tau-1}})\,{}^UY}{P^{s_\tau \mid \underline{s}^k_{\tau-1}}}]$
$P^{s_\tau \mid \underline{s}^k_{\tau-1}} = \frac{{}^UY - O^{\underline{s}^l_{\tau-1}}(\underline{s}^k_{\tau-1})}{{}^UY - {}_LY}$	$[B^{(\underline{s}^k_{\tau-1},s)} - {}_LY - \frac{(1 - P^{s_\tau \mid \underline{s}^k_{\tau-1}})({}^UY - {}_LY)}{P^{s_\tau \mid \underline{s}^k_{\tau-1}}},$ $B^{(\underline{s}^k_{\tau-1},s)} - {}_LY]$	$[B^{(\underline{s}^k_{\tau-1},s)} - {}^UY, B^{(\underline{s}^k_{\tau-1},s)} - {}_LY]$ *)	$[B^{(\underline{s}^k_{\tau-1},s)} - {}^UY, B^{(\underline{s}^k_{\tau-1},s)} - {}_LY]$
$P^{s_\tau \mid \underline{s}^k_{\tau-1}} < \frac{{}^UY - O^{\underline{s}^l_{\tau-1}}(\underline{s}^k_{\tau-1})}{{}^UY - {}_LY}$	$[B^{(\underline{s}^k_{\tau-1},s)} - \frac{O^{\underline{s}^l_{\tau-1}}(\underline{s}^k_{\tau-1}) - (1 - P^{s_\tau \mid \underline{s}^k_{\tau-1}})\,{}_LY}{P^{s_\tau \mid \underline{s}^k_{\tau-1}}},$ $B^{(\underline{s}^k_{\tau-1},s)} - {}_LY]$	$[B^{(\underline{s}^k_{\tau-1},s)} - {}^UY, B^{(\underline{s}^k_{\tau-1},s)} - {}_LY]$	$[B^{(\underline{s}^k_{\tau-1},s)} - {}^UY, B^{(\underline{s}^k_{\tau-1},s)} - {}_LY]$

Note: $P^{s_\tau \mid \underline{s}^k_{\tau-1}} = P(S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}^k_{\tau-1})$, $O^{\underline{s}^l_{\tau-1}}(\underline{s}^k_{\tau-1}) = E(Y^{\underline{s}^l_{\tau-1}} \mid \underline{S}_{\tau-1} = \underline{s}^k_{\tau-1})$, $B^{\underline{s}^k_{\tau-1}} = E(Y \mid \underline{S}_{\tau-1} = \underline{s}^k_{\tau-1})$.

*) This case is only possible if $P^{s_\tau \mid \underline{s}^k_{\tau-1}} = 0.5$ and $O^{\underline{s}^l_{\tau-1}}(\underline{s}^k_{\tau-1}) = 0.5({}^UY - {}_LY)$. Note also that cases defined by an inequality converge smoothly towards the cases defined by an equality.

Table 3b: Width of the bounds for $\theta^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}(\underline{s}_{\tau-1}^{k},s)$

Cases	$P^{s_\tau\mid\underline{s}_{\tau-1}^{k}} > \frac{O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{k}) - {}_{L}Y}{{}^{U}Y - {}_{L}Y}$	$P^{s_\tau\mid\underline{s}_{\tau-1}^{k}} = \frac{O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{k}) - {}_{L}Y}{{}^{U}Y - {}_{L}Y}$	$P^{s_\tau\mid\underline{s}_{\tau-1}^{k}} < \frac{O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{k}) - {}_{L}Y}{{}^{U}Y - {}_{L}Y}$
$P^{s_\tau\mid\underline{s}_{\tau-1}^{k}} > \frac{{}^{U}Y - O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{k})}{{}^{U}Y - {}_{L}Y}$	$\frac{1-P^{s_\tau\mid\underline{s}_{\tau-1}^{k}}}{P^{s_\tau\mid\underline{s}_{\tau-1}^{k}}}({}^{U}Y - {}_{L}Y)$	$\frac{1-P^{s_\tau\mid\underline{s}_{\tau-1}^{k}}}{P^{s_\tau\mid\underline{s}_{\tau-1}^{k}}}({}^{U}Y - {}_{L}Y)$	$\frac{{}^{U}Y - O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{k})}{P^{s_\tau\mid\underline{s}_{\tau-1}^{k}}}$
$P^{s_\tau\mid\underline{s}_{\tau-1}^{k}} = \frac{{}^{U}Y - O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{k})}{{}^{U}Y - {}_{L}Y}$	$\frac{1-P^{s_\tau\mid\underline{s}_{\tau-1}^{k}}}{P^{s_\tau\mid\underline{s}_{\tau-1}^{k}}}({}^{U}Y - {}_{L}Y)$	${}^{U}Y - {}_{L}Y$ *	${}^{U}Y - {}_{L}Y$
$P^{s_\tau\mid\underline{s}_{\tau-1}^{k}} < \frac{{}^{U}Y - O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{k})}{{}^{U}Y - {}_{L}Y}$	$\frac{O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{k}) - {}_{L}Y}{P^{s_\tau\mid\underline{s}_{\tau-1}^{k}}}$	${}^{U}Y - {}_{L}Y$	${}^{U}Y - {}_{L}Y$

Note: See note below Table 3a.

The explicit derivation of the bounds is shown in Appendix A.8.

THEOREM 5-3/4 is only concerned with cases where the treatments in the last period of both sequences, i.e. the period in which no randomisation occurs, coincide. Otherwise, the OPS versions of the assumptions are not informative at all. Without any information about the assignment mechanism in that period, we know nothing about the effects of the final treatment (other than the information implied by the *worst-case* bounds). In particular, the OPS assumption would not exclude the possibility that the effects of the last treatment dominate the effects of all previous treatments in the sequences. On the other hand, the bounds for these effects are also non-trivial, although one might reason that because the effects of sequences up to $\tau-1$ are identified, and the final element of each sequence is the same, there should be identification at least for populations defined by treatment status up to period $\tau-1$. This is not true however, because such an argument requires the assumption that the effect of the last treatment on the potential outcomes is independent of the treatments received in previous periods. Such an assumptions does not appear to be in the spirit of this paper.

The bounds for the potential outcomes obtained under these assumptions cannot be larger than the worst-case bounds, therefore we get different expressions for four different cases. It complicates the presentation of the bounds, but there is no conceptional problem in an application, because the quantities necessary to determine the applicable case are identified. However, inference on the bounds is of course complicated by this fact.

Table 3b shows that the sharpness of the bounds depend mainly on the probability of being observed in the final period in state *s*, that is the same for both sequences. In fact this a very intuitive result: If the whole population would be assigned to that state in last period of the sequence, then the OPS assumptions would obviously identify $\theta^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}(\underline{s}_{\tau-1}^{k},s)$. The less likely it is to be observed in the state common to both sequences the more matters the extra uncertainty for the final period that is induced by OPS.

Example (continued - 8)

Let us again consider the case of a binary outcome variable with two treatments each period as well as the comparison between sequences (0,0,0) and (0,1,0). Tables 3a-E and 3b-E show that the quantities appearing in Table 3a and Table 3b simplify considerably in this case:

Table 3a-E: Upper and lower bounds for $\theta^{(0,0,0),(0,1,0)}(0,0,0)$

Cases	$P^{(0,0,0)\|(0,0)} > O^{(0,1)}(0,0)$	$P^{(0,0,0)\|(0,0)} \leq O^{(0,1)}(0,0)$
$P^{(0,0,0)\|(0,0)} > 1-O^{(0,1)}(0,0)$	$[B^{(0,0,0)} - \frac{O^{(0,1)}(0,0)}{P^{(0,0,0)\|(0,0)}},\ B^{(0,0,0)} - \frac{O^{(0,1)}(0,0)-1+P^{(0,0,0)\|(0,0)}}{P^{(0,0,0)\|(0,0)}}]$	$[B^{(0,0,0)}-1,\ B^{(0,0,0)} - \frac{O^{(0,1)}(0,0)-(1-P^{(0,0,0)\|(0,0)})}{P^{(0,0,0)\|(0,0)}}]$
$P^{(0,0,0)\|(0,0)} \leq 1-O^{(0,1)}(0,0)$	$\left[B^{(0,0,0)} - \frac{O^{(0,1)}(0,0)}{P^{(0,0,0)\|(0,0)}}, B^{(0,0,0)}\right]$	$[B^{(0,0,0)}-1, B^{(0,0,0)}]$

Table 3b-E: Width of the bounds for $\theta^{(0,0,0),(0,1,0)}(0,0,0)$

	$P^{(0,0,0)\|(0,0)} > O^{(0,1)}(0,0)$	$P^{(0,0,0)\|(0,0)} \leq O^{(0,1)}(0,0)$
$P^{(0,0,0)\|(0,0)} > 1-O^{(0,1)}(0,0)$	$\frac{1-P^{(0,0,0)\|(0,0)}}{P^{(0,0,0)\|(0,0)}}$	$\frac{1-O^{(0,1)}(0,0)}{P^{(0,0,0)\|(0,0)}}$
$P^{(0,0,0)\|(0,0)} \leq 1-O^{(0,1)}(0,0)$	$\frac{O^{(0,1)}(0,0)}{P^{(0,0,0)\|(0,0)}}$	1

In Appendix B we derive also expressions for bounds based on different conditioning sets $\theta^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}(\underline{s}_{\tau-1}^{j},s)$ as well as coarser conditioning sets $\theta^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}(\underline{s}_{\tau-1}^{k})$. They have a similar structure as the bounds presented in THEOREM 5-3/4. Therefore, we refrain from discussing them here explicitly and refer the reader interested in the exact expressions to Appendix B.

It remains to establish bounds for an OPS version of Assumption 5. We show in Section 3 that ASSUMPTION 5 provides less information about the effects than ASSUMPTION 3 or 4. Therefore bounds derived under an OPS version of ASSUMPTION 5 cannot be sharper than the ones derived under ASSUMPTIONS 5-3 or 5-4. In particular, when the sequences do not coincide in the final period, we do not get more than the no-information bounds.

Generally, the derivation of these bounds is again analogous to the ones obtained in THEOREM 5-3/4, although sometimes more tedious. Therefore, we refrain from presenting them in the main body of this paper and refer the interested reader to Appendix C. We obtain as a by-product of the considerations in this appendix, also a couple of bounds for effects that are not identified by ASSUMPTION 5, but would be identified by ASSUMPTION 4. These considerations are independent of OPS and thus carry additional information about the different informational content of those two assumptions.

5 Conclusion

This paper approaches the problem of an econometric evaluation of dynamic programme sequences from an potential outcome perspective. We discuss the identifying power of several different stylised assumptions about the connection between the dynamic selection process and the potential outcomes of the different sequences of programmes. These assumptions invoke different sorts of randomisation compatible with different types of selection regimes. All assumptions are framed in a way so that they need to be, and potentially can be, justified by sufficient knowledge about the selection process in conjunction with sufficiently rich data. Parametric forms are not involved. In our most general case participation in the sequences is building up in the sense that the decision in what programme 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 prototypical for the selection mechanism in many European active labour market programmes. However, due to the endogeneity problem of selection on outcomes of past treatments, not all parameters of interest are identified, if the information set does not follow the same type of randomisation process as the outcome variables. 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 still identified.

We also present a bounds analysis to see what information is revealed by assumptions that are plausible but do not allow an exact identification of the treatment parameter of interest. Since the bounds are complex, we study only some examples to investigate some specific situations considered to be of particular practical interest.

We left many topics open for future research: First, with respect to estimation this paper does not explore the identifying power of assumption similar to the ones coming from static instrumental variable and static nonparametric selection models. Furthermore, the issue of estimation is yet unresolved, although we conjecture that the form of our identification results allows a straightforward application of nonparametric methods extensively discussed in the static evaluation literature, particularly matching methods. However, the dimensionality problem due to many different sequences will become central in designing estimators for this kind of framework. Nevertheless, reversing to a completely parametric framework for the estimation stage does not appear to be attractive.

References

Angrist, J. D. and A. B. Krueger (1999): "Empirical Strategies in Labor Economics," in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Vol. III A, chapter 23, 1277-1366.

Ashenfelter, O. (1978): "Estimating the Effect of Training Programs on Earnings ", *The Review of Economics and Statistics*, 58, 47-57.

Balke, A., and J. Pearl (1997): "Bounds on Treatment Effects from Studies with Imperfect Compliance", *Journal of the American Statistical Association*, 92, 1171-1176.

Baltagi, B.H. (1995), *Econometric Analysis of Panel Data*, Chichester: Wiley.

Brodaty, T., B. Crepon, and D. Fougere (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.

Hahn, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects", *Econometrica*, 66, 315-331.

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs", forthcoming in O. Ashenfelter and D. Card (eds.): *Handbook of Labor Economics*, Vol. III A, , 1865-2097, North-Holland, Amsterdam.

Heckman, J. J., H. Ichimura, and P. E. Todd (1998): "Matching as an Econometric Evaluation Estimator", *Review of Economic Studies*, 65, 261-294.

Heckman, J.J., and E. Vytlacil (2001a): "Instrumental variables, selection models, and tight bounds on the average treatment effects", in M. Lechner and F. Pfeiffer (eds.), *Econometric Evaluation of Active Labour Market Policies*, Heidelberg: Physica, 1-15.

Heckman, J.J., and E. Vytlacil (2001b): "Econometric Evaluation of Social Programs", forthcoming in J. J. Heckman and E. Leamer, eds., *Handbook of Econometrics, Volume 5*, North-Holland: Amsterdam.

Hirano, K., G. W. Imbens, and G. Ridder (2000): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score", *NBER Technical Working Papers*, 251.

Imbens, G. W. (2000): "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*, 87, 706-710.

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

Lalive, R., J. C. van Ours, and J. Zweimüller (2000): "The Impact of Active Labor Market Policies and Benefit Entitlement Rules on the Duration of Unemployment", *mimeo*.

Lechner, M. (1999): "Nonparametric Bounds on Employment and Income Effects of Continuous Vocational Training in East Germany," *The Econometric Journal*, 2, 1-28.

Lechner, M. (2001a): "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. (2001b): "Programme Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labour Market Policies", forthcoming in *The Review of Economics and Statistics*.

Lechner, M. (2001c): "A note on the common support assumption in microeconometric evaluation studies", *Discussion paper 2001-01, University of St. Gallen.*

Lechner, M., and F. Pfeiffer, (eds., 2001), *Econometric Evaluation of Active Labour Market Policies*, Heidelberg: Physica.

Manski, C.F. (1989): "The Anatomy of the Selection Problem", *Journal of Human Resources*, 24, 343-360.

Manski, C.F. (1990): "Nonparametric Bounds on Treatment Effects", *American Economic Review, Papers and Proceedings*, 80, 319-323.

Murphy, S. A., M. van der Laan, J. M. Robins and CPPRG (2000): "Marginal Mean Models for Dynamic Regimes", mimeo, presented at the Panel 2000 conference in Geneva.

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.

Pearl, J. (2000), *Causality - Models, Reasoning, and Inference*, Cambridge: Cambridge University Press, Chapter 8.

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", 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),. NY: Springer Verlag, pp. 69-117.

Robins, J. M., S. Greenland, and F. Hu (1999): "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, with discussion.

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

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.

van den Berg, G. J. (2000): "Duration Models: Specification, Identification, and Multiple Durations". forthcoming in Heckman, J. J., and E. Leamer (eds.), *Handbook of Econometrics*, vol. V, Amsterdam: North-Holland.

Vytlacil, E. (1999): "Independence, monotonicity and latent variable models: an equivalence result“, forthcoming in *Econometrica*.

Appendix A: Proofs of Lemmas and Theorems

A.1 Proof of Lemma 2-I

To simplify the notation, first consider the case $\delta = 1$ only. Once the connection from one period to another is established, all periods can be connected the same way. For the potential outcomes we have the following relations:

$$E(Y_t^{\underline{s}_\tau^k} \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j) = \sum_{m=1}^{M} E[\underline{1}(S_{\tau+1} = m) Y_t^{(\underline{s}_\tau^k, m)} \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j] =$$

$$= \sum_{m=1}^{M} \underset{S_{\tau+1}}{E} \{E[\underline{1}(S_{\tau+1} = m) Y_t^{(\underline{s}_\tau^k, m)} \mid S_{\tau+1} = m, \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j] \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j\} =$$

$$= \sum_{m=1}^{M} \underset{S_{\tau+1}}{E} \{\underline{1}(S_{\tau+1} = m)[E(Y_t^{(\underline{s}_\tau^k, m)} \mid S_{\tau+1} = m, \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j)] \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j\} =$$

$$= \sum_{m=1}^{M} P(S_{\tau+1} = m \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j) E(Y_t^{(\underline{s}_\tau^k, m)} \mid S_{\tau+1} = m, \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j) =: E(Y_t^{(\underline{s}_\tau^k, \bullet)} \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j); \qquad (10)$$

$$(E[\underline{1}(S_{\tau+1} = m) \mid S_{\tau+1} = a, \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j] = \begin{cases} 0 & \forall a \neq m \\ 1 & a = m \end{cases}).$$

These results can be extended to the more general case of δ periods ahead:

$$E(Y_t^{\underline{s}_\tau^k} \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j) = \sum_{m_1=1}^{M} \cdots \sum_{m_\delta=1}^{M} E\{[\underline{1}(S_{\tau+1} = m_1) \ \dots \ \underline{1}(S_{\tau+\delta} = m_\delta)] Y_t^{(\underline{s}_\tau^k, m_1, \dots, m_\delta)} \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j\} =$$

$$= \sum_{m_1=1}^{M} \cdots \sum_{m_\delta=1}^{M} \underset{S_{\tau+1},\dots,S_{\tau+\delta}}{E} \{E\{[\underline{1}(S_{\tau+1} = m_1) \dots \underline{1}(S_{\tau+\delta} = m_\delta)] Y_t^{(\underline{s}_\tau^k, m_1, \dots, m_\delta)} \mid S_{\tau+1} = s_{\tau+1}, \dots, S_{\tau+\delta} = s_{\tau+\delta}, \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j\} \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j\} =$$

$$= \sum_{m_1=1}^{M} \cdots \sum_{m_\delta=1}^{M} E\{[\underline{1}(S_{\tau+1} = m_1) \dots \underline{1}(S_{\tau+\delta} = m_\delta)] E[Y_t^{(\underline{s}_\tau^k, m_1, \dots, m_\delta)} \mid S_{\tau+1} = s_{\tau+1}, \dots, S_{\tau+\delta} = s_{\tau+\delta}, \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j] \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j\} =$$

$$= \sum_{m_1=1}^{M} \cdots \sum_{m_\delta=1}^{M} P(S_{\tau+1} = m_1, \dots, S_{\tau+\delta} = m_\delta \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j) E(Y_t^{(\underline{s}_\tau^k, m_1, \dots, m_\delta)} \mid S_{\tau+1} = m_1, \dots, S_{\tau+\delta} = m_\delta, \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j)$$

$$=: E[Y_t^{(\underline{s}_\tau^k,\bullet,\dots,\bullet)} \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j]. \tag{11}$$

Using the result of equations (10) and (11) we obtain the desired results for the connection of $\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_{\tilde{\tau}}^j)$ and $\theta_t^{\underline{s}_{\tau+1}^k,\underline{s}_{\tau+1}^l}(\underline{s}_{\tilde{\tau}}^j)$ as well as for the more general case of δ periods ahead ($\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_{\tilde{\tau}}^j)$, $\theta_t^{\underline{s}_{\tau+\delta}^k,\underline{s}_{\tau+\delta}^l}(\underline{s}_{\tilde{\tau}}^j,\cdot,s_{\tau+1},\dots,s_{\tau+\delta})$).

$$\theta_t^{\underline{s}_\tau^k,\underline{s}_\tau^l}(\underline{s}_{\tilde{\tau}}^j) = \sum_{m=1}^{M} P(S_{\tau+1} = m \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j)\, \theta_t^{(\underline{s}_\tau^k,m),(\underline{s}_\tau^l,m)}(\underline{s}_{\tilde{\tau}}^j,\cdot,s_{\tau+1} = m) =$$

$$= \sum_{m_1=1}^{M}\dots\sum_{m_\delta=1}^{M} P(S_{\tau+1} = m_1,\dots,S_{\tau+\delta} = m_\delta \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j)\, \theta_t^{(\underline{s}_\tau^k,m_1,\dots,m_\delta),(\underline{s}_\tau^l,m_1,\dots,m_\delta)}(\underline{s}_{\tilde{\tau}}^j,\cdot,s_{\tau+1} = m_1,\dots,s_{\tau+\delta} = m_\delta).$$ q.e.d.

A.2 Proof of Lemma 3

We show that ASSUMPTION 2-II in fact implies independence as defined in 2-I:[22]

$$F(Y^{\underline{s}_t^k},\underline{S}_t \mid X_0) = F(Y^{\underline{s}_t^k},S_t \mid \underline{S}_{t-1},X_0)F(\underline{S}_{t-1} \mid X_0) =$$

$$\overset{2-II}{=} F(Y^{\underline{s}_t^k} \mid \underline{S}_{t-1},X_0)F(S_t \mid \underline{S}_{t-1},X_0)F(\underline{S}_{t-1} \mid X_0) = F(Y^{\underline{s}_t^k} \mid \underline{S}_{t-1},X_0)F(\underline{S}_t \mid X_0).$$

Repeating the factorization of $F(Y^{\underline{s}_t^k},\underline{S}_{t-1} \mid X_0)$ until period 0, we obtain the desired result:

$$F(Y^{\underline{s}_t^k},\underline{S}_t \mid X_0) = F(Y^{\underline{s}_t^k} \mid X_0)F(\underline{S}_t \mid X_0).$$ q.e.d.

A.3 Proof of Lemma 4

In the first part, we show that $X_{t-1} \amalg S_{t-1} \mid \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2} = \underline{x}_{t-2}$ implies the noncausality condition. Then the reverse is shown.[23]

---

[22] 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*.

[23] For brevity, we concentrate on the case *t>2*. The proof for *t=1* follows exactly the same lines as the proof given.

Using the observation rules for X_{t-1} conditional on the observed history until period $t-2$ ($\underline{S}_{t-2} = \underline{s}_{t-2}$) ($X_{t-1} = \sum_{m=1}^{M} \underline{1}(S_{t-1} = m) X_{t-1}^{(\underline{s}_{t-2}, m)}$) it follows:

$$F(X_{t-1} \mid S_{t-1} = s_{t-1}^{l}, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2}) =$$

$$= F(\sum_{m=1}^{M} \underline{1}(S_{t-1} = m) X_{t-1}^{(\underline{s}_{t-2}, m)} \mid S_{t-1} = s_{t-1}^{l}, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2})$$

$$= F(X_{t-1}^{(\underline{s}_{t-2}, s_{t-1}^{l})} \mid S_{t-1} = s_{t-1}^{l}, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2})$$

$$= F(X_{t-1} \mid S_{t-1} = s_{t-1}^{l}, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2})$$

$$\stackrel{3b}{=} F(X_{t-1} \mid S_{t-1} = s_{t-1}^{k}, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2})$$

$$\stackrel{3b}{=} F(X_{t-1}^{(\underline{s}_{t-2}, s_{t-1}^{k})} \mid S_{t-1} = s_{t-1}^{k}, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2}).$$

Therefore, we have shown that ASSUMPTION 3b implies:

$$F(X_{t-1}^{(\underline{s}_{t-2}, s_{t-1}^{k})} \mid S_{t-1} = s_{t-1}^{k}, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2}) = F(X_{t-1}^{(\underline{s}_{t-2}, s_{t-1}^{l})} \mid S_{t-1} = s_{t-1}^{l}, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2}). \qquad (12)$$

This proofs the first part. For the second part, we reverse the argument. Substituting in equation (12) the observed outcomes, gives us the desired result:

$$F(X_{t-1} \mid S_{t-1} = s_{t-1}^{k}, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2}) = F(X_{t-1}^{(\underline{s}_{t-2}, s_{t-1}^{k})} \mid S_{t-1} = s_{t-1}^{k}, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2})$$

$$F(X_{t-1} \mid S_{t-1} = s_{t-1}^{l}, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2}) = F(X_{t-1}^{(\underline{s}_{t-2}, s_{t-1}^{l})} \mid S_{t-1} = s_{t-1}^{l}, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2})$$

Therefore, we obtain using (12):

$$F(X_{t-1} \mid S_{t-1} = s_{t-1}^{k}, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2}) = F(X_{t-1} \mid S_{t-1} = s_{t-1}^{l}, \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2}) =$$

$$= F(X_{t-1} \mid \underline{S}_{t-2} = \underline{s}_{t-2}, \underline{X}_{t-2}). \qquad \text{q.e.d.}$$

Thus, both conditions are indeed equivalent.

A.4 Proof of Theorem 1

THEOREM 1 states that if ASSUMPTIONS 1 and 2 (I or II) hold, then there is identification. ASSUMPTIONS 1 and 2b) ensure that the moments exist. Next, we show that $F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^j)$ is identified. If $F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^j)$ is identified for all values of *j* and *k*, then all treatment effects are identified, because cases where the conditioning set is less coarse as well as cases where the sequences considered are shorter are identified as well by the relations derived in LEMMA 1 and 2.

$$F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^j) = \underset{X_0}{E}[F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^j, X_0) \mid \underline{S}_t = \underline{s}_t^j] \overset{2}{=} \underset{X_0}{E}[F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^k, X_0) \mid \underline{S}_t = \underline{s}_t^j]$$

$$= \underset{X_0}{E}[F(Y \mid \underline{S}_t = \underline{s}_t^k, X_0) \mid \underline{S}_t = \underline{s}_t^j].$$

Thus, $F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^j)$ is identified from observables if ASSUMPTION 2-I or 2-II is true.

Furthermore, ASSUMPTIONS 1 and 3 also identify $F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^j)$:

$$F(Y^{\underline{s}_t^k} \mid \underline{X}_{t-1}, \underline{S}_t = \underline{s}_t^j) \overset{3a}{=} F(Y^{\underline{s}_t^k} \mid \underline{X}_{t-1}, \underline{S}_{t-1} = \underline{s}_{t-1}^j) \quad [\overset{3a}{=} F(Y^{\underline{s}_t^k} \mid \underline{X}_{t-1}, \underline{S}_{t-1} = \underline{s}_{t-1}^j, S_t = s_t^k)]$$

$$F(Y^{\underline{s}_t^k} \mid \underline{X}_{t-1}, \underline{S}_{t-1} = \underline{s}_{t-1}^j) = \frac{F(Y^{\underline{s}_t^k}, X_{t-1} \mid \underline{X}_{t-2}, \underline{S}_{t-1} = \underline{s}_{t-1}^j)}{F(X_{t-1} \mid \underline{X}_{t-2}, \underline{S}_{t-1} = \underline{s}_{t-1}^j)}$$

$$\overset{3a,3b}{=} \frac{F(Y^{\underline{s}_t^k}, X_{t-1} \mid \underline{X}_{t-2}, \underline{S}_{t-2} = \underline{s}_{t-2}^j, S_{t-1} = s_{t-1}^k)}{F(X_{t-1} \mid \underline{X}_{t-2}, \underline{S}_{t-2} = \underline{s}_{t-2}^j, S_{t-1} = s_{t-1}^k)} = F(Y^{\underline{s}_t^k} \mid \underline{X}_{t-1}, \underline{S}_{t-2} = \underline{s}_{t-2}^j, S_{t-1} = s_{t-1}^k).$$

$$[= F(Y^{\underline{s}_t^k} \mid \underline{X}_{t-1}, \underline{S}_{t-2} = \underline{s}_{t-2}^j, S_{t-1} = s_{t-1}^k, S_t = s_t^k)]$$

By repeating this argument up to period 0, we obtain:

$$F(Y^{\underline{s}_t^k} \mid \underline{X}_{t-1}, \underline{S}_t = \underline{s}_t^j) = F(Y^{\underline{s}_t^k} \mid \underline{X}_{t-1}, \underline{S}_t = \underline{s}_t^k) = F(Y \mid \underline{X}_{t-1}, \underline{S}_t = \underline{s}_t^k).$$

Hence, we have identification because of $F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^j) = \underset{\underline{X}_{t-1}}{E}[F(Y \mid \underline{S}_t = \underline{s}_t^k, \underline{X}_{t-1}) \mid \underline{S}_t = \underline{s}_t^j]$.

A.5 Proof of Theorem 2

As before we show that $F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^j)$ is identified.

$$F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^j) = \underset{\underline{H}_{t-1}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^j, \underline{H}_{t-1}) \mid \underline{S}_t = \underline{s}_t^j\}$$

$$\overset{4}{=} \underset{\underline{H}_{t-1}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_{t-1} = \underline{s}_{t-1}^j, S_t = s_t^k, \underline{H}_{t-1}) \mid \underline{S}_t = \underline{s}_t^j\}$$

$$\overset{4}{=} \underset{\underline{H}_{t-1}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_{t-1} = \underline{s}_{t-1}^j, \underline{H}_{t-1}) \mid \underline{S}_t = \underline{s}_t^j\}.$$

Using the observation rules, $H_{t-1} = H_{t-1}^{\underline{s}_t^j}$ if $\underline{S}_t = \underline{s}_t^j$. Therefore, we have:

$$\underset{\underline{H}_{t-1}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_{t-1} = \underline{s}_{t-1}^j, \underline{H}_{t-1}) \mid \underline{S}_t = \underline{s}_t^j\} = \underset{\underline{H}_{t-2}, H_{t-1}^{\underline{s}_t^j}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_{t-1} = \underline{s}_{t-1}^j, \underline{H}_{t-2}, H_{t-1}^{\underline{s}_t^j}) \mid \underline{S}_t = \underline{s}_t^j\}.$$

$$F(Y^{\underline{s}_t^k} \mid \underline{S}_{t-1} = \underline{s}_{t-1}^j, \underline{H}_{t-2}, H_{t-1}^{\underline{s}_t^j}) = \frac{F(Y^{\underline{s}_t^k}, H_{t-1}^{\underline{s}_t^j} \mid \underline{S}_{t-1} = \underline{s}_{t-1}^j, \underline{H}_{t-2})}{F(H_{t-1}^{\underline{s}_t^j} \mid \underline{S}_{t-1} = \underline{s}_{t-1}^j, \underline{H}_{t-2})}$$

$$\overset{4}{=} \frac{F(Y^{\underline{s}_t^k}, H_{t-1}^{\underline{s}_t^j} \mid \underline{S}_{t-2} = \underline{s}_{t-2}^j, S_{t-1} = \underline{s}_{t-1}^k, \underline{H}_{t-2})}{F(H_{t-1}^{\underline{s}_t^j} \mid \underline{S}_{t-2} = \underline{s}_{t-2}^j, S_{t-1} = \underline{s}_{t-1}^k, \underline{H}_{t-2})}$$

$$\overset{4}{=} \frac{F(Y^{\underline{s}_t^k}, H_{t-1}^{\underline{s}_t^j} \mid \underline{S}_{t-2} = \underline{s}_{t-2}^j, \underline{H}_{t-2})}{F(H_{t-1}^{\underline{s}_t^j} \mid \underline{S}_{t-2} = \underline{s}_{t-2}^j, \underline{H}_{t-2})}$$

$$= F(Y^{\underline{s}_t^k} \mid \underline{S}_{t-2} = \underline{s}_{t-2}^j, \underline{H}_{t-2}, H_{t-1}^{\underline{s}_t^j}),$$

$$\Rightarrow \underset{\underline{H}_{t-1}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^j, \underline{H}_{t-1}) \mid \underline{S}_t = \underline{s}_t^j\} = \underset{\underline{H}_{t-2}, H_{t-1}^{\underline{s}_t^j}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_{t-2} = \underline{s}_{t-2}^j, \underline{H}_{t-2}, H_{t-1}^{\underline{s}_t^j}) \mid \underline{S}_t = \underline{s}_t^j\}.$$

Now the same steps are repeated for H_{t-2} ($H_{t-2} = H_{t-2}^{\underline{s}_t^j}$, given $\underline{S}_t = \underline{s}_t^j$):

$$\underset{\underline{H}_{t-1}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^j, \underline{H}_{t-1}) \mid \underline{S}_t = \underline{s}_t^j\} = \underset{\underline{H}_{t-3}, H_{t-2}^{\underline{s}_t^j}, H_{t-1}^{\underline{s}_t^j}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_{t-2} = \underline{s}_{t-2}^j, \underline{H}_{t-3}, H_{t-2}^{\underline{s}_t^j}, H_{t-1}^{\underline{s}_t^j}) \mid \underline{S}_t = \underline{s}_t^j\}$$

$$\overset{4}{=} \underset{\underline{H}_{t-3}, H_{t-2}^{\underline{s}_t^j}, H_{t-1}^{\underline{s}_t^j}}{E} \{F(Y^{\underline{s}_t^k} \mid \underline{S}_{t-3} = \underline{s}_{t-3}^j, \underline{H}_{t-3}, H_{t-2}^{\underline{s}_t^j}, H_{t-1}^{\underline{s}_t^j}) \mid \underline{S}_t = \underline{s}_t^j\}\,.$$

Iterating this procedure, we get:

$$\underset{\underline{H}_{t-1}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^j, \underline{H}_{t-1}) \mid \underline{S}_t = \underline{s}_t^j\} = \underset{\underline{H}_{t-1}^{\underline{s}_t^j}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^k, \underline{H}_{t-1}^{\underline{s}_t^j}) \mid \underline{S}_t = \underline{s}_t^j\}\,.$$

But we still consider the law of $\underline{H}_{t-1}^{\underline{s}_t^j}$ given $\underline{S}_t = \underline{s}_t^j$, therefore we get the following equality:

$$\underset{\underline{H}_{t-1}^{\underline{s}_t^j}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^k, \underline{H}_{t-1}^{\underline{s}_t^j}) \mid \underline{S}_t = \underline{s}_t^j\} = \underset{\underline{H}_1, \underline{H}_{2:t-1}^{\underline{s}_t^j}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^k, \underline{H}_1, \underline{H}_{2:t-1}^{\underline{s}_t^j}) \mid \underline{S}_t = \underline{s}_t^j\}$$

$$\overset{4}{=} \underset{\underline{H}_1, \underline{H}_{2:t-1}^{\underline{s}_t^j}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^k, S_2 = s_2^k, \underline{H}_1, \underline{H}_{2:t-1}^{\underline{s}_t^j}) \mid \underline{S}_t = \underline{s}_t^j\}$$

$$= \dots$$

$$\overset{4}{=} \underset{\underline{H}_{t-1}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^k, \underline{H}_{t-1}) \mid \underline{S}_t = \underline{s}_t^j\}$$

$$= \underset{\underline{H}_{t-1}}{E}\{F(Y \mid \underline{S}_t = \underline{s}_t^k, \underline{H}_{t-1}) \mid \underline{S}_t = \underline{s}_t^j\}\,.$$

Therefore, $F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^j) = \underset{\underline{H}_{t-1}}{E}\{F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^k, \underline{H}_{t-1}) \mid \underline{S}_t = \underline{s}_t^j\} = \underset{\underline{H}_{t-1}}{E}\{F(Y \mid \underline{S}_t = \underline{s}_t^k, \underline{H}_{t-1}) \mid \underline{S}_t = \underline{s}_t^j\}$

and $F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^j)$ is identified.

q.e.d.

A.6 Proof of Theorem 3

First note that ASSUMPTION 5 implies the following restrictions:

$$F(Y^{\underline{s}_t^k} \mid S_1 = s_1^j, X_0) = F(Y^{\underline{s}_t^k} \mid S_1 = s_1^k, X_0)\,; \tag{13}$$

$$F(Y^{\underline{s}_t^k} \mid \underline{S}_\tau = \underline{s}_\tau^j, \underline{H}_{\tau-1}) = F(Y^{\underline{s}_t^k} \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^j, S_\tau = s_\tau^k, \underline{H}_{\tau-1})\,,\ \forall t, \tau : 1 < \tau \le t \le T. \tag{14}$$

A.6.1 Part a)

We need to show that $F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^{\,j})$ is identified. Let us consider the starting point for the proof in greater detail. We relate $F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^{\,j})$ to some function of the observable outcomes by sequentially applying the equations (13) and (14) to conditional expectation versions of $F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^{\,j})$:

$$F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^{\,j}) = \underset{X_0}{E}[F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^{\,k}, X_0) \mid \underline{S}_1 = \underline{s}_1^{\,j}];$$

$$F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^{\,k}, X_0) = \underset{H_1, S_2}{E}[F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^{\,k}, S_2, \underline{H}_1) \mid \underline{S}_1 = \underline{s}_1^{\,k}, X_0]$$

$$= \underset{H_1}{E}\{\underset{S_2 \mid H_1}{E}[F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^{\,k}, S_2, \underline{H}_1) \mid \underline{S}_1 = \underline{s}_1^{\,k}, \underline{H}_1] \mid \underline{S}_1 = \underline{s}_1^{\,k}, X_0\} =$$

$$= \underset{H_1}{E}\{\sum_{m=1}^{M}[F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^{\,k}, S_2 = m, \underline{H}_1) P(S_2 = m \mid \underline{S}_1 = \underline{s}_1^{\,k}, \underline{H}_1)] \mid \underline{S}_1 = \underline{s}_1^{\,k}, X_0\}$$

$$\overset{A5}{=} \underset{H_1}{E}\{\sum_{m=1}^{M}[F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^{\,k}, S_2 = s_2^k, \underline{H}_1) P(S_2 = m \mid \underline{S}_1 = \underline{s}_1^{\,k}, \underline{H}_1)] \mid \underline{S}_1 = \underline{s}_1^{\,k}, X_0\}$$

$$= \underset{H_1}{E}[F(Y^{\underline{s}_t^k} \mid \underline{S}_2 = \underline{s}_2^{\,k}, \underline{H}_1) \mid \underline{S}_1 = \underline{s}_1^{\,k}, X_0].$$

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

$$F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^{\,j}) = \underset{X_0}{E}\{\underset{H_1}{E}[F(Y^{\underline{s}_t^k} \mid \underline{S}_2 = \underline{s}_2^{\,k}, \underline{H}_1) \mid \underline{S}_1 = \underline{s}_1^{\,k}, X_0] \mid \underline{S}_1 = \underline{s}_1^{\,j}\}.$$

Continuing in this fashion until period *t* gives us the general proof of identification.

First, the relation between conditioning sets of two consecutive periods is established:

$$F(Y^{\underline{s}_t^k} \mid \underline{S}_\tau = \underline{s}_\tau^{\,k}, \underline{H}_{\tau-1}) = \underset{H_\tau, S_{\tau+1}}{E}[F(Y^{\underline{s}_t^k} \mid \underline{S}_\tau = \underline{s}_\tau^{\,k}, S_{\tau+1}, \underline{H}_\tau) \mid \underline{S}_\tau = \underline{s}_\tau^{\,k}, \underline{H}_{\tau-1}],\ t > \tau.$$

Note that $F(Y^{\underline{s}_t^k} \mid \underline{S}_\tau = \underline{s}_\tau^{\,k}, S_{\tau+1}, \underline{H}_\tau)$ does not depend on $S_{\tau+1}$ by ASSUMPTION 5. Thus, we have:

$$F(Y^{\underline{s}_t^k} \mid \underline{S}_\tau = \underline{s}_\tau^{\,k}, \underline{H}_{\tau-1}) = \underset{H_\tau, S_{\tau+1}}{E}[F(Y^{\underline{s}_t^k} \mid \underline{S}_\tau = \underline{s}_\tau^{\,k}, S_{\tau+1} = s_{\tau+1}^k, \underline{H}_\tau) \mid \underline{S}_\tau = \underline{s}_\tau^{\,k}, \underline{H}_{\tau-1}] =$$

$$= \underset{H_\tau}{E}[F(Y^{\underline{s}_t^k} \mid \underline{S}_{\tau+1} = \underline{s}_{\tau+1}^{\,k}, \underline{H}_\tau) \mid \underline{S}_\tau = \underline{s}_\tau^{\,k}, \underline{H}_{\tau-1}].$$

Applying this result sequentially completes the proof of identification of $\theta_t^{(\underline{s}_\tau^k),(\underline{s}_\tau^l)}$ and $\theta_t^{(\underline{s}_\tau^k),(\underline{s}_\tau^l)}(\underline{s}_1^j)$:

$$F(Y^{\underline{s}_t^k} \mid \underline{S}_1 = \underline{s}_1^j) = \underset{X_0}{E}\underset{H_1}{E}\{\cdots\{\underset{\underline{H}_{t-1}}{E}[F(Y^{\underline{s}_t^k} \mid \underline{S}_t = \underline{s}_t^k, \underline{H}_{t-1}) \mid \underline{S}_{t-1} = \underline{s}_{t-1}^k, \underline{H}_{t-2}]\} \mid \cdots \mid \underline{S}_1 = \underline{s}_1^k, X_0\} \mid \underline{S}_1 = \underline{s}_1^j\} =$$

$$= \underset{X_0}{E}\underset{H_1}{E}\{\cdots\{\underset{H_{t-1}}{E}[F(Y \mid \underline{S}_t = \underline{s}_t^k, \underline{H}_{t-1}) \mid \underline{S}_{t-1} = \underline{s}_{t-1}^k, \underline{H}_{t-2}]\} \mid \cdots \mid \underline{S}_1 = \underline{s}_1^k, X_0\} \mid \underline{S}_1 = \underline{s}_1^j\}. \quad \text{q.e.d.}$$

A.6.2 Part b)

It is sufficient to show that $\theta_t^{(\underline{s}_{\tau-1}^k, s_\tau^k),(\underline{s}_{\tau-1}^k, s_\tau^l)}(\underline{s}_{\tau-1}^k, s_\tau^j)$ is identified, because of $\theta_t^{(\underline{s}_{\tau-1}^k, s_\tau^k),(\underline{s}_{\tau-1}^k, s_\tau^l)}(\underline{s}_{\tau-1}^k)$ = $\sum_{m=1}^{M}\theta_t^{(\underline{s}_{\tau-1}^k, s_\tau^k),(\underline{s}_{\tau-1}^k, s_\tau^l)}(\underline{s}_{\tau-1}^k, m)P(S_\tau = m \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k)$. In the following we show that $F(Y^{\underline{s}_{t-1}^k, s_t^k} \mid \underline{S}_{t-1} = \underline{s}_{t-1}^k, S_t = s_t^j)$ is identified. If this is true then the respective conditional expectations used to define the effects are identified as well, and thus the proof is complete.

$$F(Y^{\underline{s}_{t-1}^k, s_t^k} \mid \underline{S}_{t-1} = \underline{s}_{t-1}^k, S_t = s_t^j) = \underset{\underline{H}_{t-1}}{E}[F(Y^{\underline{s}_{t-1}^k, s_t^k} \mid \underline{S}_{t-1} = \underline{s}_{t-1}^k, S_t = s_t^j, \underline{H}_{t-1}) \mid \underline{S}_{t-1} = \underline{s}_{t-1}^k, S_t = s_t^j]$$

$$\overset{A.5}{=} \underset{\underline{H}_{t-1}}{E}[F(Y^{\underline{s}_{t-1}^k, s_t^k} \mid \underline{S}_{t-1} = \underline{s}_{t-1}^k, S_t = s_t^k, \underline{H}_{t-1}) \mid \underline{S}_{t-1} = \underline{s}_{t-1}^k, S_t = s_t^j] =$$

$$= \underset{\underline{H}_{t-1}}{E}[F(Y \mid \underline{S}_{t-1} = \underline{s}_{t-1}^k, S_t = s_t^k, \underline{H}_{t-1}) \mid \underline{S}_{t-1} = \underline{s}_{t-1}^k, S_t = s_t^j]. \quad \text{q.e.d.}$$

A.6.3 Part c)

Denote the vector ($\underline{s}_{\tau-v}^j, s_{\tau-v+1}^k, ..., s_\tau^k$) as ($\underline{s}_{\tau-v}^j, \underline{s}_{\tau-v+1:\tau}^k$). We need to show that $F(Y^{\underline{s}_{\tau-v}^j, \underline{s}_{\tau-v+1:\tau}^k} \mid \underline{S}_{\tau-v} = \underline{s}_{\tau-v}^j)$ is identified. Note that if $F(Y^{\underline{s}_{\tau-v}^j, \underline{s}_{\tau-v+1:\tau}^k} \mid \underline{S}_{\tau-v} = \underline{s}_{\tau-v}^j)$ is identified, $F(Y^{\underline{s}_{\tau-v}^j, \underline{s}_{\tau-v+1:\tau}^k} \mid \underline{S}_{\tau-v} = \underline{s}_{\tau-w}^j)$, $\tau \geq w \geq v$, is identified as well, because it is a courser conditioning set (based on observables).

$$F(Y^{\underline{s}_{\tau-v}^j, \underline{s}_{\tau-v+1:\tau}^k} \mid \underline{S}_{\tau-v} = \underline{s}_{\tau-v}^j) = \underset{S_{\tau-v+1}, \underline{H}_{\tau-v}}{E}[F(Y^{\underline{s}_{\tau-v}^j, \underline{s}_{\tau-v+1:\tau}^k} \mid \underline{S}_{\tau-v} = \underline{s}_{\tau-v}^j, S_{\tau-v+1}, \underline{H}_{\tau-v}) \mid \underline{S}_{\tau-v} = \underline{s}_{\tau-v}^j].$$

Because of ASSUMPTION 5 $F(Y^{\underline{s}_{\tau-v}^j, \underline{s}_{\tau-v+1:\tau}^k} \mid \underline{S}_{\tau-v} = \underline{s}_{\tau-v}^j, S_{\tau-v+1}, \underline{H}_{\tau-v})$ does not depend on $S_{\tau-v+1}$. Hence, we get:

$$F(Y^{\underline{s}_{\tau-v}^j, \underline{s}_{\tau-v+1:\tau}^k} \mid \underline{S}_{\tau-v} = \underline{s}_{\tau-v}^j) = \underset{S_{\tau-v+1}, \underline{H}_{\tau-v}}{E}[F(Y^{\underline{s}_{\tau-v}^j, \underline{s}_{\tau-v+1:\tau}^k} \mid \underline{S}_{\tau-v} = \underline{s}_{\tau-v}^j, S_{\tau-v+1} = s_{\tau-v+1}^k, \underline{H}_{\tau-v}) \mid \underline{S}_{\tau-v} = \underline{s}_{\tau-v}^j] =$$

= $\underset{\underline{H}_{\tau-v}}{E}[F(Y^{\underline{s}^j_{\tau-v},\underline{s}^k_{\tau-v+1:\tau}} \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, S_{\tau-v+1}=s^k_{\tau-v+1}, \underline{H}_{\tau-v}) \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}]$.

$F(Y^{\underline{s}^j_{\tau-v},\underline{s}^k_{\tau-v+1:\tau}} \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, S_{\tau-v+1}=s^k_{\tau-v+1}, \underline{H}_{\tau-v})$ =

= $\underset{S_{\tau-v+2},H_{\tau-v+1}}{E}[F(Y^{\underline{s}^j_{\tau-v},\underline{s}^k_{\tau-v+1:\tau}} \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, S_{\tau-v+1}=s^k_{\tau-v+1}, S_{\tau-v+2}, \underline{H}_{\tau-v+1}) \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, S_{\tau-v+1}=s^k_{\tau-v+1}, \underline{H}_{\tau-v}]$

= $\underset{H_{\tau-v+1}}{E}[F(Y^{\underline{s}^j_{\tau-v},\underline{s}^k_{\tau-v+1:\tau}} \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, \underline{S}_{\tau-v+1:\tau-v+2}=\underline{s}^k_{\tau-v+1:\tau-v+2}, \underline{H}_{\tau-v+1}) \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, S_{\tau-v+1}=s^k_{\tau-v+1}, \underline{H}_{\tau-v}]$

Finally, for the last period relevant, we obtain a similar expression:

$F(Y^{\underline{s}^j_{\tau-v},\underline{s}^k_{\tau-v+1:\tau}} \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, \underline{S}_{\tau-v+1:\tau-1}=\underline{s}^k_{\tau-v+1:\tau-1}, \underline{H}_{\tau-2})$ =

= $\underset{H_{\tau-1}}{E}[F(Y^{\underline{s}^j_{\tau-v},\underline{s}^k_{\tau-v+1:\tau}} \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, \underline{S}_{\tau-v+1:\tau}=\underline{s}^k_{\tau-v+1:\tau}, \underline{H}_{\tau-1}) \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, \underline{S}_{\tau-v+1:\tau-1}=\underline{s}^k_{\tau-v+1:\tau-1}, \underline{H}_{\tau-2}]$

= $\underset{H_{\tau-1}}{E}[F(Y \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, \underline{S}_{\tau-v+1:\tau}=\underline{s}^k_{\tau-v+1:\tau}, \underline{H}_{\tau-1}) \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, \underline{S}_{\tau-v+1:\tau-1}=\underline{s}^k_{\tau-v+1:\tau-1}, \underline{H}_{\tau-2}]$.

Recursive substitution of these expectations shows that $F(Y^{\underline{s}^j_{\tau-v},\underline{s}^k_{\tau-v+1:\tau}} \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v})$ is identified.

$F(Y^{\underline{s}^j_{\tau-v},\underline{s}^k_{\tau-v+1:\tau}} \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}) =$

$$= \underset{\underline{H}_{\tau-v}}{E}\{\underset{H_{\tau-v+1}}{E}\{\ldots \underset{H_{\tau-1}}{E}\left\{F(Y^{\underline{s}^j_{\tau-v},\underline{s}^k_{\tau-v+1:\tau}} \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, \underline{S}_{\tau-v+1:\tau}=\underline{s}^k_{\tau-v+1:\tau}, \underline{H}_{\tau-1}) \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, \underline{S}_{\tau-v+1:\tau-1}=\underline{s}^k_{\tau-v+1:\tau-1}, \underline{H}_{\tau-2}\right\}$$
$$\ldots \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, S_{\tau-v+1}=s^k_{\tau-v+1}, \underline{H}_{\tau-v}\} \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}\}$$

$$= \underset{\underline{H}_{\tau-v}}{E}\{\underset{H_{\tau-v+1}}{E}\{\ldots \underset{H_{\tau-1}}{E}\left\{F(Y \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, \underline{S}_{\tau-v+1:\tau}=\underline{s}^k_{\tau-v+1:\tau}, \underline{H}_{\tau-1}) \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, \underline{S}_{\tau-v+1:\tau-1}=\underline{s}^k_{\tau-v+1:\tau-1}, \underline{H}_{\tau-2}\right\}$$
$$\ldots \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}, S_{\tau-v+1}=s^k_{\tau-v+1}, \underline{H}_{\tau-v}\} \mid \underline{S}_{\tau-v}=\underline{s}^j_{\tau-v}\}$$
. q.e.d.

A.6.4 A note on Robin's g-formula

The well known *G-computation algorithm* which appeared first in Robins (1986), is in fact related to parts of the proofs for part a) of Theorem 3. The G-computation algorithm is given by the following equation:

$$\Pr(Y^{\underline{s}^k_T}_{t+1}=1) = \sum_{all\ \underline{y}_t,\underline{x}_t} \{\Pr(Y_{t+1}=1 \mid \underline{Y}_t=\underline{y}_t, \underline{X}_t=\underline{x}_t, \underline{S}_t=\underline{s}^k_t)$$

$$\prod_{\tau=1}^{t}[\Pr(X_\tau=x_\tau \mid \underline{Y}_\tau=\underline{y}_\tau, \underline{X}_{\tau-1}=\underline{x}_{\tau-1}, \underline{S}_{\tau-1}=\underline{s}^k_{\tau-1})$$

$$\Pr(Y_\tau = y_\tau \mid \underline{Y}_{\tau-1} = \underline{y}_{\tau-1}, \underline{X}_{\tau-1} = \underline{x}_{\tau-1}, \underline{S}_{\tau-1} = \underline{s}^k_{\tau-1})]\}.$$

If treatment received at each time is randomly allocated conditional on past treatment and covariate history (like in ASSUMPTION 5)[24], this equation holds and the algorithm describes the distribution of the potential outcome at time $t+1$ as a function of the distribution of the observables. Thus it identifies treatment effects of the form $\theta_t^{\underline{s}^k_\tau, \underline{s}^l_\tau}$ after suitable redefining time indices. It has however no direct implication for the other treatment effects considered here.

A.7 Proof of Theorem 4

A.7.1 Part a)

Part a) of THEOREM 4 concerns the case in which the sequences used in the conditioning set has no elements in common with the two sequences to be evaluated. The effects are defined as:

$$\theta_t^{\underline{s}^k_\tau, \underline{s}^l_\tau}(\underline{s}^j_\tau) = E(Y_t^{\underline{s}^k_\tau} \mid \underline{s}^j_\tau) - E(Y_t^{\underline{s}^l_\tau} \mid \underline{s}^j_\tau).$$

In case of part a) the upper and lower bounds of the expectations are given directly by ASSUMPTION 1-II, because no sample counterparts can be observed. Thus we derive the following upper and lower bounds for those effects:

$$\theta_t^{\underline{s}^k_\tau, \underline{s}^l_\tau}(\underline{s}^j_\tau) \in [{}_LY - {}^UY, {}^UY - {}_LY].$$

The bounds have a width of $2({}^UY - {}_LY)$.

A.7.2 Part b)

Part b) concerns the opposite case compared to part a): one of the sequences used to define the effects is exactly equivalent to the sequences used in the conditioning set. It is sufficient to consider $\theta_t^{\underline{s}^k_\tau, \underline{s}^l_\tau}(\underline{s}^k_\tau)$ only, because of $\theta_t^{\underline{s}^k_\tau, \underline{s}^l_\tau}(\underline{s}^k_\tau) = -\theta_t^{\underline{s}^l_\tau, \underline{s}^k_\tau}(\underline{s}^k_\tau)$:

$$\theta_t^{\underline{s}^k_\tau, \underline{s}^l_\tau}(\underline{s}^k_\tau) = \underbrace{E(Y_t^{\underline{s}^k_\tau} \mid \underline{s}^k_\tau)}_{B_t^{\underline{s}^k_\tau}} - E(Y_t^{\underline{s}^l_\tau} \mid \underline{s}^k_\tau) = B_t^{\underline{s}^k_\tau} - E(Y_t^{\underline{s}^l_\tau} \mid \underline{s}^k_\tau) \Rightarrow \theta_t^{\underline{s}^k_\tau, \underline{s}^l_\tau}(\underline{s}^k_\tau) \in [B_t^{\underline{s}^k_\tau} - {}^UY, B_t^{\underline{s}^k_\tau} - {}_LY].$$

[24] See for example Robins, Greenland, and Hu (1999, assumption 1).

A.7.3 Part c)

Now we consider intermediate cases for which the first part of one of the sequences under investigation are the same as the conditioning set. The remainder part of the conditioning sequence is left unspecified:

$$\theta_t^{(\underline{s}_{\tau-1}^k,s),\underline{s}_\tau^l}(\underline{s}_{\tau-1}^k) = E(Y_t^{(\underline{s}_{\tau-1}^k,s)} \mid \underline{s}_{\tau-1}^k) - E(Y_t^{\underline{s}_\tau^l} \mid \underline{s}_{\tau-1}^k).$$

Note that the sample is not informative about $E(Y_t^{\underline{s}_\tau^l} \mid \underline{s}_{\tau-1}^k)$. However, the sample is informative about $E(Y_t^{(\underline{s}_{\tau-1}^k,s)} \mid \underline{s}_{\tau-1}^k)$:

$$E(Y_t^{(\underline{s}_{\tau-1}^k,s)} \mid \underline{s}_{\tau-1}^k) = \sum_{m=1}^{M} E(Y_t^{(\underline{s}_{\tau-1}^k,s)} \mid \underline{s}_{\tau-1}^k, m) P[S_\tau = m \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k]$$

$$= \sum_{m=1}^{M} \underline{1}(m \neq s) E(Y_t^{(\underline{s}_{\tau-1}^k,s)} \mid \underline{s}_{\tau-1}^k, m) P[S_\tau = m \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k] + E(Y_t^{(\underline{s}_{\tau-1}^k,s)} \mid \underline{s}_{\tau-1}^k, s) P[S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k].$$

➔ $$\max E(Y_t^{(\underline{s}_{\tau-1}^k,s)} \mid \underline{s}_{\tau-1}^k) = \sum_{m=1}^{M} \underline{1}(m \neq s)\, {}^{U}Y P[S_\tau = m \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k] + B_t^{(\underline{s}_{\tau-1}^k,s)} P[S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k]$$

$$= {}^{U}Y(1 - P[S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k]) + B_t^{(\underline{s}_{\tau-1}^k,s)} P[S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k]$$

➔ $$\min E(Y_t^{(\underline{s}_{\tau-1}^k,s)} \mid \underline{s}_{\tau-1}^k) = \sum_{m=1}^{M} \underline{1}(m \neq s)\, {}_{L}Y P[S_\tau = m \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k] + B_t^{(\underline{s}_{\tau-1}^k,s)} P[S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k]$$

$$= {}_{L}Y(1 - P[S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k]) + B_t^{(\underline{s}_{\tau-1}^k,s)} P[S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k].$$

Combining these results and using the shortcut notation $P^{s|\underline{s}_{\tau-1}^k} := P(S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k)$, we obtain the desired bounds for the effects:

$$\theta_t^{(\underline{s}_{\tau-1}^k,s),\underline{s}_\tau^l}(\underline{s}_{\tau-1}^k) \in [\,{}_{L}Y(1 - P^{s|\underline{s}_{\tau-1}^k}) + B_t^{(\underline{s}_{\tau-1}^k,s)} P^{s|\underline{s}_{\tau-1}^k} - {}^{U}Y,\; {}^{U}Y(1 - P^{s|\underline{s}_{\tau-1}^k}) + B_t^{(\underline{s}_{\tau-1}^k,s)} P^{s|\underline{s}_{\tau-1}^k} - {}_{L}Y].$$

Hence, the width is given by $2({}^{U}Y - {}_{L}Y) - ({}^{U}Y - {}_{L}Y)P^{s|\underline{s}_{\tau-1}^k}$. Its informational content depends on the probabilities $P^{s|\underline{s}_{\tau-1}^k}$. The larger $P^{s|\underline{s}_{\tau-1}^k}$, the sharper the bounds. It is comprised between $({}^{U}Y - {}_{L}Y)$ and $2({}^{U}Y - {}_{L}Y)$.

The same arguments can be used to derive the bounds for coarser conditioning sets. q.e.d.

A.8 Proof of Theorem 5-3/4

The treatment effects have the following representation:

$$\theta^{(\underline{s}_{\tau-1}^k,s),(\underline{s}_{\tau-1}^l,s)}(\underline{s}_{\tau-1}^k,s)=E(Y^{(\underline{s}_{\tau-1}^k,s)}\mid \underline{s}_{\tau-1}^k,S_\tau=s)-E(Y^{(\underline{s}_{\tau-1}^l,s)}\mid \underline{s}_{\tau-1}^k,S_\tau=s).$$

The information we have is $E(Y^{(\underline{s}_{\tau-1}^k,s)}\mid \underline{s}_{\tau-1}^k,S_\tau=s)=E(Y\mid \underline{s}_{\tau-1}^k,S_\tau=s)$ as well as that ASSUMPTIONS 6-3 and 6-4 identify $O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k):=E(Y^{\underline{s}_{\tau-1}^l}\mid \underline{s}_{\tau-1}^k)$. The following relation holds between the expectations of the potential outcomes:

$$E(Y^{\underline{s}_{\tau-1}^l}\mid \underline{s}_{\tau-1}^k)=\sum_{m=1}^{M}E(Y^{\underline{s}_{\tau-1}^l,m}\mid \underline{s}_{\tau-1}^k,m)P(S_\tau=m\mid \underline{S}_{\tau-1}=\underline{s}_{\tau-1}^k). \tag{15}$$

This gives the following bounds on the potential outcomes using ASSUMPTIONS 6-3 or 6-4. To simplify the notation, denote $P^{s_\tau\mid\underline{s}_{\tau-1}^k}:=P(S_\tau=s\mid \underline{S}_{\tau-1}=\underline{s}_{\tau-1}^k)$. Reformulating equation (15) and taking account of the fact that each expectation has to stay within its bounds, we obtain the following upper and lower bounds for the potential outcomes:

$$U^{Y^{(\underline{s}_{\tau-1}^l,s)}}(\underline{s}_{\tau-1}^k,s)=\min\left\{{}^{U}Y;\frac{O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)-\sum_{m=1}^{M}{}_{L}Y\underline{1}(m\neq s)P(S_\tau=m\mid \underline{S}_{\tau-1}=\underline{s}_{\tau-1}^k)}{P^{s_\tau\mid\underline{s}_{\tau-1}^k}}\right\}=$$

$$=\min\left\{{}^{U}Y;\frac{O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)-{}_{L}Y(1-P^{s_\tau\mid\underline{s}_{\tau-1}^k})}{P^{s_\tau\mid\underline{s}_{\tau-1}^k}}\right\}.$$

$$L^{Y^{(\underline{s}_{\tau-1}^l,s)}}(\underline{s}_{\tau-1}^k,s)=\max\left\{{}_{L}Y;\frac{O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)-{}^{U}Y(1-P^{s_\tau\mid\underline{s}_{\tau-1}^k})}{P^{s_\tau\mid\underline{s}_{\tau-1}^k}}\right\}.$$

Thus, the bounds on the effects can be expressed as follows:

$$\theta^{(\underline{s}_{\tau-1}^k,s),(\underline{s}_{\tau-1}^l,s)}(\underline{s}_{\tau-1}^k,s)\in[B^{(\underline{s}_{\tau-1}^k,s)}-U^{Y^{(\underline{s}_{\tau-1}^l,s)}}(\underline{s}_{\tau-1}^k,s),B^{(\underline{s}_{\tau-1}^k,s)}-L^{Y^{(\underline{s}_{\tau-1}^l,s)}}(\underline{s}_{\tau-1}^k,s)]$$

Taking account of the different cases explicitly, we get the following bounds:

U1) $P^{s_\tau\mid\underline{s}_{\tau-1}^k}>\dfrac{{}^{U}Y-O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)}{{}^{U}Y-{}_{L}Y}$: $B^{(\underline{s}_{\tau-1}^k,s)}-\dfrac{O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)-(1-P^{s_\tau\mid\underline{s}_{\tau-1}^k})^{U}Y}{P^{s_\tau\mid\underline{s}_{\tau-1}^k}}$

U2) $P^{s_\tau|\underline{s}_{\tau-1}^k} < \frac{{}^UY - O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)}{{}^UY - {}_LY}$: $B^{(\underline{s}_{\tau-1}^k, s)} - {}_LY$

L1) $P^{s_\tau|\underline{s}_{\tau-1}^k} > \frac{O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k) - {}_LY}{{}^UY - {}_LY}$: $B^{(\underline{s}_{\tau-1}^k, s)} - \frac{O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k) - (1 - P^{s_\tau|\underline{s}_{\tau-1}^k})\,{}_LY}{P^{s_\tau|\underline{s}_{\tau-1}^k}}$

L2) $P^{s_\tau|\underline{s}_{\tau-1}^k} < \frac{O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k) - {}_LY}{{}^UY - {}_LY}$: $B^{(\underline{s}_{\tau-1}^k, s)} - {}^UY$

Note that for the derivation of the bounds for the cases $P^{s_\tau|\underline{s}_{\tau-1}^k} = \frac{{}^UY - O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)}{{}^UY - {}_LY}$ and

$P^{s_\tau|\underline{s}_{\tau-1}^k} = \frac{O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k) - {}_LY}{{}^UY - {}_LY}$ $O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)$ is substituted using the respective equalities. q.e.d.

Appendix B: Additional bounds under ASSUMPTION 6-3 or 6-4

B.1 More general conditioning sets

Instead of bounds for $\theta^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}(\underline{s}_{\tau-1}^{k},s)$, we allow an unrestricted conditioning set in all but the last period of the treatment. For $\theta^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}(\underline{s}_{\tau-1}^{j},s)$, we obtain 16 different cases for the bounds' widths.

The treatment effects have the following representation:

$$\theta^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}(\underline{s}_{\tau-1}^{j},s)=E(Y^{(\underline{s}_{\tau-1}^{k},s)}\mid \underline{s}_{\tau-1}^{j},S_\tau=s)-E(Y^{(\underline{s}_{\tau-1}^{l},s)}\mid \underline{s}_{\tau-1}^{j},S_\tau=s).$$

ASSUMPTIONS 6-3 and 6-4 identify $O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{j}):=E(Y^{\underline{s}_{\tau-1}^{l}}\mid \underline{s}_{\tau-1}^{j})$ and $O^{\underline{s}_{\tau-1}^{k}}(\underline{s}_{\tau-1}^{j}):=E(Y^{\underline{s}_{\tau-1}^{k}}\mid \underline{s}_{\tau-1}^{j})$. Using LEMMA 2-II, the following relation holds between the expectations of the potential outcomes:

$$E(Y^{\underline{s}_{\tau-1}^{k}}\mid \underline{s}_{\tau-1}^{j})=\sum_{m=1}^{M}E(Y^{\underline{s}_{\tau-1}^{k},m}\mid \underline{s}_{\tau-1}^{j},m)P(S_\tau=m\mid \underline{S}_{\tau-1}=\underline{s}_{\tau-1}^{j}). \tag{16}$$

This gives the following bounds on the potential outcomes using ASSUMPTIONS 6-3 or 6-4. To simplify the notation, denote $P^{s_\tau|\underline{s}_{\tau-1}^{j}}:=P(S_\tau=s\mid \underline{S}_{\tau-1}=\underline{s}_{\tau-1}^{j})$. Reformulating equation (16) and taking into account that the expectations have to stay within their bounds, we obtain the following upper and lower bounds for the potential outcomes:

$$U^{Y^{(\underline{s}_{\tau-1}^{k},s)}}(\underline{s}_{\tau-1}^{j},s)=\min\left\{{}^{U}Y;\frac{O^{\underline{s}_{\tau-1}^{k}}(\underline{s}_{\tau-1}^{j})-\sum_{m=1}^{M}{}_{L}Y\underline{1}(m\neq s)P(S_\tau=m\mid \underline{S}_{\tau-1}=\underline{s}_{\tau-1}^{j})}{P^{s_\tau|\underline{s}_{\tau-1}^{j}}}\right\}$$

$$=\min\left\{{}^{U}Y;\frac{O^{\underline{s}_{\tau-1}^{k}}(\underline{s}_{\tau-1}^{j})-{}_{L}Y(1-P^{s_\tau|\underline{s}_{\tau-1}^{j}})}{P^{s_\tau|\underline{s}_{\tau-1}^{j}}}\right\}.$$

$$L^{Y^{(\underline{s}_{\tau-1}^{k},s)}}(\underline{s}_{\tau-1}^{j},s)=\max\left\{{}_{L}Y;\frac{O^{\underline{s}_{\tau-1}^{k}}(\underline{s}_{\tau-1}^{j})-{}^{U}Y(1-P^{s_\tau|\underline{s}_{\tau-1}^{j}})}{P^{s_\tau|\underline{s}_{\tau-1}^{j}}}\right\}.$$

Thus, the bounds on the effects can be expressed as follows:

$$\theta^{(\underline{s}^k_{\tau-1},s),(\underline{s}^l_{\tau-1},s)}(\underline{s}^j_{\tau-1},s)\in[L^{Y^{(\underline{s}^k_{\tau-1},s)}}(\underline{s}^j_{\tau-1},s)-U^{Y^{(\underline{s}^l_{\tau-1},s)}}(\underline{s}^j_{\tau-1},s),U^{Y^{(\underline{s}^k_{\tau-1},s)}}(\underline{s}^j_{\tau-1},s)-L^{Y^{(\underline{s}^l_{\tau-1},s)}}(\underline{s}^j_{\tau-1},s)]$$

Taking account of the different cases explicitly, we get:

L1) $$P^{s_\tau|\underline{s}^k_{\tau-1}}>\frac{{}^UY-O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})}{{}^UY-{}_LY}:\quad L^{Y^{(\underline{s}^k_{\tau-1},s)}}(\underline{s}^j_{\tau-1},s)=\frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-(1-P^{s_\tau|\underline{s}^j_{\tau-1}})\,{}^UY}{P^{s_\tau|\underline{s}^j_{\tau-1}}}$$

L2) $$P^{s_\tau|\underline{s}^j_{\tau-1}}<\frac{{}^UY-O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})}{{}^UY-{}_LY}:\quad L^{Y^{(\underline{s}^k_{\tau-1},s)}}(\underline{s}^j_{\tau-1},s)={}_LY$$

U1) $$P^{s_\tau|\underline{s}^j_{\tau-1}}>\frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-{}_LY}{{}^UY-{}_LY}:\quad U^{Y^{(\underline{s}^k_{\tau-1},s)}}(\underline{s}^j_{\tau-1},s)=\frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-(1-P^{s_\tau|\underline{s}^j_{\tau-1}})\,{}_LY}{P^{s_\tau|\underline{s}^j_{\tau-1}}}$$

U2) $$P^{s_\tau|\underline{s}^j_{\tau-1}}<\frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-{}_LY}{{}^UY-{}_LY}:\quad U^{Y^{(\underline{s}^k_{\tau-1},s)}}(\underline{s}^j_{\tau-1},s)={}^UY$$

In the following, L1k refers to the case L1 for $L^{Y^{(\underline{s}^k_{\tau-1},s)}}(\underline{s}^j_{\tau-1},s)$, L1l to the case L1 for $L^{Y^{\underline{s}^l_{\tau-1}}}(\underline{s}^j_{\tau-1},s)$, and so on.

Table B.1a: Upper and lower bounds for $\theta^{(\underline{s}^k_{\tau-1},s),(\underline{s}^l_{\tau-1},s)}(\underline{s}^j_{\tau-1},s)$

Cases	L2k and U1l	L2k and U2l	L1k and U1l	L1k and U2l
U1k and L2l	$[\frac{{}_LY-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}, \frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-{}_LY}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}]$	$[-({}^UY-{}_LY), \frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-{}_LY}{P^{s_\tau\|\underline{s}^k_{\tau-1}}}]$	$[\frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}-\frac{(1-P^{s_\tau\|\underline{s}^j_{\tau-1}})({}^UY-{}_LY)}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}, \frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-{}_LY}{P^{s_\tau\|\underline{s}^k_{\tau-1}}}]$	$[-\frac{{}^UY-O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}, \frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-{}_LY}{P^{s_\tau\|\underline{s}^k_{\tau-1}}}]$
U1k and L1l	$[\frac{{}_LY-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}, \frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}+\frac{(1-P^{s_\tau\|\underline{s}^j_{\tau-1}})({}^UY-{}_LY)}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}]$	$[-({}^UY-{}_LY), \frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}+\frac{(1-P^{s_\tau\|\underline{s}^j_{\tau-1}})({}^UY-{}_LY)}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}]$	$[\frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}-\frac{(1-P^{s_\tau\|\underline{s}^j_{\tau-1}})({}^UY-{}_LY)}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}, \frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}+\frac{(1-P^{s_\tau\|\underline{s}^j_{\tau-1}})({}^UY-{}_LY)}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}]$	$[-\frac{{}^UY-O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}, \frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}+\frac{(1-P^{s_\tau\|\underline{s}^j_{\tau-1}})({}^UY-{}_LY)}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}]$
U2k and L2l	$[\frac{{}_LY-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}, {}^UY-{}_LY]$	$[-({}^UY-{}_LY), {}^UY-{}_LY]$	$[\frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}-\frac{(1-P^{s_\tau\|\underline{s}^j_{\tau-1}})({}^UY-{}_LY)}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}, {}^UY-{}_LY]$	$[-\frac{{}^UY-O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}, {}^UY-{}_LY]$
U2k and L1l	$[\frac{{}_LY-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}, \frac{{}^UY-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}]$	$[-({}^UY-{}_LY), \frac{{}^UY-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}]$	$[\frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}-\frac{(1-P^{s_\tau\|\underline{s}^j_{\tau-1}})({}^UY-{}_LY)}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}, \frac{{}^UY-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}]$	$[-\frac{{}^UY-O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}, \frac{{}^UY-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}]$

Table B.1b: Widths for $\theta^{(\underline{s}^k_{\tau-1},s),(\underline{s}^l_{\tau-1},s)}(\underline{s}^j_{\tau-1},s)$

Cases	L2k and U1l	L2k and U2l	L1k and U1l	L1k and U2l
U1k and L2l	$\frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})+O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})-2{}_LY}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}$	$({}^UY-{}_LY)+\frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-{}_LY}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}$	$\frac{O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})-{}_LY+(1-P^{s_\tau\|\underline{s}^j_{\tau-1}})({}^UY-{}_LY)}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}$	$\frac{{}^UY-{}_LY}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}$
U1k and L1l	$\frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-{}_LY+(1-P^{s_\tau\|\underline{s}^j_{\tau-1}})({}^UY-{}_LY)}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}$	$\frac{O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}+\frac{({}^UY-{}_LY)}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}$	$2\frac{(1-P^{s_\tau\|\underline{s}^j_{\tau-1}})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}({}^UY-{}_LY)$	$\frac{{}^UY-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})+(1-P^{s_\tau\|\underline{s}^j_{\tau-1}})({}^UY-{}_LY)}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}$
U2k and L2l	${}^UY-{}_LY-\frac{{}_LY-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}$	$2({}^UY-{}_LY)$	$\frac{({}^UY-{}_LY)-(O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1}))}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}$	$\frac{{}^UY-O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}+{}^UY-{}_LY$
U2k and L1l	$\frac{{}^UY-{}_LY}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}$	$({}^UY-{}_LY)+\frac{{}^UY-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}$	$[\frac{{}^UY-O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}+\frac{(1-P^{s_\tau\|\underline{s}^j_{\tau-1}})({}^UY-{}_LY)}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}$	$\frac{2\,{}^UY-O^{\underline{s}^k_{\tau-1}}(\underline{s}^j_{\tau-1})-O^{\underline{s}^l_{\tau-1}}(\underline{s}^j_{\tau-1})}{P^{s_\tau\|\underline{s}^j_{\tau-1}}}$

B.2 Coarser conditioning sets

Instead of bounds for $\theta^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}(\underline{s}_{\tau-1}^{k},s)$, we consider bounds for effect that are unconditional in the last period. For $\theta_t^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}(\underline{s}_{\tau-1}^{k})$ we obtain 4 different possible cases. Using LEMMA 1, we get:

$$\theta_t^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}(\underline{s}_{\tau-1}^{k}) = \sum_{m=1}^{M} \theta_t^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}(\underline{s}_{\tau-1}^{k},m)P(S_\tau = m \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^{k})$$

$$= \sum_{\substack{m=1 \\ m\neq s}}^{M} \theta_t^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}(\underline{s}_{\tau-1}^{k},m)P(S_\tau = m \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^{k}) + \theta_t^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}(\underline{s}_{\tau-1}^{k},s)P(S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^{k}).$$

Using the information obtained in section B.1 with $j = k$ and the noninformative bounds ${}^{U}Y - {}_{L}Y$, we can directly compute the bounds for $\theta_t^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}(\underline{s}_{\tau-1}^{k})$:

U1) $P^{s_\tau|\underline{s}_{\tau-1}^{k}} > \dfrac{{}^{U}Y - O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{k})}{{}^{U}Y - {}_{L}Y}$:

$$U^{\theta^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}}(\underline{s}_{\tau-1}^{k}) = \sum_{\substack{m=1 \\ m\neq s}}^{M} ({}^{U}Y - {}_{L}Y)P(S_\tau = m \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^{k}) + P(S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^{k})\left(B^{(\underline{s}_{\tau-1}^{k},s)} - \frac{O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{k}) - (1-P^{s_\tau|\underline{s}_{\tau-1}^{k}})\,{}^{U}Y}{P^{s_\tau|\underline{s}_{\tau-1}^{k}}} \right)$$

$$= ({}^{U}Y - {}_{L}Y)(1-P^{s_\tau|\underline{s}_{\tau-1}^{k}}) + P^{s_\tau|\underline{s}_{\tau-1}^{k}} B^{(\underline{s}_{\tau-1}^{k},s)} - O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{k}) + (1-P^{s_\tau|\underline{s}_{\tau-1}^{k}})\,{}^{U}Y$$

U2) $P^{s_\tau|\underline{s}_{\tau-1}^{k}} < \dfrac{{}^{U}Y - O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{k})}{{}^{U}Y - {}_{L}Y}$: $U^{\theta^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}}(\underline{s}_{\tau-1}^{k}) = ({}^{U}Y - {}_{L}Y)(1-P^{s_\tau|\underline{s}_{\tau-1}^{k}}) + P^{s_\tau|\underline{s}_{\tau-1}^{k}} B^{(\underline{s}_{\tau-1}^{k},s)} - P^{s_\tau|\underline{s}_{\tau-1}^{k}}\,{}_{L}Y$

L1) $P^{s_\tau|\underline{s}_{\tau-1}^{k}} > \dfrac{O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{k}) - {}_{L}Y}{{}^{U}Y - {}_{L}Y}$:

$$L^{\theta^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}}(\underline{s}_{\tau-1}^{k}) = -({}^{U}Y - {}_{L}Y)(1-P^{s_\tau|\underline{s}_{\tau-1}^{k}}) + P^{s_\tau|\underline{s}_{\tau-1}^{k}} B^{(\underline{s}_{\tau-1}^{k},s)} - O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{k}) + (1-P^{s_\tau|\underline{s}_{\tau-1}^{k}})\,{}_{L}Y$$

L2) $P^{s_\tau|\underline{s}_{\tau-1}^{k}} < \dfrac{O^{\underline{s}_{\tau-1}^{l}}(\underline{s}_{\tau-1}^{k}) - {}_{L}Y}{{}^{U}Y - {}_{L}Y}$: $L^{\theta^{(\underline{s}_{\tau-1}^{k},s),(\underline{s}_{\tau-1}^{l},s)}}(\underline{s}_{\tau-1}^{k}) = -({}^{U}Y - {}_{L}Y)(1-P^{s_\tau|\underline{s}_{\tau-1}^{k}}) + P^{s_\tau|\underline{s}_{\tau-1}^{k}} B^{(\underline{s}_{\tau-1}^{k},s)} - P^{s_\tau|\underline{s}_{\tau-1}^{k}}\,{}^{U}Y$

Table B.2a: Upper and lower bounds for $\theta_t^{(\underline{s}_{\tau-1}^k,s),(\underline{s}_{\tau-1}^l,s)}(\underline{s}_{\tau-1}^k)$

Cases	$P^{s_\tau\|\underline{s}_{\tau-1}^k} > \frac{O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)-{}_LY}{{}^UY-{}_LY}$	$P^{s_\tau\|\underline{s}_{\tau-1}^k} < \frac{O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)-{}_LY}{{}^UY-{}_LY}$
$P^{s_\tau\|\underline{s}_{\tau-1}^k} > \frac{{}^UY-O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)}{{}^UY-{}_LY}$	$[-({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})+P^{s_\tau\|\underline{s}_{\tau-1}^k}B^{(\underline{s}_{\tau-1}^k,s)}-O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)$ $+(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})_LY,({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})+$ $P^{s_\tau\|\underline{s}_{\tau-1}^k}B^{(\underline{s}_{\tau-1}^k,s)}-O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)+(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})^UY]$	$[-({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})+P^{s_\tau\|\underline{s}_{\tau-1}^k}B^{(\underline{s}_{\tau-1}^k,s)}-P^{s_\tau\|\underline{s}_{\tau-1}^k}\,{}^UY,$ $({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})+P^{s_\tau\|\underline{s}_{\tau-1}^k}B^{(\underline{s}_{\tau-1}^k,s)}-O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)$ $+(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})^UY]$
$P^{s_\tau\|\underline{s}_{\tau-1}^k} < \frac{{}^UY-O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)}{{}^UY-{}_LY}$	$[-({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})+P^{s_\tau\|\underline{s}_{\tau-1}^k}B^{(\underline{s}_{\tau-1}^k,s)}-O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)$ $+(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})_LY,({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})+P^{s_\tau\|\underline{s}_{\tau-1}^k}B^{(\underline{s}_{\tau-1}^k,s)}-P^{s_\tau\|\underline{s}_{\tau-1}^k}\,{}_LY]$	$[-({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})+P^{s_\tau\|\underline{s}_{\tau-1}^k}B^{(\underline{s}_{\tau-1}^k,s)}-P^{s_\tau\|\underline{s}_{\tau-1}^k}\,{}^UY,$ $({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})+P^{s_\tau\|\underline{s}_{\tau-1}^k}B^{(\underline{s}_{\tau-1}^k,s)}-P^{s_\tau\|\underline{s}_{\tau-1}^k}\,{}_LY]$

Table B.2b: Widths for $\theta_t^{(\underline{s}_{\tau-1}^k,s),(\underline{s}_{\tau-1}^l,s)}(\underline{s}_{\tau-1}^k)$

Cases	$P^{s_\tau\|\underline{s}_{\tau-1}^k} > \frac{O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)-{}_LY}{{}^UY-{}_LY}$	$P^{s_\tau\|\underline{s}_{\tau-1}^k} < \frac{O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)-{}_LY}{{}^UY-{}_LY}$
$P^{s_\tau\|\underline{s}_{\tau-1}^k} > \frac{{}^UY-O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)}{{}^UY-{}_LY}$	$3({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})$	$2({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})+{}^UY-O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)$
$P^{s_\tau\|\underline{s}_{\tau-1}^k} < \frac{{}^UY-O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)}{{}^UY-{}_LY}$	$2({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})+O^{\underline{s}_{\tau-1}^l}(\underline{s}_{\tau-1}^k)-{}_LY$	$({}^UY-{}_LY)[2(1-P^{s_\tau\|\underline{s}_{\tau-1}^k})+P^{s_\tau\|\underline{s}_{\tau-1}^k}]$

Appendix C: Additional bounds for an OPS version of ASSUMPTION 5

C.1 Preliminaries

In this appendix we examine bounds related to the effects presented in THEOREM 3 under the OPS assumption as defined in the following. As a by-product we obtain for a couple of cases bounds for effects that are not identified by ASSUMPTION 5, but would be identified by ASSUMPTION 4. These considerations are independent of OPS and thus have a value on their own.

ASSUMPTION 6.5 defines formally the OPS version of ASSUMPTION 5.

Assumption 6-5 (partial sequential endogenous conditional independence assumption OPS, PSE-CIA-OPS)

a) $Y^{\underline{s}_{\tau}},...,Y^{\underline{s}_{\tau}} \coprod S_t \mid \underline{H}_{t-1} = \underline{h}_{t-1}, \underline{S}_{t-1} = \underline{s}_{t-1}$;

b) $1 > P(S_t = s_t \mid \underline{H}_{t-1} = \underline{h}_{t-1}, \underline{S}_{t-1} = \underline{s}_{t-1}) > 0$; $\forall \underline{h}_{t-1} \in \underline{\mathbf{H}}_{t-1}$, $\forall \underline{s}_t \in \mathbb{S}_t$, $\forall t: 1 \leq t \leq \tau - 1$, $\tau \leq T$.

We show in Section 3 that ASSUMPTION 5 provides less information about the effects than ASSUMPTION 3 or 4. Therefore bounds derived under ASSUMPTION 6-5 cannot be sharper than the ones derived under ASSUMPTIONS 5-3 or 5-4. In particular, when the sequences do not coincide in the final period, we do not get more than the no-information bounds presented in the section 4.2 of the main body of the paper.

Establishing the following lemma simplifies some of the arguments used in this appendix.

Lemma 1-II (connection of treatment effects defined for different lengths of treatments)

$$Y_t^{\underline{s}_{\tau}^k}(\underline{s}_{\tilde{\tau}}^j) = \sum_{m=1}^{M} P(S_{\tilde{\tau}+1} = m \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j)\, Y_t^{\underline{s}_{\tau}^k}(\underline{s}_{\tilde{\tau}}^j, s_{\tilde{\tau}+1} = m) =$$

$$= \sum_{m_1=1}^{M} ... \sum_{m_{\delta}=1}^{M} P(S_{\tilde{\tau}+1} = m_1, ..., S_{\tilde{\tau}+\delta} = m_{\delta} \mid \underline{S}_{\tilde{\tau}} = \underline{s}_{\tilde{\tau}}^j) Y_t^{\underline{s}_{\tau}^k}(\underline{s}_{\tilde{\tau}}^j, s_{\tilde{\tau}+1} = m_1, ..., s_{\tilde{\tau}+\delta} = m_{\delta}) .$$

The proof of LEMMA 1-II closely resembles the proof of LEMMA 2-II, and thus need not be repeated.

C.1 Effects similar to those defined in part a) of THEOREM 3 ($\theta_t^{\underline{s}_{\tau-1}^k,\underline{s}_{\tau-1}^l}(\underline{s}_1^j)$)

C.1.1 Finer conditioning set

Instead of the identified effect $\theta_t^{\underline{s}_{\tau-1}^k,\underline{s}_{\tau-1}^l}(\underline{s}_1^j)$ identified by ASSUMPTION 6-5, we now consider bounds for $\theta_t^{\underline{s}_{\tau-1}^k,\underline{s}_{\tau-1}^l}(\underline{s}_1^j, s_2)$. The uncertainty considered is solely related to different lengths of conditioning sets.

We obtain 16 different possible cases for the bounds' widths. The simplification used earlier, namely to restrict our presentation to effects conditioning on the same letter as one of the potential treatment ($j=k$), does not lead here to a reduction of the cases. Nevertheless for consistency reasons we present the bounds for $j=k$.

The treatment effects have the following representation:

$$\theta_t^{\underline{s}_{\tau-1}^k,\underline{s}_{\tau-1}^l}(\underline{s}_1^k, s) = E(Y^{\underline{s}_{\tau-1}^k} \mid \underline{s}_1^k, S_2 = s) - E(Y^{\underline{s}_{\tau-1}^l} \mid \underline{s}_1^k, S_2 = s) .$$

ASSUMPTION 6-5 identifies $O^{\underline{s}_{\tau-1}^k}(\underline{s}_1^k) := E(Y^{\underline{s}_{\tau-1}^k} \mid \underline{s}_1^k)$ and $O^{\underline{s}_{\tau-1}^l}(\underline{s}_1^k) := E(Y^{\underline{s}_{\tau-1}^l} \mid \underline{s}_1^k)$. Using LEMMA 1-II, we have the following connection between the above identified expectations and the interested one:

$$E(Y^{\underline{s}_{\tau-1}^k} \mid \underline{s}_1^k) = \sum_{m=1}^{M} E(Y^{\underline{s}_{\tau-1}^k} \mid \underline{s}_1^k, m) P(S_2 = m \mid \underline{S}_1 = \underline{s}_1^k) . \qquad (17)$$

This gives the following bounds on the potential outcomes. To simplify the notation, denote $P^{s_2|\underline{s}_1^k} := P(S_2 = s \mid \underline{S}_1 = \underline{s}_1^k)$. Reformulating equation (17) and taking into account that the expectations stay within their bounds, we obtain the upper and lower bounds for the potential outcomes:

$$U^{Y^{\underline{s}_{\tau-1}^k}}(\underline{s}_1^k, s) = \min\left\{ {}^{U}Y; \frac{O^{\underline{s}_{\tau-1}^k}(\underline{s}_1^k) - \sum_{m=1}^{M} {}_{L}Y \underline{1}(m \neq s) P(S_2 = m \mid \underline{S}_1 = \underline{s}_1^k)}{P^{s_2|\underline{s}_1^k}} \right\}$$

$$= \min\left\{ {}^{U}Y; \frac{O^{\underline{s}_{\tau-1}^k}(\underline{s}_1^k) - {}_{L}Y(1 - P^{s_2|\underline{s}_1^k})}{P^{s_2|\underline{s}_1^k}} \right\} .$$

$$L^{Y^{s^k_{\tau-1}}}(\underline{s}_1^k, s) = \max\left\{ {}_LY; \frac{O^{s^k_{\tau-1}}(\underline{s}_1^k) - {}^UY(1-P^{s_2|\underline{s}_1^k})}{P^{s_2|\underline{s}_1^k}} \right\}.$$

Thus, the bounds on the effects can be expressed as follows:

$$\theta^{s^k_{\tau-1}, s^l_{\tau-1}}(\underline{s}_1^k, s) \in [L^{Y^{s^k_{\tau-1}}}(\underline{s}_1^k, s) - U^{Y^{s^l_{\tau-1}}}(\underline{s}_1^k, s), U^{Y^{s^k_{\tau-1}}}(\underline{s}_1^k, s) - L^{Y^{s^l_{\tau-1}}}(\underline{s}_1^k, s)]$$

Taking account of the different cases explicitly, we get:

L1) $$P^{s_2|\underline{s}_1^k} > \frac{{}^UY - O^{s^k_{\tau-1}}(\underline{s}_1^k)}{{}^UY - {}_LY}: \quad L^{Y^{s^k_{\tau-1}}}(\underline{s}_1^k, s) = \frac{O^{s^k_{\tau-1}}(\underline{s}_1^k) - (1-P^{s_2|\underline{s}_1^k})\,{}^UY}{P^{s_2|\underline{s}_1^k}}$$

L2) $$P^{s_2|\underline{s}_1^k} < \frac{{}^UY - O^{s^k_{\tau-1}}(\underline{s}_1^k)}{{}^UY - {}_LY}: \quad L^{Y^{s^k_{\tau-1}}}(\underline{s}_1^k, s) = {}_LY$$

U1) $$P^{s_2|\underline{s}_1^k} > \frac{O^{s^k_{\tau-1}}(\underline{s}_1^k) - {}_LY}{{}^UY - {}_LY}: \quad U^{Y^{s^k_{\tau-1}}}(\underline{s}_1^k, s) = \frac{O^{s^k_{\tau-1}}(\underline{s}_1^k) - (1-P^{s_2|\underline{s}_1^k})\,{}_LY}{P^{s_2|\underline{s}_1^k}}$$

U2) $$P^{s_2|\underline{s}_1^k} < \frac{O^{s^k_{\tau-1}}(\underline{s}_1^k) - {}_LY}{{}^UY - {}_LY}: \quad U^{Y^{s^k_{\tau-1}}}(\underline{s}_1^k, s) - {}^UY$$

In the following, L1k, L1l , ..., refer respectively to case L1 for $L^{Y^{s^k_{\tau-1}}}(\underline{s}_1^k, s)$, to case L1 for $L^{Y^{s^l_{\tau-1}}}(\underline{s}_1^k, s)$ and so on.

Table C.1a: Upper and lower bounds for $\theta^{s^k_{\tau-1},s^l_{\tau-1}}(\underline{s}_1^k, s)$

Cases	L2k and U1l	L2k and U2l	L1k and U1l	L1k and U2l
U1k and L2l	$[\frac{{}_LY-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}, \frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)-{}_LY}{P^{s_2\mid\underline{s}_1^k}}]$	$[-({}^UY-{}_LY), \frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)-{}_LY}{P^{s_2\mid\underline{s}_1^k}}]$	$[\frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}} - \frac{(1-P^{s_2\mid\underline{s}_1^k})({}^UY-{}_LY)}{P^{s_2\mid\underline{s}_1^k}}, \frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)-{}_LY}{P^{s_2\mid\underline{s}_1^k}}]$	$[-\frac{{}^UY-O^{s^k_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}, \frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)-{}_LY}{P^{s_2\mid\underline{s}_1^k}}]$
U1k and L1l	$[\frac{{}_LY-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}, \frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}} + \frac{(1-P^{s_2\mid\underline{s}_1^k})({}^UY-{}_LY)}{P^{s_2\mid\underline{s}_1^k}}]$	$[-({}^UY-{}_LY), \frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}} + \frac{(1-P^{s_2\mid\underline{s}_1^k})({}^UY-{}_LY)}{P^{s_2\mid\underline{s}_1^k}}]$	$[\frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}} - \frac{(1-P^{s_2\mid\underline{s}_1^k})({}^UY-{}_LY)}{P^{s_2\mid\underline{s}_1^k}}, \frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}} + \frac{(1-P^{s_2\mid\underline{s}_1^k})({}^UY-{}_LY)}{P^{s_2\mid\underline{s}_1^k}}]$	$[-\frac{{}^UY-O^{s^k_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}, \frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}} + \frac{(1-P^{s_2\mid\underline{s}_1^k})({}^UY-{}_LY)}{P^{s_2\mid\underline{s}_1^k}}]$
U2k and L2l	$[\frac{{}_LY-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}, {}^UY-{}_LY]$	$[-({}^UY-{}_LY), {}^UY-{}_LY]$	$[\frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}} - \frac{(1-P^{s_2\mid\underline{s}_1^k})({}^UY-{}_LY)}{P^{s_2\mid\underline{s}_1^k}}, {}^UY-{}_LY]$	$[-\frac{{}^UY-O^{s^k_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}, {}^UY-{}_LY]$
U2k and L1l	$[\frac{{}_LY-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}, \frac{{}^UY-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}]$	$[-({}^UY-{}_LY), \frac{{}^UY-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}]$	$[\frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}} - \frac{(1-P^{s_2\mid\underline{s}_1^k})({}^UY-{}_LY)}{P^{s_2\mid\underline{s}_1^k}}, \frac{{}^UY-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}]$	$[-\frac{{}^UY-O^{s^k_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}, \frac{{}^UY-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}]$

Table C.1b: Widths for $\theta^{s^k_{\tau-1},s^l_{\tau-1}}(\underline{s}_1^k, s)$

Cases	L2k and U1l	L2k and U2l	L1k and U1l	L1k and U2l
U1k And L2l	$\frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)+O^{s^l_{\tau-1}}(\underline{s}_1^k)-2{}_LY}{P^{s_2\mid\underline{s}_1^k}}$	$({}^UY-{}_LY)+\frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)-{}_LY}{P^{s_2\mid\underline{s}_1^k}}$	$\frac{O^{s^l_{\tau-1}}(\underline{s}_1^k)-{}_LY+(1-P^{s_2\mid\underline{s}_1^k})({}^UY-{}_LY)}{P^{s_2\mid\underline{s}_1^k}}$	$\frac{{}^UY-{}_LY}{P^{s_2\mid\underline{s}_1^k}}$
U1k And L1l	$\frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)-{}_LY+(1-P^{s_2\mid\underline{s}_1^k})({}^UY-{}_LY)}{P^{s_2\mid\underline{s}_1^k}}$	$\frac{O^{s^k_{\tau-1}}(\underline{s}_1^k)-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}+\frac{({}^UY-{}_LY)}{P^{s_2\mid\underline{s}_1^k}}$	$2\frac{(1-P^{s_2\mid\underline{s}_1^k})}{P^{s_2\mid\underline{s}_1^k}}({}^UY-{}_LY)$	$\frac{{}^UY-O^{s^l_{\tau-1}}(\underline{s}_1^k)+(1-P^{s_2\mid\underline{s}_1^k})({}^UY-{}_LY)}{P^{s_2\mid\underline{s}_1^k}}$
U2k And L2l	${}^UY-{}_LY-\frac{{}_LY-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}$	$2({}^UY-{}_LY)$	$\frac{({}^UY-{}_LY)-(O^{s^k_{\tau-1}}(\underline{s}_1^k)-O^{s^l_{\tau-1}}(\underline{s}_1^k))}{P^{s_2\mid\underline{s}_1^k}}$	$\frac{{}^UY-O^{s^k_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}+{}^UY-{}_LY$
U2k And L1l	$\frac{{}^UY-{}_LY}{P^{s_2\mid\underline{s}_1^k}}$	$({}^UY-{}_LY)+\frac{{}^UY-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}$	$[\frac{{}^UY-O^{s^k_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}+\frac{(1-P^{s_2\mid\underline{s}_1^k})({}^UY-{}_LY)}{P^{s_2\mid\underline{s}_1^k}}$	$\frac{2{}^UY-O^{s^k_{\tau-1}}(\underline{s}_1^k)-O^{s^l_{\tau-1}}(\underline{s}_1^k)}{P^{s_2\mid\underline{s}_1^k}}$

C.1.2 The effect of OPS

For $\theta_t^{(\underline{s}_{\tau-1}^k,s),(\underline{s}_{\tau-1}^l,s)}(\underline{s}_1^j,\bullet,s)$, we obtain 16 different possible cases for the bounds' widths. Again, the simplification used earlier, namely to restrict our presentation to effects conditioning on the same letter as one of the potential treatment ($j=k$), does not lead here to a reduction of the cases. But for reasons of consistency we present the bounds for $j=k$.

The treatment effects have the following representation:

$$\theta_t^{(\underline{s}_{\tau-1}^k,s),(\underline{s}_{\tau-1}^l,s)}(\underline{s}_1^k,\bullet,s)=E(Y^{(\underline{s}_{\tau-1}^k,s)}\mid \underline{s}_1^k,\bullet,S_\tau=s)-E(Y^{(\underline{s}_{\tau-1}^l,s)}\mid \underline{s}_1^k,\bullet,S_\tau=s).$$

ASSUMPTION 6-5 identifies $O^{\underline{s}_{\tau-1}^k}(\underline{s}_1^k):=E(Y^{\underline{s}_{\tau-1}^k}\mid \underline{s}_1^k)$ and $O^{\underline{s}_{\tau-1}^l}(\underline{s}_1^k):=E(Y^{\underline{s}_{\tau-1}^l}\mid \underline{s}_1^k)$. Using LEMMA 2-II, we have the following connection between the above identified expectations and the interested one:

$$E(Y^{\underline{s}_{\tau-1}^k}\mid \underline{s}_1^k)=\sum_{m=1}^{M}E(Y^{(\underline{s}_{\tau-1}^k,m)}\mid \underline{s}_1^k,\bullet,m)P(S_\tau=m\mid \underline{S}_1=\underline{s}_1^k). \qquad (18)$$

This gives the following bounds on the potential outcomes. To simplify the notation, denote $P^{s_\tau|\underline{s}_1^k}:=P(S_\tau=s\mid \underline{S}_1=\underline{s}_1^k)$. Reformulating equation (18) and taking into account that the expectations stay within their bounds, we obtain the upper and lower bounds for the potential outcomes:

$$U^{Y^{(\underline{s}_{\tau-1}^k,s)}}(\underline{s}_1^k,\bullet,s)=\min\left\{{}^{U}Y;\frac{O^{\underline{s}_{\tau-1}^k}(\underline{s}_1^k)-\sum_{m=1}^{M}{}_{L}Y\underline{1}(m\neq s)P(S_\tau=m\mid \underline{S}_1=\underline{s}_1^k)}{P^{s_\tau|\underline{s}_1^k}}\right\}$$

$$=\min\left\{{}^{U}Y;\frac{O^{\underline{s}_{\tau-1}^k}(\underline{s}_1^k)-{}_{L}Y(1-P^{s_\tau|\underline{s}_1^k})}{P^{s_\tau|\underline{s}_1^k}}\right\}.$$

$$L^{Y^{(\underline{s}_{\tau-1}^k,s)}}(\underline{s}_1^k,\bullet,s)=\max\left\{{}_{L}Y;\frac{O^{\underline{s}_{\tau-1}^k}(\underline{s}_1^k)-{}^{U}Y(1-P^{s_\tau|\underline{s}_1^k})}{P^{s_\tau|\underline{s}_1^k}}\right\}.$$

Thus, the bounds on the effects can be expressed as follows:

$$\theta^{(\underline{s}_{\tau-1}^k,s),(\underline{s}_{\tau-1}^l,s)}(\underline{s}_1^k,\bullet,s)\in[L^{Y^{(\underline{s}_{\tau-1}^k,s)}}(\underline{s}_1^k,\bullet,s)-U^{Y^{(\underline{s}_{\tau-1}^l,s)}}(\underline{s}_1^k,\bullet,s),U^{Y^{(\underline{s}_{\tau-1}^k,s)}}(\underline{s}_1^k,\bullet,s)-L^{Y^{(\underline{s}_{\tau-1}^l,s)}}(\underline{s}_1^k,\bullet,s)]$$

The different cases are exactly the same as in C.1.1 with $P(S_\tau=s\mid \underline{S}_1=\underline{s}_1^k)$ replacing $P(S_2=s\mid \underline{S}_1=\underline{s}_1^k)$.

Now consider the particular case $\tau = 2$ (and $j = k$): There are only 4 different widths. $L^{Y^{(\underline{s}_{\tau-1}^k,s)}}(\underline{s}_1^k,\bullet,s)$ and $U^{Y^{(\underline{s}_{\tau-1}^k,s)}}(\underline{s}_1^k,\bullet,s)$ are now replaced by the observable $B^{(\underline{s}_1^k,s)} = E(Y^{(\underline{s}_1^k,s)} \mid \underline{s}_1^k, s)$.

Table C.2a: Upper and lower bounds for $\theta_t^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}(\underline{s}_1^k,s)$

Cases	$P^{s_2\|\underline{s}_1^k} > \frac{O^{\underline{s}_1^l}(\underline{s}_1^k) - {}_LY}{{}^UY - {}_LY}$	$P^{s_2\|\underline{s}_1^k} = \frac{O^{\underline{s}_1^l}(\underline{s}_1^k) - {}_LY}{{}^UY - {}_LY}$	$P^{s_2\|\underline{s}_1^k} < \frac{O^{\underline{s}_1^l}(\underline{s}_1^k) - {}_LY}{{}^UY - {}_LY}$
$P^{s_2\|\underline{s}_1^k} > \frac{{}^UY - O^{\underline{s}_1^l}(\underline{s}_1^k)}{{}^UY - {}_LY}$	$[B^{(\underline{s}_1^k,s)} - \frac{O^{\underline{s}_1^l}(\underline{s}_1^k) - (1-P^{s_2\|\underline{s}_1^k})\,{}_LY}{P^{s_2\|\underline{s}_1^k}},$ $B^{(\underline{s}_1^k,s)} - \frac{O^{(\underline{s}_1^l)}(\underline{s}_1^k) - (1-P^{s_2\|\underline{s}_1^k})\,{}^UY}{P^{s_2\|\underline{s}_1^k}}]$	$[B^{(\underline{s}_1^k,s)} - {}^UY,$ $B^{(\underline{s}_1^k,s)} - {}^UY + \frac{(1-P^{s_2\|\underline{s}_1^k})({}^UY - {}_LY)}{P^{s_2\|\underline{s}_1^k}}]$	$[B^{(\underline{s}_1^k,s)} - {}^UY,$ $B^{(\underline{s}_1^k,s)} - \frac{O^{\underline{s}_1^l}(\underline{s}_1^k) - (1-P^{s_2\|\underline{s}_1^k})\,{}^UY}{P^{s_2\|\underline{s}_1^k}}]$
$P^{s_2\|\underline{s}_1^k} = \frac{{}^UY - O^{\underline{s}_1^l}(\underline{s}_1^k)}{{}^UY - {}_LY}$	$[B^{(\underline{s}_1^k,s)} - {}_LY - \frac{(1-P^{s_2\|\underline{s}_1^k})({}^UY - {}_LY)}{P^{s_2\|\underline{s}_1^k}},$ $B^{(\underline{s}_1^k,s)} - {}_LY]$	$[B^{(\underline{s}_1^k,s)} - {}^UY, B^{(\underline{s}_1^k,s)} - {}_LY]$ *)	$[B^{(\underline{s}_1^k,s)} - {}^UY, B^{(\underline{s}_1^k,s)} - {}_LY]$
$P^{s_2\|\underline{s}_1^k} < \frac{{}^UY - O^{\underline{s}_1^l}(\underline{s}_1^k)}{{}^UY - {}_LY}$	$[B^{(\underline{s}_1^k,s)} - \frac{O^{\underline{s}_1^l}(\underline{s}_1^k) - (1-P^{s_2\|\underline{s}_1^k})\,{}_LY}{P^{s_2\|\underline{s}_1^k}},$ $B^{(\underline{s}_1^k,s)} - {}_LY]$	$[B^{(\underline{s}_1^k,s)} - {}^UY, B^{(\underline{s}_1^k,s)} - {}_LY]$	$[B^{(\underline{s}_1^k,s)} - {}^UY, B^{(\underline{s}_1^k,s)} - {}_LY]$

Note: $P^{s_2|\underline{s}_1^k} = P(S_2 = s \mid \underline{S}_1 = \underline{s}_1^k)$, $O^{\underline{s}_1^l}(\underline{s}_1^k) = E(Y^{\underline{s}_1^l} \mid \underline{S}_1 = \underline{s}_1^k)$, $B^{(\underline{s}_1^k,s)} = E(Y \mid \underline{S}_1 = \underline{s}_1^k, S_2 = s)$.

*) This case is only possible if $P^{s_2|\underline{s}_1^k} = 0.5$ and $O^{\underline{s}_1^l}(\underline{s}_1^k) = 0.5({}^UY - {}_LY)$. Note also that cases defined by an inequality converge smoothly towards the cases defined by an equality.

Table C.2b: Widths for $\theta_t^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}(\underline{s}_1^k,s)$

Cases	$P^{s_2\|\underline{s}_1^k} > \frac{O^{\underline{s}_1^l}(\underline{s}_1^k) - {}_LY}{{}^UY - {}_LY}$	$P^{s_2\|\underline{s}_1^k} = \frac{O^{\underline{s}_1^l}(\underline{s}_1^k) - {}_LY}{{}^UY - {}_LY}$	$P^{s_2\|\underline{s}_1^k} < \frac{O^{\underline{s}_1^l}(\underline{s}_1^k) - {}_LY}{{}^UY - {}_LY}$
$P^{s_2\|\underline{s}_1^k} > \frac{{}^UY - O^{\underline{s}_1^l}(\underline{s}_1^k)}{{}^UY - {}_LY}$	$\frac{(1-P^{s_2\|\underline{s}_1^k})}{P^{s_2\|\underline{s}_1^k}}({}^UY - {}_LY)$	$\frac{(1-P^{s_2\|\underline{s}_1^k})}{P^{s_2\|\underline{s}_1^k}}({}^UY - {}_LY)$	$\frac{{}^UY - O^{\underline{s}_1^l}(\underline{s}_1^k)}{P^{s_2\|\underline{s}_1^k}}$
$P^{s_2\|\underline{s}_1^k} = \frac{{}^UY - O^{(\underline{s}_1^l)}(\underline{s}_1^k)}{{}^UY - {}_LY}$	$\frac{(1-P^{s_2\|\underline{s}_1^k})({}^UY - {}_LY)}{P^{s_2\|\underline{s}_1^k}}$	${}^UY - {}_LY$ *)	${}^UY - {}_LY$
$P^{s_2\|\underline{s}_1^k} < \frac{{}^UY - O^{\underline{s}_1^l}(\underline{s}_1^k)}{{}^UY - {}_LY}$	$\frac{O^{\underline{s}_1^l}(\underline{s}_1^k) - {}_LY}{P^{s_2\|\underline{s}_1^k}}$	${}^UY - {}_LY$	${}^UY - {}_LY$

Note: $P^{s_2|\underline{s}_1^k} = P(S_2 = s \mid \underline{S}_1 = \underline{s}_1^k)$, $O^{\underline{s}_1^l}(\underline{s}_1^k) = E(Y^{\underline{s}_1^l} \mid \underline{S}_1 = \underline{s}_1^k)$, $B^{(\underline{s}_1^k,s)} = E(Y \mid \underline{S}_1 = \underline{s}_1^k, S_2 = s)$.
*) See note on Table C.2a.

Consider now instead of $\theta_t^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}(\underline{s}_1^k,s)$ a coarser conditioning, as in $\theta_t^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}(\underline{s}_1^k)$. For all treatment effects of the form $\theta_t^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}(\underline{s}_1^k)$, we obtain 4 different possible cases for the bounds' widths. Using LEMMA 1 we obtain.

$$\theta_t^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}(\underline{s}_1^k)=\sum_{m=1}^{M}\theta_t^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}(\underline{s}_1^k,m)P(S_2=m\mid \underline{S}_1=\underline{s}_1^k)$$

$$=\sum_{\substack{m=1\\ m\neq s}}^{M}\theta_t^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}(\underline{s}_1^k,m)P(S_2=m\mid \underline{S}_1=\underline{s}_1^k)+\theta_t^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}(\underline{s}_1^k,s)P(S_2=s\mid \underline{S}_1=\underline{s}_1^k)\,.$$

Using the information in Table C.2a and the non informative bounds ${}^{U}Y-{}_{L}Y$, we can easily compute the bounds for $\theta_t^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}(\underline{s}_1^k)$:

U1) $P^{s_2|\underline{s}_1^k}>\dfrac{{}^{U}Y-O^{\underline{s}_1^l}(\underline{s}_1^k)}{{}^{U}Y-{}_{L}Y}$:

$$U^{\theta^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}}(\underline{s}_1^k)=\sum_{\substack{m=1\\ m\neq s}}^{M}({}^{U}Y-{}_{L}Y)P(S_2=m\mid \underline{S}_1=\underline{s}_1^k)+P(S_2=s\mid \underline{S}_1=\underline{s}_1^k)\left(B^{(\underline{s}_1^k,s)}-\frac{O^{\underline{s}_1^l}(\underline{s}_1^k)-(1-P^{s_2|\underline{s}_1^k})\,{}^{U}Y}{P^{s_2|\underline{s}_1^k}}\right)$$

$$=({}^{U}Y-{}_{L}Y)(1-P^{s_2|\underline{s}_1^k})+P^{s_2|\underline{s}_1^k}B^{(\underline{s}_1^k,s)}-O^{\underline{s}_1^l}(\underline{s}_1^k)+(1-P^{s_2|\underline{s}_1^k})\,{}^{U}Y$$

U2) $P^{s_2|\underline{s}_1^k}<\dfrac{{}^{U}Y-O^{\underline{s}_1^l}(\underline{s}_1^k)}{{}^{U}Y-{}_{L}Y}$: $\quad U^{\theta^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}}(\underline{s}_1^k)=({}^{U}Y-{}_{L}Y)(1-P^{s_2|\underline{s}_1^k})+P^{s_2|\underline{s}_1^k}B^{(\underline{s}_1^k,s)}-P^{s_2|\underline{s}_1^k}\,{}_{L}Y$

L1) $P^{s_2|\underline{s}_1^k}>\dfrac{O^{\underline{s}_1^l}(\underline{s}_1^k)-{}_{L}Y}{{}^{U}Y-{}_{L}Y}$:

$$L^{\theta^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}}(\underline{s}_1^k)=-({}^{U}Y-{}_{L}Y)(1-P^{s_2|\underline{s}_1^k})+P^{s_2|\underline{s}_1^k}B^{(\underline{s}_1^k,s)}-O^{\underline{s}_1^l}(\underline{s}_1^k)+(1-P^{s_2|\underline{s}_1^k})\,{}_{L}Y$$

L2) $P^{s_2|\underline{s}_1^k}<\dfrac{O^{\underline{s}_1^l}(\underline{s}_1^k)-{}_{L}Y}{{}^{U}Y-{}_{L}Y}$: $\quad L^{\theta^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}}(\underline{s}_1^k)=-({}^{U}Y-{}_{L}Y)(1-P^{s_2|\underline{s}_1^k})+P^{s_2|\underline{s}_1^k}B^{(\underline{s}_1^k,s)}-P^{s_2|\underline{s}_1^k}\,{}^{U}Y$

Table C.3a: Upper and lower bounds for $\theta_t^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}(\underline{s}_1^k)$ *:*

Cases	$P^{s_2\|\underline{s}_1^k} > \frac{O^{\underline{s}_1^l}(\underline{s}_1^k) - {}_LY}{{}^UY - {}_LY}$	$P^{s_2\|\underline{s}_1^k} < \frac{O^{\underline{s}_1^l}(\underline{s}_1^k) - {}_LY}{{}^UY - {}_LY}$
$P^{s_2\|\underline{s}_1^k} > \frac{{}^UY - O^{\underline{s}_1^l}(\underline{s}_1^k)}{{}^UY - {}_LY}$	$[-({}^UY - {}_LY)(1-P^{s_2\|\underline{s}_1^k}) + P^{s_2\|\underline{s}_1^k} B^{(\underline{s}_1^k,s)} - O^{\underline{s}_1^l}(\underline{s}_1^k) + (1-P^{s_2\|\underline{s}_1^k}){}_LY, ({}^UY - {}_LY)(1-P^{s_2\|\underline{s}_1^k}) + P^{s_2\|\underline{s}_1^k} B^{(\underline{s}_1^k,s)} - O^{\underline{s}_1^l}(\underline{s}_1^k) + (1-P^{s_2\|\underline{s}_1^k}){}^UY]$	$[-({}^UY - {}_LY)(1-P^{s_2\|\underline{s}_1^k}) + P^{s_2\|\underline{s}_1^k} B^{(\underline{s}_1^k,s)} - P^{s_2\|\underline{s}_1^k}\,{}^UY, ({}^UY - {}_LY)(1-P^{s_2\|\underline{s}_1^k}) + P^{s_2\|\underline{s}_1^k} B^{(\underline{s}_1^k,s)} - O^{\underline{s}_1^l}(\underline{s}_1^k) + (1-P^{s_2\|\underline{s}_1^k}){}^UY]$
$P^{s_2\|\underline{s}_1^k} < \frac{{}^UY - O^{\underline{s}_1^l}(\underline{s}_1^k)}{{}^UY - {}_LY}$	$[-({}^UY - {}_LY)(1-P^{s_2\|\underline{s}_1^k}) + P^{s_2\|\underline{s}_1^k} B^{(\underline{s}_1^k,s)} - O^{\underline{s}_1^l}(\underline{s}_1^k) + (1-P^{s_2\|\underline{s}_1^k}){}_LY, ({}^UY - {}_LY)(1-P^{s_2\|\underline{s}_1^k}) + P^{s_2\|\underline{s}_1^k} B^{(\underline{s}_1^k,s)} - P^{s_2\|\underline{s}_1^k}\,{}_LY]$	$[-({}^UY - {}_LY)(1-P^{s_2\|\underline{s}_1^k}) + P^{s_2\|\underline{s}_1^k} B^{(\underline{s}_1^k,s)} - P^{s_2\|\underline{s}_1^k}\,{}^UY, ({}^UY - {}_LY)(1-P^{s_2\|\underline{s}_1^k}) + P^{s_2\|\underline{s}_1^k} B^{(\underline{s}_1^k,s)} - P^{s_2\|\underline{s}_1^k}\,{}_LY]$

Table C.3b: Widths for $\theta_t^{(\underline{s}_1^k,s),(\underline{s}_1^l,s)}(\underline{s}_1^k)$ *:*

Cases	$P^{s_2\|\underline{s}_1^k} > \frac{O^{\underline{s}_1^l}(\underline{s}_1^k) - {}_LY}{{}^UY - {}_LY}$	$P^{s_2\|\underline{s}_1^k} < \frac{O^{\underline{s}_1^l}(\underline{s}_1^k) - {}_LY}{{}^UY - {}_LY}$
$P^{s_2\|\underline{s}_1^k} > \frac{{}^UY - O^{\underline{s}_1^l}(\underline{s}_1^k)}{{}^UY - {}_LY}$	$3({}^UY - {}_LY)(1-P^{s_2\|\underline{s}_1^k})$	$2({}^UY - {}_LY)(1-P^{s_2\|\underline{s}_1^k}) + {}^UY - O^{\underline{s}_1^l}(\underline{s}_1^k)$
$P^{s_2\|\underline{s}_1^k} < \frac{{}^UY - O^{\underline{s}_1^l}(\underline{s}_1^k)}{{}^UY - {}_LY}$	$2({}^UY - {}_LY)(1-P^{s_2\|\underline{s}_1^k}) + O^{\underline{s}_1^l}(\underline{s}_1^k) - {}_LY$	$({}^UY - {}_LY)[2(1-P^{s_2\|\underline{s}_1^k}) + P^{s_2\|\underline{s}_1^k}]$

C.2 The implication of OPS on effects similar to those defined in part b) of THEOREM 3 ($\theta_t^{(\underline{s}_{\tau-1}^k,s_\tau^k),(\underline{s}_{\tau-1}^k,s_\tau^l)}(\underline{s}_{\tau-1}^k, s_\tau^j)$ and $\theta_t^{(\underline{s}_{\tau-1}^k,s_\tau^k),(\underline{s}_{\tau-1}^k,s_\tau^l)}(\underline{s}_{\tau-1}^k)$)

C.2.1 Bounds for $\theta_t^{(\underline{s}_{\tau-2}^k,s_{\tau-1}^k,s),(\underline{s}_{\tau-2}^k,s_{\tau-1}^l,s)}(\underline{s}_{\tau-1}^k, s)$ and $\theta_t^{(\underline{s}_{\tau-2}^k,s_{\tau-1}^k,s),(\underline{s}_{\tau-2}^k,s_{\tau-1}^l,s)}(\underline{s}_{\tau-1}^k)$

For treatment effects of the form $\theta_t^{(\underline{s}_{\tau-2}^k,s_{\tau-1}^k,s),(\underline{s}_{\tau-2}^k,s_{\tau-1}^l,s)}(\underline{s}_{\tau-1}^k, s)$, we obtain 4 different possible cases for the bounds' widths. Note that beyond OPS also the deviation of the treatments in the second to last period is a difference to the identified $\theta_t^{(\underline{s}_{\tau-1}^k,s_\tau^k),(\underline{s}_{\tau-1}^k,s_\tau^l)}(\underline{s}_{\tau-1}^k, s_\tau^j)$.[25]

The treatment effects have the following representation:

[25] The bounds for a more general conditioning set, $\theta_t^{(\underline{s}_{\tau-2}^j,s_{\tau-1}^k,s),(\underline{s}_{\tau-2}^j,s_{\tau-1}^l,s)}(\underline{s}_{\tau-1}^j, s)$, are easily derivable using the same procedure as in C.1.1 with $P(S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^j)$, $O^{(\underline{s}_{\tau-2}^j,s_{\tau-1}^k)}(\underline{s}_{\tau-1}^j) = E(Y^{(\underline{s}_{\tau-2}^j,s_{\tau-1}^k)} \mid \underline{s}_{\tau-1}^j)$, and $O^{(\underline{s}_{\tau-2}^j,s_{\tau-1}^l)}(\underline{s}_{\tau-1}^j) = E(Y^{(\underline{s}_{\tau-2}^j,s_{\tau-1}^l)} \mid \underline{s}_{\tau-1}^j)$.

$$\theta_t^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^k, s),(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}(\underline{s}_{\tau-1}^k, s) = E(Y^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^k, s)} \mid \underline{s}_{\tau-1}^k, s) - E(Y^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)} \mid \underline{s}_{\tau-1}^k, s).$$

ASSUMPTION 6-5 identifies $O^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)}(\underline{s}_{\tau-1}^k) := E(Y^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)} \mid \underline{s}_{\tau-1}^k)$. Using LEMMA 2-II, we have the following connection between the above identified expectations and the one of interest:

$$E(Y^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)} \mid \underline{s}_{\tau-1}^k) = \sum_{m=1}^{M} E(Y^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, m)} \mid \underline{s}_{\tau-1}^k, m) P(S_\tau = m \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k). \tag{19}$$

This gives us the bounds on the potential outcomes. To simplify the notation, denote
$P^{s_\tau \mid \underline{s}_{\tau-1}^k} := P(S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k)$. Reformulating equation (19) and taking into account that the expectations stay within their bounds, we obtain the upper and lower bounds for the potential outcomes:

$$U^{Y^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}}(\underline{s}_{\tau-1}^k, s) = \min\left\{ {}^{U}Y; \frac{O^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)}(\underline{s}_{\tau-1}^k) - \sum_{m=1}^{M} {}_{L}Y \underline{1}(m \neq s) P(S_\tau = m \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k)}{P^{s_\tau \mid \underline{s}_{\tau-1}^k}} \right\}$$

$$= \min\left\{ {}^{U}Y; \frac{O^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)}(\underline{s}_{\tau-1}^k) - {}_{L}Y(1 - P^{s_\tau \mid \underline{s}_{\tau-1}^k})}{P^{s_\tau \mid \underline{s}_{\tau-1}^k}} \right\}.$$

$$L^{Y^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}}(\underline{s}_{\tau-1}^k, s) = \max\left\{ {}_{L}Y; \frac{O^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)}(\underline{s}_{\tau-1}^k) - {}^{U}Y(1 - P^{s_\tau \mid \underline{s}_{\tau-1}^k})}{P^{s_\tau \mid \underline{s}_{\tau-1}^k}} \right\}.$$

Thus, the bounds on the effects can be expressed as follows:

$$\theta^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^k, s),(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}(\underline{s}_{\tau-1}^k, s) \in [B^{(\underline{s}_{\tau-1}^k, s)} - U^{Y^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}}(\underline{s}_{\tau-1}^k, s), B^{(\underline{s}_{\tau-1}^k, s)}(\underline{s}_{\tau-1}^k, s) - L^{Y^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}}(\underline{s}_{\tau-1}^k, s)].$$

Taking account of the different cases explicitly, we get:

L1) $\quad P^{s_\tau \mid \underline{s}_{\tau-1}^k} > \dfrac{{}^{U}Y - O^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)}(\underline{s}_{\tau-1}^k)}{{}^{U}Y - {}_{L}Y}: \; L^{Y^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}}(\underline{s}_{\tau-1}^k, s) = \dfrac{O^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)}(\underline{s}_{\tau-1}^k) - (1 - P^{s_\tau \mid \underline{s}_{\tau-1}^k})\, {}^{U}Y}{P^{s_\tau \mid \underline{s}_{\tau-1}^k}}$

L2) $\quad P^{s_\tau \mid \underline{s}_{\tau-1}^k} < \dfrac{{}^{U}Y - O^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)}(\underline{s}_{\tau-1}^k)}{{}^{U}Y - {}_{L}Y}: \quad L^{Y^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}}(\underline{s}_{\tau-1}^k, s) = {}_{L}Y$

U1) $\quad P^{s_\tau \mid \underline{s}_{\tau-1}^k} > \dfrac{O^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)}(\underline{s}_{\tau-1}^k) - {}_{L}Y}{{}^{U}Y - {}_{L}Y}: \quad U^{Y^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}}(\underline{s}_{\tau-1}^k, s) = \dfrac{O^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)}(\underline{s}_{\tau-1}^k) - (1 - P^{s_\tau \mid \underline{s}_{\tau-1}^k})\, {}_{L}Y}{P^{s_\tau \mid \underline{s}_{\tau-1}^k}}$

U2) $\quad P^{s_\tau|\underline{s}^k_{\tau-1}} < \dfrac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - {}_LY}{{}^UY - {}_LY}: \quad U^{Y^{(\underline{s}^k_{\tau-2},s^l_{\tau-1},s)}}(\underline{s}^k_{\tau-1}, s) = {}^UY$

Table C.4a: Upper and lower bounds for $\theta_t^{(\underline{s}^k_{\tau-2},s^k_{\tau-1},s),(\underline{s}^k_{\tau-2},s^l_{\tau-1},s)}(\underline{s}^k_{\tau-1}, s)$

Cases	$P^{s_\tau\|\underline{s}^k_{\tau-1}} > \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - {}_LY}{{}^UY - {}_LY}$	$P^{s_\tau\|\underline{s}^k_{\tau-1}} = \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - {}_LY}{{}^UY - {}_LY}$	$P^{s_\tau\|\underline{s}^k_{\tau-1}} < \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - {}_LY}{{}^UY - {}_LY}$
$P^{s_\tau\|\underline{s}^k_{\tau-1}} > \frac{{}^UY - O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})}{{}^UY - {}_LY}$	$[B^{(\underline{s}^k_{\tau-1},s)} - \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - (1-P^{s_\tau\|\underline{s}^k_{\tau-1}})_LY}{P^{s_\tau\|\underline{s}^k_{\tau-1}}},$ $B^{(\underline{s}^k_{\tau-1},s)} - \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - (1-P^{s_\tau\|\underline{s}^k_{\tau-1}})^UY}{P^{s_\tau\|\underline{s}^k_{\tau-1}}}]$	$[B^{(\underline{s}^k_{\tau-1},s)} - {}^UY,$ $B^{(\underline{s}^k_{\tau-1},s)} - {}^UY + \frac{(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})({}^UY - {}_LY)}{P^{s_\tau\|\underline{s}^k_{\tau-1}}}]$	$[B^{(\underline{s}^k_{\tau-1},s)} - {}^UY,$ $B^{(\underline{s}^k_{\tau-1},s)} - \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - (1-P^{s_\tau\|\underline{s}^k_{\tau-1}})^UY}{P^{s_\tau\|\underline{s}^k_{\tau-1}}}]$
$P^{s_\tau\|\underline{s}^k_{\tau-1}} = \frac{{}^UY - O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})}{{}^UY - {}_LY}$	$[B^{(\underline{s}^k_{\tau-1},s)} - {}_LY - \frac{(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})({}^UY - {}_LY)}{P^{s_\tau\|\underline{s}^k_{\tau-1}}},$ $B^{(\underline{s}^k_{\tau-1},s)} - {}_LY]$	$[B^{(\underline{s}^k_{\tau-1},s)} - {}^UY, B^{(\underline{s}^k_{\tau-1},s)} - {}_LY]$ *)	$[B^{(\underline{s}^k_{\tau-1},s)} - {}^UY, B^{(\underline{s}^k_{\tau-1},s)} - {}_LY]$
$P^{s_\tau\|\underline{s}^k_{\tau-1}} < \frac{{}^UY - O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})}{{}^UY - {}_LY}$	$[B^{(\underline{s}^k_{\tau-1},s)} - \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - (1-P^{s_\tau\|\underline{s}^k_{\tau-1}})_LY}{P^{s_\tau\|\underline{s}^k_{\tau-1}}},$ $B^{(\underline{s}^k_{\tau-1},s)} - {}_LY]$	$[B^{(\underline{s}^k_{\tau-1},s)} - {}^UY, B^{(\underline{s}^k_{\tau-1},s)} - {}_LY]$	$[B^{(\underline{s}^k_{\tau-1},s)} - {}^UY, B^{(\underline{s}^k_{\tau-1},s)} - {}_LY]$

Note: $P^{s_\tau|\underline{s}^k_{\tau-1}} = P(S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}^k_{\tau-1})$, $O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) = E(Y^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})} \mid \underline{S}_{\tau-1} = \underline{s}^k_{\tau-1})$,
$B^{(\underline{s}^k_{\tau-1},s)} = E(Y \mid \underline{S}_{\tau-1} = \underline{s}^k_{\tau-1}, S_\tau = s_\tau)$.
*) This case is only possible if $P^{s_\tau|\underline{s}^k_{\tau-1}} = 0.5$ and $O^{\underline{s}^l_{\tau-1}}(\underline{s}^k_{\tau-1}) = 0.5({}^UY - {}_LY)$. Note also that cases defined by an inequality converge smoothly towards the cases defined by an equality.

Table C.4b: Widths for $\theta_t^{(\underline{s}^k_{\tau-2},s^k_{\tau-1},s),(\underline{s}^k_{\tau-2},s^l_{\tau-1},s)}(\underline{s}^k_{\tau-1}, s)$

Cases	$P^{s_\tau\|\underline{s}^k_{\tau-1}} > \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - {}_LY}{{}^UY - {}_LY}$	$P^{s_\tau\|\underline{s}^k_{\tau-1}} = \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - {}_LY}{{}^UY - {}_LY}$	$P^{s_\tau\|\underline{s}^k_{\tau-1}} < \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - {}_LY}{{}^UY - {}_LY}$
$P^{s_\tau\|\underline{s}^k_{\tau-1}} > \frac{{}^UY - O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})}{{}^UY - {}_LY}$	$\frac{(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})}{P^{s_\tau\|\underline{s}^k_{\tau-1}}}({}^UY - {}_LY)$	$\frac{(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})}{P^{s_\tau\|\underline{s}^k_{\tau-1}}}({}^UY - {}_LY)$	$\frac{{}^UY - O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})}{P^{s_\tau\|\underline{s}^k_{\tau-1}}}]$
$P^{s_\tau\|\underline{s}^k_{\tau-1}} = \frac{{}^UY - O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})}{{}^UY - {}_LY}$	$\frac{(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})}{P^{s_\tau\|\underline{s}^k_{\tau-1}}}({}^UY - {}_LY)$	${}^UY - {}_LY$ *)	${}^UY - {}_LY$
$P^{s_\tau\|\underline{s}^k_{\tau-1}} < \frac{{}^UY - O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})}{{}^UY - {}_LY}$	$\frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - {}_LY}{P^{s_\tau\|\underline{s}^k_{\tau-1}}}$	${}^UY - {}_LY$	${}^UY - {}_LY$

Note: *) This case is only possible if $P^{s_\tau|\underline{s}^k_{\tau-1}} = 0.5$ and $O^{\underline{s}^l_{\tau-1}}(\underline{s}^k_{\tau-1}) = 0.5({}^UY - {}_LY)$. Note also that cases defined by an inequality converge smoothly towards the cases defined by an equality.

Next we show how the result change when we consider a conditioning set that is only defined up to period $\tau-1$. For $\theta_t^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^k, s),(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}(\underline{s}_{\tau-1}^k)$, we obtain 4 different possible cases for the bounds' widths. Using LEMMA 1, we get:

$$\theta_t^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^k, s),(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}(\underline{s}_{\tau-1}^k) = \sum_{m=1}^{M} \theta_t^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^k, s),(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}(\underline{s}_{\tau-1}^k, m) P(S_\tau = m \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k)$$

$$= \sum_{\substack{m=1 \\ m\neq s}}^{M} \theta_t^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^k, s),(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}(\underline{s}_{\tau-1}^k, m) P(S_\tau = m \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k) + \theta_t^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^k, s),(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}(\underline{s}_{\tau-1}^k, s) P(S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k).$$

Using the information in Table C.4a and the non informative bounds ${}^U Y - {}_L Y$, we can easily compute the bounds for $\theta_t^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^k, s),(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}(\underline{s}_{\tau-1}^k)$:

U1) $P^{s_\tau|\underline{s}_{\tau-1}^k} > \dfrac{{}^U Y - O^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)}(\underline{s}_{\tau-1}^k)}{{}^U Y - {}_L Y}$:

$$U^{\theta^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^k, s),(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}}(\underline{s}_{\tau-1}^k) = \sum_{\substack{m=1 \\ m\neq s}}^{M} ({}^U Y - {}_L Y) P(S_\tau = m \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k) + P(S_\tau = s \mid \underline{S}_{\tau-1} = \underline{s}_{\tau-1}^k)\left(B^{(\underline{s}_{\tau-1}^k, s)} - \frac{O^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)}(\underline{s}_{\tau-1}^k) - (1-P^{s_\tau|\underline{s}_{\tau-1}^k})\,{}^U Y}{P^{s_\tau|\underline{s}_{\tau-1}^k}} \right)$$

$$= ({}^U Y - {}_L Y)(1-P^{s_\tau|\underline{s}_{\tau-1}^k}) + P^{s_\tau|\underline{s}_{\tau-1}^k} B^{(\underline{s}_{\tau-1}^k, s)} - O^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)}(\underline{s}_{\tau-1}^k) + (1-P^{s_\tau|\underline{s}_{\tau-1}^k})\,{}^U Y$$

U2) $P^{s_\tau|\underline{s}_{\tau-1}^k} < \dfrac{{}^U Y - O^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)}(\underline{s}_{\tau-1}^k)}{{}^U Y - {}_L Y}$:

$$U^{\theta^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^k, s),(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}}(\underline{s}_{\tau-1}^k) = ({}^U Y - {}_L Y)(1-P^{s_\tau|\underline{s}_{\tau-1}^k}) + P^{s_\tau|\underline{s}_{\tau-1}^k} B^{(\underline{s}_{\tau-1}^k, s)} - P^{s_\tau|\underline{s}_{\tau-1}^k}\,{}_L Y$$

L1) $P^{s_\tau|\underline{s}_{\tau-1}^k} > \dfrac{O^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)}(\underline{s}_{\tau-1}^k) - {}_L Y}{{}^U Y - {}_L Y}$:

$$L^{\theta^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^k, s),(\underline{s}_{\tau-2}^k, s_{\tau-1}^l, s)}}(\underline{s}_{\tau-1}^k) = -({}^U Y - {}_L Y)(1-P^{s_\tau|\underline{s}_{\tau-1}^k}) + P^{s_\tau|\underline{s}_{\tau-1}^k} B^{(\underline{s}_{\tau-1}^k, s)} - O^{(\underline{s}_{\tau-2}^k, s_{\tau-1}^l)}(\underline{s}_{\tau-1}^k) + (1-P^{s_\tau|\underline{s}_{\tau-1}^k})\,{}_L Y$$

L2) $\quad P^{s_\tau|\underline{s}^k_{\tau-1}} < \dfrac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - {}_LY}{{}^UY - {}_LY}$:

$$L^{\theta^{(\underline{s}^k_{\tau-2},s^k_{\tau-1},s),(\underline{s}^k_{\tau-2},s^l_{\tau-1},s)}}(\underline{s}^k_{\tau-1}) = -({}^UY - {}_LY)(1 - P^{s_\tau|\underline{s}^k_{\tau-1}}) + P^{s_\tau|\underline{s}^k_{\tau-1}} B^{(\underline{s}^k_{\tau-1},s)} - P^{s_\tau|\underline{s}^k_{\tau-1}}\,{}^UY$$

Table C.5a: Upper and lower bounds for $\theta_t^{(\underline{s}^k_{\tau-2},s^k_{\tau-1},s),(\underline{s}^k_{\tau-2},s^l_{\tau-1},s)}(\underline{s}^k_{\tau-1})$

Cases	$P^{s_\tau\|\underline{s}^k_{\tau-1}} > \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - {}_LY}{{}^UY - {}_LY}$	$P^{s_\tau\|\underline{s}^k_{\tau-1}} < \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - {}_LY}{{}^UY - {}_LY}$
$P^{s_\tau\|\underline{s}^k_{\tau-1}} > \frac{{}^UY - O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})}{{}^UY - {}_LY}$	$[-({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})+P^{s_\tau\|\underline{s}^k_{\tau-1}}B^{(\underline{s}^k_{\tau-1},s)}-O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})$ $+(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})\,{}_LY, ({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})+$ $P^{s_\tau\|\underline{s}^k_{\tau-1}}B^{(\underline{s}^k_{\tau-1},s)}-O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})+(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})\,{}^UY]$	$[-({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})+P^{s_\tau\|\underline{s}^k_{\tau-1}}B^{(\underline{s}^k_{\tau-1},s)}-P^{s_\tau\|\underline{s}^k_{\tau-1}}\,{}^UY,$ $({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})+P^{s_\tau\|\underline{s}^k_{\tau-1}}B^{(\underline{s}^k_{\tau-1},s)}-O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})$ $+(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})\,{}^UY]$
$P^{s_\tau\|\underline{s}^k_{\tau-1}} < \frac{{}^UY - O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})}{{}^UY - {}_LY}$	$[-({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})+P^{s_\tau\|\underline{s}^k_{\tau-1}}B^{(\underline{s}^k_{\tau-1},s)}-O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})$ $+(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})\,{}_LY, ({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})+P^{s_\tau\|\underline{s}^k_{\tau-1}}B^{(\underline{s}^k_{\tau-1},s)}-P^{s_\tau\|\underline{s}^k_{\tau-1}}\,{}_LY]$	$[-({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})+P^{s_\tau\|\underline{s}^k_{\tau-1}}B^{(\underline{s}^k_{\tau-1},s)}-P^{s_\tau\|\underline{s}^k_{\tau-1}}\,{}^UY,$ $({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})+P^{s_\tau\|\underline{s}^k_{\tau-1}}B^{(\underline{s}^k_{\tau-1},s)}-P^{s_\tau\|\underline{s}^k_{\tau-1}}\,{}_LY]$

Table C.5b: Widths for $\theta_t^{(\underline{s}^k_{\tau-2},s^k_{\tau-1},s),(\underline{s}^k_{\tau-2},s^l_{\tau-1},s)}(\underline{s}^k_{\tau-1})$

Cases	$P^{s_\tau\|\underline{s}^k_{\tau-1}} > \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - {}_LY}{{}^UY - {}_LY}$	$P^{s_\tau\|\underline{s}^k_{\tau-1}} < \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1}) - {}_LY}{{}^UY - {}_LY}$
$P^{s_\tau\|\underline{s}^k_{\tau-1}} > \frac{{}^UY - O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})}{{}^UY - {}_LY}$	$3({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})$	$2({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})+{}^UY-O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})$
$P^{s_\tau\|\underline{s}^k_{\tau-1}} < \frac{{}^UY - O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})}{{}^UY - {}_LY}$	$2({}^UY-{}_LY)(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})+O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-1})-{}_LY$	$({}^UY-{}_LY)[2(1-P^{s_\tau\|\underline{s}^k_{\tau-1}})+P^{s_\tau\|\underline{s}^k_{\tau-1}}]$

C.2.2 Bounds for $\theta_t^{(\underline{s}^k_{\tau-2},s^k_{\tau-1},s),(\underline{s}^k_{\tau-2},s^l_{\tau-1},s)}(\underline{s}^k_{\tau-2},\bullet,s)$

Relating to the second part of THEOREM 3b), for treatment effects of the form $\theta_t^{(\underline{s}^k_{\tau-2},s^k_{\tau-1},s),(\underline{s}^k_{\tau-2},s^l_{\tau-1},s)}(\underline{s}^k_{\tau-2},\bullet,s)$ we obtain 16 different possible cases for the bounds' widths.

The treatment effects have the following representation:

$$\theta_t^{(\underline{s}^k_{\tau-2},s^k_{\tau-1},s),(\underline{s}^k_{\tau-2},s^l_{\tau-1},s)}(\underline{s}^k_{\tau-2},\bullet,s) = E(Y^{(\underline{s}^k_{\tau-2},s^k_{\tau-1},s)} \mid \underline{s}^k_{\tau-2},\bullet,s) - E(Y^{(\underline{s}^k_{\tau-2},s^l_{\tau-1},s)} \mid \underline{s}^k_{\tau-2},\bullet,s)\,.$$

ASSUMPTION 6-5 identifies $O^{(\underline{s}^k_{\tau-2},s^k_{\tau-1})}(\underline{s}^k_{\tau-2}) := E(Y^{(\underline{s}^k_{\tau-2},s^k_{\tau-1})} \mid \underline{s}^k_{\tau-2})$ and $O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-2}) := E(Y^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})} \mid \underline{s}^k_{\tau-2})$. Using LEMMA 2-II, we have the following connection between the above identified expectations and the interesting ones:

$$E(Y^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})} \mid \underline{s}^k_{\tau-2}) = \sum_{m=1}^{M} E(Y^{(\underline{s}^k_{\tau-2},s^l_{\tau-1},m)} \mid \underline{s}^k_{\tau-2},\bullet,m) P(S_\tau = m \mid \underline{S}_{\tau-2} = \underline{s}^k_{\tau-2}) . \qquad (20)$$

This gives the following bounds on the potential outcomes. To simplify the notation, denote $P^{s_\tau \mid \underline{s}^k_{\tau-2}} := P(S_\tau = s \mid \underline{S}_{\tau-2} = \underline{s}^k_{\tau-2})$. Reformulating equation (20) and taking into account that the expectations stay within their bounds, we obtain the upper and lower bounds for the potential outcomes:

$$U^{Y^{(\underline{s}^k_{\tau-2},s^l_{\tau-1},s)}}(\underline{s}^k_{\tau-2},\bullet,s) = \min\left\{ {}^{U}Y; \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-2}) - \sum_{m=1}^{M} {}_{L}Y \underline{1}(m \neq s) P(S_\tau = m \mid \underline{S}_{\tau-2} = \underline{s}^k_{\tau-2})}{P^{s_\tau \mid \underline{s}^k_{\tau-2}}} \right\}$$

$$= \min\left\{ {}^{U}Y; \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-2}) - {}_{L}Y(1 - P^{s_\tau \mid \underline{s}^k_{\tau-2}})}{P^{s_\tau \mid \underline{s}^k_{\tau-2}}} \right\}.$$

$$L^{Y^{(\underline{s}^k_{\tau-2},s^l_{\tau-1},s)}}(\underline{s}^k_{\tau-2},\bullet,s) = \max\left\{ {}_{L}Y; \frac{O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-2}) - {}^{U}Y(1 - P^{s_\tau \mid \underline{s}^k_{\tau-2}})}{P^{s_\tau \mid \underline{s}^k_{\tau-2}}} \right\}.$$

Thus, the bounds on the effects can be expressed as follows:

$$\theta^{(\underline{s}^k_{\tau-2},s^k_{\tau-1},s),(\underline{s}^k_{\tau-2},s^l_{\tau-1},s)}(\underline{s}^k_{\tau-2},\bullet,s) \in [L^{Y^{(\underline{s}^k_{\tau-2},s^k_{\tau-1},s)}}(\underline{s}^k_{\tau-2},\bullet,s) - U^{Y^{(\underline{s}^k_{\tau-2},s^l_{\tau-1},s)}}(\underline{s}^k_{\tau-2},\bullet,s),$$
$$U^{Y^{(\underline{s}^k_{\tau-2},s^k_{\tau-1},s)}}(\underline{s}^k_{\tau-2},\bullet,s) - L^{Y^{(\underline{s}^k_{\tau-2},s^l_{\tau-1},s)}}(\underline{s}^k_{\tau-2},\bullet,s)]$$

The different cases are exactly the same as in Table C.1a with $P(S_\tau = s \mid \underline{S}_{\tau-2} = \underline{s}^k_{\tau-2})$ replacing $P(S_2 = s \mid \underline{S}_1 = \underline{s}^k_1)$, $O^{(\underline{s}^k_{\tau-2},s^k_{\tau-1})}(\underline{s}^k_{\tau-2})$ and $O^{(\underline{s}^k_{\tau-2},s^l_{\tau-1})}(\underline{s}^k_{\tau-2})$ replacing $O^{\underline{s}^k_{\tau-1}}(\underline{s}^k_1)$ and $O^{\underline{s}^l_{\tau-1}}(\underline{s}^k_1)$.

C.3 *Effects similar to those defined in part c) of THEOREM 3* $(\theta_t^{(\underline{s}^j_{\tau-v}, s^k_{\tau-v+1},...,s^k_\tau),(\underline{s}^j_{\tau-v}, s^l_{\tau-v+1},...,s^l_\tau)}(\underline{s}^j_{\tau-w}))$

C.3.1 A finer conditioning set

To analyse the bounds for treatment effects of the form $\theta_t^{(\underline{s}^j_{\tau-1-v}, s^k_{\tau-v},...,s^k_{\tau-1}),(\underline{s}^j_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})}(\underline{s}^j_{\tau-1-w}, s)$ we obtain 16 different possible cases. The simplification used earlier, namely to restrict our presentation to effects conditioning on the same letter as one of the potential treatment ($j = k$), does not conduce here to a reduction of the cases. But for consistency reasons we present the bounds for $j = k$.

The treatment effects have the following representation:

$$\theta_t^{(\underline{s}^k_{\tau-1-v}, s^k_{\tau-v},...,s^k_{\tau-1}),(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})}(\underline{s}^k_{\tau-1-w}, s) = E(Y^{(\underline{s}^k_{\tau-1-v}, s^k_{\tau-v},...,s^k_{\tau-1})} \mid \underline{s}^k_{\tau-1-w}, s) - E(Y^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})} \mid \underline{s}^k_{\tau-1-w}, s)$$

ASSUMPTION 6-5 identifies $O^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})}(\underline{s}^k_{\tau-1-w}) := E(Y^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})} \mid \underline{s}^k_{\tau-1-w})$ and $O^{(\underline{s}^k_{\tau-1-v}, s^k_{\tau-v},...,s^k_{\tau-1})}(\underline{s}^k_{\tau-1-w}) := E(Y^{(\underline{s}^k_{\tau-1-v}, s^k_{\tau-v},...,s^k_{\tau-1})} \mid \underline{s}^k_{\tau-1-w})$. Using LEMMA 1-II, we have the following connection between the above identified expectations and the interesting ones:

$$E(Y^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})} \mid \underline{s}^k_{\tau-1-w}) = \sum_{m=1}^{M} E(Y^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})} \mid \underline{s}^k_{\tau-1-w}, m) P(S_{\tau-w} = m \mid \underline{S}_{\tau-1-w} = \underline{s}^k_{\tau-1-w}) . \qquad (21)$$

This gives the following bounds on the potential outcomes. To simplify the notation, denote $P^{s_{\tau-w}|\underline{s}^k_{\tau-1-w}} := P(S_{\tau-w} = s \mid \underline{S}_{\tau-1-w} = \underline{s}^k_{\tau-1-w})$. Reformulating equation (21) and taking into account that the expectations stay within their bounds, we obtain the upper and lower bounds for the potential outcomes:

$$U^{Y(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})}(\underline{s}^k_{\tau-1-w}, s) = \min\left\{ {}^{U}Y; \frac{O^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})}(\underline{s}^k_{\tau-1-w}) - \sum_{m=1}^{M} {}_{L}Y\underline{1}(m \neq s) P(S_{\tau-w} = m \mid \underline{S}_{\tau-1-w} = \underline{s}^k_{\tau-1-w})}{P^{s_{\tau-w}|\underline{s}^k_{\tau-1-w}}} \right\}$$

$$= \min\left\{ {}^{U}Y; \frac{O^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})}(\underline{s}^k_{\tau-1-w}) - {}_{L}Y(1 - P^{s_{\tau-w}|\underline{s}^k_{\tau-1-w}})}{P^{s_{\tau-w}|\underline{s}^k_{\tau-1-w}}} \right\}.$$

$$L^{Y^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})}}(\underline{s}^k_{\tau-1-w}, s) = \max\left\{ {}_{L}Y; \frac{O^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})}(\underline{s}^k_{\tau-1-w}) - {}^{U}Y(1 - P^{s_{\tau-w}|\underline{s}^k_{\tau-1-w}})}{P^{s_{\tau-w}|\underline{s}^k_{\tau-1-w}}} \right\}.$$

Thus, the bounds on the effects can be expressed as follows:

$$\theta^{(\underline{s}^k_{\tau-1-v}, s^k_{\tau-v},...,s^k_{\tau-1}),(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})}(\underline{s}^k_{\tau-1-w}, s) \in [L^{Y^{(\underline{s}^k_{\tau-1-v}, s^k_{\tau-v},...,s^k_{\tau-1})}}(\underline{s}^k_{\tau-1-w}, s) - U^{Y^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})}}(\underline{s}^k_{\tau-1-w}, s),$$
$$U^{Y^{(\underline{s}^k_{\tau-1-v}, s^k_{\tau-v},...,s^k_{\tau-1})}}(\underline{s}^k_{\tau-1-w}, s) - L^{Y^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})}}(\underline{s}^k_{\tau-1-w}, s)]$$

The different cases are exactly the same as in Table C.1a with $P(S_{\tau-w} = s \mid \underline{S}_{\tau-1-w} = \underline{s}^k_{\tau-1-w})$ replacing $P(S_2 = s \mid \underline{S}_1 = \underline{s}_1^k)$, $O^{(\underline{s}^k_{\tau-1-v}, s^k_{\tau-v},...,s^k_{\tau-1})}(\underline{s}^k_{\tau-1-w})$ and $O^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})}(\underline{s}^k_{\tau-1-w})$ replacing $O^{s^k_{\tau-1}}(\underline{s}_1^k)$ and $O^{s^l_{\tau-1}}(\underline{s}_1^k)$.

C.3.2 The effect of OPS

For all treatment effects of the form $\theta_t^{(\underline{s}^j_{\tau-1-v}, s^k_{\tau-v},...,s^k_{\tau-1},s),(\underline{s}^j_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1},s)}(\underline{s}^j_{\tau-1-w}, \bullet, s)$, we obtain 16 different possible cases for the bounds' widths. The simplification used earlier, namely to restrict our presentation to effects conditioning on the same letter as one of the potential treatment ($j = k$), does not conduce here to a reduction of the cases. But for consistency we present the bounds for $j = k$.

The treatment effects have the following representation:

$$\theta_t^{(\underline{s}^k_{\tau-1-v}, s^k_{\tau-v},...,s^k_{\tau-1},s),(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1},s)}(\underline{s}^k_{\tau-1-w}, \bullet, s) = E(Y^{(\underline{s}^k_{\tau-1-v}, s^k_{\tau-v},...,s^k_{\tau-1},s)} \mid \underline{s}^k_{\tau-1-w}, \bullet, s) - E(Y^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1},s)} \mid \underline{s}^k_{\tau-1-w}, \bullet, s).$$

ASSUMPTION 6-5 identifies $O^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})}(\underline{s}^k_{\tau-1-w}) := E(Y^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})} \mid \underline{s}^k_{\tau-1-w})$ and $O^{(\underline{s}^k_{\tau-1-v}, s^k_{\tau-v},...,s^k_{\tau-1})}(\underline{s}^k_{\tau-1-w}) := E(Y^{(\underline{s}^k_{\tau-1-v}, s^k_{\tau-v},...,s^k_{\tau-1})} \mid \underline{s}^k_{\tau-1-w})$. Using LEMMA 2-II, we have the following connection between the above identified expectations and the interesting ones:

$$E(Y^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1})} \mid \underline{s}^k_{\tau-1-w}) = \sum_{m=1}^{M} E(Y^{(\underline{s}^k_{\tau-1-v}, s^l_{\tau-v},...,s^l_{\tau-1},m)} \mid \underline{s}^k_{\tau-1-w}, \bullet, m) P(S_\tau = m \mid \underline{S}_{\tau-1-w} = \underline{s}^k_{\tau-1-w}). \qquad (22)$$

This gives the following bounds on the potential outcomes. To simplify the notation, denote $P^{s_\tau \mid \underline{s}^k_{\tau-1-w}} := P(S_\tau = s \mid \underline{S}_{\tau-1-w} = \underline{s}^k_{\tau-1-w})$. Reformulating equation (22) and taking into account that the expectations stay within their bounds, we obtain the upper and lower bounds for the potential outcomes:

$$U^{Y(\underline{s}^k_{\tau-1-v},s^l_{\tau-v},\dots,s^l_{\tau-1},s)}(\underline{s}^k_{\tau-1-w},\bullet,s)=\min\left\{{}^{U}Y;\frac{O^{(\underline{s}^k_{\tau-1-v},s^l_{\tau-v},\dots,s^l_{\tau-1})}(\underline{s}^k_{\tau-1-w})-\sum_{m=1}^{M}{}_{L}Y\underline{1}(m\neq s)P(S_\tau=m\,|\,\underline{S}_{\tau-1-w}=\underline{s}^k_{\tau-1-w})}{P^{s_\tau|\underline{s}^k_{\tau-1-w}}}\right\}$$

$$=\min\left\{{}^{U}Y;\frac{O^{(\underline{s}^k_{\tau-1-v},s^l_{\tau-v},\dots,s^l_{\tau-1})}(\underline{s}^k_{\tau-1-w})-{}_{L}Y(1-P^{s_\tau|\underline{s}^k_{\tau-1-w}})}{P^{s_\tau|\underline{s}^k_{\tau-1-w}}}\right\}.$$

$$L^{Y^{(\underline{s}^k_{\tau-1-v},s^l_{\tau-v},\dots,s^l_{\tau-1},s)}}(\underline{s}^k_{\tau-1-w},\bullet,s)=\max\left\{{}_{L}Y;\frac{O^{(\underline{s}^k_{\tau-1-v},s^l_{\tau-v},\dots,s^l_{\tau-1})}(\underline{s}^k_{\tau-1-w})-{}^{U}Y(1-P^{s_\tau|\underline{s}^k_{\tau-1-w}})}{P^{s_\tau|\underline{s}^k_{\tau-1-w}}}\right\}.$$

Thus, the bounds on the effects can be expressed as follows:

$$\begin{aligned}\theta^{(\underline{s}^k_{\tau-1-v},s^k_{\tau-v},\dots,s^k_{\tau-1},s),(\underline{s}^k_{\tau-1-v},s^l_{\tau-v},\dots,s^l_{\tau-1},s)}(\underline{s}^k_{\tau-1-w},\bullet,s)\in[&L^{Y^{(\underline{s}^k_{\tau-1-v},s^k_{\tau-v},\dots,s^k_{\tau-1},s)}}(\underline{s}^k_{\tau-1-w},\bullet,s)-U^{Y^{(\underline{s}^k_{\tau-1-v},s^l_{\tau-v},\dots,s^l_{\tau-1},s)}}(\underline{s}^k_{\tau-1-w},\bullet,s),\\ &U^{Y^{(\underline{s}^k_{\tau-1-v},s^k_{\tau-v},\dots,s^k_{\tau-1},s)}}(\underline{s}^k_{\tau-1-w},\bullet,s)-L^{Y^{(\underline{s}^k_{\tau-1-v},s^l_{\tau-v},\dots,s^l_{\tau-1},s)}}(\underline{s}^k_{\tau-1-w},\bullet,s)]\end{aligned}.$$

The different cases are exactly the same as in C.3.1 with $P(S_\tau=s\,|\,\underline{S}_{\tau-1-w}=\underline{s}^k_{\tau-1-w})$.